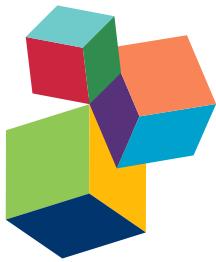


# **THE PROCEEDINGS FROM HALOPHILES 2013, THE INTERNATIONAL CONGRESS ON HALOPHILIC MICROORGANISMS**

**EDITED BY: R. Thane Papke, Aharon Oren, Antonio Ventosa and**

**Jesse G. Dillon**

**PUBLISHED IN: Frontiers in Microbiology**



# frontiers

## **Frontiers Copyright Statement**

© Copyright 2007-2015 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

**ISSN 1664-8714**

**ISBN 978-2-88919-570-1**

**DOI 10.3389/978-2-88919-570-1**

## **About Frontiers**

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## **Frontiers Journal Series**

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## **Dedication to Quality**

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## **What are Frontiers Research Topics?**

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)

# THE PROCEEDINGS FROM HALOPHILES 2013, THE INTERNATIONAL CONGRESS ON HALOPHILIC MICROORGANISMS

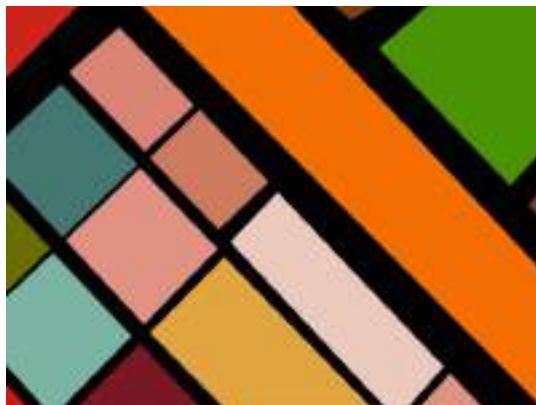
Topic Editors:

**R. Thane Papke**, University of Connecticut, USA

**Aharon Oren**, The Hebrew University of Jerusalem, Israel

**Antonio Ventosa**, University of Sevilla, Spain

**Jesse G. Dillon**, California State University, Long Beach, USA



Thank you to Austin Wood, Dale Thompson and the Great Salt Lake Institute for providing the cover art.

The Halophiles 2013 meeting is a multidisciplinary international congress, with a strong history of regular triennial meetings since 1978. Our mission is to bring researchers from a wide diversity of investigation interests (e.g., protein and species evolution; niche adaptation, ecology, taxonomy, genomics, metagenomics, horizontal gene transfer, gene regulation; DNA replication, repair and recombination; signal transduction; community assembly and species distribution; astrobiology; biotechnological applications; adaptation to radiation,

desiccation, osmotic stress) into a single forum for the integration and synthesis of ideas and data from all three domains of life, and their viruses, yet from a single environment; salt concentrations greater than seawater. This cross-section of research informs our understanding of the microbiological world in many ways. The halophilic environment is extreme, especially above 10% NaCl, restricting life solely to microbes. The microorganisms that live there are adapted to extreme conditions, and are notable for their ability to survive high doses of radiation and desiccation. Therefore, the hypersaline environment is a model system (both the abiotic, and biologic factors) for insightful understanding regarding conditions and life in the absence of plant and animals (e.g., life on the early earth, and other solar system bodies like Mars and Europa). Lower salinity conditions (e.g., 6-10% NaCl) form luxuriant microbial mats considered modern analogues of fossilized stromatolites, which are enormous microbially produced structures fashioned during

the Precambrian (and still seen today in places like Shark's Bay, Australia). Hypersaline systems are island-like habitats spread patchily across the earth's surface, and similar to the Galapagos Islands represent unique systems excellent for studying the evolutionary pressures that shape microbial community assembly, adaptation, and speciation. The unique adaptations to this extreme environment produce valuable proteins, enzymes and other molecules capable of remediating harsh human instigated environments, and are useful for the production of biofuels, vitamins, and retinal implants, for example. This research topic is intended to capture the breadth and depth of these topics.

**Citation:** Papke, R. T., Oren, A., Ventosa, A., Dillon, J. G., eds. (2015). The Proceedings from Halophiles 2013, the International Congress on Halophilic Microorganisms. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-570-1

# Table of Contents

## 07 *Preface to the proceedings of Halophiles 2013*

R. Thane Papke

### A league of their own

#### 12 *Salty sisters: the women of halophiles*

Bonnie K. Baxter, Nina Gunde-Cimerman and Aharon Oren

### Communities, diversity and, evolution

#### 25 *Going from microbial ecology to genome data and back: studies on a haloalkaliphilic bacterium isolated from Soap Lake, Washington State*

Melanie R. Mormile

#### 35 *Recent studies in microbial degradation of petroleum hydrocarbons in hypersaline environments*

Babu Z. Fathepure

#### 51 *Mining a database of single amplified genomes from Red Sea brine pool extremophiles—improving reliability of gene function prediction using a profile and pattern matching algorithm (PPMA)*

Stefan W. Grötzinger, Intikhab Alam, Wail Ba Alawi, Vladimir B. Bajic, Ulrich Stingl and Jörg Eppinger

#### 65 *Population and genomic analysis of the genus Halorubrum*

Matthew S. Fullmer, Shannon M. Soucy, Kristen S. Swithers, Andrea M. Makkay, Ryan Wheeler, Antonio Ventosa, J. Peter Gogarten and R. Thane Papke

#### 80 *Patterns of microbial diversity along a salinity gradient in the Guerrero Negro solar saltern, Baja CA Sur, Mexico*

Jesse G. Dillon, Mark Carlin, Abraham Gutierrez, Vivian Nguyen and Nathan McLain

#### 93 *Evidence from phylogenetic and genome fingerprinting analyses suggests rapidly changing variation in Halorubrum and Haloarcula populations*

Nikhil Ram Mohan, Matthew S. Fullmer, Andrea M. Makkay, Ryan Wheeler, Antonio Ventosa, Adit Naor, J. Peter Gogarten and R. Thane Papke

#### 102 *Inteins as indicators of gene flow in the halobacteria*

Shannon M. Soucy, Matthew S. Fullmer, R. Thane Papke and Johann Peter Gogarten

#### 116 *Comparison of prokaryotic community structure from Mediterranean and Atlantic saltern concentrator ponds by a metagenomic approach*

Ana B. Fernández, Blanca Vera-Gargallo, Cristina Sánchez-Porro, Rohit Ghai, R. Thane Papke, Francisco Rodriguez-Valera and Antonio Ventosa

- 128 Raman spectroscopy in halophile research**  
Jan Jehlička and Aharon Oren
- 135 A Glimpse of the genomic diversity of haloarchaeal tailed viruses**  
Ana Senčilo and Elina Roine
- 141 Halophilic Aspergillus penicilliodes from athalassohaline, thalassohaline, and polyhaline environments**  
Sarita W. Nazareth and Valerie Gonsalves

### **Adaptations and metabolism**

- 146 Extracellular DNA metabolism in Haloferax volcanii**  
Scott Chimileski, Kunal Dolas, Adit Naor, Uri Gophna and R. Thane Papke
- 158 Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes**  
Aharon Oren
- 164 Adaptation to high salt concentrations in halotolerant/halophilic fungi: a molecular perspective**  
Ana Plemenitaš, Metka Lenassi, Tilen Konte, Anja Kejžar, Janja Zajc, Cene Gostinčar and Nina Gunde-Cimerman
- 176 Polyploidy in haloarchaea: advantages for growth and survival**  
Karolin Zerulla and Jörg Soppa
- 184 Light-dependent expression of four cryptic archaeal circadian gene homologs**  
Michael Maniscalco, Jennifer Nannen, Valerie Sodi, Gillian Silver, Phillip L. Lowrey and Kelly A. Bidle
- 194 Dihydroxyacetone metabolism in Haloferax volcanii**  
Matthew Ouellette, Andrea M. Makkay and R. Thane Papke
- 205 Protective role of salt in catalysis and maintaining structure of halophilic proteins against denaturation**  
Rajeshwari Sinha and Sunil K. Khare

### **Biochemistry and molecular biology**

- 211 An experimental point of view on hydration/solvation in halophilic proteins**  
Romain Talon, Nicolas Coquelle, Dominique Madern and Eric Girard
- 219 The haloarchaeal MCM proteins: bioinformatic analysis and targeted mutagenesis of the  $\beta$ 7- $\beta$ 8 and  $\beta$ 9- $\beta$ 10 hairpin loops and conserved zinc binding domain cysteines**  
Tatjana P. Kristensen, Reeja Maria Cherian, Fiona C. Gray and Stuart A. MacNeill
- 232 Identification of carotenoids from the extremely halophilic archaeon Haloarcula japonica**  
Rie Yatsunami, Ai Ando, Ying Yang, Shinichi Takaichi, Masahiro Kohno, Yuriko Matsumura, Hiroshi Ikeda, Toshiaki Fukui, Kaoru Nakasone, Nobuyuki Fujita, Mitsuo Sekine, Tomonori Takashina and Satoshi Nakamura
- 237 Prebiotic protein design supports a halophile origin of foldable proteins**  
Liam M. Longo and Michael Blaber
- 240 Glass-forming property of hydroxyectoine is the cause of its superior function as a desiccation protectant**  
Christoph Tanne, Elena A. Golovina, Folkert A. Hoekstra, Andrea Meffert and Erwin A. Galinski

**253 *Glutamine synthetase 2 is not essential for biosynthesis of compatible solutes in Halobacillus halophilus***

Anna Shiyan, Melanie Thompson, Saskia Köcher, Michaela Tausendschön,  
Helena Santos, Inga Hänelt and Volker Müller

**264 *N-glycosylation in Haloferax volcanii: adjusting the sweetness***

Jerry Eichler, Adi Arbiv, Chen Cohen-Rosenzweig, Lina Kaminski, Lina Kandiba and  
Zvia Konrad

# Preface to the proceedings of Halophiles 2013

R. Thane Papke \*

Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT, USA

**Keywords:** *Haloferax volcanii*, halophilic and halotolerant microorganisms, halophile biochemistry, halophile molecular biology, halophile metabolism, halophile adaptations, halophile communities, halophile evolution

## Introduction

Halophiles are organisms, and their viruses, adapted to thriving and reproducing in high salt concentrations, representatives of which can be found in all domains of life. To flourish in these environments, an individual must overcome many obstacles including high osmotic pressure, low water availability, high and low pH conditions, high temperatures, and high cell densities that enforce stiff competition for limited resources due to the low solubility of gasses and other nutrients. Therefore, all halophiles can be considered poly-extremophiles.

These interesting life forms do not cause infectious diseases or cancer, impact human lifestyles or infect our food supply. So why study them? Often halophiles are studied because the salty environment is a model system for uncovering basic principles of microbial life. This extreme environment selects against any organism not able to cope with high salt, which reduces the overall community structure and function. The expectation is that this decrease in complexity allows for more easily achieved insights into fundamental microbial adaptations, and ecological, biogeochemical and evolutionary processes. Because the extreme hypersaline environments are limited to only microbial life, the habitat is analogous to that which existed on Earth before the Cambrian Explosion. Thus investigations of hypersaline habitats can deliver insight into the longest epoch of life. Modern microbial mats growing in hypersaline habitats are similar in structure to ancient stromatolites found in the Precambrian fossil record. Related to this is the search for life in the cosmos: the long period of dominant microbial life on Earth suggests a higher likelihood of finding microbial life on another world than finding advanced life, sentient beings, or even trees. Additionally, the adaptations required for life in high salt can produce enzymes that are interesting for biotechnology, industrial processes, and bioremediation.

Every 3 years since the late 1970s halophile researchers who focus on microbial life in hypersaline environments have gathered to present their latest exciting research to like-minded souls. This eBook is a compendium of research written by many of the presenters at Halophiles 2013, the international congress held at the University of Connecticut in Storrs, CT (to view all of the conference oral and poster presentation abstracts, please visit [https://www.regonline.com/custImages/250000/250066/Halophiles2013Program\\_final.pdf](https://www.regonline.com/custImages/250000/250066/Halophiles2013Program_final.pdf)). Its range of subject matter is extensive reflecting the 4-day event, and the breadth of research interests. To address this diversity of topics, the chapters are arranged into slightly narrower areas of research interests, but even within these defined partitions the reader will find a wide range of interesting research. Here, we explore the world of halophiles.

## OPEN ACCESS

**Edited and reviewed by:**

Andreas Teske,  
University of North Carolina at Chapel  
Hill, USA

**\*Correspondence:**

R. Thane Papke,  
[thane@uconn.edu](mailto:thane@uconn.edu)

**Specialty section:**

This article was submitted to  
*Extreme Microbiology*,  
a section of the journal  
*Frontiers in Microbiology*

**Received:** 16 March 2015

**Accepted:** 07 April 2015

**Published:** 22 April 2015

**Citation:**

Papke RT (2015) Preface to the  
proceedings of Halophiles 2013.  
*Front. Microbiol.* 6:341.  
doi: 10.3389/fmicb.2015.00341

## A League of Their Own

The lead article by Baxter et al. (2014) is based on the keynote talk of the conference, which is a reflection and commentary on gender bias in science and more specifically in the field of halophiles.

The authors conducted a thorough study of women's success in halophile research through time by measuring who gives invited lectures at the Halophile conferences, an honor that connotes peer recognition obtained through individual hard work and dedication. The authors found that the field of halophiles is "an example of progress" achieving gender equitability not found in many fields. Further, many significant women researchers and their contributions to the field are highlighted in this paper: it is a long, and compelling list.

## Communities, Diversity, and Evolution

The first chapter in the section on Communities, Diversity, and Evolution is the most frequently viewed article from the conference proceedings; it has been viewed several 100 times more often than its closest competitor, and more than a 1000 compared to the 3rd most viewed contribution. There is no surprise it was written by Mormile (2014), one of the featured women in the Baxter et al., study. This manuscript highlights the metabolism and adaptations discovered through sequencing the *Halanaerobium hydrogeniformans* genome, a hydrogen producing haloalkaliphile cultivated from Soap Lake, WA.

The second chapter in this section by Fathepure (2014) is a review on the bioremediation of petroleum compounds in hypersaline environments. Oil production generates hypersaline waters due to the extraction process, and many petroleum rich regions are surrounded by natural hypersaline systems like sabkhas or coastal salt marshes that experience crude oil pollution. Therefore, remediation of "production water" and polluted natural brines requires for decontamination microbes and enzymes that are adapted to function at elevated NaCl concentrations. This manuscript gives an up to date review on the state of the field.

Grotzinger et al. (2014) used cell sorting followed by whole genome amplification from single cells and genome sequencing to examine communities that live in hypersaline pools existing at the bottom of the Red Sea. They were particularly interested in discovering genes from halophiles that might be of commercial value (e.g., hydrolases, dehydrogenases). To aid their cultivation independent search, they developed a bioinformatic tool called a profile and pattern matching algorithm to find genes of interest. In this manuscript they detail their efforts, successes and failures, at mining the single amplified genome data.

Fullmer et al. (2014) sequenced genomes from 19 *Halorubrum* strains all cultivated from the hypersaline lake Aran Bidgol located in Iran and compared them to *Halorubrum* genomes available at the NCBI database. Multilocus sequence analysis was used to establish a phylogeny among strains and the robustness of phylogenetic clusters was tested by average nucleotide identity and G + C content, which showed conformity unless groups were comprised of strains cultivated from different geographic locations. Inteins and clustered regularly interspaced short palindromic repeats (CRISPRs) within and between phylogenetic assemblages were patchily distributed among all strains including those that were >99.5% identical for DNA sequence across five protein coding loci.

In the next chapter, Dillon et al. (2013) examined different salinity ponds from the Guerrero Negro solar saltern in Baja California Sur, Mexico using 16S rRNA and bacteriorhodopsin genes as molecular markers for assessing microbial diversity. This study showed that contrary to expectations, ponds of similar salinity had variable community structure; bacterial exceeded archaeal diversity; one of the ponds was dominated by clones that were unrelated to *Salinibacter ruber*, the typically reported bacterial inhabitant of saturated brines. The authors also showed that the haloarchaeal diversity in lower salinities was largely previously uncharacterized.

Multilocus sequence analysis and genome fingerprinting of 43 *Halorubrum* and *Haloarcula* strains cultivated from the hypersaline lake Aran Bidgol in Iran (Mohan et al., 2014) showed that these haloarchaeal genomes were exceedingly dynamic: nearly every strain examined had a unique fingerprint, even ones with 100% DNA sequence identity for five protein coding genes (~2500 nt). As a result, the authors concluded that the accumulation of this dramatic genomic variance was due to extensive gene gain and loss, which occurred faster than the neutral mutation rate.

Inteins are selfish genetic elements that insert themselves into highly conserved proteins. Through extensive haloarchaeal genome analysis (118 genomes, from 26 genera) Soucy et al. (2014) showed that most inteins had invaded only a fraction of the available insertion sites, though some were more adept than others. The absence of inteins despite available invasion sites indicates an inability to be mobilized likely based on the capacity of cells to exchange DNA with other community members. Therefore, the authors suggested gene transfer is not random in the haloarchaea, but instead exhibits extensive biases.

The study by Fernandez et al. (2014) used metagenomics to compare the community structure and functional properties between two saltern systems located on different coasts of Spain. Of significance was the observation that middle salinity concentrator ponds can have highly variable communities. The community from the Atlantic saltern 21% NaCl pond was more similar in structure to the Mediterranean saltern 33% pond, rather than a similar salinity pond (19%) at the same location. Additional analyses indicated carbon and phosphate availability might have a greater effect than NaCl concentrations in determining community structure.

Raman spectroscopy has become a useful tool in assessing biomolecules and minerals in strains and communities. Here, Jehlicka and Oren (2013) review how this technique has been applied to halophile communities living in gypsum crusts, evaporitic sediments, halite inclusions and endoliths, as well as to cultures for the purpose of describing the detection and distribution of important microbiological and geochemical markers.

Sencilo and Roine (2014) review what is known about the genomes of tailed viruses that infect haloarchaeal cells. Perhaps it is not surprising that these viruses are highly adapted to their hosts as their DNA G + C percentage is very high, like that of their hosts. Unexpectedly however, it appears they have much in common with their bacterial counterparts including genome content and organization, and similar capsid architecture and assembly. Thus, these strong commonalities suggest deep evolutionary

relationships and a possible common ancestor for all tailed viruses/phage.

Nazareth and Gonsalves (2014) cultivated halophilic *Aspergillus* strains from many different hypersaline environments and characterized their ability to grow in different concentrations of salt, and on different carbon sources. Conidia germination and morphological changes in response to different salt concentrations were also examined. As no growth or conidia germination was detected in media without salt, and optimum growth was determined to be around 10%, these fungal strains were considered truly halophilic.

## Adaptations and Metabolism

Despite being available in all habitats, little is known about DNA as a nutrient. In this publication by Chimileski et al. (2014) *Haloferax volcanii* was used as a model organism for exploring extracellular DNA metabolism. *Hfx. volcanii* grew best on DNA as a phosphate source, only slightly as a nitrogen source, and not at all as a carbon source. Furthermore, these cells were fussy about the sources of DNA they consumed and the bias was based on DNA methylation. These authors also identified and confirmed the gene HVO\_1477 is required for growth on DNA, and that its homologs are widespread in archaea.

This review by Oren (2013) examines the dogma surrounding the linkage between excessive acidic amino acids in the proteomes of cells and the presence of high intracellular KCl concentrations used for osmotic balance. While the canonical examples *Halobacteriales* and *Salinibacter ruber* demonstrate both the salt-in strategy and an acidic proteome, recent genomic and metagenomic analyses revealed the decoupling of those two phenotypes. These new findings are unexplained but it is clear that our current understanding is too simplistic.

Plemenitas et al. (2014) review what is known about fungal adaptations to high salt concentrations from a molecular and genomic perspective, by comparing the salt tolerant *Hordeum werneckii* and the obligate halophile *Wallemia ichthyophaga*. They show that though signaling pathways necessary for sensing and responding to increasing salt concentrations are conserved between them, the observed structural and regulatory differences could account for their overall salt adaptations. Further, genomic analyses show substantial evolutionary or adaptation strategy differences between them.

The presence of multiple chromosome copies offers many advantages to the survival of cells that have this phenotype. Zerulla and Soppa (2014) review these advantages for species of haloarchaea, which have demonstrated high copy numbers, even in stationary phase. Most evolutionary explanations for the presence and origin of polyploidy are based on repair of damaged or mutated DNA and require the precondition of homologous recombination. Recent work on copy number and DNA as phosphate source however suggest polyploidy could stem from a need for intracellular phosphate storage.

Circadian rhythm has been studied in two of the three domains of life but nothing was known about the subject in Archaea, except for the notable presence of cyanobacterial *Kai*-family genes. Maniscalco et al. (2014) studied *cir* gene

expression in *Haloferax volcanii* and demonstrated that those homologs are upregulated during 12 h diurnal cycles compared to dark conditions alone: they also showed that gene knockouts disrupted rhythmic gene expression. This groundbreaking work should cast bright light on archaeal circadian rhythms.

Metabolism of dihydroxyacetone (DHA) in the haloarchaea was thought only to occur in the species *Haloquadratum walsbyi*, and requires kinases. This manuscript by Ouellette et al. (2013) demonstrated *Haloferax volcanii* also grows on DHA as a sole carbon source, and that though phosphorylated by a DHA kinase, phosphorylation of DHA primarily occurs by a glycerol kinase. Further, genomic analyses unexpectedly showed that DHA and glycerol kinases are widespread throughout the haloarchaea, suggesting DHA is an important nutrient for all species to metabolize.

In this review chapter by Sinha and Khare (2014), the role of salt in controlling the stability of proteins is explored. Though it has been known that proteins adapted to hypersaline conditions typically are not functional without salt, the presence of salt also provides proteins with protection against the denaturing effects of temperature, chaotropic agents, organic solvents, and mutations. Understanding how this effect works could lead to the development of better biocatalysts.

## Biochemistry and Molecular Biology

Talon et al. (2014) used malate dehydrogenases from *Chloroflexus aurantiacus* and *Salinibacter ruber* as models for understanding the adaptations required for the solvation of proteins under hypersaline conditions. They demonstrated that water molecules have indirect and direct hydrogen bonds with the *C. aurantiacus* and *S. ruber* proteins respectively, which stabilized the particular versions. The substitution of non-polar amino acids in *C. aurantiacus* by acidic ones on the surface of the *S. ruber* protein was noted and thermodynamic arguments indicated these were the appropriate adaptations to high internal salt concentration experienced by *S. ruber*, and by extension all haloarchaea.

The DNA replication helicase catalytic core is homologous between eukaryotes and archaea, suggesting an archaeal model organism can provide deeper understanding into its structure and function. In this chapter by Kristensen et al. (2014) the first extensive *in vivo* genetic manipulation of the MCM complex for an archaeon is reported. Guided by multiple sequence alignments and a crystal structure many conserved amino acids –singletons, small clusters and larger clusters– were deleted and assessed for impact. Results indicate that *Haloferax volcanii* is an excellent model organism for reverse genetic analysis of MCM, and other key eukaryote homologs.

Yatsunami et al. (2014) report here the discovery, the cellular composition, and the antioxidant potential of carotenoids produced by the previously uninvestigated species *Haloarcula japonica*. Their results suggest that *H. japonica* may have very high carotenoid content compared to other haloarchaeal species, which may confer higher resistance to damaging radiation.

The origin and evolution of amino acids is largely inferred but a consensus of 10 are suggested to have been available in the prebiotic Earth. Here Longo and Blaber (2014) report the

analysis of enriching a protein with the set of prebiotic amino acids and determining its folding potential. They noticed that proteins remained stable when the core was hydrophobic and the surface had a high negative charge (i.e., acidic amino acids). Both of these characteristics are found in proteins adapted to being soluble in high salt concentrations leading the authors to suggest the prebiotic early earth environment may have been very salty.

Compatible solutes are organic molecules intracellularly accumulated in many halophiles to balance the osmotic pressure of their external environment. Hydroxyectoine is a common compatible solute and it also protects cells and proteins from desiccation, and heat. Tanne et al. (2014) show that hydroxyectoine produced by *Chromohalobacter salexigens*, from the family *Halomonadaceae*, has physical properties that allows the biological processes of cells to continue functioning even in a dehydrated condition.

The moderate halophile *Halobacillus halophilus* produces glutamate and glutamine as compatible solutes. Shiyan et al. (2014) hypothesized that the annotated glutamine synthetase A2 found in the *H. halophilus* genome was key to the biosynthesis of glutamate and glutamine as a compatible solute. To their surprise, knock out analysis revealed this enzyme was not involved, indicating some unknown enzyme must be responsible for generating these compatible solutes.

## References

- Baxter, B. K., Gunde-Cimerman, N., and Oren, A. (2014). Salty sisters: the women of halophiles. *Front. Microbiol.* 5:192. doi: 10.3389/fmicb.2014.00192
- Chimileski, S., Dolas, K., Naor, A., Gophna, U., and Papke, R. T. (2014). Extracellular DNA metabolism in *Halofexax volcanii*. *Front. Microbiol.* 5:57. doi: 10.3389/fmicb.2014.00057
- Dillon, J. G., Carlin, M., Gutierrez, A., Nguyen, V., and McLain, N. (2013). Patterns of microbial diversity along a salinity gradient in the Guerrero Negro solar saltern, Baja CA Sur, Mexico. *Front. Microbiol.* 4:399. doi: 10.3389/fmicb.2013.00399
- Eichler, J., Arbib, A., Cohen-Rosenzweig, C., Kaminski, L., Kandiba, L., and Konrad, Z. (2013). N-glycosylation in *Halofexax volcanii*: adjusting the sweetness. *Front. Microbiol.* 4:403. doi: 10.3389/fmicb.2013.00403
- Fathepure, B. Z. (2014). Recent studies in microbial degradation of petroleum hydrocarbons in hypersaline environments. *Front. Microbiol.* 5:173. doi: 10.3389/fmicb.2014.00173
- Fernandez, A. B., Vera-Garcia, B., Sanchez Perez, G., Ghai, R., Papke, R. T., Rodriguez-Valera, F., et al. (2014). Comparison of prokaryotic community structure from Mediterranean and Atlantic saltern concentrator ponds by a metagenomic approach. *Front. Microbiol.* 5:196. doi: 10.3389/fmicb.2014.00196
- Fullmer, M. S., Soucy, S. M., Swithers, K. S., Makkay, A. M., Wheeler, R., Ventosa, A., et al. (2014). Population and genomic analysis of the genus Halorubrum. *Front. Microbiol.* 5:140. doi: 10.3389/fmicb.2014.00140
- Grotzinger, S. W., Alam, I., Ba Alawi, W., Bajic, V. B., Stingl, U., and Eppinger, J. (2014). Mining a database of single amplified genomes from Red Sea brine pool extremophiles-improving reliability of gene function prediction using a profile and pattern matching algorithm (PPMA). *Front. Microbiol.* 5:134. doi: 10.3389/fmicb.2014.00134
- Jehlicka, J., and Oren, A. (2013). Raman spectroscopy in halophile research. *Front. Microbiol.* 4:380. doi: 10.3389/fmicb.2013.00380
- Kristensen, T. P., Maria Cherian, R., Gray, F. C., and Macneill, S. A. (2014). The haloarchaeal MCM proteins: bioinformatic analysis and targeted mutagenesis of the beta7-beta8 and beta9-beta10 hairpin loops and conserved zinc binding domain cysteines. *Front. Microbiol.* 5:123. doi: 10.3389/fmicb.2014.00123
- Longo, L. M., and Blaber, M. (2014). Prebiotic protein design supports a halophile origin of foldable proteins. *Front. Microbiol.* 4:418. doi: 10.3389/fmicb.2013.00418
- Maniscalco, M., Nannen, J., Sodi, V., Silver, G., Lowrey, P. L., and Bidle, K. A. (2014). Light-dependent expression of four cryptic archaeal circadian gene homologs. *Front. Microbiol.* 5:79. doi: 10.3389/fmicb.2014.00079
- Mohan, N. R., Fullmer, M. S., Makkay, A. M., Wheeler, R., Ventosa, A., Gogarten, J. P., et al. (2014). Evidence from phylogenetic and genome fingerprining analyses suggests rapidly changing variation in Halorubrum and Haloarcula populations. *Front. Microbiol.* 5:143. doi: 10.3389/fmicb.2014.00143
- Mormile, M. R. (2014). Going from microbial ecology to genome data and back: studies on a haloalkaliphilic bacterium isolated from Soap Lake, Washington State. *Front. Microbiol.* 5:628. doi: 10.3389/fmicb.2014.00628
- Nazareth, S. W., and Gonsalves, V. (2014). Halophilic *Aspergillus penicilliodes* from athalassohaline, thalassohaline, and polyhaline environments. *Front. Microbiol.* 5:412. doi: 10.3389/fmicb.2014.00412
- Oren, A. (2013). Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. *Front. Microbiol.* 4:315. doi: 10.3389/fmicb.2013.00315
- Ouellette, M., Makkay, A. M., and Papke, R. T. (2013). Dihydroxyacetone metabolism in *Halofexax volcanii*. *Front. Microbiol.* 4:376. doi: 10.3389/fmicb.2013.00376
- Plemenitas, A., Lenassi, M., Konte, T., Kejzar, A., Zajc, J., Gostincar, C., et al. (2014). Adaptation to high salt concentrations in halotolerant/halophilic fungi: a molecular perspective. *Front. Microbiol.* 5:199. doi: 10.3389/fmicb.2014.00199
- Sencilo, A., and Roine, E. (2014). A Glimpse of the genomic diversity of haloarchaeal tailed viruses. *Front. Microbiol.* 5:84. doi: 10.3389/fmicb.2014.00084
- Shiyan, A., Thompson, M., Kocher, S., Tausendschon, M., Santos, H., Hanelt, I., et al. (2014). Glutamine synthetase 2 is not essential for biosynthesis of compatible solutes in *Halobacillus halophilus*. *Front. Microbiol.* 5:168. doi: 10.3389/fmicb.2014.00168
- Sinha, R., and Khare, S. K. (2014). Protective role of salt in catalysis and maintaining structure of halophilic proteins against denaturation. *Front. Microbiol.* 5:165. doi: 10.3389/fmicb.2014.00165

The post translational addition of glycans to proteins (N-glycosylation), though found in all domains of life, is an important adaptation for organisms living in the high salt environment, as modification in response to changing salinity conditions provides flexibility to a protein's ability to remain soluble and functioning. This chapter by Eichler et al. (2013) reviews the how's, what's, where's, and why's of N-glycosylation for *Halofexax volcanii* in response to salinity fluctuations.

As the lead organizer of the conference, and co-editor of this eBook, it was a great pleasure for me to serve the community and orchestrate this meeting: It was a labor of love and I would gladly do it again. It was truly an unforgettable conference, everyone had a great time, and we are all looking forward to the next one, Halophiles 2016, to be held in San Juan, Puerto Rico.

## Acknowledgments

I would like to thank the halophile community, those who attended the Halophiles 2013 conference, the conference organizers, and all the authors who contributed to this eBook. I would also like to thank my co-editors for their efforts in making this project run smoothly and effortlessly. This is only possible because of all of you!

- Soucy, S. M., Fullmer, M. S., Papke, R. T., and Gogarten, J. P. (2014). Inteins as indicators of gene flow in the halobacteria. *Front. Microbiol.* 5:299. doi: 10.3389/fmicb.2014.00299
- Talon, R., Coquelle, N., Madern, D., and Girard, E. (2014). An experimental point of view on hydration/solvation in halophilic proteins. *Front. Microbiol.* 5:66. doi: 10.3389/fmicb.2014.00066
- Tanne, C., Golovina, E. A., Hoekstra, F. A., Meffert, A., and Galinski, E. A. (2014). Glass-forming property of hydroxyectoine is the cause of its superior function as a desiccation protectant. *Front. Microbiol.* 5:150. doi: 10.3389/fmicb.2014.00150
- Yatsunami, R., Ando, A., Yang, Y., Takaichi, S., Kohno, M., Matsumura, Y., et al. (2014). Identification of carotenoids from the extremely halophilic archaeon *Haloarcula japonica*. *Front. Microbiol.* 5:100. doi: 10.3389/fmicb.2014.00100
- Zerulla, K., and Soppa, J. (2014). Polyploidy in haloarchaea: advantages for growth and survival. *Front. Microbiol.* 5:274. doi: 10.3389/fmicb.2014.00274

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Papke. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Salty sisters: the women of halophiles

Bonnie K. Baxter<sup>1\*</sup>, Nina Gunde-Cimerman<sup>2,3</sup> and Aharon Oren<sup>4</sup>

<sup>1</sup> Great Salt Lake Institute, Westminster College, Salt Lake City, UT, USA

<sup>2</sup> Molecular Genetics and Microbiology, University of Ljubljana, Ljubljana, Slovenia

<sup>3</sup> Centre of Excellence for Integrated Approaches in Chemistry and Biology of Proteins, Ljubljana, Slovenia

<sup>4</sup> Department of Plant and Environmental Sciences, The Institute of Life Sciences, The Edmond J. Safra Campus, The Hebrew University of Jerusalem, Givat Ram, Israel

**Edited by:**

R. Thane Papke, University of Connecticut, USA

**Reviewed by:**

Jocelyne DiRuggiero, The Johns Hopkins University, USA

Meral Birbir, Marmara University, Turkey

**\*Correspondence:**

Bonnie K. Baxter, Great Salt Lake Institute, Westminster College, 1840 S 1300 E, Salt Lake City, UT 84105, USA

e-mail: bbaxter@westminstercollege.edu

A history of halophile research reveals the commitment of scientists to uncovering the secrets of the limits of life, in particular life in high salt concentration and under extreme osmotic pressure. During the last 40 years, halophile scientists have indeed made important contributions to extremophile research, and prior international halophiles congresses have documented both the historical and the current work. During this period of salty discoveries, female scientists, in general, have grown in number worldwide. But those who worked in the field when there were small numbers of women sometimes saw their important contributions overshadowed by their male counterparts. Recent studies suggest that modern female scientists experience gender bias in matters such as conference invitations and even representation among full professors. In the field of halophilic microbiology, what is the impact of gender bias? How has the participation of women changed over time? What do women uniquely contribute to this field? What are factors that impact current female scientists to a greater degree? This essay emphasizes the "her story" (not "history") of halophile discovery.

**Keywords:** halophiles, women in science, diversity, history of science, Nobel Prize

## INTRODUCTION

Women's participation in science was historically limited as the knowledge-creators were traditionally men in most civilizations. Women's societal roles made careers more challenging in general, and science was seen as a demanding career. But many of us working in science today are surprised to learn the level of gender bias that persists even now, and perhaps we have reached a point to challenge the alarming statistics.

Though representation varies across fields, women are awarded about half of science doctorates in the US, but only 21% of full science professors are women (National Science Foundation, National Center for Science and Engineering Statistics, 2013). This issue affects not only job opportunities, but also one's purse as female scientists earn only 82% of that of males in the US, and this figure is even lower in Europe (European Gender Summit, 2012). Other studies, such as one among chemistry graduate students in the UK, have looked at the social factors that disproportionately affect women such as family choices (Royal Society of Chemistry, 2008). Furthermore, disparities in promotion, grant-funding and tenure persist (Shen, 2013).

In Europe, a commission of scientists is working on this issue. A recent congress identified priority areas that may impact the level of participation of women in these fields (European Gender Summit, 2012). These include active recruitment and retention, assessment and value systems. Much focus of this commission is on changing institutional practices and processes to result in more balance in human resources.

Whatever the factors that impact women's participation, it is clear that bias also exists, and there is evidence that this bias, at least in conference participation, may be at the front

end: the invitation. Skewed ratios of men to women lecturers at conferences have been shown in several recent studies. Isbell et al. (2012) published an analysis of speakers in the female-heavy field of primatology, Schroeder et al. (2013) quantified the number of female invitees and speakers at evolutionary biology conferences, and Eisen (2012) reported on the number of talks given by women at quantitative biology meetings. All three studies used powerful statistics to compare the speakers to baseline populations of participants in their field of study. Which factors caused the biases observed? Schroeder et al. (2013) concluded that that fewer women were among the invitees and/or women turned down invitations at a higher rate. A preference for invited male scientist authors was also discovered in a review of "News & Views" articles in *Nature* and "Perspectives" in *Science* for 2010 through 2011 (Conley and Stadmark, 2012), prompting the journals to analyze the underrepresentation in their practices of author invitation.

If there is bias at the level of invitation, then we should look closely at those who do the inviting. The organizing committee of any conference typically is charged with the structure of the meeting and the invited speaker list. An analysis of 460 symposia supported by the American Society for Microbiology demonstrated that at least one female scientist on the organizing committee resulted in a greater number of invited speakers who were women (Casadevall and Handelsman, 2014).

Thus, the issue of the participation of women in many aspects of scientific life has garnered some attention in a number of forums. The authors of this study are halophile scientists and participants in the triennial international halophile congresses. We became curious about the gender balance within this field and

set out on a study to analyze the representation and participation of women in high-salt microbiology. The results were presented at the Halophiles 2013 conference in Storrs, CT, USA, and they are now reported here. Most importantly, we also highlight some of the most prominent female scientists in the field of halophilic microorganisms.

## MATERIALS AND METHODS

We obtained the programs from each conference on halophiles from 1978 to the recent meeting in 2013 (**Table 1**; Oren, 2011). In looking for historically important and noteworthy women halophile scientists, we looked for early participants, multiple conference participants, contributions to the field, and female lectures who were invited to speak. We also collected information about important early discoveries in the field that were attributed to women.

For examination of current researchers, female lecturers, who had a track record in publications in halophile science and were invited to speak at conferences from 2001 forward, were asked to self-report their main achievements in the field of halophilic microorganisms and to select references that they consider most significant. In some cases, we edited the text for length or consistency.

To get an idea of participation of women at these conferences, in the programs of meetings before 2001, we counted the number of female participants as the programs did not distinguish between oral versus poster presentations and the numbers of total participants were low. In 2001, the programs were organized to show oral talks, and for these conferences through the most recent one, we counted the numbers of women who gave oral lectures. Comparing this number of women to the total number of presentations, we computed a percentage of females for each conference. If the gender was not discernable (e.g., initials were used), then we did not count the person in the total.

## RESULTS

### HISTORICAL CONTRIBUTORS

When, in the days of Abraham the Patriarch during the destruction of the cities of Sodom and Gomorrah, Lot's wife turned into a pillar of salt on the shore of the Dead Sea, as described in Genesis 19: 26, she sure did not have much opportunity to marvel at the interesting halophilic microorganisms inhabiting the lake. The interesting properties of organisms such as *Haloarcula marismortui*, recovered from the Dead Sea by a woman scientist (Ginzburg et al., 1970; Oren et al., 1990), were discovered only three millennia later.

Here is a selection of women-scientists who devoted efforts to the elucidation of the properties of halophilic life. It should be noted that these women were pioneers and worked in an age of extreme underrepresentation of women in science. Our selection is based on the historical importance of their contributions to the field in the period up to the end of the 20th century.

### **Clara Hamburger**

Flagellated unicellular algae colored orange-red by their high content of  $\beta$ -carotene were first documented in saltern brines by Dunal in 1838, who named the organism *Haematococcus salinus*. Two detailed descriptions of the organism were published in 1905, one by Clara Hamburger from Heidelberg and one by E. C. Teodoresco in Bucharest. Clara Hamburger's study is by far the most detailed. She provided a summary of the debate, still unresolved at the time, about the cause of the red-purple or red-orange coloration of saltern ponds approaching salt saturation. Unfortunately for her, the formal description of the genus *Dunaliella* and the species *Dunaliella salina* by Teodoresco (1905) preceded the publication of Hamburger's (1905) study by a few months, as she describes herself in her article. Hamburger's observations about the location of the  $\beta$ -carotene in small globules within the single chloroplast were surprisingly exact, and she criticized the incorrect statements of Teodoresco on this subject: "It occurs in the form of small droplets, and is, as seems sure to me, only deposited in the outer alveolar

**Table 1 | A compilation of all international Halophile conferences, locations and an indication of the participation of women.**

Year	Conference title	Location	#Women	#Total
1978	European Molecular Biology Organization (EMBO) Workshop on Halophilism	Rehovot, Israel	*8	*56
1985	**The Molecular Basis of Haloadaptation in Microorganisms	Obermarchtal, Germany	*2	*21
1986	Aspects of Halophilism	Jerusalem, Israel	*8	*39
1989 (March)	Modern Aspects of Halophilism	Neve Ilan and Rehovot, Israel	*22	*57
1989 (September)	**General and Applied Aspects of Halophilic Microorganisms	Alicante, Spain	*23	*65
1992	American Society for Microbiology (ASM): Halophilic Bacteria	Williamsburg, VA, USA	*6	*34
1997	Microbiology and Biogeochemistry of Hypersaline Environments	Jerusalem, Israel	*13	*69
2001	International Conference on Halophilic Microorganisms	Seville, Spain	8	61
2004	**Halophiles 2004	Ljubljana, Slovenia	17	57
2007	**Halophiles 2007: International Congress Exploring Life at High Salinity	Colchester, UK	16	49
2010	Halophiles 2010: International Conference on Halophilic Microorganisms	Beijing, China	15	48
2013	**Halophiles 2013	Storrs, CT, USA	27	59

\*The value is numbers of female participants and total participants. From 2001 forward, all values are numbers of female oral presenters and total lectures.

\*\*The conferences had at least one female convener on the organizing committee.

layer of the plasma, while the chromatophore [= chloroplast] is the bearer of the green pigment. The remark by Teodoresco ‘the blood pigment that impregnates not only the chromatophore, but also the whole body of adult individuals’ does not correspond with my observations.” (translation: Aharon Oren).

#### **Helena Petter and Trijntje Hof**

Among the group of microbiologists associated with the “Delft School” in the 1920s–1930s were two women who wrote their Ph.D. thesis on halophilic microorganisms and contributed much to our understanding of the halophiles: Helena Petter and Trijntje Hof. In her thesis (University of Utrecht) “Over roode en andere bacterieën van gezouten visch” (On red and other bacteria of salted fish; Petter, 1931, 1932), Helena Petter studied a variety of different halophilic prokaryotes, most of them red pigmented members of the *Halobacteriaceae*, isolated from salted fish and from “Trapani” salt used to preserve fish, which she obtained from a cannery in Bergen, Norway. Her isolates included rod-shaped bacteria as well as coccoid and sarcina-shaped form. Her studies include a description of “Bacterium halobium” (currently known as *Halobacterium salinarum*). The name bacterioruberin for the carotenoid pink pigment of the *Halobacteriaceae* was first introduced by Petter. She made drawings of the gas vacuoles within the cells, and she was the first to explain the possible ecological advantages of these gas vacuoles and of the buoyancy these gas vesicles bestow upon the cells: in hypersaline lakes in which only little oxygen can dissolve, there may be a selective advantage for oxygen-dependent microorganisms to float toward the surface of the brine when oxygen becomes limiting in the deeper layers.

Trijntje Hof’s thesis (University of Leiden) on “Investigations concerning bacterial life in strong brines” (Hof, 1935) included a description of a novel type of motile halophilic rod that caused a purple discoloration of salted beans. She named this organism *Pseudomonas beijerinckii* in honor of Martinus Beijerinck, the founding father of the Delft school of microbiology. *P. beijerinckii* was renamed as *Chromohalobacter beijerinckii* in 2006.

#### **Arlette Danon**

Arlette Danon of the Weizmann Institute of Science, Rehovot, Israel, started work on the function of the purple pigment bacteriorhodopsin (BR) of *Halobacterium* very soon after the chemical structure of the pigment was discovered in the early 1970s. The discovery that light energy absorbed by BR in the purple membrane of *Halobacterium* is conserved by photophosphorylation with the formation of ATP (Danon and Stoeckenius, 1974; Danon and Caplan, 1976) is the key to the understanding of the function of the pigment. She also discovered that in the light *Halobacterium halobium* (*salinarum*) cells can fix a certain amount of CO<sub>2</sub>, mediated by BR (Danon and Caplan, 1977). The mechanism of this light-dependent CO<sub>2</sub> fixation, later studied by another woman halophile scientist – Barbara Javor – is still not completely clear.

Danon passed away at an early age, ending a short but very promising career in halophile science.

#### **Margot Kogut**

“In the early 1950s . . . I can still see in my mind’s eye the lady who worked at the bench adjacent to mine, Margot Kogut, standing there in white lab coat and scarf huddled over a cup of hot tea.”

Thus described John Ormerod (2003) his memories of Margot Kogut of University of London Kings College. Margot studied the salt relationships of moderately halophilic Bacteria, using *Salinivibrio costicola* as the model organism. Her studies of in vitro protein synthesis by its ribosomes, in collaboration with Donn Kushner’s group in Canada (Wydro et al., 1977) led to a proper assessment of the true intracellular environment of a moderate halophile that can adapt to a broad range of salt concentrations. Her studies of the lipids in the cell membrane and the changes in lipid composition as a function of the salinity of the medium (Russell and Kogut, 1985; Russell et al., 1986) were used to confirm and to illustrate her ideas about the difference between “adaptation” to extreme environments and “adaptability” toward changes in the parameters determining the extreme habitat (Kogut, 1980a,b).

Together with Bill Grant (Leicester), Kogut organized the 1985 symposium on ‘The molecular basis of haloadaptation in microorganisms’ in Obermarchtal, Germany (Grant and Kogut, 1986). This meeting can be considered as the first congress on all aspects of halophilic microorganisms, the Rehovot 1978 symposium being almost entirely dedicated to the biochemistry of BR.

#### **Barbara Javor**

Barbara Javor received her Bachelor of Arts degree from the University of California Santa Barbara, and her Ph.D. in Biology at the University of Oregon. She spent a post-doctoral period to study the salterns of Eilat at the Red Sea coast of Israel, and has worked on halophilic microorganisms at Scripps Institution of Oceanography, University of California. Later she joined different biotechnological companies in the San Diego, CA, area.

Among halophile scientists she is renowned for her textbook “Hypersaline environments.” Microbiology and biogeochemistry (Javor, 1989). This famous and important monograph explores all aspects of natural hypersaline lakes, salterns, saline soils and other salt-stressed environments and the microorganisms inhabiting them. She has explored the prokaryote diversity of salterns in different locations worldwide (Javor, 1983, 1984), and isolated interesting “box-shaped” halophilic Archaea, which turned out to belong to the genus *Haloarcula* (Javor et al., 1982). She also continued the study of the light-dependent CO<sub>2</sub> fixation (Javor, 1988), a topic initiated by Danon a decade earlier (see above).

Some of her experiences as a consultant for salt production companies and on the links between the biology of the saltern ponds and the quantity and quality of salt produced are summarized in her paper on industrial microbiology of solar salt production (Javor, 2002).

#### **Margaret Ginzburg**

Margaret Ginzburg (the Hebrew University of Jerusalem) has published a large number of papers, many of which were prepared jointly with her husband Ben-Zion Ginzburg, both on halophilic Archaea and on *Dunaliella*.

After the original isolate of *Halobacterium marismortui* retrieved from the Dead Sea by Benjamin Elazari-Volcani had been lost, Ginzburg et al. (1970) isolated and characterized a very similar strain from the lake, and described it as *Haloarcula marismortui*

(Oren et al., 1990). This isolate served as the object for many studies on the ion metabolism, the permeability properties of the membranes, and the bioenergetics of halophilic Archaea, many of which also included comparisons with *Halobacterium salinarum*, found to behave differently in certain aspects (Ginzburg, 1969, 1978; Ginzburg et al., 1970). Ion metabolism and the metabolism of the osmotic solute glycerol were studied in different *Dunaliella* isolates, including strains obtained from the Dead Sea (Ginzburg and Ginzburg, 1985), and her studies about *Dunaliella* led her to write a long review about the adaptation of *Dunaliella* to life at high salt concentrations (Ginzburg, 1987). Caplan and Ginzburg (1978) edited the proceedings volume of the above-mentioned symposium on “Energetics and structure of halophilic microorganisms” held in Rehovot and dedicated mainly to the properties of BR.

#### Rita Colwell

Rita Colwell, the first female Director of the U.S. National Science Foundation, is well known for her efforts to increase the participation of underrepresented groups in science. During her career as a microbiologist, she studied *Vibrio cholerae* and other slightly halophilic *Vibrio* species, focusing on the role of the environment and climate change in disease dynamics. But some of Colwell's studies were devoted to prokaryotes living at higher salt concentrations. Her 1979 taxonomic study on red halophilic Archaea, performed jointly with Carol Litchfield (see below) and other coworkers, provided a polyphasic comparative study of the then-known, limited diversity of the *Halobacteriaceae* (Colwell et al., 1979). Rita also is a member of the International Committee on Systematics of Prokaryotes (ICSP) Subcommittee on the Taxonomy of *Halobacteriaceae*. More recently her group has described a new species of *Halobacillus*, *Halobacillus thailandensis*, from fish sauce (Chaiyanan et al., 1999).

#### Carol Litchfield

Carol Litchfield had a great fascination for salt lakes and the organisms inhabiting them. “Red – the magic color for solar salt production” (Litchfield, 1991) was the starting point for many of her studies on halophiles in solar salterns world-wide, in Great Salt Lake, Utah, and in other hypersaline aquatic environments. With her student, Russell Vreeland, she discovered and described *Halomonas elongata* from a saltern pond on Bonaire (Vreeland et al., 1980), an organism that since has become the model organism for the study of moderate halophiles, and that, thanks to the production of ectoine, has also a great biotechnological interest. Litchfield extended her interest in halophiles on Earth to the possibility that such organisms may inhabit or may have inhabited other planets in the past (Litchfield, 1998). She applied techniques such as polar lipid analysis and pigment analysis to the comparative characterization of the biota of saltern ponds and salt lakes (Litchfield et al., 2000), and she also pioneered combining cultivation of the halophiles along with molecular techniques (Litchfield and Gillivet, 2002). She was an active member of the ICSP Subcommittee on the Taxonomy of *Halobacteriaceae*. Interestingly, Litchfield also was an amateur salt industry historian, and she co-edited a book of essays on the history of salt and salt-making (Litchfield et al., 2001). She was an avid collector of old

books and documents on salt as well as artifacts connected with salt production. Her extensive collection is now at the Hagley American Industry Museum in Delaware.

#### Ada Yonath

In 2009, Ada Yonath was the first woman in 45 years to win the Nobel Prize in Chemistry and the first Israeli woman to win a Nobel Prize. A member of the faculty of the Weizmann Institute of Science in Rehovot, Israel, she is a crystallographer best known for her pioneering work on the structure and function of ribosomes. In many of her studies that gained her the Nobel Prize, the Dead Sea archaeon *Haloarcula marismortui* served as the model organism (Makowski et al., 1987; von Böhlen et al., 1991; Schlunzen et al., 1999; Yonath, 2002a,b). In a short essay recently published in *Nature* (Yonath, 2011), Yonath expressed her feelings and ideas about being a female scientist and about the long way to success. She also discussed there the special position of women in science: “*I think science is gender-independent*,” and “*I'm trying to change the image of scientists, especially of female scientists, so there will not be so many anti-female sentiments*.” We all hope that Yonath's example will be followed by many other women who will discover the fascinating world of the halophilic microorganism and the opportunities these organisms offer as model systems to answer basic questions in biology.

#### Ada Zamir

Ada Zamir, who joined the faculty of the Weizmann Institute of Science in Rehovot, Israel, in 1964, used *Dunaliella salina* as the model organism for her studies on defense mechanisms of algae against high light intensities and high salinity. She identified a number of membrane proteins induced by *Dunaliella* as a reaction to salt stress: a special carbonic anhydrase and a novel transferrin-like protein (Sadka et al., 1991; Fischer et al., 1996, 1997; Bageshwar et al., 2004). Zamir also studied the molecular mechanisms responsible for the photoinduction of massive β-carotene accumulation by *D. salina* (Lers et al., 1989).

#### EXAMPLES OF ACTIVE AND NOTEWORTHY CONTRIBUTORS

To represent the breadth of current female researchers in halophilic microbiology, we relied on the programs from conferences listed in Table 1. We recognize the important work in contributed oral presentations and posters by all women since 1978, but for this analysis we focused only on those women who were listed in the programs as “invited speakers.” In addition, these women were filtered for those who had a long-standing commitment to the halophile science demonstrated by years of involvement and publication record. We asked these scientists, who were still working in the field, to self-report their achievements. Those who responded are listed below, in alphabetical order, as examples. Included are their research interests and references for their work. This method resulted in what is certainly not an exhaustive list, but it should underscore the participation of women in this area of science.

#### Josefa Antón

One of Antón's main achievements is the development of fluorescence in situ hybridization (FISH) protocols for analyzing microbial communities inhabiting hypersaline environments.

This technique was instrumental in the discovery of the abundance of extremely halophilic bacteria in close-to-saturation environments, and, more specifically, in the discovery of *Salinibacter ruber*, that turned out to be an ecologically relevant extremely halophile. This discovery was followed by a wealth of studies on the diversity and biogeography of this species, as well as genomic studies. Antón is still working on the microdiversity of this and other extremely halophilic Bacteroidetes. Another focus of her research is the study of yet uncultured halophilic viruses by cloning their complete genomes into fosmids. This work initiated the work on environmental haloviruses, which is continuously growing incorporating new approaches and technologies (Representative references: Antón et al., 1999, 2000; Santos et al., 2007).

#### **Bonnie K. Baxter**

The microbial foundation of Great Salt Lake, an iconic hyper saline environment, lay virtually unexplored until Baxter began working at the lake in 1998. Baxter reached out to halophile scientists and created collaborations to build an understanding of lake microbial communities. Her projects range from microbial diversity (including bacteria, archaea, viruses, algae, and fungi) to the biogeochemistry of stromatolites to ancient biological molecules in halite. What we now know is that this enormous lake has many micro-niches, all of them with their own microbial communities. Her work has shown that this lake is stratified vertically, horizontally, and temporally. Weaving public outreach into her research, Dr. Baxter turned this unique approach to science into “Great Salt Lake Institute,” which facilitates research and education on this unique body of water (Representative references: Baxter et al., 2005; Griffith et al., 2008; Meuser et al., 2013).

#### **Kathleen C. Benison**

Benison has made contributions to the understanding of extremophile life and fossilization in acid saline lake and ground-water systems, with a focus on ephemeral lake settings in Western Australia, as well as their ancient counterparts in the Permian redbeds and evaporites of the North American midcontinent. She applied the perspective of a geologist and geochemist, linking water chemistry, chemical sediments, and microorganisms and their temporal dynamics in relation to flooding, evaporation, and desiccation stages of these extreme lakes and shallow groundwaters. Benison showed that prokaryotes, algae, suspect fungi, and certain organic compounds can be easily trapped and well-preserved in halite and gypsum, both as solid or fluid inclusions. Together with Melanie Mormile, she has described, for the first time, diverse microbiological communities that thrive in acid brines, with pH as low as 1.5 and salinities up to 32% total dissolved solids. Most of the very diverse organisms are novel. This work has also implications for the understanding of life on early Earth and on other planets, in particular on Mars (Representative references: Benison et al., 2008; Mormile et al., 2009; Conner and Benison, 2013).

#### **Angela Corcelli**

Corcelli's studies on halophiles at the University of Bari, Italy, center on two topics: the membrane lipids of extreme halophiles and the properties of BR and the lipids interacting with the BR

proton pump in the purple membrane of *Halobacterium*. She discovered a variety of cardiolipins in halophilic Archaea and investigated their function. She also elucidated the structure of the unique sulfonolipids of *Salinibacter ruber* and related members of the *Bacteroidetes*. Her “lipidomics” studies extend from pure culture studies of model halophilic organisms to the characterization of the lipids present in complex communities of halophiles in their natural environment (Representative references: Corcelli et al., 2004; Lopalco et al., 2011; Lobasso et al., 2012).

#### **Jocelyne DiRuggiero**

Dr. DiRuggiero's scientific interests are in the adaptations of extremophiles to environmental stresses and in the microbial ecology of extreme environments, in particular environments with extremes in temperature, high salt concentrations, and hyper-arid conditions. Her major contributions to the field of halophiles have been in the elucidation of stress responses of model halophilic archaea to radiation and oxidative stress. Using a combination of functional genomic and genetics she discovered a shift in the archaea away from the eukaryotic model of homologous recombination repair of DNA double-strand breaks. Investigating oxidative stress, and the deleterious effect of ionization radiation, she discovered that non-enzymatic antioxidant processes are essential for the high level of radiation resistance found in *Halobacterium*. More recently, her work in environmental microbiology revealed the ecology of one of the most halophilic environments on Earth. DiRuggiero found that halite pinnacles (NaCl rocks) from the Atacama Desert are inhabited by a photosynthetic-based, archaea-dominated community; she discovered that the diverse prokaryotic assemblages are associated with novel algae related to oceanic picoplankton (Representative references: Kish and DiRuggiero, 2008; Robinson et al., 2011, 2014).

#### **Christine Ebel**

Ebel's studies focused on the stability and composition of halophilic enzymes. Her interdisciplinary approach is unique as she combines molecular biology, cellular biochemistry, structural studies, biophysical chemistry, and thermodynamics. Enzymes from halophilic archaea are also halophilic as they have a requirement for high salt. Ebel explores the effects of salts on halophilic enzymes including activity, solubility and stability. In addition, she investigates enzyme kinetics under these conditions. She has shown that water and ion binding to halophilic proteins is significant in their function. Also, Ebel was involved in elucidating crystal structures of several halophilic enzymes (Representative references: (Ebel et al., 1999, 2002; Madern et al., 2000)).

#### **Sabrina Fröls**

Complex microbial communities, i.e., biofilms formed by archaea and bacteria are recognized to be the predominant microbial mode of life in nature and found in a remarkable spectrum of habitats. Halophilic biofilm forming archaea were identified from sediments around an underwater fresh spring in the Dead Sea. Selected haloarchaeal strains of five different genera were tested in regard to surface adhesion. Fröls et al. (2012) showed, also by microscopic analyses, that this ability is widely distributed in halophilic archaea. The observed biofilms varied in architecture. Biofilm composition analyses revealed extracellular polymeric substances (extracellular

DNA and glycoconjugates). By transmission electron microscopy studies of attached *Halobacterium salinarum* they observed multiple pili structures which might be involved in the haloarchaeal biofilm formation (Representative references: Fröls et al., 2012; Ionescu et al., 2012; Fröls, 2013).

### Nina Gunde-Cimerman

Gunde-Cimerman discovered halophilic and halotolerant fungi, particularly melanized black yeasts and related fungi as inhabitants of solar salterns and other hypersaline environments around the world. Besides describing their biodiversity, including new species, she focused on molecular mechanisms of adaptation to hypersalinity. Along with Ana Plemenitaš she studied physiological and molecular mechanisms of three model organisms: halotolerant *Aureobasidium pullulans*, extremely halotolerant *Hortaea werneckii* and the obligate halophile *Wallemia ichthyophaga*, at the level of cell wall, membranes, compatible solutes and expression of selected genes and lately also by whole genome sequencing (Representative references: Gunde-Cimerman et al., 2000; Plemenitaš et al., 2008; Zajc et al., 2013).

### Inga Hänelt

The first response to an osmotic upshift and the immediate loss of water usually is the fast accumulation of K<sup>+</sup> via channels, pumps and transporters. Hänelt's research focuses on the functional role and molecular architecture of K<sup>+</sup> translocating systems, in particular the bacterial Ktr system belonging to the superfamily of potassium transporters (Ktr/Trk/HKT family). Hänelt and others identified a unique flexible linker (loop) within the translocating subunit of Ktr/Trk systems that is crucial for the controlled uptake of potassium. The loop was shown to function as molecular gate that opens for the passage of K<sup>+</sup> and is missing in classical K<sup>+</sup> channels (Representative references: Hänelt et al., 2010a,b, 2011).

### Julie Maupin-Furlow

Prior to completion of any genome sequences of halophilic archaea, Maupin-Furlow and coworkers performed the groundbreaking experiments that demonstrated 20S proteasomes that catalyze proteolysis are expressed in the haloarchaeon *Haloferax volcanii*. They were purified, biochemically characterized and sequenced. Maupin-Furlow and co-workers showed that 20S proteasomes were not restricted to species of *Thermoplasma* and that they could be disassembled and reassembled *in vitro* based on salt concentration. This was the first observation of this kind for archaea. Maupin-Furlow also contributed to the characterization of the ubiquitin-fold proteins of the model archaeon *Haloferax volcanii*. Distribution of proteins of the ubiquitin-fold superfamily suggests sumpylation is universal to Archaea. This study provided a fundamental insight into the diverse cellular functions of the ubiquitin-fold superfamily and the capacity for an archaeal ubiquitin-activating enzyme E1 homolog to have broad substrate specificity (Representative references: Wilson et al., 1999; Humbard et al., 2010; Miranda et al., 2011).

### Noha Mesbah

Mesbah's research is focused on poly extremophiles: halophiles, alkaliphiles, and thermophiles – on their isolation, identification,

and characterization of adaptive strategies that allow them to survive and grow in the presence of multiple environmental extremes. She described a novel order, *Natranaeriales*, with three obligately anaerobic, halophilic alkalithermophilic bacteria: *Natranaerobius thermophilus*, *N. Trueperi*, and *Natronovirga wadinatruncensis*. *N. thermophilus* has been used as the model microorganisms. It acidifies its cytoplasm, maintaining a constant transmembrane pH gradient of approximately 1 unit, acid inside. The genome of *N. thermophilus* was sequenced. Analysis showed an unusually large number of cation-proton antiporters, which aid to withstand multiple environmental extremes (Representative references: (Mesbah et al., 2007; Mesbah and Wieg, 2011; Zhao et al., 2011)).

### Melanie Mormile

Mormile's research on saline systems has ranged from studying the microbial ecology in hypersaline lakes to the development of methods for the retrieval of viable bacteria from ancient salt crystals. She isolated *Halomonas campisalis* from the region around Soap Lake in Washington State, with the aim to treat saline, alkaline waste. Further isolates from Soap Lake included a new genus, *Nitrincola lacisaponensis*, and a bacterium, *Halanaerobium hydrogeniformans*, capable of significant hydrogen production from alkali-treated cellulose biomass. Along with K. Benison and F. Oboh-Ikuenobe she studied the microbial communities in the acidic saline lakes of western Australia. She also retrieved a viable *Halobacterium salinarum* cell from a 97 kyr halite crystal fluid inclusion. This work has led others to further refine the techniques used to retrieve microorganisms from evaporative crystals much older than 97 kyr (Representative references: Mormile et al., 1999, 2003, 2009).

### Francisca Oboh-Ikuenobe

Oboh-Ikuenobe studied sediments and waters of different acid, neutral, and alkaline hypersaline ephemeral lakes in Western Australia, utilizing sedimentological, geochemical, mineralogical, microbiological and palynological techniques. The aim was to identify microorganisms trapped in evaporites and/or sediments, to gather paleoenvironmental and paleoclimatic information and make analogies with Mars. Her group identified novel Bacteria/Archaea and algae and found microfossils in halite, gypsum, and hematite precipitates. Oboh-Ikuenobe's focus was mainly on palynology. She identified *Dunaliella* as a proxy of arid climatic conditions and high concentration of salt in the region's geologic record (Representative references: Benison et al., 2007; Bowen et al., 2008; Sanchez Botero et al., 2013).

### Felicitas Pfeifer

Pfeifer worked on *Halobacterium salinarum* (*halobium*) already in the late 1970s and characterized the genetic variability of the large plasmids that incurred insertions and deletions due to insertion elements. She isolated many of these ISH elements from purple membrane (*bop*) mutants, recognized a transposition burst of the ISH27 insertion element family, and presented data on an AT-rich island in the genome of *Halobacterium salinarum* PHH1 harboring many insertion elements. Pfeifer was involved in the construction of vector plasmids to manipulate haloarchaea, and a vector carrying the origin of replication of plasmid pHH1 was used to demonstrate that a fragment as large as 11 kbp is required for gas vesicle formation. These proteinaceous structures enable the

cells to increase buoyancy and float to the surface of the brine. Pfeifer and coworkers demonstrated that a total of 14 *gvp* genes arranged in two clusters are required for gas vesicle formation in *Halobacterium salinarum* and *Haloferax mediterranei*. Gas vesicle formation is influenced by environmental factors such as salt concentration, oxygen availability, and light and thus serves as a model system to study signal transduction pathways in haloarchaea (Representative references: Pfeifer and Betlach, 1985; Pfeifer and Blaseio, 1990; Pfeifer, 2012).

### Ana Plemenitaš

Plemenitaš' lab together with Gunde-Cimerman lab focused on molecular mechanisms of adaptation to hypersalinity, particularly in the extremotolerant model organism black yeast *Hortaea werneckii*. The adaptations were studied on the level of membrane composition and fluidity, ion pumps, differential gene expression at high and low salinity and selection of target genes (for example HAL) for transformation of yeast and plants to increase their halotolerance. The main focus of her lab is the study of high osmolarity glycerol signaling pathway (HOG) – identification of the components, their interconnectivity and differences in selected halotolerant and halophilic fungal model organisms (Representative references: Lenassi et al., 2007, 2013; Vaupotiè and Plemenitaš, 2007).

### Mecky Pohlschroder

A significant portion of the prokaryotic proteome is composed of secreted proteins, which play crucial roles in a variety of vital cellular processes, including nutrient uptake, cell wall biosynthesis, conjugation and motility. Pohlschroder focuses on characterizing the protein translocation pathways that transport these proteins to the cell surface. Computational searches performed by Pohlschroder suggested that several aspects of the universally conserved Sec translocation pathway are unique to archaeal species. Although most prokaryotic proteins are secreted in an unfolded conformation via the Sec pathway, nearly half of haloarchaeal secreted proteins are transported in a folded conformation via the twin arginine transport (Tat) pathway, as an adaptation to the high salt environments. Pohlschroder's lab has developed several computational programs that can accurately predict the transport pathways used by specific substrates as well as the subcellular localization of secreted archaeal proteins ([singalfind.org](http://singalfind.org)). Most recently, her lab identified a previously unknown cell surface anchoring mechanism. These investigations will also elucidate molecular mechanisms that support crucial biological processes in prokaryotes, as illustrated by the variety of phenotypic defects identified by the Pohlschroder lab in an *Haloferax volcanii* archaeosortase deletion mutant, which include deficiencies in cell wall biosynthesis, as well as surface adhesion, mating and motility (Representative references: Pohlschroder et al., 1997; Rose et al., 2002; Abdul Halim et al., 2013).

### Emilia Quesada

Quesada was one of the pioneers, together with Antonio Ventosa, in the study of halophilic bacteria inhabiting hypersaline waters. They developed different techniques for their cultivation and identification. She also made the first studies of saline soils and described its microbiota (bacteria and halophilic archaea).

She discovered many species of halophilic bacteria and became a member of the International subcommittee on the *Halomonadaceae* family. Her lab described for the first time systems of quorum sensing in halophilic bacteria. They also found and developed a number of exopolysaccharides produced by halophilic bacteria of biotechnological interest (Representative references: Mata et al., 2008; Amjres et al., 2011; Luque et al., 2012).

### Elina Roine

The general objective of Roine's research is to characterize viruses from scarcely studied niches, to increase information of the functions of genes residing in prokaryotic viral genomes and virus related functions in their respective hosts. Her main impact in haloarchaeal research has been detailed characterization of new viruses, such as the description of the first ssDNA archaeal virus and a new family of pleomorphic viruses. She showed that the genomes of haloarchaeal pleomorphic viruses differ and that the structure of their major N-glycan modifying spike protein is involved in the host interaction. On the basis of these studies Roine and co-workers suggested a wider evolutionary connections between the pleomorphic viruses and a group of bacteriophages. She also studied the icosahedral virus SH1 and, together with Angela Corcelli, the lipids of haloarchaeal viruses (Representative references: Pietilä et al., 2009; Kandiba et al., 2012; Seneilo et al., 2012).

### Shereen Sabet

Sabet was the first to describe the viral assemblage in the hyper-saline Mono Lake with a metagenomic approach. Although her goal was to clone an entire viral genome from an environmental sample, she was nevertheless able to clone several fragments. She was the first to isolate viruses from Exportadora de Sal in Baja California, Mexico. The genome of one of those viruses, GNF2, has been sequenced. She also undertook a comparative metabolic analysis on the archaeal and bacterial isolates from ESSA, similar to cultures that had been isolated from the same site 24 years prior by Barbara Javor. She showed adaptive micro-environmental signatures of substrate usage within the halophile population. Shereen's most recent contribution is isolation of viral plaques from the Great Salt Lake and the Cargill solar salterns in San Francisco Bay (Representative references: Sabet et al., 2006, 2009; Sabet, 2012).

### Helena Santos

Santos's work is focused on the understanding how marine hyperthermophilic microorganisms adapt to life at temperatures around 100°C, which are lethal to so many other cells. Microorganisms isolated from marine habitats are slightly halophilic and, like many halophiles, accumulate compatible solutes for osmoregulation. Santos and others found that hyperthermophiles accumulate exquisite organic solutes, such as mannosylglycerate and di-myoinositol-phosphate, which are used for osmoprotection and also for thermoprotection. They identified several novel solutes, characterized their biosynthetic pathways, determined the 3D-structures of biosynthetic enzymes, studied their physiological role using suitable mutants, probed the molecular basis for protein stabilization, and searched for novel applications (five international

patents). Amongst Santos' recent achievements is the establishment of the role of ionic solutes in thermoprotection using mutants and evolution of the biosynthesis of di-*myo*-inositol phosphate. Her group also established the role of mannosylglycerate in the stabilization of proteins, via restriction of the slow motions of specific structural elements (Representative references: Rodrigues et al., 2007; Faria et al., 2008; Pais et al., 2012).

### Amy Schmid

Schmid's overarching research goal involves understanding the molecular mechanisms underlying the transcriptional response of halophilic archaea to environmental changes and how this results in physiological adaptation. Central to this process are gene regulatory networks (GRNs) composed of groups of interacting transcription factors (TFs) and their target gene promoters. Upon sensing a change in the environment, signal transduction cascades propagate the information to GRNs, where TFs proteins induce genes encoding proteins that restore the cell to a stable state and prepare for future stress. In halophilic archaea, and archaea generally, the molecular function of TFs is unclear relative to the other domains of life. Schmid uses *Halobacterium salinarum* as a model system for understanding how GRNs function dynamically to survive during strong daily fluctuations in conditions within arid hypersaline environments, which can result in extreme oxidative damage to macromolecules. In response, TFs induce genes encoding enzymes to neutralize oxidants, repair damaged molecules, and restore redox balance in the cell. In more recent work, the Schmid group has focused on validating subsets of the GRN composed of 2–3 interacting TFs using integrated genome-scale TF-DNA binding and gene expression analyses (Representative references: Schmid et al., 2011; Sharma et al., 2012; Todor et al., 2013).

### Helga Stan-Lotter

Stan-Lotter initially focused on the ATPase enzymes of halophiles for energy transduction. Archaeal enzymes were found to have a similar structure as those from non-Archaea, but there were enough differences to put them into a separate class. In the Austrian alps, she isolated viable extremely halophilic Archaea from rock salt deposits of Triassic and Permian age. Taxonomic characterization showed that some properties of the isolates were similar to those of known haloarchaea; however, numerous differences suggested that the strains were novel species. These microorganisms may have survived enclosed in the fluid inclusions of rock salt since the evaporation of ancient brines. Repeated re-isolation of the same strains from the same sites suggested authenticity of the isolates. In preparation for the search for extraterrestrial life in halite on planets like Mars, Stan-Lotter provided a halophile strain isolated from Permian salt, *Halococcus dombrowskii*, for exposure of extremophiles on the outside of the International Space Station. This strain was chosen due to its resistance to desiccation and irradiation (Representative references: Stan-Lotter et al., 1999; McGenity et al., 2000; Fendrihan et al., 2009).

### Nicole Wagner

Wagner initially studied optimization of retinal containing proteins for application in devices and played an integral role in the proof-of-concept studies that helped to found in 2009 LambdaVision Incorporated. Via site-directed mutagenesis, site-specific

saturation mutagenesis, and directed evolution she was able to genetically engineer BR, for application in a number of device architectures. Most recently, a protein-based retinal implant, which is targeted at restoring vision to patients with age-related macular degeneration and retinitis pigmentosa. BR, located in the outer membrane of *Halobacterium salinarum*, functions as a light-transducing proton pump. The native protein has a photocycle, which is made up of a series of transient photochemical and conformational intermediates. It is this unique photocycle that makes BR one of the most studied proteins for use in bioelectronics and biomimetic devices (Representative references: Greco et al., 2012; Wagner et al., 2013a,b).

### Tatjana Zhilina

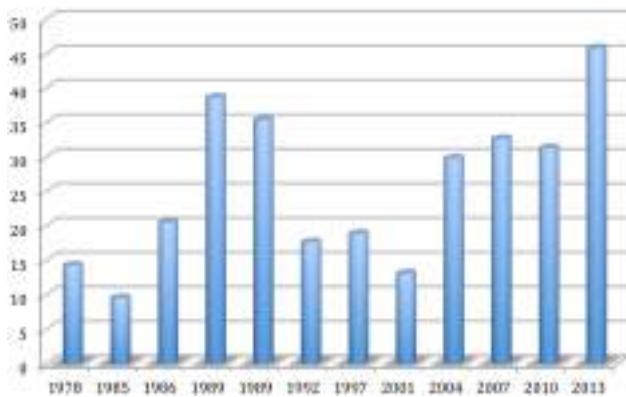
Zhilina, with coworkers was involved in the isolation, identification and description of new species of halophilic bacteria from sediments of soda lake Magadi (Kenya), amongst them a novel genus and species, *Natrananerobaculum magadiense* gen. nov., sp. nov. and from sediments of the soda-depositing soda lake Tanatar III (Altay, Russia) of the order *Halanaerobiales*, which represented a new branch within the family Halobacteroidaceae. They described a novel species in a new genus with the name *Fuchsella alkaliacetigena* gen. nov., sp. nov. They also obtained new isolates from denitrifying enrichments with various electron donors using sediment samples from hypersaline soda lakes. They were identified as members of the *Gammaproteobacteria* closely associated with the *Alkalispirillum-Alkalilimnicola* group and demonstrated much higher metabolic diversity of haloalkaliphilic *Gammaproteobacteria* than was originally anticipated (Representative references: Sorokin et al., 2006; Zhilina et al., 2012; Zavarzina et al., 2013).

## ANALYSIS OF HALOPHILE CONFERENCE PARTICIPATION 1978–2013

International symposia for halophile microbiologists have been held with some regularity since 1978 (Oren, 2011). Conference programs for 12 international Halophiles meetings (Table 1) were analyzed. For all meetings, we noted female presenters as well as female membership of the organizing committee. We examined the rosters of *all participants* for conferences held 1978 through 1997 since the numbers of attendees was rather low and the programs were arranged in a way that did not distinguish the type or presentation. However, for the meetings during the years 2001–2013, we only looked at the oral presentations. This count revealed a steadily growing number of women over time, with the exception of the two conferences in 1989, which seemed heavy in female participation when compared to the years before and after this date (Table 1 and Figure 1). The high numbers of female speakers for 2013, 46%, was an unanticipated result and deserves some analysis as this is out of sync with reports from other fields (Eisen, 2012; Isbell et al., 2012; Schroeder et al., 2013).

The halophile conferences that had at least one woman as a member of the organizing committee (1985, September, 1989, 2004, 2007, and 2013) show higher numbers of female presenters with the exception of the earlier date, 1985, where the women gave only 9% of the presentations. If we leave out the data from 1985, these other conferences show an average of 36% women giving talks. The number drops to 30% if the data set for 1985

### Percent Women Participants/Presenters at Halophile Meetings



**FIGURE 1 |** Participation of women as speakers in international halophile conferences from 1978 until 2013.

is included. In contrast, the conferences that did not have female convener membership had an average of 20% female presenters.

## DISCUSSION

The underrepresentation of women in science has been well documented in the literature. Several studies show societal factors that are at play (Shen, 2013), and gender biases in various fields have been identified (Conley and Stadmark, 2012; Eisen, 2012; Schroeder et al., 2013). Indeed, women have gained ground in science, but there are many gaps still to close (Shen, 2013).

Halophile science is rather interdisciplinary as participants may work in the lab and/or field and be a part of microbial ecology, biochemistry, geology, or even planetary science communities. Therefore, this population of scientists should represent a broad sampling with which to ask the question of gender bias. Given the male-dominated sub-fields, we predicted that women would be underrepresented in halophile science.

## HALOPHILE SCIENTISTS: AN EXAMPLE OF PROGRESS

Historically, many women scientists have been engaged in halophilic microbiology. A few notable examples, including Yonath's Nobel Prize and Colwell as the President of the U.S. National Science Foundation, indicate that there were indeed amazing female contributors as this field was developing. In recent years, the numbers of women participating in the international congresses has generally increased.

Among the successes we encountered, the authors of this study also uncovered stories of discrimination as we polled women currently working in halophile research. Several very accomplished female scientists noted that their discoveries had been overshadowed by those of men in the field. Certainly no field of science is without bias and discrimination as it is an endeavor of humans. Halophile science is no exception, but by and large, here we report some very positive outcomes.

The most significant moment in this study was the end of our presentation for opening event at the 2013 Halophiles Meeting in

CT, USA, where we announced that 46% of the speakers for the conference would be women. This was unexpected for us as we calculated the results, and it was unanticipated for our audience. It seems the population of salty scientists are doing better than other fields if representation of women in lectures is any measure (Conley and Stadmark, 2012; Eisen, 2012; Schroeder et al., 2013). Why would this particular field of study lead to less gender bias than other fields have observed?

Discussions among colleagues revealed an important point: halophile microbiologists who attend the international congresses understand that they are in a field where mentorship is valued. This has historically been a small group of scientists, committed to holding conferences despite no professional organization and no secure funding. Strong friendships were formed, and many of the participants began collaborations that spanned years, working side by side in the laboratory or the field (Figure 2). Also, in more than a decade, organizing committees have made the invitation of young scientists and in particular, women, a mission of the meetings. This philosophy of inclusion and even actively providing travel funds for underrepresented speakers has ensured an almost equal gender representation.

Perhaps placement of women on the organizing committees for the international halophile congresses has allowed for more inclusion. When we analyzed the conference data we saw a 10–16% increase in female speakers when at least one woman was included among the conveners. This echoes the study by Casadevall and Handelsman (2014) that showed as much when looking at a much broader sample of American Society of Microbiology symposia. Indeed, the gender balance and role of the organizing committee for any conference cannot be overlooked and may provide a significant impact on representation of women. Halophile scientists have employed this strategy often, and we suggest that this becomes a continued practice. We are a community that encourages the participation of women by inclusion among the conveners, resulting on average a higher number of female speakers on average than in other fields.



**FIGURE 2 | Cooperative Fieldwork.** Carol Litchfield, center, leads a sampling brigade at Great Salt Lake, Utah, USA.

We propose that the model of inclusion and mentorship experienced in our field should continue through concerted efforts, and it should be applied more broadly to all fields of science. To extrapolate from a famous quote by Marie Curie: after all, science is without genders, and it is only through lack of the historical sense that gender qualities have been attributed to it.

## ACKNOWLEDGMENTS

This chapter is dedicated to Carol Litchfield (**Figure 2**) who mentored many men and women of halophiles. Also, many thanks to Bill Grant, Thane Papke, Antonio Ventosa, and Terry McGenity for helping collect programs, photographs and historical information to support this project.

## REFERENCES

- Abdul Halim, M. F., Zou, J., Pfeiffer, F., Haft, D., and Pohlschroder, M. (2013). *Haloraxus volcanii* archaeosortase is required for motility, mating and C-terminal processing of the S-layer glycoprotein. *Mol. Microbiol.* 88, 1164–1175. doi: 10.1111/mmi.12248
- Amjres, H., Béjar, V., Quesada, E., Abrini, J., and Llamas, I. (2011). *Halomonas rissenensis* sp. nov., an exopolysaccharide-producing, halophilic bacterium isolated from a solar saltern. *Int. J. Syst. Evol. Microbiol.* 61, 2600–2605. doi: 10.1099/ijss.0.027268-0
- Antón, J., Llobet-Brossa, E., Rodríguez-Valera, F., and Amann, R. (1999). Fluorescent in situ hybridization analysis of the prokaryotic community inhabiting crystallizer ponds. *Environ. Microbiol.* 1, 517–523. doi: 10.1046/j.1462-2920.1999.00065.x
- Antón, J., Rosselló-Mora, R., Rodriguez-Valera, F., and Amann, R. (2000). Extremely halophilic bacteria in crystallizer ponds from solar salterns. *Appl. Environ. Microbiol.* 66, 3052–3057. doi: 10.1128/AEM.66.7.3052-3057.2000
- Bageshwar, U. K., Premkumar, L., Gokhman, I., Savchenko, T., Sussman, J. L., and Zamir, A. (2004). Natural protein engineering: a uniquely salt-tolerant, but not halophilic,  $\alpha$ -type carbonic anhydrase from algae proliferating in low- to hyper-saline environments. *Protein Eng. Des. Sel.* 17, 191–200. doi: 10.1093/protein/gzh022
- Baxter, B. K., Litchfield, C. D., Sowers, K., Griffith, J. D., DasSarma, P. A., and DasSarma, S. (2005). “Microbial diversity of Great Salt Lake,” in *Adaptation to Life in High Salt Concentrations in Archaea Bacteria, and Eukarya*, eds N. Gundee-Cimerman, A. Oren, and A. Plemenitaš (Dordrecht, the Netherlands: Springer), 9–25. doi: 10.1007/1-4020-3633-7\_2
- Benison, K. C., Bowen, B. B., Oboh-Ikuuenobe, F. E., Jagniecki, E., LaClair, D. A., Story, S. L., et al. (2007). The sedimentology of acid saline lakes in southern Western Australia: newly described processes and products of an extreme environment. *J. Sediment. Res.* 77, 366–388. doi: 10.2110/jsr.2007.038
- Benison, K. C., Jagniecki, E. A., Edwards, T. B., Mormile, M. C., and Storrie-Lombardi, M. C. (2008). “Hairy blobs”: microbial suspects from modern and ancient ephemeral acid saline evaporites. *Astrobiology* 8, 807–821. doi: 10.1089/ast.2006.0034
- Bowen, B. B., Benison, K. C., Oboh-Ikuuenobe, F. E., Story, S., and Mormile, M. (2008). Active hematite concretion formation in modern acid saline lake sediments, Lake Brown, Western Australia. *Earth Planet. Sci. Lett.* 268, 52–63. doi: 10.1016/j.epsl.2007.12.023
- Caplan, S. R., and Ginzburg, M. (eds.) (1978). *Energetics and Structure of Halophilic Microorganisms*. Amsterdam: Elsevier – North Holland Biomedical Press.
- Casadevall, A., and Handelsman, J. (2014). The presence of female conveners correlates with a higher proportion of female speakers at scientific symposia. *mBio* 5, e00846–e00913. doi: 10.1128/mBio.00846-13.
- Chaiyanan, S., Chaiyanan, S., Maugel, T., Huq, A., Robb, F. R., and Colwell, R. R. (1999). Polyphasic taxonomy of a novel *Halobacillus*, *Halobacillus thailandensis* sp. nov. isolated from fish sauce. *Syst. Appl. Microbiol.* 22, 360–365. doi: 10.1016/S0723-2020(99)80043-5
- Colwell, R. R., Litchfield, C. D., Vreeland, R. H., Kiefer, L. A., and Gibbons, N. E. (1979). Taxonomic studies of red halophilic bacteria. *Int. J. Syst. Bacteriol.* 29, 379–399. doi: 10.1099/00207713-29-4-379
- Conley, D., and Stadmark, J. (2012). Gender matters: a call to commission more women writers. *Nature* 488, 590. doi: 10.1038/488590a
- Conner, A. J., and Benison, K. C. (2013). Acidophilic halophilic microorganisms in fluid inclusions in halite from Lake Magic, Western Australia. *Astrobiology* 9, 850–860. doi: 10.1089/ast.2012.0956
- Corcelli, A., Lattanzio, V. M. T., Mascolo, G., Babudri, F., Oren, A., and Kates, M. (2004). Novel sulfonolipid in the extremely halophilic bacterium *Salinibacter ruber*. *Appl. Environ. Microbiol.* 70, 6678–6685. doi: 10.1128/AEM.70.11.6678-6685.2004
- Danon, A., and Caplan, S. R. (1976). Stimulation of ATP synthesis in *Halobacterium halobium* R1 by light-induced or artificially created proton electrochemical potential gradients across the cell membrane. *Biochim. Biophys. Acta* 423, 133–140. doi: 10.1016/0005-2728(76)90107-9
- Danon, A., and Caplan, S. R. (1977). CO<sub>2</sub> fixation by *Halobacterium halobium*. *FEBS Lett.* 74, 255–258. doi: 10.1016/0014-5793(77)80858-2
- Danon, A., and Stoeckenius, W. (1974). Photophosphorylation in *Halobacterium halobium*. *Proc. Natl. Acad. Sci. U.S.A.* 71, 1234–1238. doi: 10.1073/pnas.71.4.1234
- Ebel, C., Costenaro, L., Pascu, M., Faou, P., Kernel, B., Proust-De Martin, F., et al. (2002). Solvent interactions of halophilic malate dehydrogenase. *Biochemistry* 41, 13234–13244. doi: 10.1021/bi0258290
- Ebel, C., Faou, P., and Zaccai, G. (1999). Protein–solvent and weak protein–protein interactions in halophilic malate dehydrogenase. *J. Crystal Growth* 196, 395–402. doi: 10.1016/S0022-0248(98)00822-7
- Eisen, J. (2012). Q-Bio Conference in Hawaii, Bring your Surfboard and your Y Chromosome b/c they Don’t Take a XX. Available at: <http://phylogenomics.blogspot.com/2012/09/q-bio-conference-in-hawaii-bring-your.html> (accessed October 12, 2013).
- European Gender Summit. (2012). “Developing systematic implementation strategy to advance EU policy on gender equality in science,” in *Proceedings of the Presentations given to European Union*, Brussels. doi: 10.1089/ast.2007.0234
- Faria, T. Q., Mingote, A., Siopa, F., Ventura, R., Maycock, C., and Santos, H. (2008). Design of new enzyme stabilizers inspired by glycosides of hyperthermophilic microorganisms. *Carbohydr. Res.* 343, 3025–3033. doi: 10.1016/j.carres.2008.08.030
- Fendrihan, S., Berces, A., Lammer, H., Musso, M., Ronto, G., Polacksek, T. K., et al. (2009). Investigating the effects of simulated Martian ultraviolet radiation on *Halococcus dombrowskii* and other extremely halophilic archaeabacteria. *Astrobiology* 9, 104–112. doi: 10.1089/ast.2007.0234
- Fischer, M., Gokhman, I., Pick, U., and Zamir, A. (1996). A salt-resistant plasma membrane carbonic anhydrase is induced by salt in *Dunaliella salina*. *J. Biol. Chem.* 271, 17718–17723. doi: 10.1074/jbc.271.30.17718
- Fischer, M., Gokhman, I., Pick, U., and Zamir, A. (1997). A structurally novel transferrin-like protein accumulates in the plasma membrane of the unicellular green alga *Dunaliella salina* grown in high salinities. *J. Biol. Chem.* 272, 1565–1570. doi: 10.1074/jbc.272.3.1565
- Fröls, S. (2013). Archaeal biofilms: widespread and complex. *Biochem. Soc. Trans.* 41, 393–398. doi: 10.1042/BST20120304
- Fröls, S., Dyall-Smith, M., and Pfeifer, F. (2012). Biofilm formation by haloarchaea. *Environ. Microbiol.* 14, 3159–3174. doi: 10.1111/j.1462-2920.2012.02895.x
- Ginzburg, M. (1969). The unusual membrane permeability of two halophilic unicellular organisms. *Biochim. Biophys. Acta* 173, 370–376. doi: 10.1016/0005-2736(69)90002-9
- Ginzburg, M. (1978). “Ion metabolism in whole cells of *Halobacterium halobium* and *H. marismortui*,” in *Energetics and Structure of Halophilic Microorganisms*, eds S. R. Caplan and M. Ginzburg (Amsterdam: Elsevier/North Holland Biomedical Press), 561–577.
- Ginzburg, M. (1987). *Dunaliella*: a green alga adapted to salt. *Adv. Bot. Res.* 14, 93–183. doi: 10.1016/S0065-2296(08)60271-2
- Ginzburg, M., and Ginzburg, B. Z. (1985). Ion and glycerol concentrations in 12 isolates of *Dunaliella*. *J. Exp. Bot.* 36, 1064–1074. doi: 10.1093/jxb/36.7.1064
- Ginzburg, M., Sachs, L., and Ginzburg, B. Z. (1970). Ion metabolism in a Halobacterium. I. Influence of age of culture on intracellular concentrations. *J. Gen. Physiol.* 55, 187–207. doi: 10.1085/jgp.55.2.187
- Grant, W. D., and Kogut, M. (1986). The molecular basis of haloadaptation in microorganisms. *FEMS Microbiol. Rev.* 39, 1. doi: 10.1111/j.1574-6968.1986.tb01834.x

- Greco, J. A., Wagner, N. L., and Birge, R. R. (2012). Fourier Transform holographic associative processors based on bacteriorhodopsin. *Int. J. Unconvent. Comput.* 8, 433–457.
- Griffith, J. D., Willcox, S., Powers, D. W., Nelson, R., and Baxter, B. K. (2008). Discovery of abundant cellulose microfibers encased in 250 Ma permian halite: a macromolecular target in the search for life on other planets. *Astrobiology* 8, 215–228. doi: 10.1089/ast.2007.2007
- Gunde-Cimerman, N., Zalar, P., de Hoog, S., and Plemenitaš, A. (2000). Hypersaline waters in salterns: natural ecological niches for halophilic black yeasts. *FEMS Microbiol. Ecol.* 32, 235–240.
- Hamburger, C. (1905). Zur Kenntnis der *Dunaliella salina* und einer Amöbe aus Salinenwasser von Cagliari. *Arch. Protistenkdunde* 6, 111–131.
- Hänelt, I., Löchte, S., Sundermann, L., Elbers, K., Vor der Brüggen, M., and Bakker, E. P. (2010a). Gain of function mutations in membrane region M2C2 of KtrB open a gate controlling K<sup>+</sup> transport by the KtrAB system from *Vibrio alginolyticus*. *J. Biol. Chem.* 285, 10318–10327. doi: 10.1074/jbc.M109.089870
- Hänelt, I., Wunnicke, D., Müller-Trimbusch, M., Vor der Brüggen, M., Kraus, I., Bakker, E. P., et al. (2010b). Membrane region M2C2 in subunit KtrB of the K<sup>+</sup> uptake system KtrAB from *Vibrio alginolyticus* forms a flexible gate controlling K<sup>+</sup> flux: an electron paramagnetic resonance study. *J. Biol. Chem.* 285, 28210–28219. doi: 10.1074/jbc.M110.139311
- Hänelt, I., Tholema, N., Kröning, N., Vor der Brüggen, M., Wunnicke, D., and Bakker, E. P. (2011). KtrB, a member of the superfamily of K<sup>+</sup> transporters. *Eur. J. Cell Biol.* 90, 696–704. doi: 10.1016/j.ejcb.2011.04.010
- Hof, T. (1935). Investigations concerning bacterial life in strong brines. *Rev. Trav. Bot. Neerl.* 32, 92–173.
- Humbard, M. A., Miranda, H. V., Lim, J.-M., Krause, D. J., Pritz, J. R., Zhou, G., et al. (2010). Ubiquitin-like small archaeal modifier proteins (SAMPs) in *Haloferax volcanii*. *Nature* 463, 54–60. doi: 10.1038/nature08659
- Ionescu, D., Siebert, C., Polerecky, L., Munwes, Y. Y., Lott, C., Häusler, S., et al. (2012). Microbial and chemical characterization of underwater fresh water springs in the Dead Sea. *PLoS ONE* 7:e38319. doi: 10.1371/journal.pone.0038319
- Isbell, L. A., Young, T. P., and Harcourt, A. H. (2012). Stag parties linger: continued gender bias in a female-rich scientific discipline. *PLoS ONE* 7:e49682. doi: 10.1371/journal.pone.0049682
- Javor, B. J. (1983). Planktonic standing crop and nutrients in a saltern ecosystem. *Limnol. Oceanogr.* 28, 153–159. doi: 10.4319/lo.1983.28.1.0153
- Javor, B. J. (1984). Growth potential of halophilic bacteria isolated from solar salt environments: carbon sources and salt requirements. *Appl. Environ. Microbiol.* 48, 352–360.
- Javor, B. J. (1988). CO<sub>2</sub> fixation in halobacteria. *Arch. Microbiol.* 149, 433–440. doi: 10.1007/BF00425584
- Javor, B. (1989). *Hypersaline Environments: Microbiology and Biogeochemistry*. Berlin: Springer-Verlag. doi: 10.1007/978-3-642-74370-2
- Javor, B. J. (2002). Industrial microbiology of solar salt production. *J. Ind. Microbiol. Biotechnol.* 28, 42–47. doi: 10.1038/sj/jim/7000173
- Javor, B., Requadt, C., and Stoeckenius, W. (1982). Box-shaped halophilic bacteria. *J. Bacteriol.* 151, 1532–1542.
- Kandiba, L., Aitio, O., Helin, J., Guan, Z., Permi, P., Bamford, D. H., et al. (2012). Diversity in prokaryotic glycosylation: an archaeal-derived N-linked glycan contains legionaminic acid. *Mol. Microbiol.* 84, 578–593. doi: 10.1111/j.1365-2958.2012.08045.x
- Kish, A., and DiRuggiero, J. (2008). Rad50 is not essential for the mre11-dependent repair of DNA double strand breaks in *Halobacterium* sp. str. NRC-1. *J. Bacteriol.* 190, 5210–5216. doi: 10.1128/JB.00292-08
- Kogut, M. (1980a). Are there strategies of microbial adaptation to extreme environments? *Trends Biochem. Sci.* 5, 11–14. doi: 10.1016/S0968-0004(80)80070-3
- Kogut, M. (1980b). Microbial strategies of adaptability. *Trends Biochem. Sci.* 5, 47–50. doi: 10.1016/S0968-0004(80)80095-8
- Lers, A., Bigner, Y., and Zamir, A. (1989). Photoinduction of massive β-carotene accumulation by the alga *Dunaliella bardawil*. Kinetics and dependence on gene action. *Plant Physiol.* 93, 389–395. doi: 10.1104/pp.93.2.389
- Lenassi, M., Gostinéar, C., Jackman, S., Turk, M., Sadowski, I., Nislow, C., et al. (2013). Whole genome duplication and enrichment of metal cation transporters revealed by de novo genome sequencing of extremely halotolerant black yeast *Hortaea werneckii*. *PLoS ONE* 8:e71328. doi: 10.1371/journal.pone.0071328
- Lenassi, M., Vaupotière, T., Gunde-Cimerman, N., and Plemenitaš, A. (2007). The MAP kinase HwHog1 from the halophilic black yeast *Hortaea werneckii*: coping with stresses in solar salterns. *Saline Syst.* 3, 3. doi: 10.1186/1746-1448-3-3
- Litchfield, C. D. (1991). “Red – the magic color for solar salt production,” in *Das Salz in der Rechts- und Handelsgeschichte*, eds J.-C. Hocquet and R. Palme (Hall in Tirol: Berenkamp), 403–412.
- Litchfield, C. D. (1998). Survival strategies for microorganisms in hypersaline environments and their relevance to life on early Mars. *Meteorit. Planet. Sci.* 33, 813–819. doi: 10.1111/j.1945-5100.1998.tb01688.x
- Litchfield, C. D., and Gillivet, P. M. (2002). Microbial diversity and complexity in hypersaline environments: a preliminary assessment. *J. Ind. Microbiol. Biotechnol.* 28, 48–55. doi: 10.1038/sj/jim/7000175
- Litchfield, C. D., Irby, A., Kis-Papo, T., and Oren, A. (2000). Comparisons of the polar lipid and pigment profiles of two solar salterns located in Newark, California, USA, and Eilat, Israel. *Extremophiles* 4, 259–265. doi: 10.1007/s007920070011
- Litchfield, C. D., Palme, R., and Piasecki, P. (eds) (2001). *Le monde du sel. Mélanges offerts à Jean-Claude Hocquet*. Hall in Tirol: Berenkamp.
- Lobasso, S., Lopalco, P., Angelini, R., Pollice, A., Laera, G., Milano, F., et al. (2012). Isolation of Squarebop I bacteriorhodopsin from biomass of coastal salterns. *Protein Expr. Purif.* 84, 73–79. doi: 10.1016/j.pep.2012.04.017
- Lopalco, P., Lobasso, S., Baroni, M., Angelini, R., and Corcelli, A. (2011). “Impact of lipidomics on the microbial world of hypersaline environments,” in *Halophiles and Hypersaline Environments: Current Research and Future Trends*, eds A. Ventosa, A. Oren, and Y. Ma (Berlin: Springer), 123–135. doi: 10.1007/978-3-642-20198-6\_6
- Luque, R., Béjar, V., Quesada, E., Martínez-Checa, F., and Llamas, I. (2012). *Halomonas ramblicola* sp. nov., a moderately halophilic bacterium from Rambla Salada, a Mediterranean hypersaline rambla in south-east Spain. *Int. J. Syst. Evol. Microbiol.* 62, 2903–2909. doi: 10.1099/ijss.0.039453-0
- Madern, D., Ebel, C., and Zaccai, G. (2000). Halophilic adaptations of enzymes. *Extremophiles* 4, 91–98. doi: 10.1007/s007920050142
- Makowski, I., Frolow, F., Saper, M. A., Shoham, M., Wittmann, H. G., and Yonath, A. (1987). Single crystals of large ribosomal particles from *Halobacterium marismortui* diffract to 6 Å. *J. Mol. Biol.* 193, 819–822. doi: 10.1016/0022-2836(87)90362-7
- Mata, J. A., Béjar, V., Bressollier, P., Tallon, R., Urdaci, M. C., Quesada, E., et al. (2008). Characterization of exopolysaccharides produced by three moderately halophilic bacteria belonging to the family Alteromonadaceae. *J. Appl. Microbiol.* 105, 521–528. doi: 10.1111/j.1365-2672.2008.03789.x
- McGenity, T. J., Gemmell, R. T., Grant, W. D., and Stan-Lotter, H. (2000). Origins of halophilic microorganisms in ancient salt deposits (MiniReview). *Environ. Microbiol.* 2, 243–250. doi: 10.1046/j.1462-2920.2000.00105.x
- Mesbah, N. M., Hedrick, D. B., Peacock, A., Rohde, M., and Wiegel, J. (2007). *Natrananaerobius thermophilus* gen. nov. sp. nov., a halophilic, alkalithermophile bacterium from soda lakes of the Wadi An Natrun, Egypt, and proposal of *Natrananaerobiaceae* fam. nov. and *Natrananaerobiales* ord. nov. *Int. J. Syst. Evol. Microbiol.* 57, 2507–2512. doi: 10.1099/ijss.0.65068-0
- Mesbah, N. M., and Wiegel, J. (2011). Biochemical characterization of the Na<sup>+</sup>-translocating F1Fo-ATPase from the halophilic alkalithermophile *Natrananaerobius thermophilus*. *Biochim. Biophys. Acta* 1807, 1133–1142. doi: 10.1016/j.bbabi.2011.05.001
- Mesbah, N. M., Cook, G. M., and Wiegel, J. (2009). The halophilic alkalithermophile *Natrananaerobius thermophilus* adapts to multiple environmental extremes using a large repertoire of Na<sup>+</sup>(K<sup>+)/H<sup>+</sup>) antiporters. *Mol. Microbiol.* 74, 270–281. doi: 10.1111/j.1365-2958.2009.06845.x</sup>
- Meuser, J. E., Baxter, B. K., Spear, J. R., Peters, J. W., Posewitz, M. C., and Boyd, E. S. (2013). Contrasting patterns of community assembly in the stratified water column of great salt lake, Utah. *Microb. Ecol.* 66, 268–280. doi: 10.1007/s00248-013-0180-9
- Miranda, H. V., Nembharkar, N., Su, D., Hepowitz, N., Krause, D. J., Pritz, J. R., et al. (2011). E1- and ubiquitin-like proteins provide a direct link between protein conjugation and sulfur transfer in archaea. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4417–4422. doi: 10.1073/pnas.1018151108
- Mormile, M. R., Biesen, M. A., Gutierrez, M. C., Ventosa, A., Pavlovich, J. B., Onstott, T. C., et al. (2003). Isolation of *Halobacterium salinarum* retrieved directly from halite brine inclusions. *Env. Microbiol.* 5, 1094–1102. doi: 10.1046/j.1462-2920.2003.00509.x

- Mormile, M. R., Hong, B.-Y., and Benison, K. C. (2009). Molecular analysis of the microbial communities of Mars-analog lakes in Western Australia. *Astrobiology* 9, 919–930. doi: 10.1089/ast.2008.0293
- Mormile, M. R., Romine, M. F., Garcia, M. T., Ventosa, A., Bailey, T. J., and Peyton, B. M. (1999). *Halomonas campisalis* sp. nov., a denitrifying, moderately haloalkaliphilic bacterium. *Syst. Appl. Microbiol.* 22, 551–558. doi: 10.1016/S0723-2020(99)80008-3
- National Science Foundation, National Center for Science and Engineering Statistics. (2013). *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2013*. Special Report NSF 13-304. Arlington, VA. Available at: <http://www.nsf.gov/statistics/wmpd/>
- Oren, A. (2011). “A short history of the symposia on halophilic microorganisms: from Rehovot 1978 to Beijing 2010,” in *Halophiles and Hypersaline Environments: Current Research and Future Trends*, eds A. Ventosa, A. Oren, and Y. Ma (Berlin: Springer), 373–382.
- Oren, A., Ginzburg, M., Ginzburg, B. Z., Hochstein, L. I., and Volcani, B. E. (1990). *Haloarcula marismortui* (Volcani) sp. nov., nom. rev., an extremely halophilic bacterium from the Dead Sea. *Int. J. Syst. Bacteriol.* 40, 209–210. doi: 10.1099/00207713-40-2-209
- Ormerod, J. (2003). Every dogma has its day: a personal look at carbon metabolism in photosynthetic bacteria. *Photosynth. Res.* 76, 135–143. doi: 10.1023/A:1024938531382
- Pais, T. M., Lamosa, P., Matzapetakis, M., Turner, D. L., and Santos, H. (2012). Mannosylglycerate stagilizes staphylococcal nuclease with restriction of slow β-sheet motions. *Protein Sci.* 21, 1126–1137. doi: 10.1002/pro.2100
- Petter, H. F. M. (1931). On bacteria of salted fish. *Proc. Kon. Akad. Wetensch Ser. B* 34, 1417–1423.
- Petter, H. F. M. (1932). *Over Roode en Andere Bacterieën van Gezouten Visch*. Ph.D. thesis, University of Utrecht, Utrecht.
- Pfeifer, F. (2012). Distribution, formation and regulation of gas vesicles. *Nat. Rev. Microbiol.* 10, 705–715. doi: 10.1038/nrmicro2834
- Pfeifer, F., and Betlach, M. (1985). Genome organization of *Halobacterium halobium*: a 70 kbp island of more (AT) rich DNA in the chromosome. *Mol. Gen. Genet.* 198, 449–455. doi: 10.1007/BF00332938
- Pfeifer, F., and Blaejoe, U. (1990). Transposition burst of the ISH27 insertion element family in *Halobacterium halobium*. *Nucleic Acids Res.* 18, 6921–6925. doi: 10.1093/nar/18.23.6921
- Pietila, M. K., Roine, E., Paulin, L., Kalkkinen, N., and Bamford, D. H. (2009). An ssDNA virus infecting archaea: a new lineage of viruses with a membrane envelope. *Mol. Microbiol.* 72, 307–319. doi: 10.1111/j.1365-2958.2009.06642.x
- Plemenitaš, A., Vaupotić, T., Lenassi, M., Kogej, T., and Gunde-Cimerman, N. (2008). Adaptation of extremely halotolerant black yeast *Hortaea werneckii* to increased osmolarity: a molecular perspective at a glance. *Stud. Mycol.* 61, 67–75. doi: 10.3114/sim.2008.61.06
- Pohlschröder, M., Prinz, W., Hartmann, E., and Beckwith, J. (1997). Protein translocation in the three domains of life: variations on a theme. *Cell* 91, 563–566. doi: 10.1016/S0092-8674(00)80443-2
- Robinson, C. K., Webb, K. M., Kaur, A., Jaruga, P., Dizdaroglu, M., Baliga, N., et al. (2011). A major role for non-enzymatic antioxidant processes in the radioresistance of *Halobacterium salinarum*. *J. Bacteriol.* 193, 1653–1662. doi: 10.1128/JB.01310-10
- Robinson, C. K., Wierzchos, J., Black, C., Crits-Christoph, A., Ma, B., Ravel, J., et al. (2014). Microbial diversity and the presence of algae in halite endolithic communities are correlated to atmospheric moisture in the hyper-arid zone of the Atacama Desert. *Environ. Microbiol.* doi: 10.1111/1462-2920.12364 [Epub ahead of print].
- Rodrigues, M. V., Borges, N., Henriques, M., Lamosa, P., Ventura, R., Fernandes, C., et al. (2007). Bifunctional CTP: inositol-1-phosphate cytidylyltransferase/CDP-inositol:inositol-1-phosphate transferase, the key enzyme for di-myo-inositol-phosphate synthesis in several (hyper)thermophiles. *J. Bacteriol.* 189, 5405–5412. doi: 10.1128/JB.00465-07
- Rose, R. W., Brüser, T., Kissinger, J. C., and Pohlschröder, M. (2002). Adaptation of protein secretion to extremely high salt concentrations by extensive use of the twin arginine translocation pathway. *Mol. Microbiol.* 5, 943–950. doi: 10.1046/j.1365-2958.2002.03090.x
- Royal Society of Chemistry. (2008). *Change of Heart: Career Intentions and the Chemistry*. Ph.D. thesis, Royal Society of Chemistry, London.
- Russell, N. J., Adams, R., Bygraves, J., and Kogut, M. (1986). Cell envelope phospholipid changes in a moderate halophile during phenotypic adaptation to altered salinity and osmotic stress. *FEMS Microbiol. Rev.* 39, 103–107. doi: 10.1111/j.1574-6968.1986.tb01849.x
- Russell, N. J., and Kogut, M. (1985). Haloadaptation: salt sensing and cell-envelope changes. *Microbiol. Sci.* 2, 345–350.
- Sabet, S. (2012). “Halophilic viruses,” in *Advances in Understanding the Biology of Halophilic Microorganisms*, ed. R. H. Vreeland (New York, NY: Springer Publishing Company), 81–116. doi: 10.1007/978-94-007-5539-0\_4
- Sabet, S., Chu, W., and Jiang, S. C. (2006). Isolation and genetic analysis of haloalkaliphilic bacteriophages in a North American soda lake. *Microb. Ecol.* 51, 543–554. doi: 10.1007/s00248-006-9069-1
- Sabet, S., Diallo, L., Hays, L., Jung, W., and Dillon, J. (2009). Characterization of halophiles isolated from solar salterns in Baja California, Mexico. *Extremophiles* 13, 643–656. doi: 10.1007/s00792-009-0247-1
- Sadka, A., Himmelhoch, S., and Zamir, A. (1991). A 150 kilodalton cell surface protein is induced by salt in the halotolerant green alga *Dunaliella salina*. *Plant Physiol.* 95, 822–831. doi: 10.1104/pp.95.3.822
- Sanchez Botero, C., Oboh-Ikuenobe, F. E., and Macphail, M. (2013). First fossil pollen record of the Northern Hemisphere species *Aglaoreidia cyclops* Erdtman, 1960 in Australia. *Alcheringa* 37, 1–5. doi: 10.1080/03115518.2013.785336
- Santos, F., Meyer-Dierks, A., Peña, A., Rosselló-Mora, R., Amann, R., and Antón, J. (2007). Metagenomic approach to the study of halophages: the environmental halophage 1. *Environ. Microbiol.* 9, 1711–1723. doi: 10.1111/j.1462-2920.2007.01289.x
- Schlunzen, F., Kölln, I., Janell, D., Gluhmann, M., Levin, I., Bashan, A., et al. (1999). The identification of selected components in electron density maps of prokaryotic ribosomes at 7 Å resolution. *J. Synchrotron Rad.* 6, 928–941. doi: 10.1107/S0909049599006664
- Schmid, A. K., Pan, M., Sharma, K., and Baliga, N. S. (2011). Two transcription factors are necessary for iron homeostasis in a salt-dwelling archaeon. *Nucleic Acids Res.* 39, 2519–2533. doi: 10.1093/nar/gkq1211
- Schroeder, J., Dugdale, H. L., Radersma, R., Hinsch, M., Buehler, D. M., Saul, J., et al. (2013). Fewer invited talks by women in evolutionary biology symposia. *J. Evol. Biol.* 26, 2063–2069. doi: 10.1111/jeb.12198
- Senélio, A., Paulin, L., Kellner, S., Helm, M., and Roine, E. (2012). Related haloarchaeal pleomorphic viruses contain different genome types. *Nucleic Acids Res.* 40, 5523–5534. doi: 10.1093/nar/gks215
- Sharma, K., Gillum, N., Boyd, J. L., and Schmid, A. (2012). The RosR transcription factor is required for gene expression dynamics in response to extreme oxidative stress in a hypersaline-adapted archaeon. *BMC Genomics* 13:351. doi: 10.1186/1471-2164-13-351
- Shen, H. (2013). Mind the gender gap. *Nature* 495, 22–24. doi: 10.1038/495022a
- Sorokin, Y. D., Zhilina, T. N., Lysenko, A. M., Tourova, T. P., and Spiridonova, E. M. (2006). Metabolic versatility of haloalkaliphilic bacteria from soda lakes belonging to the *Alkalipsirillum-Alkalimimicola* group. *Extremophiles* 10, 213–220. doi: 10.1007/s00792-005-0487-7
- Stan-Lotter, H., McGinity, T. J., Legat, A., Denner, E. B. M., Glaser, K., Stetter, K. O., et al. (1999). Very similar strains of *Halococcus salifodinae* are found in geographically separated Permo-Triassic salt deposits. *Microbiology* 145, 3565–3574.
- Teodorescu, E. C. (1905). Organisation et développement du *Dunaliella*, nouveau genre de Volvocacée-Polyblepharidée. *Beih. Z. Bot. Centralbl.* 18, 215–232.
- Todor, H., Sharma, K., Pittman, A. M., and Schmid, A. K. (2013). Protein-DNA binding dynamics predict transcriptional response to nutrients in archaea. *Nucleic Acids Res.* 41, 8546–8558. doi: 10.1093/nar/gkt659
- Vaupotić, T., and Plemenitaš, A. (2007). Differential gene expression and Hog1 interaction with osmoreponsive genes in the extremely halotolerant black yeast *Hortaea werneckii*. *BMC Genomics* 8:280. doi: 10.1186/1471-2164-8-280
- von Böhlen, K., Makowski, I., Hansen, H. A. S., Bartels, H., Berkovitch-Yellin, Z., Zaytsev-Bashan, A., et al. (1991). Characterization and preliminary attempts for derivatization of crystals of large ribosomal subunits from *Haloarcula marismortui* diffracting to 3 Å resolution. *J. Mol. Biol.* 222, 11–15. doi: 10.1016/0022-2836(91)90730-T
- Vreeland, R. H., Litchfield, C. D., Martin, E. L., and Elliot, E. (1980). Halomonas elongata, a new genus and species of extremely salt-tolerant bacteria. *Int. J. Syst. Bacteriol.* 30, 485–495. doi: 10.1099/00207713-30-2-485

- Wagner, N. L., Greco, J. A., and Birge, R. R. (2013a). Directed evolution in nanotechnology. *J. R. Soc. Interface* 10, 20130197.
- Wagner, N. L., Greco, J. A., and Birge, R. R. (2013b). "Visual restoration using microbial rhodopsins," in *Bionanotechnology: Biological Self-Assembly and its Applications*, ed. B. H. A. Rehm (Norfolk, UK: Caister Academic Press), 205–240.
- Wilson, H. L., Aldrich, H. C., and Maupin-Furlow, J. A. (1999). Halophilic 20S proteasomes of the archaeon *Haloflexax volcanii*: purification, characterization, and gene sequence analysis. *J. Bacteriol.* 181, 5814–5824.
- Wydro, R. M., Madira, W., Hiramatsu, T., Kogut, M., and Kushner, D. J. (1977). Salt-sensitive in vitro protein synthesis by a moderately halophilic bacterium. *Nature* 269, 824–825. doi: 10.1038/269824a0
- Yonath, A. (2002a). The search and its outcome: high-resolution structures of ribosomal particles from mesophilic, thermophilic, and halophilic bacteria at various functional states. *Ann. Rev. Biophys. Biomol. Struct.* 31, 257–273. doi: 10.1146/annurev.biophys.31.082901.134439
- Yonath, A. (2002b). High-resolution structures of large ribosomal subunits from mesophilic eubacteria and halophilic archaea at various functional states. *Curr. Prot. Peptide Sci.* 3, 67–78. doi: 10.2174/1389203023380828
- Yonath, A. E. (2011). Tough science. *Nature* 478, S6–S7. doi: 10.1038/478S6a
- Zajc, J., Liu, Y., Dai, W., Yang, Z., Hu, J., Gostinéar, C., et al. (2013). Genome and transcriptome sequencing of the halophilic fungus *Wallemia ichthyophaga*: haloadaptations present and absent. *BMC Genomics* 14:617. doi: 10.1186/1471-2164-14-617
- Zavarzina, D. G., Zhilina, T. N., Kolganova, T. V., Kotelev, M. S., Kuznetsov, B. B., Osipov, G. A., et al. (2013). *Natranaerobaculum magadiense* gen. nov., sp. nov., an anaerobic, alkalithermophilic bacterium from soda lake sediment. *Int. J. Syst. Evol. Microbiol.* 63, 4456–4461. doi: 10.1099/ijss.0.054536-0
- Zhao, B., Mesbah, N. M., Dalin, E., Goodwin, L., Nolan, M., Pitluck, S., et al. (2011). Complete genome sequence of the anaerobic, halophilic alkathermophile *Natranaerobius thermophilus* JW/NM-WN-LFT. *J. Bacteriol.* 193, 4023–4024. doi: 10.1128/JB.05157-11
- Zhilina, T. N., Zavarzina, D. G., Kostrikina, N. A., Osipov, G. A., Panteleeva, A. N., Tourova, T. P., et al. (2012). *Fuchsella alkaliacetigena* gen. nov., sp. nov., an alkaliphilic, lithoautotrophic homoacetogen from a soda lake. *Int. J. Syst. Evol. Microbiol.* 62, 1666–1673. doi: 10.1099/ijss.0.034363-0

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Received: 07 February 2014; accepted: 10 April 2014; published online: 04 June 2014.*

*Citation: Baxter BK, Gunde-Cimerman N and Oren A (2014) Salty sisters: the women of halophiles. Front. Microbiol. 5:192. doi: 10.3389/fmicb.2014.00192*

*This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.*

*Copyright © 2014 Baxter, Gunde-Cimerman and Oren. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# Going from microbial ecology to genome data and back: studies on a haloalkaliphilic bacterium isolated from Soap Lake, Washington State

Melanie R. Mormile\*

Department of Biological Sciences, Missouri University of Science and Technology, Rolla, MO, USA

**Edited by:**

Aharon Oren, The Hebrew University of Jerusalem, Israel

**Reviewed by:**

James A. Coker, University of Maryland University College, USA

Matti Tapani Karp, Tampere University of Technology, Finland

Jean-Luc Cayol, Laboratory of Microbiology Aix-Marseille University, France

**\*Correspondence:**

Melanie R. Mormile, Department of Biological Sciences, Missouri University of Science and Technology, 400 West 11th Street, Rolla, MO 65409-1120, USA  
e-mail: mmormile@mst.edu

Soap Lake is a meromictic, alkaline ( $\sim\text{pH } 9.8$ ) and saline ( $\sim 14\text{--}140 \text{ g liter}^{-1}$ ) lake located in the semiarid area of eastern Washington State. Of note is the length of time it has been meromictic (at least 2000 years) and the extremely high sulfide level ( $\sim 140 \text{ mM}$ ) in its monimolimnion. As expected, the microbial ecology of this lake is greatly influenced by these conditions. A bacterium, *Halanaerobium hydrogeniformans*, was isolated from the mixolimnion region of this lake. *Halanaerobium hydrogeniformans* is a haloalkaliphilic bacterium capable of forming hydrogen from 5- and 6-carbon sugars derived from hemicellulose and cellulose. Due to its ability to produce hydrogen under saline and alkaline conditions, in amounts that rival genetically modified organisms, its genome was sequenced. This sequence data provides an opportunity to explore the unique metabolic capabilities of this organism, including the mechanisms for tolerating the extreme conditions of both high salinity and alkalinity of its environment.

**Keywords:** Soap Lake, *Halanaerobium hydrogeniformans*, alkaliphile, halotolerant, biohydrogen, genome analysis

## INTRODUCTION

Soap Lake is a meromictic, haloalkaline lake located in Washington State. It is thought that the aerobic and anaerobic layers of this lake have not mixed in over 2000 years (Peyton and Yonge, 2002). The lake's meromictic characteristic is due to the steep gradient in salt concentrations between the mixolimnion and the monimolimnion,  $15 \text{ gL}^{-1}$  and  $140 \text{ gL}^{-1}$ , respectively (Sorokin et al., 2007), and the shape of the lake's basin (Edmondson and Anderson, 1965). It is the terminal lake in the chain of lakes that formed in the Lower Grand Coulee during the Missoula Floods. This terminal lake has no surface inlets or outlets. The lack of outlets is the primary reason for the lake's salinity (Anderson, 1958). Soap Lake's water levels are supplied by water runoff from cliffs and plateaus surrounding the lake and from groundwater seepage, with evaporation as the main method for water loss (Anderson, 1958). The alkalinity of Soap Lake is maintained at a nearly constant pH of 9.8 in both the mixolimnion and the monimolimnion (Dimitriu et al., 2008). This alkalinity is controlled by the presence of carbonates and bicarbonates. The concentrations of carbonates in the mixolimnion of Soap Lake average around  $8,500 \text{ mgL}^{-1}$  and  $24,000 \text{ mgL}^{-1}$  in the monimolimnion. In comparison, the concentrations of bicarbonates in Soap Lake were always found to be lower than the carbonates with  $2000 \text{ mgL}^{-1}$  in the mixolimnion and  $4,800 \text{ mgL}^{-1}$  in the monimolimnion (Anderson, 1958).

This environment, due to its high salinity and alkalinity, impacts the microbial community in a number of ways. Though the pH of the environment is 9.8, it can be predicted that the internal pH values of the organisms present is lower. As such, alkaliphilic bacteria must be able to maintain homeostasis

(Krulwich, 1995). In addition, there is a greater energy cost for the production of adenosine triphosphate (ATP) via chemiosmotic means under alkaline conditions (Krulwich et al., 2011). The organisms present also have to retain water in their cells and maintain osmotic homeostasis. They can achieve this by either using a "salting in" strategy or by using organic osmoregulatory compounds. The "salting in" process is typically used by Archaea while Bacteria tend to rely on osmoprotectant compounds.

A number of interesting and novel haloalkaliphilic bacteria have been isolated from Soap Lake. These bacteria include, *Ectothiorhodospira vacuolata* strain (Chadwick and Irgens, 1991), *Halomonas campisalis* (Mormile et al., 1999), *Nitrincola lacinapponensis* (Dimitriu et al., 2005), *Thiocapsa imhoffii* (Asao et al., 2007), *Bacillus* sp. strain SFB (Pollock et al., 2007), *Alkalitalea saponilacus* (Zhao and Chen, 2012), and 'Candidatus *Heliomonas lunata*' strain SLH (Asao et al., 2012). *Halanaerobium hydrogeniformans* was isolated from an enrichment initially prepared for iron-reducing bacteria (Begemann et al., 2012). Though the bacterium is capable of iron-reduction (Paul et al., 2014), it can grow fermentatively on a variety of carbohydrates producing  $\text{H}_2$  in yields comparable to a *Clostridium paraputreficum* that was modified to overexpress a hydrogenase gene (Begemann et al., 2012). Due to *Halanaerobium hydrogeniformans*' ability to produce notable amounts of  $\text{H}_2$  from sugars and its haloalkaliphilic characteristics, its genome was sequenced and annotated (Brown et al., 2011).

*Halanaerobium hydrogeniformans* is a Gram negative, non-motile, non-sporulating rod-shaped bacterium (Begemann et al., 2012). Its genome size is 2,613,116 bp and has a 33.1% G+C content (Brown et al., 2011). It also contains 2,391 candidate protein-encoding genes. In addition to biofuel applications, the

availability of the genome sequence and annotation data of *Halanaerobium hydrogeniformans* enables the determination of the adaptations this organism possesses that facilitates it to thrive under the haloalkaline conditions found in Soap Lake.

## MATERIALS AND METHODS

*Halanaerobium hydrogeniformans*' genome data (Brown et al., 2011) was interrogated to gain information on the function of this bacterium's genome. Information on candidate protein-encoding genes and RNA genes were obtained by using the integrated microbial genomes (IMG) system (Markowitz et al., 2012). BioCyc databases and pathway tools were also used (Caspi et al., 2010). Another sequenced *Halanaerobium*, *Halanaerobium praevalens* GSL<sup>T</sup> (Ivanova et al., 2011) a non-alkaliphilic bacterium, was used as a comparator organism. *Halanaerobium praevalens* GSL<sup>T</sup> was first isolated from the sediments of the Great Salt Lake in Utah (Zeikus et al., 1983). Similar amino acid sequences were determined by performing protein BLAST searches (Altschul et al., 1997). The complete genome of *Halanaerobium hydrogeniformans* has been deposited in NCBI Genomes with accession number NC\_014654.

## RESULTS AND DISCUSSION

### GENOME PROPERTIES

Of the 2391 candidate protein-encoding genes, there are 1867 with function predictions in the genome (**Table 1**). Four 5S rRNA, 16S rRNA, and 23S rRNA genes each are present as are 57 tRNA genes. There are 2082 genes assigned to clusters of orthologous groups (COGs). Interestingly, approximately 25% of the protein-encoding genes are for transmembrane proteins. The distribution of the genes into COG functional categories is provided in **Figure 1** and **Table 2**. The gene count for the different Kyoto Encyclopedia of Genes and Genomes (KEGG) categories is similar between *Halanaerobium hydrogeniformans* and *Halanaerobium praevalens* GSL<sup>T</sup> except for a few categories (**Table 2**). *Halanaerobium praevalens* GSL<sup>T</sup> only has a gene count of 85 for amino acid metabolism while *Halanaerobium hydrogeniformans* has 138. *Halanaerobium praevalens* GSL<sup>T</sup> also has lower gene counts for the KEGG categories of metabolism and metabolism of cofactors and vitamins. On the other hand, *Halanaerobium hydrogeniformans* has a much lower gene count for KEGG category cell motility. Though both of these organisms are not considered to be motile, there are strains of *Halanaerobium praevalens* GSL<sup>T</sup> that are (Kobayashi et al., 2000 and Eder et al., 2001).

### METABOLIC CAPABILITIES

*Halanaerobium hydrogeniformans* has 20% of its genes in the COG category of metabolism and 7% of its genes in the carbohydrate category. Thus, it is not surprising that *Halanaerobium hydrogeniformans* is capable of growth on a number of sugars derived from cellulose and hemicellulose (Begemann et al., 2012). When grown on cellobiose, biomass is produced along with fermentation products, such as formate, acetate, and hydrogen (Begemann et al., 2012). By considering the annotated genome, it should be possible to determine the putative pathway from cellobiose to hydrogen. Cellobiose can be brought into the cell by a putative phosphotransferase system (PTS) lactose/cellobiose-specific

**Table 1 | Genome statistics.**

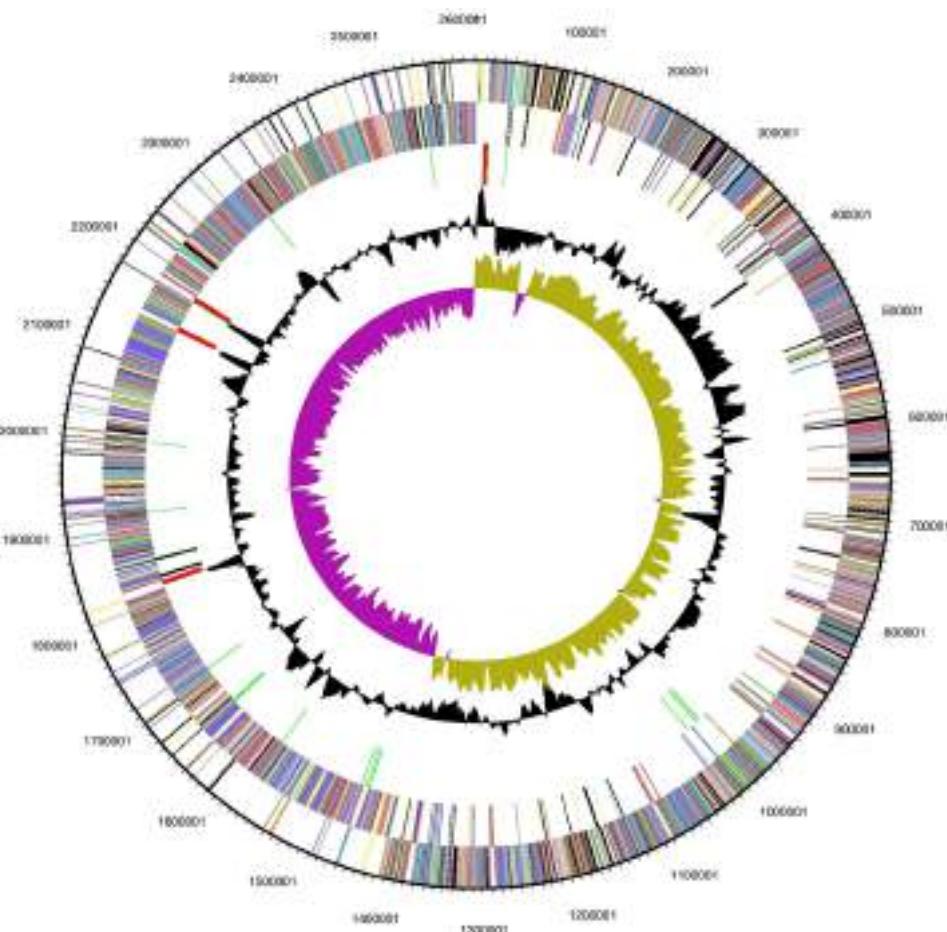
	Number
Total number of bases	2,613,117
Number of DNA coding bases	2,286,541
G+C percentage	33.16
Contigs	1
Total number of genes	2463
Number of protein coding genes	2391
Pseudo genes	96
5S rRNA genes	4
16S rRNA genes	4
23S rRNA genes	4
tRNA genes	57
Genes with function prediction	1867
Genes assigned to clusters of orthologous groups (COGs)	2082
Genes coding transmembrane proteins	624

transporter subunit IIB (gene designated as Halsa\_0653). Once inside the cell, cellobiose would be cleaved and enter the Embden-Meyerhof pathway of glycolysis with the formation of pyruvate. The enzymes for this pathway are present in *Halanaerobium Hydrogeniformans*<sup>1</sup>. Once formed, there are a number of possible fates for pyruvate. A putative pyruvate formate-lyase catalyzes pyruvate and coenzyme A to form formate and acetyl Co A (Halsa\_0723). A possible fate for formate is to be broken down into CO<sub>2</sub> and H<sub>2</sub> by formate-hydrogen lyase (**Figure 2**). However, there was no gene identified that would code for the enzyme, formate-hydrogen lyase. As reported earlier, *Halanaerobium hydrogeniformans* does accumulate formate (Begemann et al., 2012). Thus, it is unlikely that this organism is forming hydrogen from formate. *Halanaerobium praevalens* GSL<sup>T</sup> does not appear to possess this enzyme either. However, formate that is released by these fermentative organisms can be used by sulfate-reducing prokaryotes present in Soap Lake (Dimitriu et al., 2008).

*Halanaerobium hydrogeniformans*' genome possesses an *ldh* gene, indicating that lactate dehydrogenase should also be present (Halsa\_1287). However, lactate has not been detected as a metabolic product from this organism. It is interesting to note that many fermentative organisms possess *ldh* genes (Carere et al., 2012). However, only a few, such as *Bacillus cereus*, had been found to produce lactate in high yields.

*Halanaerobium hydrogeniformans* appears to possess three putative pyruvate dehydrogenase genes (Halsa\_0164, Halsa\_0919, and Halsa\_2297; **Figure 2**). Other genera, *Caldicellulosiruptor*, *Clostridia*, and *Thermoanaerobacter*, also possess putative *pdh* genes but there has been no evidence for functional enzyme production (Carere et al., 2012). *Halanaerobium hydrogeniformans* possesses a gene for the formation of pyruvate:ferredoxin oxidoreductase (Halsa\_2334) as well as two genes that encode

<sup>1</sup>[http://www.genome.jp/kegg-bin/show\\_pathway?has00010](http://www.genome.jp/kegg-bin/show_pathway?has00010)



**FIGURE 1 | Graphical circular map of *Halanaerobium hydrogeniformans*' genome.** From the outside to the center: genes of the forward strand (color by COG categories), genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew. Image generated from IMG (Markowitz et al., 2012).

a polypeptide pyruvate flavodoxin/ferredoxin oxidoreductase domain-containing protein and subunit beta (*Halsa\_0798* and *Halsa\_0799*). Furthermore, it possesses two genes, *Halsa\_1768* and *Halsa\_1862* that encode for iron hydrogenases. *Halsa\_1862* is part of a putative operon that includes a NADH dehydrogenase (*Halsa\_1863*), a ferredoxin-like protein (*Halsa\_1864*), a histidine kinase (*Halsa\_1865*), NADH-quinone oxidoreductase subunit E (*Halsa\_1866*), PHP domain-containing protein (*Halsa\_1867*), an iron-sulfur binding hydrogenase (*Halsa\_1868*), an iron-sulfur cluster domain-containing protein (*Halsa\_1869*), an anti-sigma regulatory factor, serine/threonine protein kinase (*Halsa\_1870*), and an unidentified open reading frames (ORF; *Halsa\_1871*; Figure 3). The organism's ability to produce substantial amounts of H<sub>2</sub>, 2.3 hydrogen molar yield from cellobiose, (Begemann et al., 2012) is of interest as a possible biofuel-producing organism.

It is likely that fermenters such as *Halanaerobium hydrogeniformans*, has a role in interspecies hydrogen transfer in the Soap Lake ecosystem. For example, sulfate- and iron-reducing bacteria were found in the sediments of Soap Lake (Dimitriu et al., 2008) and these organisms can serve as sinks for the H<sub>2</sub> produced

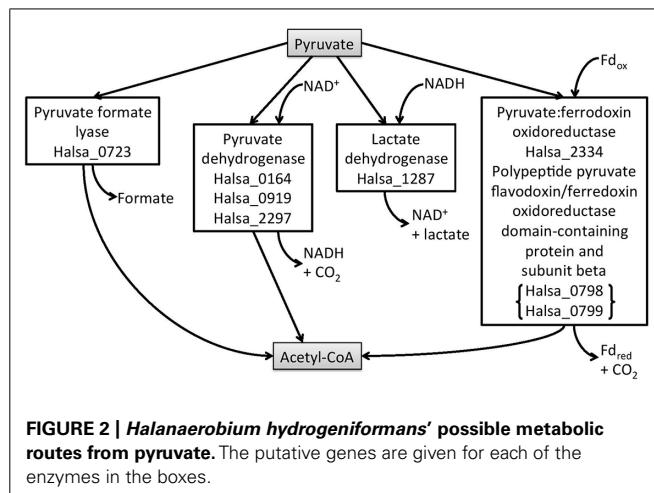
(Jones et al., 1998). However, there have been limited studies on interspecies hydrogen transfer in hypersaline environments. In our own studies, when H<sub>2</sub> and CO<sub>2</sub> were provided as substrates, low numbers of methanogens were detected in the sediments and monimolimnion of Soap Lake while no methanogens were detected in the mixolimnion and chemocline (Dimitriu et al., 2008). Due to thermodynamic constraints (-34 kJ/mol H<sub>2</sub>; Oren, 1999), autotrophic methanogenesis is unlikely to occur, especially in environments with large amounts of sulfate present, such as Soap Lake. Sulfate reduction with H<sub>2</sub> is slightly more thermodynamically favorable than methanogenesis in hypersaline environments (Oren, 2010). In fact, hydrogenotrophic sulfate reducers have been reported from the hypersaline soda lakes of the Kulunda Steppe in southeastern Siberia in Russia (Foti et al., 2007). The first report of interspecies hydrogen transfer possible in hypersaline soda lakes involved a hydrogenotrophic sulfate-reducing bacterium, *Desulfohalobium retbaense*, was found to utilize the H<sub>2</sub> produced by two species of *Halanaerobium*, *Halanaerobium saccharolytica* subsp. *Senegalense*, and *Halanaerobium* sp. strain FR1H from glycerol fermentation (Cayol et al.,

**Table 2 | Number of genes associated with the general COG functional categories.**

KEGG Category	<i>Halanaerobium hydrogeniformans</i> gene count	<i>Halanaerobium praevalens</i> gene count
Amino acid metabolism	138	85
Biosynthesis of other secondary metabolites	15	8
Carbohydrate metabolism	171	137
Cell motility	5	59
Energy metabolism	93	82
Folding, sorting, and degradation	32	30
Glycan biosynthesis and metabolism	25	28
Lipid metabolism	51	39
Membrane transport	95	83
Metabolism	486	375
Metabolism of cofactors and vitamins	110	83
Metabolism of other amino acids	24	32
Metabolism of terpenoids and polyketides	20	17
Nucleotide metabolism	77	81
Replication and repair	41	39
Signal transduction	39	53
Transcription	4	4
Translation	77	80
Transport and catabolism	2	1
Xenobiotics biodegradation and metabolism	25	27

2002). When *Desulfohalobium retbaense* was present as an H<sub>2</sub>-scavenger, glycerol consumption increased and H<sub>2</sub> concentrations approached or were at undetectable amounts.

From early on, it was recognized that glycerol was a major carbon source in saline lakes (Borowitzka, 1981). Glycerol is produced as an osmoregulatory solute by organisms such as green alga, *Dunaliella salina* (Oren, 1993). Not only can glycerol be released from lysed cells but can also leak from healthy cells (Bardavid et al., 2008). This source of carbon can be used by halophilic aerobic prokaryotes, such as *Halococcus* and *Salinibacter*. These aerobic bacteria oxidize glycerol incompletely with excretion of products such as acetic acid, lactic acid, and pyruvic acid (Oren, 2008). Other microorganisms present in these hypersaline environments can subsequently use these products. When a cell takes up glycerol, the glycerol can be converted into dihydroxyacetone and then integrated into pyruvate metabolism, resulting in the products listed above. Glycerol can also be

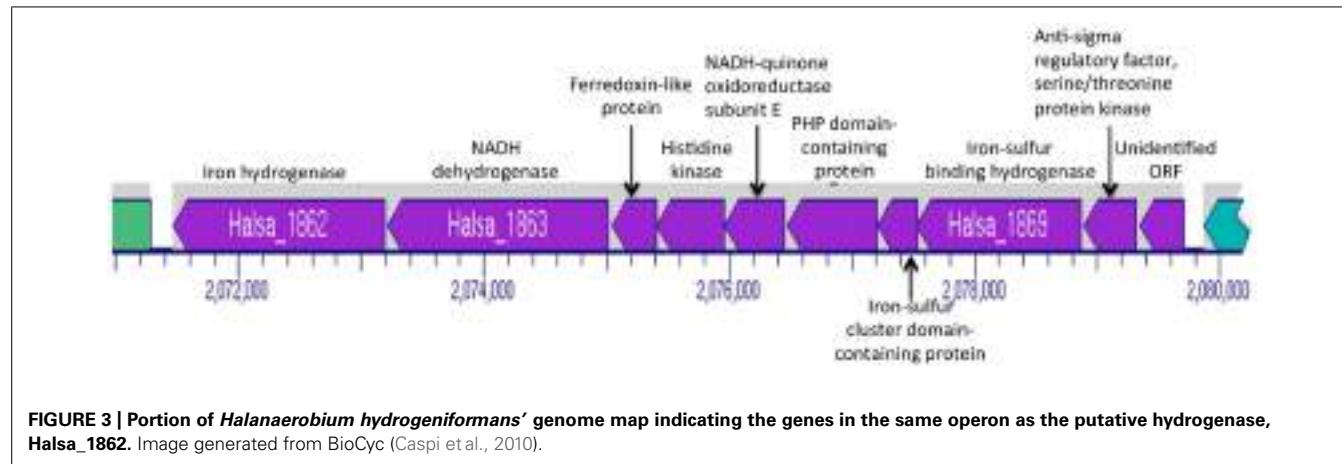
**FIGURE 2 |** *Halanaerobium hydrogeniformans'* possible metabolic routes from pyruvate. The putative genes are given for each of the enzymes in the boxes.

converted into 1,3-propanediol to replenish NAD<sup>+</sup> from NADH<sub>2</sub> resulting when glycerol is oxidized to dihydroxyacetone and dihydroxyacetone phosphate is oxidized to phosphoenolpyruvate. Much of the NADH<sub>2</sub> produced is recycled to NAD<sup>+</sup> through the formation of fermentation end products, such as ethanol, acetate, and butyrate. However, some NAD<sup>+</sup> must be replenished through an alternate pathway (Zeng, 1996). Excess glycerol can be shunted into the 1,3-propanediol production pathway where NADH<sub>2</sub> is re-oxidized to form 1,3-propanediol. This metabolism is present in *Halanaerobium hydrogeniformans* (Roush et al., 2014).

The metabolism of glycerol is of interest not only for its ecological role as a source of carbon in saline lakes but also for the formation of commodity compounds, such as 1,3-propanediol. Glycerol is formed as a byproduct during biodiesel production (Thompson and He, 2006). The first step in the conversion of glycerol to 1,3-propanediol is the removal of a water molecule from glycerol by the enzyme glycerol dehydratase. This step creates the intermediate 3-hydroxypropanal. Next, the enzyme 1,3-propanediol dehydrogenase, oxidizes NADH<sub>2</sub> to form 1,3-propanediol, replenishing the NAD<sup>+</sup> needed by the cell for normal metabolism (Zeng, 1996). The genome of *Halanaerobium hydrogeniformans* revealed that it possessed the possibility of this metabolism<sup>2</sup>. The genes that it possesses that can possibly contribute to this pathway are Halsa\_0984 (a putative glycerol dehydratase), Halsa\_0672 (a putative 1,3-propanediol dehydrogenase), and Halsa\_2285 (another putative 1,3-propanediol dehydrogenase). It was determined experimentally that *Halanaerobium hydrogeniformans* is capable of forming 1,3-propanediol from glycerol. After a 5-day incubation with 30 mM glycerol and pH 11 and 7% NaCl conditions, *Halanaerobium hydrogeniformans* was able to convert 31.5% of the glycerol to 1,3-propanediol. When B<sub>12</sub> was provided at concentrations from 25 to 100 µg/L, glycerol to 1,3-propanediol conversion ranged from 59.1 to 60.3% (Roush, 2013).

Glycine betaine is another osmoregulatory compound found in hypersaline environments (Welsh, 2000). *Halanaerobium*

<sup>2</sup>[http://www.genome.jp/kegg-bin/show\\_pathway?has00561](http://www.genome.jp/kegg-bin/show_pathway?has00561)



**FIGURE 3 | Portion of *Halanaerobium hydrogeniformans*' genome map indicating the genes in the same operon as the putative hydrogenase, Halsa\_1862.** Image generated from BioCyc (Caspi et al., 2010).

*hydrogeniformans* possesses an ATP-binding cassette (ABC) transporter, Halsa\_1783, that can possibly bring this compound into the cell. Not only can this compound be used as an osmoregulatory compound but can be a potential source of energy and carbon for the cell. Glycine betaine could possibly be used in the Stickland reaction with the amino acid, serine, as observed in *Halanaerobacter salinarius* (Mounté et al., 1999).

#### MOBILE DNA

*Halanaerobium hydrogeniformans*' genome was interrogated by using IMG to determine the most abundant COGs genes present. The most abundant COG genes in this genome were found to be transposases (Table 3). This should not come as a surprise as Aziz et al. (2010) found that transposases are both ubiquitous and abundant in both genomes and metagenome libraries. They determined the average number of transposases possessed across known genomes to be 38 per genome. *Halanaerobium hydrogeniformans* contains 72 annotated transposase genes (Table 3). In comparison, *Halanaerobium praevalens* GSL<sup>T</sup> was found to possess 20 annotated transposase genes. Transposase enzymes are responsible for the excision and movement of DNA segments within a chromosome. Transposase-encoding genes are flanked with insertion sequences (IS). These IS are short, inverted terminal repeats. Previously, it was thought that IS segments of DNA were selfish or parasitic (Orgel and Crick, 1980). However, it is now thought that transposable elements convey selective advantages to their hosts. These advantages can include the mobilization and/or activation of beneficial genes (Nowacki et al., 2009) or to generate phenotypic diversity (Brazelton and Baross, 2009). However, there are costs, such as transposon-induced mutations, that need to be balanced by the organisms (Aziz et al., 2010).

A further breakdown of the transposases in *Halanaerobium hydrogeniformans* reveals that eight IS families are present in this genome (Table 4). IS families are based upon similarities and differences in structure, organization, and the nucleotide and protein sequence relationships (Mahillon and Chandler, 1998). For example, the IS3 family is characterized by having lengths between 1,200 and 1,550 base pairs (bp) and inverted terminal repeats of 20 to 40 bp (Mahillon and Chandler, 1998). Interestingly, these

sequences generally have two consecutive and partially overlapping ORF, *orfA* and *orfB*. These mobile segments of DNA transposes through a circular intermediate. Of the IS families identified in *Halanaerobium hydrogeniformans*' genome, the only other IS family present that possesses more than one orf is IS21. The IS21 family has two orfs, a long upstream frame, *istA*, and a shorter downstream frame, *istB*. These two proteins carry several blocks of highly conserved residues (Mahillon and Chandler, 1998). Work is currently being done by Ron Frank, Missouri S&T, to determine if the putative transposases are active in *Halanaerobium hydrogeniformans*. If so, it is suspected that these genes are can become mobile and potentially activate beneficial genes to increase the fitness of this organism to tolerate environmental pressures (Aziz et al., 2010) that are present in Soap Lake.

#### CYCLIC-di-GMP

The second most numerous group of identified genes in the *Halanaerobium hydrogeniformans*' genome are the HD-GYP domain genes of COGs 2206 and 3437 (Table 3). In addition, there are eight genes identified as belonging in COG 2199 of the FOG: GGDEF domain. The GGDEF domain encodes for enzymes that produce cyclic-di-GMP, a ubiquitous second messenger in bacteria (Jenal and Malone, 2006). It is involved in cell signaling, exopolysaccharide formation, attachment, and biofilm production. The HD-GYP domain genes encode for diguanylate cyclase and metal dependent phosphohydrolase, an enzyme responsible for producing cyclic-di-GMP and it requires the presence of divalent cations, most likely Mg<sup>2+</sup> or Mn<sup>2+</sup> (Castiglione et al., 2011). Previous analysis performed indicates that both of these metals, Mg<sup>2+</sup> and Mn<sup>2+</sup>, 8,170.0 and 404.0 mg/kg dry weight, respectively, are present in the sediment of Soap Lake (Sigrid Penrod, personal communication). In comparing *Halanaerobium hydrogeniformans*' genome with *Halanaerobium praevalens* GSL<sup>T</sup>'s, only *Halanaerobium hydrogeniformans*' genome possesses genes for diguanylate cyclase with metal dependent phosphohydrolase. Thus far, only a few environmental signals have been identified that regulate cyclic di-GMP-mediated signaling pathways (Römling et al., 2013), and none are known for *Halanaerobium*. *Halanaerobium hydrogeniformans* forms mucous-like mats in cultures that are not vigorously shaken (Begemann et al., 2012). One possible

**Table 3 | Most abundant COG genes identified in *Halanaerobium hydrogeniformans*' genome.**

COG ID	COG Name	<i>Halanaerobium hydrogeniformans</i>	<i>Halanaerobium praevalens</i> GSL <sup>T</sup>
2801	Transposase and inactivated derivatives	22	6
2963	Transposase and inactivated derivatives	20	6
2206	HD-GYP domain	14	9
2826	Transposase and inactivated derivatives, IS30 family	12	2
0438	Glycosyltransferase	11	9
3328	Transposase and inactivated derivatives	10	0
1028	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	9	2
3437	Response regulator containing a CheY-like receiver domain and an HD-GYP domain	9	3
0675	Transposase and inactivated derivatives	8	6
0747	ABC-type dipeptide transport system, periplasmic component	7	6
1309	Transcriptional regulator	8	7
2199	FOG: GGDEF domain	8	9

*Halanaerobium praevalens* GSL<sup>T</sup> number of genes for each COG ID is also provided.

**Table 4 | Number of annotated genes in insertion sequence families present in *Halanaerobium hydrogeniformans*.**

Insertion sequence family	Number of genes present
IS3	22
IS30	11
IS256	8
IS200/IS605	7
IS4	3
IS6	3
IS21	2
IS91	1

role this set of putative genes may play is the formation of these mats.

### GLYCOSYLTRANSFERASES

There is evidence for the occurrence of glycosyltransferases, COG 0438 (**Table 3**). Nine of the 11 putative genes in *Halanaerobium hydrogeniformans* encode for glycosyl transferase group 1 enzymes. There is one putative sucrose-phosphate synthase (Halsa\_0772) and one hypothetical protein (Halsa\_0632). These enzymes are defined by the utilization of an activated donor sugar group substrate that contains a phosphate leaving group (Lairson et al., 2008). They are involved in the biosynthesis of cell walls, membranes, and envelop biogenesis. Specifically, these enzymes catalyze the first step in the sucrose synthesis pathway and are thought to play a role in osmotic stress protection (Chua et al., 2008). *Halanaerobium hydrogeniformans*' Halsa\_0772 gene has a 74% identity to a sucrose-phosphate synthase that is present in *Halanaerobium praevalens* GSL<sup>T</sup>, indicating a common mechanism for osmotic stress protection.

### SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDRs)

Nine putative genes in COG 1028 were found in *Halanaerobium hydrogeniformans*' genome. Only two were found in *Halanaerobium praevalens* GSL<sup>T</sup>'s genome. These genes encode for short-chain dehydrogenases/reductases (SDRs) with different specificities. This super family of enzymes catalyze a variety of NAD(P)(H) oxidation/reduction reactions (Kallberg et al., 2002). These enzymes are also recognized to catalyze the metabolism of steroids, cofactors, carbohydrates, lipids, aromatic compounds, and amino acids, and act in redox sensing. They are also associated with biotin metabolism and fatty acid biosynthesis and metabolism. There hasn't been much research performed on this family of enzymes in extremophilic bacteria. The research that has been focused on characterizing these enzymes from extremophilic organisms has been on thermophilic prokaryotes such as, *Thermus thermophiles* HB8 (Asada et al., 2009), *Sulfolobus acidocaldarius* (Pennacchio et al., 2010), and *Thermococcus sibiricus* (Stekhanova et al., 2010).

### ABC TRANSPORTERS

There are a number of ATP-binding cassettes (ABC) transporters represented in *Halanaerobium hydrogeniformans*' genome. Of these, seven COG 0747 putative genes have been identified, Halsa\_0302, Halsa\_0968, Halsa\_1628, Halsa\_1745, Halsa\_2053, Halsa\_2146, and Halsa\_2227 (**Table 3**). These genes encode for ABC-type nickel/dipeptide/oligopeptide periplasmic transport systems (Tam and Saier, 1993). Nickel is required for five types of enzymes; urease, hydrogenase, carbon monoxide dehydrogenase, methyl-S-coenzyme M reductase, and one class of superoxide dismutase (Hausinger, 1997). *Halanaerobium hydrogeniformans* does not appear to possess any of these enzymes. However, there are 524 genes that have been identified as hypothetical proteins and have no assigned functions. Thus far, only two possible hydrogenases, Halsa\_1768 and Halsa\_1862, have been identified. These are both Fe-only hydrogenases. It will be interesting to determine

the concentration of nickel that is required by the organism as well as to determine if there are nickel-requiring enzymes present.

The protein-coding genes that were connected to membrane transport KEGG pathways were explored through IMG. These genes can indicate what is needed and utilized by the bacterium. For example, there are numerous genes that encode for iron uptake proteins. Iron III can possibly be taken up by proteins encoded by *AfuA* (Halsa\_2074), *AfuB* (Halsa\_2073), and *AfuC* (Halsa\_2072). Siderophore-mediated transport of iron complexes are likely in this bacterium. These proteins can possibly be encoded by *FhuD* (Halsa\_2140, Halsa\_2186, Halsa\_2212, and Halsa\_2233), *FhuB* (Halsa\_1986, Halsa\_2185, Halsa\_2211, and Halsa\_2232), and *FhuC* (Halsa\_1985, Halsa\_2184, Halsa\_2210, and Halsa\_2231). *FhuD* is a periplasmic protein and *FhuB* and *FhuC* are cytoplasmic membrane-associated proteins responsible for siderophore-mediated iron transport (Katoh et al., 2001). It appears that *Halanaerobium hydrogeniformans* also possesses genes for proteins responsible to taking up another metal, tungstate. *TupA* (Halsa\_2175), *TupB* (Halsa\_2174), and *TupC* (Halsa\_2173) were each found to be present. These genes do not appear to be present in *Halanaerobium praevalens* GSL<sup>T</sup>. Zinc is another metal that is possibly taken up by *Halanaerobium hydrogeniformans*. *ZnuA* (Halsa\_0273), *ZnuB* (Halsa\_0275), and *ZnuC* (Halsa\_0274) were found in the bacterium's genome.

*Halanaerobium hydrogeniformans*' ability to utilize various carbon sources can be inferred by the transporters that it contains. Halsa\_1981 was identified as possibly being involved with uptake of glucose/mannose (*MalK*), maltose/maltodextrin (*MalK*), galactose oligomer/maltooligosaccharide (*MsmX*), arabinoooligosaccharide (*MsmX*), raffinose/stachyose/melibiose (*MsmK*), sorbitol/mannitol (*SmoK*),  $\alpha$ -glucoside (*AglK*), cellobiose (*MsiK*), and chitobiose (*MsiK*). In addition to Halsa\_1981, other genes are present that could encode for other carbon-intake ABC transporters. A total of 10 putative genes for ABC transporters for ribose/autoinducer 2/D-xylose, *RbsB*, *RbsC*, and *RbsA*, were identified. The genes, *UgpB*, *UgpA*, and *UgpE*, responsible for sn-glycerol 3-phosphate uptake were also found. Currently, the range of sources of carbon is unknown for *Halanaerobium hydrogeniformans*. Previous studies have demonstrated that the bacterium can use glucose, cellobiose, ribose, xylose, arabinose, galactose, and mannose (Begemann et al., 2012).

Glycerol can be used as either a carbon source or as an osmoprotectant (Oren, 1993). *Halanaerobium hydrogeniformans* possesses the genes, *OpuBB* and *OpuBA*, that are putative osmoprotectant ABC transport genes. In addition, it has putative trehalose/maltose ABC transport genes, *ThuE*, *ThuF*, and *ThuG*. Trehalose is considered a universal stress molecule and can serve as an osmoprotectant and in *Chromohalobacter salexigens*, it can serve to protect against temperature extremes (Reina-Bueno et al., 2012). However, trehalose was not confirmed to protect against desiccation. *Halanaerobium hydrogeniformans* does appear to have a mechanism to protect itself against desiccation. When grown with little or no agitation, it grows in an opaque mass (Begemann et al., 2012). It possesses an operon that contains a capsular exopolysaccharide family protein (Halsa\_0553), a lipopolysaccharide biosynthesis protein (Halsa\_0554), a polysaccharide export protein (Halsa\_0555), and a PHP domain-containing protein

(Halsa\_0556). Thus, *Halanaerobium hydrogeniformans* appears to be capable of protecting itself against osmotic and desiccation pressures.

### PHOSPHOTRANSFERASE SYSTEMS (PTSs)

In addition to the ABC transport systems, *Halanaerobium hydrogeniformans* has numerous PTSs to bring in sources of carbon. Glucose, maltose, arbutin/salicin, *N*-acetyl muramic acid, and trehalose can be brought into a cell with the Crr kinase protein (Halsa\_0150 and Halsa\_1861). *N*-acetyl-D-glucosamine can possibly be brought into the cell with NagE (Halsa\_0149). Proteins encoded by *CelA* (Halsa\_0141), *CelB* (Halsa\_0142), and *CelC* (Halsa\_0143), could bring cellobiose into the cell. Putative genes for mannitol (*MtlA*), sorbitol (*SrlA*, *SrlE*, *SrlB*), galactitol (*GatA*, *GatB*, *GatC*), and fructose (*FruA*) are also present.

*Halanaerobium hydrogeniformans* has two putative nitrogen-related PTS genes, Halsa\_0019 and Halsa\_2283. Nitrogen-related PTS genes are found in Gram-negative bacteria, can regulate carbon and nitrogen metabolism, are required for virulence by some bacteria, and can play a role in potassium homeostasis (Pflüger-Grau and Görke, 2010). Halsa\_0019 is likely to be involved with the regulation of fructose metabolism. Halsa\_0020 is a putative gene for *FruA*, a fructose PTS, and Halsa\_0018 is a putative 1-phosphofructokinase. In addition, when a BLAST search was performed on the amino acid sequence encoded by Halsa\_0019, a 79 and 76% identity was found with a fructose-specific PTS from *Halanaerobium saccharolyticum* and *Halanaerobium praevalens*, respectively. The role for Halsa\_2283 isn't as apparent as for Halsa\_0019. The gene in the same operon, Halsa\_2284, was not identified. In addition, when a BLAST search was performed on the amino acid sequence encoded by Halsa\_2283, only a 54% identity was found for a fructose-specific PTS from *Halanaerobium saccharolyticum*.

### OTHER TRANSPORT SYSTEMS

Being bacterial and not archaeal, one of the intriguing aspects of the Halanaerobiales order is that they use a "salting in" mechanism to protect themselves against osmotic shock (Detkova and Boltyanskaya, 2007). *Halanaerobium hydrogeniformans* possesses putative genes that possibly encode for TrkA-C domain containing proteins (Halsa\_0281, Halsa\_0709, and Halsa\_1061) and TrkA-N domain containing proteins (Halsa\_0737, Halsa\_1057, Halsa\_1056, Halsa\_1352, and Halsa\_1257). These genes are responsible for potassium ion transport into the cell. In addition, there are a number of putative symporters for the cell. These include a putative sodium/dicarboxylate symporter (Halsa\_0959), sodium/sulfate symporter (Halsa\_1097), and sodium/proline symporter (Halsa\_1726). It is interesting to note that these symporters would bring sodium into the cell. There are also putative Na<sup>+</sup>/H<sup>+</sup> antiporters present. These antiporters would remove sodium from the cell while bringing in protons and contributing to the pH homeostasis of the cell (Janto et al., 2011). Halsa\_0468, Halsa\_1158, Halsa\_1560, and Halsa\_2086 possibly encode for putative Na<sup>+</sup>/H<sup>+</sup> antiporter NhaC-like proteins. In addition, Halsa\_0689 and Halsa\_0691 possibly code for cation/proton antiporters. The gene that is present between these two, Halsa\_0690, is a putative multiple resistance and

pH regulation protein F gene. The two genes after Halsa\_0691, (Halsa\_0692 and Halsa\_0693) possibly encode for subunits of a multicomponent  $\text{Na}^+/\text{H}^+$  antiporter. Thus, many of these genes are likely involved with the maintenance of osmotic pressure and Halsa\_0690 might be involved with pH regulation of the cell.

Besides potassium and sodium, other cations need to be transported into the cell. There are three copies, (Halsa\_0666, Halsa\_1667, and Halsa\_2286) of a magnesium transporter for *Halanaerobium hydrogeniformans*. There is one putative gene for a cobalt transport protein (Halsa\_1890). Halsa\_1241 is a putative gene for a chromate transporter. Two putative zinc/iron permease genes are next to each other on the genome (Halsa\_2161 and Halsa\_2162). These cations, along with iron, would need to be taken up into the cell to serve as co-factors for enzymatic activity. Furthermore, one possible way that ammonium can enter the cell is through putative cation transporter Halsa\_1351.

Another aspect that needs to be balanced between the cell and its haloalkaline environment is the anions, especially chloride. For example, *Halobacillus halophilus*, a low G+C, Gram-positive, moderately halophilic bacterium, has an absolute requirement for chloride (Saum et al., 2013). *Halanaerobium hydrogeniformans* possesses a putative  $\text{Cl}^-$  channel voltage-gated family protein (Halsa\_0736) and an anion transporter (Halsa\_0628) that can possibly transport chloride into the cell and help to achieve an anionic balance.

## SUMMARY

*Halanaerobium hydrogeniformans* is a unique bacterium that is ideally adapted to its haloalkaliphilic lake environment. It is capable of utilizing a variety of carbon sources and appears to possess the cell membrane transport systems to bring them into the cell. Once inside the cell, there is a complete Embden-Meyerhof pathway of glycolysis. However, the Kreb's cycle is not complete. The organism relies on a number of fermentative metabolisms. It has been found to form acetate, formate, and hydrogen as fermentation products from simple sugars. It can also ferment glycerol, a widespread carbon source in saline environments. The bacterium also possesses transporters to bring in required metals and other ions. In addition to the metals required for enzymatic activity, the organism also possesses a variety of transporters that can bring in potassium and remove sodium to help to regulate the osmotic pressure. The  $\text{Na}^+/\text{H}^+$  antiporters are important for both maintaining osmotic pressure and the pH of the cell. The organism also possesses a number of transposases. The transposases enable the organism to mobilize genes and affect gene regulation.

*Halanaerobium hydrogeniformans* has a number of similarities to *Halanaerobium praevalens*  $\text{GSL}^\text{T}$ . Both organisms do not appear to possess formate-hydrogen lyase while they do appear to possess glycosyl transferases and fructose-specific phosphotransferase. On the other hand, the two organisms have a number of differences that are likely related to the environments, hypersaline vs. haloalkaline, where they were isolated from. *Halanaerobium praevalens*  $\text{GSL}^\text{T}$  possesses fewer genes for metabolism, such as the genes required for amino acid metabolism and cofactor and vitamin production. *Halanaerobium praevalens*  $\text{GSL}^\text{T}$  does not possess diguanylate cyclase with metal dependent phosphohydrolase genes

or many of the metal-uptake proteins that *Halanaerobium hydrogeniformans* possesses. Furthermore, *Halanaerobium praevalens*  $\text{GSL}^\text{T}$  possesses less than a third of the number of transposase genes that *H. hydrogeniformans* does. The presence of these genes in *Halanaerobium hydrogeniformans* likely enables the organism to better tolerate the alkaline conditions, in addition to the saline conditions, and the metal content present in the sediments of Soap Lake. Furthermore, the transposases could provide genetic diversity that can lead to adaptive advantages for *Halanaerobium hydrogeniformans*.

## ACKNOWLEDGMENTS

These sequence data were produced by the US Department of Energy Joint Genome Institute <http://www.jgi.doe.gov/> in collaboration with the user community. I thank my former undergraduate students, Jill Wildhaber and Sarah Rommelfanger, and my former graduate student, Daniel Roush, who helped to look through the genome data with me.

## REFERENCES

- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Anderson, G. C. (1958). Seasonal characteristics of two saline lakes in Washington. *Limnol. Oceanogr.* 3, 51–68. doi: 10.4319/lo.1958.3.1.0051
- Asada, Y., Endo, S., Inoue, Y., Mamiya, H., Hara, A., Kunishima, N., et al. (2009). Biochemical and structural characterization of a short-chain dehydrogenase/reductase of *Thermus thermophiles* HB8 a hyperthermostable aldose-1-dehydrogenase with broad substrate specificity. *Chem. Biol. Interact.* 178, 117–126. doi: 10.1016/j.cbi.2008.09.018
- Asao, M., Takaichi, S., and Madigan, M. T. (2007). *Thiocapsa inhoffii*, sp. nov., an alkaliphilic purple sulfur bacterium of the family Chromatiaceae from Soap Lake, Washington (USA). *Arch. Microbiol.* 188, 665–675. doi: 10.1007/s00203-007-0287-9
- Asao, M., Takaichi, S., and Madigan, M. T. (2012). Amino acid-assimilating phototrophic heliobacteria from soda lake environments: *Heliolestis acidaminovorans* sp. nov. and 'Candidatus *Heliomonas lunata*'. *Extremophiles* 16, 585–595. doi: 10.1007/s00792-012-0458-8
- Aziz, R. K., Breitbart, M., and Edwards, R. A. (2010). Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res.* 38, 4207–4217. doi: 10.1093/nar/gkq140
- Bardavid, R. E., Khristo, P., and Oren, A. (2008). Interrelationships between *Dunaliella* and halophilic prokaryotes in saltern crystallizer ponds. *Extremophiles* 12, 5–14. doi: 10.1007/s00792-006-0053-y
- Begemann, M. B., Mormile, M. R., Sitton, O. C., Wall, J. D., and Elias, D. A. (2012). A streamlined strategy for biohydrogen production with *Halanaerobium hydrogeniformans*, an alkiliphilic bacterium. *Front. Microbiol.* 3:93. doi: 10.3389/fmicb.2012.00093
- Borowitzka, L. J. (1981). The microflora. Adaptation to life in extremely saline lakes. *Hydrobiologia* 81, 33–46. doi: 10.1007/BF00048704
- Brazelton, W. J., and Baross, J. A. (2009). Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. *ISME J.* 3, 1420–1424. doi: 10.1038/ismej.2009.79
- Brown, S. D., Begemann, M. B., Mormile, M. R., Wall, J. D., Han, C. S., Goodwin, L. A., et al. (2011). Complete genome sequence of the haloalkaliphilic, hydrogen-producing bacterium *Halanaerobium hydrogeniformans*. *J. Bacteriol.* 193, 3682–3683. doi: 10.1128/JB.05209-11
- Carere, C. R., Rydzak, T., Verbeke, T. J., Cicek, N., Levine, D. B., and Sparling, R. (2012). Linking genome content to biofuel production yields: a meta-analysis of major catabolic pathways among select H2 and ethanol-producing bacteria. *BMC Microbiol.* 12:295. doi: 10.1186/1471-2180-12-295
- Caspi, R., Altman, T., Dale, J. M., Dreher, K., Fulcher, C. A., Gilham, F., et al. (2010). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 38, D473–D479. doi: 10.1093/nar/gkp875

- Castiglione, N., Stelitano, V., Rinaldo, S., Giardina, G., Caruso, M., and Cutruzzolà, F. (2011). Metabolism of cyclic-di-GMP in bacterial biofilms: from a general overview to biotechnological applications. *Indian J. Biotechnol.* 10, 423–431.
- Cayol, J. L., Fardeau, M. L., Garcia, J. L., and Ollivier, B. (2002). Evidence of inter-species hydrogen transfer from glycerol in saline environments. *Extremophiles* 6, 131–134. doi: 10.1007/s007920100229
- Chadwick, L. J., and Irgens, R. L. (1991). Hydrogen gas production by an *Ectothiorhodospira vacuolata* strain. *Appl. Environ. Microbiol.* 57, 594–596.
- Chua, T. K., Bujnicki, J. M., Tan, T.-C., Huynh, E., Patel, B. K., and Sivaraman, J. (2008). The structure of sucrose phosphate synthase from *Halothermothrix orenii* reveals its mechanism of action and binding mode. *Plant Cell* 20, 1059–1072. doi: 10.1105/tpc.107.051193
- Detkova, E. N., and Boltyanskaya, Y. V. (2007). Osmoadaptation of haloalkaliphilic bacteria: role of osmoregulators and their possible practical application. *Mikrobiologija* 76, 511–522. doi: 10.1134/S0026261707050013
- Dimitriu, P. A., Pinkart, H. C., Peyton, B. M., and Mormile, M. R. (2008). Spatial and temporal patterns in the microbial diversity of a meromictic soda lake in Washington State. *Appl. Environ. Microbiol.* 74, 4877–4888. doi: 10.1128/AEM.00455-08
- Dimitriu, P. A., Shukla, S. K., Conradt, J., Márquez, C., Ventosa, A., Maglia, A., et al. (2005). *Nitrincola lacisaponensis* gen. nov., sp. nov., a novel alkaliphilic bacterium isolated from an alkaline, saline lake. *Int. J. Syst. Evol. Microbiol.* 55, 2273–2278. doi: 10.1099/ijss.0.63647-0
- Eder, W., Jahneke, L. L., Schmidt, M., and Huber, R. (2001). Microbial diversity of the brine-seawater interface of the Kebrit Deep, Red Sea, studied via 16S rRNA gene sequences and cultivation methods. *Appl. Environ. Microbiol.* 67, 3077–3085. doi: 10.1128/AEM.67.7.3077-3085.2001
- Edmondson, W. T., and Anderson, G. C. (1965). Some features of saline lakes in central Washington. *Limnol. Oceanogr.* 10, 87–96.
- Foti, M., Sorokin, D. Y., Lomans, B., Mussman, M., Zacharova, E. E., Pimenov, N. V., et al. (2007). Diversity, activity, and abundance of sulfate-reducing bacteria in saline and hypersaline soda lakes. *Appl. Environ. Microbiol.* 73, 2093–2100. doi: 10.1128/AEM.02622-06
- Hausinger, R. P. (1997). Metallocenter assembly in nickel-containing enzymes. *J. Biol. Inorg. Chem.* 2, 279–286. doi: 10.1007/s007750050133
- Ivanova, N., Sikorski, J., Chertkov, O., Nolan, M., Lucas, S., Hammon, N., et al. (2011). Complete genome sequence of the extremely halophilic *Halanaerobium praevalens* type strain (GSLT). *Stand. Genomic Sci.* 4, 312–321. doi: 10.4056/sigs.1824509
- Janto, B., Ahmed, A., Ito, M., Liu, J., Hicks, D. B., Pagni, S., et al. (2011). Genome of alkaliphilic *Bacillus pseudofirmus* OF4 reveals adaptations that support the ability to grow in an external pH range from 7.5 to 11.4. *Environ. Microbiol.* 13, 3289–3309. doi: 10.1111/j.1462-2920.2011.02591.x
- Jenal, U., and Malone, J. (2006). Mechanisms of cyclic-di-gmp signaling in bacteria. *Annu. Rev. Genet.* 40, 385–407. doi: 10.1146/annurev.genet.40.110405.090423
- Jones, B. E., Grant, W. D., Duckworth, A. W., and Owenson, G. G. (1998). Microbial diversity of soda lakes. *Extremophiles* 2, 191–200. doi: 10.1007/s007920050060
- Kallberg, Y., Oppermann, U., Jörnvall, H., and Persson, B. (2002). Short-chained dehydrogenases/reductases (SDRs): coenzyme-based functional assignments in completed genomes. *Eur. J. Biochem.* 269, 4409–4417. doi: 10.1046/j.1432-1033.2002.03130.x
- Katoh, H., Hagino, N., Grossman, A. R., and Ogawa, T. (2001). Genes essential to iron transport in the cyanobacterium *Synechocystis* sp. Strain PCC 6803. *J. Bacteriol.* 183, 2779–2784. doi: 10.1128/JB.183.9.2779-2784.2001
- Kobayashi, T., Kimura, B., and Fujii, T. (2000). Strictly anaerobic halophiles isolated from canned Swedish fermented herrings (Surströmming). *Int. J. Food Microbiol.* 54, 81–89. doi: 10.1016/S0168-1605(99)00172-5
- Krulwich, T. A. (1995). Alkaliphiles: ‘basic’ molecular problems of pH tolerance and bioenergetics. *Mol. Microbiol.* 15, 403–410. doi: 10.1111/j.1365-2958.1995.tb02253.x
- Krulwich, T. A., Sachs, G., and Padan, E. (2011). Molecular aspects of bacterial pH sensing and homeostasis. *Nat. Rev. Microbiol.* 9, 330–343. doi: 10.1038/nrmicro2549
- Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008). Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* 77, 521–555. doi: 10.1146/annurev.biochem.76.061005.092322
- Mahillon, J., and Chandler, M. (1998). Insertion sequences. *Microbiol. Mol. Biol. Rev.* 62, 725–774.
- Markowitz, V. M., Chen, I.-M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., et al. (2012). IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* 40, D115–D122. doi: 10.1093/nar/gkr1044
- Mormile, M. R., Romine, M. F., Garcia, M. T., Ventosa, A., Bailey, T. J., and Peyton, B. M. (1999). *Halomonas campialis* sp. nov., a denitrifying, moderately haloalkaliphilic bacterium. *Syst. Appl. Microbiol.* 22, 551–558. doi: 10.1016/S0723-2020(99)80008-3
- Mounté, S., Manac'h, N., Hirschler, A., Caumette, P., Willison, J. C., and Matheron, R. (1999). *Haloanaerobacter salinarus* sp. nov., a novel halophilic fermentative bacterium that reduces glycine-betaine to trimethylamine with hydrogen or serine as electron donors: emendation of the genus *Haloanaerobacter*. *Int. J. Syst. Bacteriol.* 49, 103–112. doi: 10.1099/00207713-49-1-103
- Nowacki, M., Higgins, B. P., Maquilan, G. M., Swart, E. C., Doak, T. G., and Landweber, L. F. (2009). A functional role for transposases in a large eukaryotic genome. *Science* 324, 935–938. doi: 10.1126/science.1170023
- Oren, A. (1993). Availability, uptake and turnover of glycerol in hypersaline environments. *FEMS Environ. Microbiol.* 12, 15–23. doi: 10.1111/j.1574-6941.1993.tb00012.x
- Oren, A. (1999). Bioenergetic aspects of halophilism. *Microbiol. Mol. Biol. Rev.* 63, 334–348.
- Oren, A. (2008). Interrelationships between *Dunaliella* and halophilic prokaryotes in saltern crystallizer ponds. *Extremophiles* 12, 5–14. doi: 10.1007/s00792-006-0053-y
- Oren, A. (2010). Thermodynamic limits to microbial life at high salt concentrations. *Environ. Microbiol.* 13, 1908–1923. doi: 10.1111/j.1462-2920.2010.02365.x
- Orgel, L. E., and Crick, F. H. (1980). Selfish DNA: the ultimate parasite. *Nature* 284, 604–607. doi: 10.1038/284604a0
- Paul, V. G., Minteer, S. D., Treu, B. L., and Mormile, M. R. (2014). Ability of a haloalkaliphilic bacterium isolated from Soap Lake, Washington to generate electricity at pH 11.0 and 7% salinity. *Environ. Technol.* 35, 1003–1011. doi: 10.1080/09593330.2013.858186
- Pennacchio, A., Giordano, A., Pucci, B., Rossi, M., and Raia, C. A. (2010). Biochemical characterization of a recombinant short-chain NAD(H)-dependent dehydrogenase/reductase from *Sulfolobus acidocaldarius*. *Extremophiles* 14, 193–204. doi: 10.1007/s00792-009-0298-3
- Peyton, B. M., and Yonge, D. R. (2002). *Biodegradation of Non-Point Source Pollutants in Soap Lake, Washington*. Project Completion Report. State of Washington Water Research Report WRR-11. Pullman, WA: State of Washington Water Research Center.
- Pflüger-Grau, K., and Görke, B. (2010). Regulatory roles of the bacterial nitrogen-related phosphotransferase system. *Trends Microbiol.* 18, 205–214. doi: 10.1016/j.tim.2010.02.003
- Pollock, J., Weber, K. A., Lack, J., Achenbach, L. A., Mormile, M. R., and Coates, J. D. (2007). Alkaline iron (III) reduction by a novel alkaliphilic, halotolerant, *Bacillus* sp. isolated from salt flat sediments of Soap Lake. *Appl. Microbiol. Biotechnol.* 77, 927–934. doi: 10.1007/s00253-007-1220-5
- Reina-Bueno, M., Argandoña, M., Salvador, M., Rodríguez-Moya, J., Iglesias-Guerra, F., Csonka, L. N., et al. (2012). Role of trehalose in salinity and temperature tolerance in the model halophilic bacterium, *Chromohalobacter salexigens*. *PLoS ONE* 7:e33587. doi: 10.1371/journal.pone.0033587
- Römling, U., Galperin, M. Y., and Gomelsky, M. (2013). Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol. Mol. Biol. Rev.* 77, 1–52. doi: 10.1128/MMBR.00043-12
- Roush, D. W. (2013). *Production of 1,3-propanediol from glycerol under haloalkaline conditions by Halanaerobium hydrogeniformans*. Ph.D. thesis, Missouri University of Science and Technology, Rolla, MO.
- Roush, D. W., Elias, D. A., and Mormile, M. R. (2014). Metabolic capabilities of the members of the order Halanaerobiales and their potential biotechnological applications. *Curr. Biotechnol.* 3, 3–9. doi: 10.2174/221155010301140303102741
- Saum, S. H., Pfeiffer, F., Palm, P., Rampp, M., Schuster, S. C., Müller, V., et al. (2013). Chloride and organic osmolytes: a hybrid strategy to cope with elevated salinities by the moderately halophilic, chloride-dependent bacterium

- Halobacillus halophilus*. *Environ. Microbiol.* 15, 1619–1633. doi: 10.1111/j.1462-2920.2012.02770.x
- Sorokin, D., Foti, M., Pinkart, H. C., and Muyzer, G. (2007). Sulfur-oxidizing bacteria in Soap Lake (Washington State), a meromictic, haloalkaline lake with an unprecedented high sulfide content. *Appl. Environ. Microbiol.* 73, 451–455. doi: 10.1128/AEM.02087-06
- Stekhanova, T. N., Mardanov, A. V., Bezsdunova, E. Y., Gumerov, V. M., Ravin, N. V., Skryabin, K. G., et al. (2010). Characterization of a thermostable short-chain alcohol dehydrogenase from the hyperthermophilic archaeon *Thermococcus sibiricus*. *Appl. Environ. Microbiol.* 76, 4096–4098. doi: 10.1128/AEM.02797-09
- Tam, R., and Saier, M. H. (1993). Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol. Rev.* 57, 320–346.
- Thompson, J. C., and He, B. B. (2006). Characterization of crude glycerol from biodiesel production from multiple feedstocks. *Appl. Eng. Agric.* 22, 261–265. doi: 10.13031/2013.20272
- Welsh, D. T. (2000). Ecological significance of compatible solute accumulation by micro-organisms: from single cells to global climate. *FEMS Microbiol. Rev.* 24, 263–290. doi: 10.1111/j.1574-6976.2000.tb00542.x
- Zeikus, J. G., Hegge, P. W., Thompson, T. E., Phelps, T. J., and Langworthy, T. A. (1983). Isolation and description of *Halanaerobium praevalens* gen. nov. and sp. nov., an obligately anaerobic halophile common to Great Lake sediments. *Curr. Microbiol.* 9, 225–234. doi: 10.1007/BF01567586
- Zeng, A.-P. (1996). Pathway and kinetic analysis of 1,3-propanediol production from glycerol fermentation by *Clostridium butyricum*. *Bioprocess Eng.* 14, 169–175. doi: 10.1007/BF01464731
- Zhao, B., and Chen, S. (2012). *Alkalitalea saponilacus* gen. nov., sp. nov., an obligately anaerobic, alkaliphilic, xylanolytic bacterium from a meromictic soda lake. *Int. J. Syst. Evol. Microbiol.* 62, 2618–2623. doi: 10.1099/ijs.0.038315-0

**Conflict of Interest Statement:** The author holds two patents on biohydrogen production by *Halanaerobium hydrogeniformans*. Elias, Mormile, Begemann, and Wall. “A combined fossil fuel free process of lignocellulosic pretreatment with biological production”, U.S. Patent No. US 8,148,133, Issued: April 3, 2012. Elias, Mormile, Begemann, and Wall. “Fossil Fuel-Free Process of Lignocellulosic Pretreatment with Biological Hydrogen Production”, U.S. Patent No. US 8,034,592 B2, Issued October 11, 2011.

Received: 19 August 2014; paper pending published: 10 October 2014; accepted: 03 November 2014; published online: 19 November 2014.

Citation: Mormile MR (2014) Going from microbial ecology to genome data and back: studies on a haloalkaliphilic bacterium isolated from Soap Lake, Washington State. *Front. Microbiol.* 5:628. doi: 10.3389/fmicb.2014.00628

This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.

Copyright © 2014 Mormile. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Recent studies in microbial degradation of petroleum hydrocarbons in hypersaline environments

Babu Z. Fathepure \*

Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK, USA

**Edited by:**

Antonio Ventosa, University of Sevilla, Spain

**Reviewed by:**

Ronald Orem, United States Geological Survey, USA

Marco J. L. Coolen, Woods Hole Oceanographic Institution, USA

**\*Correspondence:**

Babu Z. Fathepure, Department of Microbiology and Molecular Genetics, Oklahoma State University, 307 Life Sciences East, Stillwater, OK, USA

e-mail: babu.fathepure@okstate.edu

Many hypersaline environments are often contaminated with petroleum compounds. Among these, oil and natural gas production sites all over the world and hundreds of kilometers of coastlines in the more arid regions of Gulf countries are of major concern due to the extent and magnitude of contamination. Because conventional microbiological processes do not function well at elevated salinities, bioremediation of hypersaline environments can only be accomplished using high salt-tolerant microorganisms capable of degrading petroleum compounds. In the last two decades, there have been many reports on the biodegradation of hydrocarbons in moderate to high salinity environments. Numerous microorganisms belonging to the domain Bacteria and Archaea have been isolated and their phylogeny and metabolic capacity to degrade a variety of aliphatic and aromatic hydrocarbons in varying salinities have been demonstrated. This article focuses on our growing understanding of bacteria and archaea responsible for the degradation of hydrocarbons under aerobic conditions in moderate to high salinity conditions. Even though organisms belonging to various genera have been shown to degrade hydrocarbons, members of the genera *Halomonas Alcanivorax*, *Marinobacter*, *Haloferax*, *Haloarcula*, and *Halobacterium* dominate the published literature. Despite rapid advances in understanding microbial taxa that degrade hydrocarbons under aerobic conditions, not much is known about organisms that carry out similar processes in anaerobic conditions. Also, information on molecular mechanisms and pathways of hydrocarbon degradation in high salinity is scarce and only recently there have been a few reports describing genes, enzymes and breakdown steps for some hydrocarbons. These limited studies have clearly revealed that degradation of oxygenated and non-oxygenated hydrocarbons by halophilic and halotolerant microorganisms occur by pathways similar to those found in non-halophiles.

**Keywords:** hypersaline environments, biodegradation, oxygenated and non-oxygenated hydrocarbons, halophilic and halotolerant bacteria and archaea, molecular mechanism of degradation

## BACKGROUND

Many hypersaline environments including natural saline lakes, salt flats, saline industrial effluents, oil fields, and salt marshes are often contaminated with high levels of petroleum hydrocarbons. These systems have considerable economic, ecological and scientific value. Among the contaminated hypersaline environments, oilfields pose a special problem due to their sheer numbers all over the world and due their high salinity caused by salty brackish water (produced water) generated during oil and natural gas extraction. Produced waters are by far the largest volume byproduct or waste associated with oil and gas production. For every barrel of oil produced, roughly 10 barrels of produced waters are generated. In the United States about 20–30 billion barrels of produced waters are generated each year and the worldwide estimate is about 70 billion barrels per year (Veil et al., 2004). The primary constituents in produced water that limit its disposal or reuse are high levels of salt (1000–250,000 mg/L), oil and grease, various toxic chemicals, heavy metals, and naturally occurring radioactive materials (Veil et al., 2004; Cuadros-Orellana et al., 2006; Bonfá et al., 2011).

Remediation of produced water is costly to oil and gas producers and inappropriate management can lead to environmental problems. Presently, >95% of all produced waters are re-injected, however prior to 1965–1970 most of the produced water waste was released to the surface. Even now many small-to moderate-sized operators continue to release substantial quantities of produced waters to the surface and shallow subsurface because of leaky tanks and flow-lines and due to accidents and vandalism. Sabkhas or coastal salt marshes are ubiquitous features in arid and semi-arid regions of the world (Arabian Peninsula, Central Asia, and Australia). These habitats are characterized by high salinity and extensive crude oil contamination (Fowler et al., 1993; Al-Mueini et al., 2007; Al-Mailem et al., 2013). Understanding the fate of petroleum compounds in such environmentally and economically sensitive habitats is important.

Bioremediation technology utilizes microorganisms to degrade toxic pollutants to harmless products such as CO<sub>2</sub>, H<sub>2</sub>O, and other inorganic compounds and these processes are environmentally safe and cost efficient (Philip et al., 2005). It has been reported that roughly 25% of all petroleum-contaminated land is being bioremediated using natural attenuation processes

thus underscoring the importance of microorganisms in remediation strategies (Holden et al., 2002). However, application of microbial technologies for treating contaminated high salinity or fluctuating salinity environment is limited due to the detrimental effects of salt on microbial life including disruption cell membrane, denaturation of enzymes, low solubility of oxygen, low solubility of hydrocarbons, and desiccation (Pernetti and Di Palma, 2005). Therefore, bioremediation of saline environments without costly dilution of salt-laden soil and water requires halophilic or halotolerant organisms that tolerate high salt concentrations. Halophiles are classified into three groups according to their optimal salt concentration for growth: slightly halophilic (1–3% w/v), moderately halophilic (3–15% w/v), and extremely halophilic (15–32% w/v) (Kushner, 1978; Ventosa and Nieto, 1995; Oren, 2013).

## DEGRADATION OF HYDROCARBONS IN HYPERSALINE ENVIRONMENTS

Petroleum is a complex mixture of different hydrocarbons including aliphatic (linear or branched), cycloalkanes, mono- and polycyclic aromatics, asphaltenes and resins and majority of these compounds are stable, toxic, and carcinogenic (Philip et al., 2005; Yemashova et al., 2007). Hydrocarbons differ in their susceptibility to microbial attack and generally degrade in the following order of decreasing susceptibility: n-alkanes > branched alkanes > low molecular weight aromatics > cyclic alkanes, > polycyclic aromatic hydrocarbons (Leahy and Colwell, 1990). Although many of these compounds can be relatively easily degraded under soil and freshwater environments (Van Hamme et al., 2003; Cao et al., 2009) and low salinity marine habitats (Harayama et al., 1999; Head and Swannell, 1999; Head et al., 2006; McGenity et al., 2012), little is known about their fate in moderate to high salinity conditions (3–30% salt). In 1992 Oren (Oren et al., 1992) provided an overview of the degradation of aromatic and aliphatic hydrocarbons in saline habitats and our understanding of metabolic capabilities of halophilic and halotolerant organisms has substantially advanced since this publication (Patzelt, 2005). For example, recent excellent reviews by Le Borgne et al. (2008), Martins and Peixoto (2012), McGenity (2010), and Patzelt (2005) attest to our improved understanding of the hydrocarbon biodegradation by halophilic and halotolerant microorganisms. Nonetheless, our knowledge on biochemistry, genetics, and pathways of hydrocarbon degradation in halophiles and halotolerants is sparse. Such information is crucial for designing novel and more efficient technologies for the remediation of contaminated high salinity environments and for understanding the carbon cycle in such extreme habitats. The goal of this review is to provide an overview of our current knowledge of the biodegradation of non-oxygenated and oxygenated hydrocarbons by bacteria and archaea in wide ranging salinities (6–30% NaCl) and to highlight recent discoveries in molecular mechanisms of degradation by halophilic and halotolerant organisms.

## CRUDE OIL

Crude oil is a mixture of hydrocarbons composed of mainly oxygenated and non-oxygenated hydrocarbons (Yemashova et al.,

2007). To date many studies have reported the ability of microorganisms to utilize crude oil components as the growth substrates in moderate to high salinity environments (Table 1). Diaz et al. (2000) have enriched microbial consortia, MPD-7 and MPD-M from Cormorant oil fields in North Sea and sediments associated with mangrove roots, respectively. These cultures degraded aliphatic and aromatic hydrocarbons in crude oil at salinity ranging from 3.5 to 10% NaCl. Total oil degradation by MPD-7 ranged from 20 to 38%, while MPD-M degraded much higher amount of crude oil ranging between 45 and 48%. In a subsequent study, Diaz et al. (2002) have immobilized the MPD-M culture on polypropylene fibers and showed that the culture was able to degrade crude oil at much higher salinity up to 18% NaCl. Riis et al. (2003) were able to show the degradation of diesel fuel in the presence of salt up to 17.5% by microbial communities extracted from Argentinean saline soils. In addition, these

**Table 1 | Biodegradation of crude oil under moderate to high salinity environment.**

Hydro-carbon	Degrader	Salinity (%)	References
Crude oil	<i>Streptomyces albiaxialis</i>	3–30	Kuznetsov et al., 1992
	Enrichment culture, brines of the Kalamkass oil fields, Kazakhstan	10–25	Zvyagintseva et al., 1995
	<i>Marinobacter aquaeolei</i>	0–20	Huu et al., 1999
	Bacterial consortia MPD-7	3.5–10	Diaz et al., 2000
	<i>Rhodococcus erythropolis</i>	0–10	Zvyagintseva et al., 2001
	<i>Dietzia maris</i>		
	MPD-M culture immobilized on polypropylene fibers	0–18	Diaz et al., 2002
	<i>Fusarium lateritium</i>	5–10	Obuekwe et al., 2005
	<i>Drechslera</i> sp.		
	Microbial community, Argentinean soil	>17	Riis et al., 2003
	<i>Cellulomonas</i> sp. <i>Bacillus</i> sp.	>17	Riis et al., 2003
	<i>Dietzia</i> sp. <i>Halomonas</i> sp.		
	<i>Rhodococcus</i> sp. <i>Gordonia</i> sp.	15	Borzenkov et al., 2006
	<i>Dietzia</i> sp. <i>Pseudomonas</i> sp.		
	<i>Actinopolyspora</i> sp. DPD1	20	Al-Mueini et al., 2007
	<i>Bacillus</i> sp. strain DHT	0–10	Kumar et al., 2007
	<i>Halomonas shengliensis</i>	0–15	Wang et al., 2007
	Strain C5	12	Chamkha et al., 2008
	<i>Halomonas</i> sp. C2SS100	0–10	Mnif et al., 2009, 2011
	<i>Pseudomonas</i> sp. C450R		
	<i>Haloferax</i> sp. <i>Halobacterium</i> sp.	>26	Al-Mailem et al., 2010
	<i>Halococcus</i> sp.		
	<i>Amycolicicoccus subflavus</i> DQS3-9A1	1–12	Wang et al., 2010
	<i>Marinobacter sedimentalis</i> , <i>Marinobacter falvimaris</i>	6	Al-Mailem et al., 2013

investigators isolated several halotolerant bacteria of the genera *Cellulomonas*, *Bacillus*, *Dietzia*, and *Halomonas* with the ability to degrade crude oil as the carbon source. Similarly, many other investigators have isolated pure cultures including *Halomonas shengliensis* (Wang et al., 2007), *Halomonas* sp. strain C2SS100 (Mnif et al., 2009), *Marinobacter aquaeolei* (Huu et al., 1999), *Streptomyces albiaxialis* (Kuznetsov et al., 1992), *Rhodococcus erythropolis*, and *Dietzia maris* (Zvyagintseva et al., 2001) from oilfields, production water, and other saline environments that degrade crude oil as the source of carbon in the presence of 0–30% salt. Borzenkov et al. (2006) reported the isolation of several strains of hydrocarbon-oxidizing bacteria representing the genera *Rhodococcus*, *Gordonia*, *Dietzia*, and *Pseudomonas* from oil and stratal waters of Tatarstan, western Siberia, and Vietnam oil-fields. All these strains oxidized *n*-alkane fraction of crude oil in a medium containing 15% NaCl. A *Bacillus* sp. strain DHT, isolated from oil contaminated soil, grew and produced biosurfactant when cultured in the presence of variety of hydrocarbons including crude oil, diesel oil, hexadecane, naphthalene, pyrene, dibenzothiophene, salicylate, catechol, and phenanthrene as the sole sources of carbon in the presence of 0–10% salinity and at 30–45°C. However, no growth occurred on toluene, phenol, 2-hydroxyquinoline and carbazole (Kumar et al., 2007). Similarly, Mnif et al. (2011) have reported the isolation of several strains of thermophilic and mesophilic hydrocarbon degrading as well as biosurfactant producing organisms from Tunisian oil fields. Among these, *Pseudomonas* sp. strain C450R and *Halomonas* sp. strain C2SS100 could degrade 93–96% of the aliphatic fraction of crude oil ( $C_{13}$ – $C_{29}$ ), while producing biosurfactants in the presence of 5–10% NaCl. Such organisms could play important role in the degradation of poorly soluble high molecular weight hydrocarbons in crude oil. Chamkha et al. (2008) have isolated a strain C5 closely related to *Geobacillus pallidus* from a tyrosol degrading enrichment developed from production water from a high-temperature oil field in Tunisia. The organism degraded crude oil and diesel as the source of carbon in the presence of 0–12% NaCl. Wang et al. (2010) have isolated a moderate halophilic actinomycete, *Amycolicoccus subflavus* DQS3-9A1<sup>T</sup> from oily sludge at Daqing Oilfield, China with the ability to degrade crude oil in the presence of 1–12% NaCl. Later, Nie et al. (2013) studied the genetic capability of the DQS3-9A1<sup>T</sup> to metabolize a range of short-chain and long-chain *n*-alkanes such as propane and  $C_{10}$ – $C_{36}$  alkanes, respectively, as the sole carbon sources in the presence of 1–12% NaCl. Recently, Al-Mailem et al. (2013) have isolated *Marinobacter sedimentalis* and *Marinobacter falvimaris* from soil and pond water collected from hypersaline sabkhas (18–20% salinity) in Kuwait. Isolation of these organisms was accomplished using agar plates provided with crude oil vapor as the sole source of carbon and 6% NaCl. These studies also showed that both organisms were capable of fixing atmospheric nitrogen and such potential is beneficial for effective bioremediation of petroleum compounds at high salinity without the need of providing fertilizer.

Studies also have reported archaeal ability to degrade crude oil in hypersaline environments. Zvyagintseva et al. (1995) have reported that a significant amount of isoprenoid and *n*-alkane fractions of crude oil was degraded in the presence of 10–25% of

salt by an enrichment developed from the brines of the Kalamkass oil fields in Kazakhstan. Al-Mailem et al. (2010) have isolated extremely halophilic archaeal strains of *Haloferax*, *Halobacterium*, and *Halococcus* from a hypersaline coastal area of the Arabian Gulf in a mineral salt medium with crude oil vapor as the source of carbon in the presence of >26% NaCl and at 40–45°C. These organisms also metabolized various aliphatic and aromatic hydrocarbons as the sole sources of carbon and energy at high salinity. Undoubtedly such properties are important for the bioremediation of crude oil-impacted high salinity arid sites. In a subsequent study by some of the same authors, the impact of adding organic fertilizer (casamino acid) and illumination (light/dark) on the bioremediation of crude oil was assessed using hypersaline soil (>22% salinity) and pond water (>16% salinity) collected from a supertidal sabkha at Al-Khiran, Kuwait. Results showed a significantly increased biodegradation of crude oil in the presence of casamino acid and when incubated under continuous illumination (Al-Mailem et al., 2012). The data suggested that the observed increased degradation was mainly due to archaeal members, with little or no contribution from bacteria. The authors theorize that hypersaline environments suffer from the lack of oxygen due low solubility and archaea in such environments would use the red pigment-mediated ATP synthesis perhaps analogous to bacteriorhodopsin-like system to meet the shortage of ATP produced via oxidative phosphorylation caused by low oxygen tension. This strategy would allow archaea to utilize the available limited oxygen to initiate degradation of hydrocarbons in high salinity conditions. In addition, the authors contend that casamino acid could have been used as the source of amino acids resulting in better growth and degradation. In conclusion, the enhanced hydrocarbon degradation in the presence of light is an interesting observation and warrants further investigation into why archaea dominate hypersaline environments. In return, such knowledge could be helpful to develop strategies to enhance hydrocarbon degradation in high salinity environments.

Only few studies are available on the fungal ability to degrade hydrocarbons in high salinity environments. Obuekwe et al. (2005) are the first to report the isolation *Fusarium lateritium*, *Drechslera* sp., and *Papulaspora* sp. from a salt marsh in the Kuwaiti desert that are capable of degrading crude oil as the sole carbon source at salinity ranging from 5 to 10%. Overall, bacteria, archaea and a few eukaryotes have been shown to degrade crude oil over a broad range of salinity (0–30%). Of these, eubacteria such as *Marinobacter aquaeolei*, *Streptomyces albiaxialis*, and *Actinopolyspora* sp. and archaea such as *Haloferax*, *Halobacterium*, and *Halococcus* withstand extreme salinity (20–30%) and such organisms are important for the cleanup of oil-impacted hypersaline environments since natural attenuation in such environments is too slow (McGenity, 2010).

## ALIPHATIC COMPOUNDS

Ward and Brock (1978) carried out some of the earliest experiments on the biodegradation aliphatic compounds including mineral oil and  $^{14}\text{C}$ -hexadecane in water samples of varying salinity (3.3–28.4% salt) collected at the salt evaporation ponds near the south end of Great Salt Lake (GSL), Utah and also from the

middle part of GSL. The authors reported decreasing rates of degradation of mineral oil and  $^{14}\text{C}$ -hexadecane with increasing salinity up to 20% in natural sample as well in microbial consortium enriched from water samples from GSL. At salinity greater than 20%, degradation was severely inhibited and this lack of degradation was not due to low levels of dissolved oxygen or lack of growth promoting nutrients since both were provided in the experiments. The authors conclude that the rate limitations were probably due to high salinity. Gauthier et al. (1992) have reported the ability of type strain, *Marinobacter hydrocarbonoclasticus* (originally named *Alteromonas* strain sp -17, isolated by Al-Mallah et al. (1990) from hydrocarbon-contaminated sediments in the Mediterranean Sea) to utilize hexadecane (100%), eicosane (91%), and heneicosane (84%), in the presence of 4.6–20% NaCl. In addition, the organism also degraded phenanthrene (41%) and other aliphatics at low levels as single sources of carbon and energy. Later Fernandez-Linares et al. (1996) have studied the effect of various concentrations of NaCl on growth and degradation of eicosane by *M. hydrocarbonoclasticus* and found that an increase in salinity from 1.2 to 14.5% NaCl had no significant effect on eicosane degradation. Huu et al. (1999) have reported the isolation of *Marinobacter aquaeolei* from an oil-producing well in southern Vietnam that degrades *n*-hexadecane and pristane as the sole sources of carbon at 0–20% salinity. Plotnikova et al. (2001) reported degradation of octane as the sole source of carbon in the presence of 6% salt by several gram positive bacteria including *Rhodococcus* sp., *Arthrobacter* sp., and *Bacillus* sp., isolated from sediment samples from chemical- and salt processing plants in Russia. Abed et al. (2006) have shown the biodegradation pristane and *n*-octadecane at salinity ranging from 5 to 12% at temperatures between 15 and 40°C by microbial mats from the coastal flats of the Arabian Gulf. Al-Mueini et al. (2007) have reported the isolation of an extremely halophilic actinomycete, *Actinopolyspora* sp. DPD1 from an oil production site in the Sultanate of Oman and shown to degrade *n*-alkanes (pentadecane, eicosane, pentacosane) and fluorene at 25% salt. The organism efficiently degraded pentadecane (100% in 4 days) and eicosane (80% in 10 days). Degradation of longer chain alkanes such as pentacosane ( $\text{C}_{25}\text{H}_{52}$ ) proceeded at much slower rate resulting in only 15% degradation in 2 weeks and no triacontane ( $\text{C}_{30}\text{H}_{62}$ ) was degraded even after 20 days of incubation. Degradation of fluorine by *Actinopolyspora* sp. DPD1 resulted in several novel intermediates and appears to proceed through previously undescribed breakdown pathway. The observation that *Actinopolyspora* sp. DPD1 can degrade long chain *n*-alkanes and a polyaromatic hydrocarbon is indicative of its metabolic versatility. Sass et al. (2008) isolated a strain DS-1, closely related to *Bacillus aquimaris* from Discovery deep-sea hypersaline anoxic sediment that grew using *n*-alkanes (*n*-dodecane and *n*-hexadecane) as the sole sources of carbon in the presence of 12–20% NaCl. Mnif et al. (2009, 2011) isolated *Halomonas* sp. strain C2SS100 and *Pseudomonas* sp. strain C450R on the basis of their ability to degrade crude oil also degraded hexadecane as the sole carbon source in the presence of 5–10% NaCl. Dastgheib et al. (2011) have isolated a halotolerant *Alcanivorax* sp. strain Qtet3 from tetracosane degrading enrichments obtained from a hydrocarbon contaminated soils from Qom location in Iran. Strain Qtet3

degrades a wide range of *n*-alkanes (from  $\text{C}_{10}$  to  $\text{C}_{34}$ ) with considerable growth on  $\text{C}_{14}$  and  $\text{C}_{16}$  in the presence of 0–15% NaCl. Strain Qtet3 completely degraded tetracosane ( $\text{C}_{24}\text{H}_{50}$ ) as the sole carbon source in 20 days. In addition, the organism also degrades phytane and pristane, but not aromatic hydrocarbons such as naphthalene, phenanthrene, pyrene, and anthracene. As indicated above, two Marinobacters, *M. sedimentalis* and *M. salviniae* isolated on the basis of their ability to grow on crude oil from hypersaline sabkhas in Kuwait also utilized Tween 80 and a wide range of individual aliphatic hydrocarbons ( $\text{C}_9$ – $\text{C}_{40}$ ) as carbon sources in the presence of 6% NaCl (Al-Mailem et al., 2013).

Reports also exist on the ability of archaea that degrade aliphatic hydrocarbons at high salinity. Bertrand et al. (1990) were among the first to report the isolation of a halophilic archaea, strain EH4, which was recently classified as *Haloarcula vallis-mortis* (see Tapilatu et al., 2010) from a salt marsh near the town of Aigues-Mortes in Southern France. The EH4 was isolated using agar plates containing eicosane as the sole carbon source. Contrary to the results observed by Ward and Brock (1978) at GSL, the growth of EH4 on eicosane increased with increasing salinity. Growth and degradation was maximum at 20% salinity and non-detectable below 10% salinity. Experiments also showed that the isolate was able to degrade a mixture of aliphatic and aromatic hydrocarbons including tetradecane, hexadecane, eicosane, heneicosane, pristane, acenaphthene, phenanthrene, anthracene, and 9-methyl antracene in the presence of >20% NaCl. Kulichevskaya et al. (1992) have reported the isolation of an archaeon, *Halobacterium* that degraded *n*-alkane ( $\text{C}_{10}$ – $\text{C}_{30}$ ) in a medium containing 29% NaCl. Tapilatu et al. (2010) have reported the isolation of several strains of archaea that degrade *n*-alkanes (heptadecane and eicosane) in the presence of 22.5% NaCl from a shallow crystallizer pond (Camargue, France) with no known contamination history. Of these isolates, strain, MSNC 2 was closely related to *Haloarcula* and strains, MSNC 4, MSNC 14, and MSNC 16 to *Haloferax*. In addition, strain MSNC 14 also degraded phenanthrene. Three extremely halophilic archaeal strains, *Haloferax*, *Halobacterium* and *Halococcus* isolated on the basis of crude oil utilization also degraded *n*-alkanes and mono and polyaromatic compounds as the sole sources of carbon and energy in the presence of 26% NaCl (Al-Mailem et al., 2010). Overall, studies reveal that both bacteria and archaea have the capacity to metabolize *n*-alkanes with varying chain lengths in the presence of salt ranging from low to extremely high (Table 2).

## POLYCYCLIC AROMATIC HYDROCARBONS

Polyyclic aromatic hydrocarbons (PAHs) are ubiquitous in many oily and saline environments. Crude oil contains PAHs containing two to four and five ring-molecules. Because of their toxic, mutagenic, and carcinogenic properties, persistence of PAHs in the environment are of particular concern (Menzie et al., 1992; Gibbs, 1997; Cao et al., 2009). The persistence of PAHs in the environment depends on the number of rings in the molecule and environmental factors such as pH, temperature, and salinity. Although studies have reported the degradation PAHs by non-halophiles and in marine habitats, little is known about the fate

**Table 2 | Degradation of aliphatic hydrocarbons in moderate to high salinity conditions.**

Hydrocarbon	Structure	Degrader	NaCl (%)	References
Octane	C <sub>8</sub> H <sub>18</sub>	<i>Rhodococcus</i> sp., <i>Arthrobacter</i> sp., <i>Bacillus</i> sp.	6	Plotnikova et al., 2001, 2011
Decane	C <sub>10</sub> H <sub>22</sub>	<i>Bacillus</i> sp. strain DHT	10	Kumar et al., 2007
Tetradecane	C <sub>14</sub> H <sub>30</sub>	EH4 ( <i>Haloarcula vallismortis</i> )	>20	Bertrand et al., 1990
Pentadecane	C <sub>15</sub> H <sub>32</sub>	<i>Actinopolyspora</i> sp.	25	Al-Mueini et al., 2007
		<i>Marinobacter aquaeolei</i>	0–20	Huu et al., 1999
Hexadecane	C <sub>16</sub> H <sub>34</sub>	Enrichment, Great Salt Lake, Utah	< 20	Ward and Brock, 1978
		EH4 ( <i>Haloarcula vallismortis</i> )	>20	Bertrand et al., 1990
		<i>Marinobacter hydrocarbonoclasticus</i>	4.6–20	Gauthier et al., 1992
		<i>Marinobacter aquaeolei</i>	0–20	Huu et al., 1999
		<i>Bacillus</i> sp. strain DHT	10	Kumar et al., 2007
		Strain DS1	12–20	Sass et al., 2008
		<i>Halomonas</i> sp. C2SS100	10	Mnif et al., 2009
		<i>Pseudomonas</i> sp. C450R		
Octadecane	C <sub>18</sub> H <sub>38</sub>	Microbial mats	5–12	Abed et al., 2006
		<i>Haloferax</i> sp., <i>Halobacterium</i> sp., <i>Halococcus</i> sp.	>26	Al-Mailem et al., 2010
Heptadecane	C <sub>17</sub> H <sub>36</sub>	<i>Haloarcula</i> sp., <i>Haloferax</i> sp.	>22	Tapilatu et al., 2010
Pristane	C <sub>19</sub> H <sub>40</sub>	EH4 ( <i>Haloarcula vallismortis</i> )	>20	Bertrand et al., 1990
		<i>Marinobacter hydrocarbonoclasticus</i>	4.6–20	Gauthier et al., 1992
		<i>Marinobacter aquaeolei</i>	0–20	Huu et al., 1999
		Microbial mats, Arabian Gulf coast, Saudi Arabia	5–12	Abed et al., 2006
		<i>Alcanivorax</i> sp. Qtet3	0–15	Dastgheib et al., 2011
Eicosane	C <sub>20</sub> H <sub>42</sub>	EH4 ( <i>Haloarcula vallismortis</i> )	10–20	Bertrand et al., 1990
		<i>Marinobacter hydrocarbonoclasticus</i>	1–14	Fernandez-Linares et al., 1996
		<i>Actinopolyspora</i> sp. DPD1	25	Al-Mueini et al., 2007
		<i>Haloarcula</i> sp., <i>Haloferax</i> sp.	>22	Tapilatu et al., 2010
Phytane	C <sub>20</sub> H <sub>42</sub>	<i>Alcanivorax</i> sp. strain Qtet3	0–15	Dastgheib et al., 2011
Heneicosane	C <sub>21</sub> H <sub>44</sub>	EH4 ( <i>Haloarcula vallismortis</i> )	>20	Bertrand et al., 1990
		<i>Marinobacter hydrocarbonoclasticus</i>	4.6–20	Gauthier et al., 1992
Tetracosane	C <sub>24</sub> H <sub>50</sub>	<i>Alcanivorax</i> sp. strain Qtet3	0–15	Dastgheib et al., 2011
Pentacosane	C <sub>25</sub> H <sub>52</sub>	<i>Actinopolyspora</i> sp. DPD1	25	Al-Mueini et al., 2007
n-Alkane	C <sub>10</sub> –C <sub>30</sub>	<i>Halobacterium</i> sp.	29	Kulichevskaya et al., 1992
	C <sub>10</sub> –C <sub>34</sub>	<i>Haloferax</i> sp., <i>Halobacterium</i> sp., <i>Halococcus</i> sp.	>26	Al-Mailem et al., 2010
	C <sub>10</sub> –C <sub>34</sub>	<i>Alcanivorax</i> sp. strain Qtet3	0–15	Dastgheib et al., 2011
	C <sub>9</sub> –C <sub>40</sub>	<i>Marinobacter sedimentalis</i> , <i>Marinobacter falvimaris</i>	6	Al-Mailem et al., 2013

of these compounds in high salinity environments. Ashok et al. (1995) have isolated bacterial strains of the genus *Micrococcus*, *Pseudomonas*, and *Alcaligenes* from soil samples near an oil refinery that degraded naphthalene and anthracene as the sole sources of carbon at 7.5% salinity. Plotnikova et al. (2001, 2011) have isolated *Pseudomonas* sp., *Rhodococcus* sp., *Arthrobacter* sp., and *Bacillus* sp. from soil and sediment contaminated with waste generated by chemical and salt-producing plants. All these isolates degraded naphthalene and salicylate as the sole carbon sources in the presence of 5–9% NaCl. In addition, some of these organisms also grew on phenanthrene, biphenyl, o-phthalate, gentisate, octane, and phenol as the sole sources of carbon. Zhao et al. (2009) have shown the degradation of phenanthrene in the presence of 5–15% NaCl by a halophilic bacterial consortium developed from soil samples collected from the Shengli Oilfield in China. Phenanthrene was completely degraded by the enrichment in 8 days. Molecular analysis of the enrichment culture indicated

the presence of *alpha* and *gamma*-proteobacteria including members of the genus *Halomonas*, *Chromohalobacter*, *Alcanivorax*, *Marinobacter*, *Idiomarina*, and *Thalassospira*. Dastgheib et al. (2012) have obtained a mixed culture (Qphe-SubIV) consisting of *Halomonas* sp. and *Marinobacter* sp. from hydrocarbon-contaminated saline soil collected from five different regions in Iran. These organisms degraded several PAHs including naphthalene, phenanthrene, anthracene, fluoranthene, fluorine, pyrene, benz[a]anthracene, and benzo[a]pyrene as the sole carbon sources in the presence of 1–15% NaCl. Recently, Al-Mailem et al. (2013) have reported the ability of *Marinobacter sedimentalis* and *Marinobacter falvimaris* isolated from hypersaline sabkhas to degrade biphenyl, phenanthrene, anthracene and naphthalene as the sole sources of carbon and energy at 6% NaCl. More recently, Gao et al. (2013) have isolated *Marinobacter nanhaiticus* Strain D15-8W from a phenanthrene-degrading enrichment obtained from a sediment from the South China Sea. The strain D15-8W

degrades naphthalene, phenanthrene or anthracene as the sole source of carbon in the presence of 0.5–15% with optimum degradation in the presence of 1–5% NaCl.

Studies also show the ability of archaea to degrade PAHs in high salinity environments. As mentioned above, strain EH4 (*Haloarcula vallismortis*), not only degraded *n*-alkanes but also degraded a mixture of alkanes and aromatic compounds such as acenaphthene, anthracene, and phenanthrene at >20% NaCl (Bertrand et al., 1990). Bonfá et al. (2011) have isolated several strains of *Haloferax* that degrade a mixture of the PAHs including naphthalene, anthracene, phenanthrene, pyrene and benzo[a]anthracene at high salinity (20% NaCl). Extremely halophilic archaeal strains of *Haloferax*, *Halobacterium*, and *Halococcus* isolated from a hypersaline coastal area of the Arabian Gulf not only degraded crude oil and *n*-octadecane as the carbon sources, but also grew on phenanthrene at 26% salinity (Al-Mailem et al., 2010). Erdogmuş et al. (2013) showed the degradation of naphthalene, phenanthrene and pyrene as the sole carbon sources in the presence of 20% NaCl by several archaeal strains including *Halobacterium piscisalsi*, *Halorubrum ezzemoulene*, *Halobacterium salinarium*, *Haloarcula hispanica*, *Haloferax* sp., *Halorubrum* sp. and *Haloarcula* sp. isolated from brine samples of Camaltı Saltern in Turkey. The hydrocarbon degradation potential of *Halorubrum* sp. and *Halorubrum ezzemoulene* was documented for the first time in this study. These reports clearly demonstrate the potential of bacteria and archaea to degrade PAHs in high salinity environments (Table 3).

#### BENZENE, TOLUENE, ETHYLBENZENE, AND XYLEMES

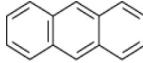
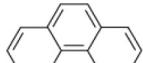
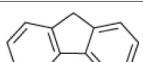
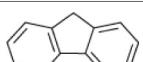
The most abundant hydrocarbons in produced water are the one-ring aromatic hydrocarbons, benzene, toluene, ethylbenzene, and xylenes (BTEX) and low molecular weight saturated hydrocarbons (Neff et al., 2011). Benzene is a category A carcinogen. Leakage from produced water storage tanks, pipelines, spills, and seepage from surface contaminated sites can cause major BTEX contamination (Philip et al., 2005). BTEX are relatively highly soluble in water and hence can contaminate large volumes of groundwater. Although there have been many recent reports on the biodegradation of non-oxygenated hydrocarbons in moderate to high salinity environments, only few reports exist on the biodegradation of BTEX compounds (Table 4). Nicholson and Fathepure (2004, 2005) have reported the degradation of BTEX at high salinity in microcosms established with soil samples from an oilfield and from an uncontaminated salt flat in Oklahoma. Subsequently, enrichment cultures were obtained from both sites on mineral salts medium containing 14.5% NaCl and benzene as the sole carbon source. The oilfield enrichment degraded BTEX in the presence of 3–14.5% NaCl, whereas the enrichment from the salt flat degraded only benzene and toluene as the sole carbon sources in the presence of 0–23% NaCl. Furthermore, these studies have demonstrated complete mineralization of  $^{14}\text{C}$ -benzene to  $^{14}\text{CO}_2$  by the enrichment cultures in the presence of 14.5% NaCl. Sei and Fathepure (2009) have developed an enrichment culture using sediment samples from Rozel point in GSL, Utah. The enrichment completely degraded benzene or toluene as the sole source of carbon within 1, 2, and 5 weeks in the presence of 14, 23, and 29% NaCl, respectively. In addition, these authors have successfully isolated two strains of *gamma-proteobacteria* identified

as *Arhodomonas* sp. strain Seminole (previously referred to as strain SEM-2) and *Arhodomonas* sp. strain Rozel from enrichments developed using a soil sample from an oilfield in Oklahoma and a sediment sample from Rozel Point, respectively (Nicholson and Fathepure, 2006; Azetsu et al., 2009). These strains rapidly degraded benzene and toluene as the sole sources of carbon in the presence of 3–23% NaCl and no degradation was seen at 0 and 30% NaCl. Li et al. (2006) have isolated a *Planococcus* sp. strain ZD22 using a contaminated soil collected from a site near the Daqing oil field in China. The strain ZD22 is a psychrotolerant and moderate haloalkaliphile and degrades BTEX in the presence of 0.5–25% salt. In addition, the strain ZD22 also degraded chlorobenzene, bromobenzene, iodobenzene, and fluorobenzene. This ability of the strain ZD22 to utilize different aromatic compounds, combined with its ability to grow under multiple extreme conditions including low temperature, high salinity, and alkaline pH make it a good candidate for the biodegradation of toxic wastes. Berlendis et al. (2010) have tested the ability of two previously isolated Marinobacters, *Marinobacter vinifirmus* and *Marinobacter hydrocarbonoclasticus* to degrade BTEX as the sole carbon sources at 3–15% salinity. *M. vinifirmus* was able to degrade all the added benzene and toluene in 3 days, while 65% of total ethylbenzene and 20% of total *p*-xylene were removed in 7 days in the presence of 6% NaCl. Similarly, *M. hydrocarbonoclasticus* degraded 10% of benzene, 20% of toluene, 60% of ethylbenzene, and 70% of the added *p*-xylene in 7 days as the sole sources of carbon at 6% salinity. Recently Al-Mailem et al. (2013) have isolated *Marinobacter sedimentalis* and *Marinobacter falvimaris* on the basis of their ability to utilize *n*-alkanes and PAHs. These bacteria were also able to degrade benzene as the sole carbon source in the presence of 6% NaCl thus extending the substrate range for this group of organisms. This is important because Marinobacters are one of the most important groups of halophiles found in a variety of ecosystems ranging from extremely cold to hot, low to high salinity and over a broad range of pH demonstrating their tremendous adaptation capabilities (Duran, 2010). Hassan et al. (2012) have reported the isolation of *Alcanivorax* sp. HA03 from soda lakes in Wadi El Natrun capable of degrading benzene, toluene, and chlorobenzene as the sole sources of carbon at salinity ranging from 3 to 15% NaCl. This observation that *Alcanivorax* can also degrade aromatic compounds expands the metabolic capability of this group of organisms because *Alcanivorax* are primarily known for their ability to degrade aliphatic hydrocarbons. Degradation of benzene was also reported in archaea. For example, the crude oil degrading *Haloferax*, *Halobacterium*, and *Halococcus* isolated from a hypersaline Arabian Gulf coast degraded benzene as the sole source of carbon at 26% salinity (Al-Mailem et al., 2010). As mentioned above, to date, only few microorganisms have been shown to degrade BTEX in moderate to high salinity conditions. This is not surprising considering that BTEX are volatile compounds and lack an activating oxygen or nitrate moiety thus making these compounds less available and resistant to biodegradation.

#### PHENOLICS AND BENZOATES

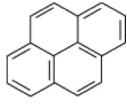
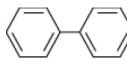
Industrial effluents generated from many food, dye, pharmaceutical, and chemical processing operations are often characterized by

**Table 3 | Biodegradation of polycyclic aromatic hydrocarbons in moderate to high salinity conditions.**

Hydrocarbon	Structure	Degrader	Salinity (%)	References
Naphthalene		<i>Micrococcus</i> sp. <i>Pseudomonas</i> sp.	7.5	Ashok et al., 1995
		<i>Alcaligenes</i> sp.		
		<i>Pseudomonas</i> sp. <i>Rhodococcus</i> sp.	6–9	Plotnikova et al., 2001, 2011
		<i>Arthrobacter</i> sp. <i>Bacillus</i> sp.		
		<i>Bacillus</i> sp strain DHT	10	Kumar et al., 2007
		<i>Haloferax</i> sp. <i>Halobacterium</i> sp.	>26	Al-Mailem et al., 2010
		<i>Halococcus</i> sp.		
		<i>Haloferax</i> spp.	20	Bonfá et al., 2011
		<i>Arthrobacter</i> spp. SN17	6–9	Plotnikova et al., 2011
		Mixed culture (Ophe-SubIV)	1–15	Dastgheib et al., 2012
		<i>Marinobacter sedimentalis</i>	6	Al-Mailem et al., 2013
		<i>Marinobacter falvimaris</i>		
		<i>Marinobacter nankaiticus</i>	0.5–15	Gao et al., 2013
		<i>Halobacterium piscisalsi</i> , <i>Halorubrum ezzemoulense</i> , <i>Halobacterium salinarium</i> , <i>Haloarcula hispanica</i>	20	Erdogmuş et al., 2013
		<i>Haloferax</i> sp. <i>Halorubrum</i> sp.		
		<i>Haloarcula</i> sp.		
Anthracene		EH4 ( <i>Haloarcula vallismortis</i> )	>20	Bertrand et al., 1990
		<i>Micrococcus</i> sp. <i>Pseudomonas</i> sp.	7.5	Ashok et al., 1995
		<i>Alcaligenes</i> sp.		
		<i>Haloferax</i> spp	20	Bonfá et al., 2011
		Mixed culture (Ophe-SubIV)	1–15	Dastgheib et al., 2012
		<i>Marinobacter sedimentalis</i>	6	Al-Mailem et al., 2013
		<i>Marinobacter falvimaris</i>		
Phenanthrene		<i>Marinobacter nankaiticus</i>	0.5–15	Gao et al., 2013
		EH4 ( <i>Haloarcula vallismortis</i> )	>20	Bertrand et al., 1990
		<i>Micrococcus</i> sp. <i>Pseudomonas</i> sp.	7.5	Ashok et al., 1995
		<i>Alcaligenes</i> sp.		
		<i>Pseudomonas</i> sp. <i>Arthrobacter</i> sp.	5–9	Plotnikova et al., 2001, 2011
		Microbial consortium, Shengli Oilfield, China	5–15	Zhao et al., 2009
		<i>Haloferax</i> sp	>22	Tapilatu et al., 2010
		<i>Haloferax</i> sp. <i>Halobacterium</i> sp.	>26	Al-Mailem et al., 2010
		<i>Halococcus</i> sp.		
		<i>Haloferax</i> spp.	20	Bonfá et al., 2011
		Mixed culture (Ophe-SubIV)	1–15	Dastgheib et al., 2012
		<i>Marinobacter sedimentalis</i>	6	Al-Mailem et al., 2013
		<i>Marinobacter falvimaris</i>		
		<i>Marinobacter nankaiticus</i>	0.5–15	Gao et al., 2013
Acenaphthene		<i>Halobacterium piscisalsi</i> , <i>Halorubrum ezzemoulense</i> , <i>Halobacterium salinarium</i> , <i>Haloarcula hispanica</i>	20	Erdogmuş et al., 2013
		<i>Haloferax</i> sp. <i>Halorubrum</i> sp.		
Fluorene		<i>Haloarcula</i> sp.		
		EH4 ( <i>Haloarcula vallismortis</i> )	>20	Bertrand et al., 1990
		<i>Actinopolyspora</i> sp. DPD1	5–20	Al-Mueini et al., 2007
		Mixed culture (Ophe-SubIV)	1–15	Dastgheib et al., 2012

(Continued)

**Table 3 | Continued**

Hydrocarbon	Structure	Degrader	Salinity (%)	References
Pyrene		<i>Bacillus</i> sp strain DHT	10	Kumar et al., 2007
		<i>Haloferax</i> spp.	20	Bonfá et al., 2011
		Mixed culture (Ophe-SubIV)	1–15	Dastgheib et al., 2012
		<i>Halobacterium piscisalsi</i> , <i>Halorubrum ezzemoulense</i> , <i>Halobacterium salinarium</i> , <i>Haloarcula hispanica</i>	20	Erdogmuş et al., 2013
		<i>Haloferax</i> sp. <i>Halorubrum</i> sp. <i>Haloarcula</i> sp.		
Biphenyl		<i>Rhodococcus</i> sp. <i>Arthrobacter</i> sp.	6–9	Plotnikova et al., 2001, 2011
		<i>Haloferax</i> sp. <i>Halobacterium</i> sp.	>26	Al-Mailem et al., 2010
		<i>Halococcus</i>		
		<i>Marinobacter sedimentalis</i>	6	Al-Mailem et al., 2013
		<i>Marinobacter falvimaris</i>		

**Table 4 | Degradation of benzene, toluene, ethylbenzene and xylenes in moderate to high salinity conditions.**

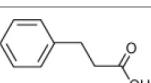
Hydrocarbon	Structure	Degrader	Salinity (%)	References
Benzene		Enrichment, Oilfield Oklahoma	3–14	Nicholson and Fathepure, 2004
		Enrichment, Great Salt Plains, Oklahoma	0–23	Nicholson and Fathepure, 2005
		<i>Planococcus</i> sp. strain ZD22	5–20	Li et al., 2006
		<i>Arhodomonas</i> sp. strain Seminole	3–17	Nicholson and Fathepure, 2006; Dalvi et al., 2012
		Enrichment, Rozel Point, Great Salt Lake, Utah	0–29	Sei and Fathepure, 2009
		<i>Arhodomonas</i> sp. strain Rozel	3–23	Azetsu et al., 2009; Dalvi et al., 2012
		<i>Marinobacter vinifirmus</i> , <i>M. hydrocarbonoclasticus</i>	3–15	Berlendis et al., 2010
		<i>Haloferax</i> sp. <i>Halobacterium</i> sp. <i>Halococcus</i> sp.	>26	Al-Mailem et al., 2010
		<i>Alcanivorax</i> sp. HA03	3–15	Hassan et al., 2012
		<i>Marinobacter sedimentalis</i> , <i>Marinobacter falvimaris</i>	6	Al-Mailem et al., 2013
Toluene		Enrichment, oilfield soil, Oklahoma	3–14	Nicholson and Fathepure, 2004
		Enrichment, Great Salt Plains, Oklahoma	0–23	Nicholson and Fathepure, 2005
		<i>Planococcus</i> sp. strain ZD22	5–20	Li et al., 2006
		<i>Arhodomonas</i> sp. strain Seminole	3–17	Nicholson and Fathepure, 2006; Dalvi et al., 2012
		Enrichment, Rozel Point, Great Salt Plains, Utah	0–29	Sei and Fathepure, 2009
		<i>Arhodomonas</i> sp. strain Rozel	3–23	Azetsu et al., 2009; Dalvi et al., 2012
		<i>Marinobacter vinifirmus</i> , <i>M. hydrocarbonoclasticus</i>	3–15	Berlendis et al., 2010
		<i>Haloferax</i> sp. <i>Halobacterium</i> sp. <i>Halococcus</i> sp.	>26	Al-Mailem et al., 2010
		<i>Alcanivorax</i> sp. HA03	3–15	Hassan et al., 2012
Ethylbenzene		Enrichment, oilfield soil, Oklahoma	3–14	Nicholson and Fathepure, 2004
		<i>Planococcus</i> sp. strain ZD22	5–20	Li et al., 2006
		<i>Marinobacter vinifirmus</i> , <i>M. hydrocarbonoclasticus</i>	3–15	Berlendis et al., 2010
Xylene		Enrichment, oilfield soil, Oklahoma	3–15	Nicholson and Fathepure, 2004
		<i>Planococcus</i> sp. strain ZD22	5–20	Li et al., 2006
		<i>Marinobacter vinifirmus</i> , <i>M. hydrocarbonoclasticus</i>	3–15	Berlendis et al., 2010

high salinity and the presence of phenolics and benzoates (Garcia et al., 2005b). In addition, compounds such as 4-hydroxybenzoic, ferulic, *p*-coumaric, vanillic, cinnamic, and syringic acids are naturally present in lignin and plant root exudates (Le Borgne et al., 2008). In recent years, many studies have successfully

isolated bacteria and archaea that degrade oxygenated aromatics in saline conditions. **Table 5** lists organisms that degrade oxygenated hydrocarbons in moderate to high salinity conditions.

Woolard and Irvine (1995) showed that a halophile isolated from a mixed culture obtained from a saltern at GSL,

**Table 5 | Biodegradation of phenolics and benzoates in moderate to high salinity conditions.**

Hydrocarbon	Structure	Degrader	Salinity (%)	References
Phenol		Halophilic isolate	1–15	Woolard and Irvine, 1995
		<i>Halomonas</i> sp.	1–14	Hinteregger and Streischa, 1997
		<i>Candida tropica</i>	15	Bastos et al., 2000
		<i>Halomonas campialis</i>	0–15	Alva and Peyton, 2003
		<i>Halomonas organivorans</i>	1.5–30	Garcia et al., 2004, 2005b
		<i>Thelassobacillus devorans</i>	7.5–10	Garcia et al., 2005a
		<i>Arthrobacter</i> sp.	6–9	Plotnikova et al., 2011
		<i>Halomonas organivorans</i> , <i>Arhodomonas aquaeolei</i> , <i>Modicisalibacter tunisiensis</i>	10	Bonfá et al., 2013
		Strain C5	12	Chamkha et al., 2008
Benzoate		<i>Halomonas halodurans</i>	15	Rosenberg, 1983
		<i>Haloferax</i> sp. D1227	15	Emerson et al., 1994
		<i>Halomonas</i> sp.	35	Kleinsteuber et al., 2001
		<i>Halomonas organivorans</i>	1.5–30	Garcia et al., 2004, 2005b
		<i>Halomonas elongate</i>	10	Garcia et al., 2005b
		<i>Halomonas eurihalina</i>		
		<i>Marinobacter lipolyticus</i>		
		<i>Halomonas campialis</i>	5–10	Oie et al., 2007
		<i>Chromohalobacter</i> sp. strain HS-2	10	Kim et al., 2008
		<i>Haloferax</i> sp.	20	Bonfá et al., 2011
Salicylate		Strain C5	12	Chamkha et al., 2008
		<i>Pseudomonas</i> sp. <i>Rhodococcus</i> sp.	5–9	Plotnikova et al., 2001
		<i>Arthrobacter</i> sp.		
		<i>Bacillus</i> sp.		
		<i>Halomonas organivorans</i>	1.5–30	Garcia et al., 2004
		<i>Halomonas organivorans</i> , <i>Salinicoccus roseus</i>	10	Garcia et al., 2005b
		<i>Halomonas venusta</i>		
		<i>Halomonas alimentaria</i>		
		<i>Halomonas campialis</i>	5–10	Oie et al., 2007
		<i>Bacillus</i> sp. strain DHT	10	Kumar et al., 2007
o-Phthalate		<i>Haloferax</i> sp.	20	Bonfá et al., 2011
		<i>Rhodococcus</i> sp. <i>Arthrobacter</i> sp.	5–9	Plotnikova et al., 2001, 2011
Gentisate		<i>Rhodococcus</i> sp. <i>Arthrobacter</i> sp.	3–29	Plotnikova et al., 2001
		<i>Bacillus</i> sp.		
4-Hydroxy-benzoate		<i>Haloarcula</i> sp. strain D1,	17	Fairley et al., 2002
		<i>Halomonas organivorans</i>	1.5–30	Garcia et al., 2004
		<i>Halomonas elongate</i>	10	Garcia et al., 2005b
		Halophilic archaeal strains	20	Cuadros-Orellana et al., 2006, 2012
		<i>Chromohalobacter</i> sp. strain HS-2	10	Kim et al., 2008
		Strain C5	12	Chamkha et al., 2008
		<i>Haloferax</i> sp.	20	Bonfá et al., 2011
		<i>Halobacterium</i> sp. <i>Haloferax</i> sp.	20	Erdogmuş et al., 2013
		<i>Halorubrum</i> sp. <i>Haloarcula</i> sp.		
Phenyl propionic acid		<i>Haloferax</i> sp. D1227	5–29	Emerson et al., 1994; Fu and Oriel, 1999

(Continued)

**Table 5 | Continued**

Hydrocarbon	Structure	Degrader	Salinity (%)	References
Cinnamic acid		<i>Halomonas organivorans</i>	10	Garcia et al., 2005b
		<i>Halomonas elongata</i>		
		<i>Halomonas glaciei</i>		
		<i>Halomonas organivorans</i>	1.5–30	Garcia et al., 2004, 2005b
p-Coumaric acid		<i>Haloferax</i> sp. D1227	5–30	Emerson et al., 1994
		<i>Halomonas organivorans</i>	1.5–30	Garcia et al., 2004
		<i>Halomonas organivorans</i>	10	Garcia et al., 2005b
		<i>Halomonas salina/Halomonas halophila</i>		
		<i>Halomonas elongate</i>		
		<i>Halomonas</i> strain IMPC	0–25	Abdelkafi et al., 2006
		Strain C5	12	Chamkha et al., 2008
Ferulic acid		<i>Halomonas organivorans</i>	1.5–30	Garcia et al., 2004
		<i>Halomonas organivorans</i>	10	Garcia et al., 2005b
		<i>Halomonas salina</i>		
		<i>Chromohalobacter israelensis</i>		
		<i>Halomonas</i> strain IMPC	0–25	Abdelkafi et al., 2006
		Strain C5	12	Chamkha et al., 2008

Utah readily degraded phenol in the presence of 1–15% NaCl. Similarly, Hinteregger and Streischsberg (1997) reported that a *Halomonas* sp. isolated from a co-culture developed from GSL degraded phenol as the sole source of carbon in the presence of 1–14% salt. Complete degradation of phenol occurred in 13 h at 5% NaCl but at higher NaCl concentrations, degradation occurred with longer lag periods. For example, at 14% NaCl, phenol was completely removed with a lag of 100 h. The degradation of phenol in this organism was accompanied by the accumulation of *cis, cis*-muconic acid, a product of *ortho*-cleavage pathway by catechol 1, 2-dioxygenase enzyme. Bastos et al. (2000) reported the isolation of a yeast, *Candida tropicalis* from an enrichment developed from Amazonian rain forest soil that degraded phenol in the presence of up to 15% NaCl. Alva and Peyton (2003) isolated a haloalkaliphile, *Halomonas campialis* near Soap Lake in central Washington and showed that this organism degraded phenol and catechol as the sole sources of carbon at pH 8–11 and salinity of 0–15%. Formation of metabolic intermediates such as catechol and *cis, cis*-muconic acid suggests that phenol was degraded by the *ortho*-cleavage pathway of the beta-ketoadipate branch. A Gram-positive halophilic bacterium, *Thalassobacillus devorans* isolated from an enrichment culture developed from saline habitats in southern Spain was shown to degrade phenol (Garcia et al., 2005a) in the presence of 7.5–10% NaCl. The strain C5, closely related to *Geobacillus pallidus* isolated from a tyrosol-utilizing enrichment also degrades a variety of other oxygenated aromatic compounds including benzoic, *p*-hydroxybenzoic, protocatechuic, vanillic, *p*-hydroxyphenylacetic, 3,4-dihydroxyphenylacetic, cinnamic,

ferulic, phenol, and *m*-cresol. However, no degradation of non-oxygenated hydrocarbons such as toluene, naphthalene, and phenanthrene was observed (Chamkha et al., 2008). Recently, Bonfá et al. (2013) have shown the degradation of phenol as the sole source of carbon in the presence of 10% NaCl by *Halomonas organivorans*, *Arhodomonas aquaeolei*, and *Modicisalibacter tunisiensis* isolated from different hypersaline environments.

Many reports also exist on the ability of halophilic and halotolerant organisms to degrade benzoates in high salinity conditions. The halotolerant, *Pseudomonas halodurans* (reclassified as *Halomonas halodurans*) degrades benzoic acid in the presence of >15% NaCl (Rosenberg, 1983). Garcia et al. (2004, 2005b) have isolated several strains of *Halomonas* spp. including the *Halomonas organivorans* from water and sediment of salterns and hypersaline soils collected in different part of the Southern Spain with salinity of the sampling site ranging from 4 to 17%. These isolates degraded a wide range of aromatic compounds including benzoic acid, *p*-hydroxybenzoic acid, phenol, salicylic acid, *p*-aminosalicylic acid, phenylacetic acid, phenylpropionic acid, cinnamic acid, ferulic acid, and *p*-coumaric acid as the sole sources of carbon in the presence of 10% NaCl. Abdelkafi et al. (2006) have reported the isolation of a *p*-coumaric acid degrading *Halomonas* strain IMPC from a *p*-coumaric acid degrading enrichment culture obtained from a table-olive fermentation rich in aromatic compounds. This strain converted *p*-coumaric acid to *p*-hydroxybenzaldehyde, *p*-hydroxybenzoic acid, and then to protocatechuic acid prior to ring cleavage in the presence of 0–25% NaCl. In addition, the strain also degraded other

lignin-related compounds such as cinnamic acid, *m*-coumaric acid, *m*- and *p*-methoxycinnamic acid, *m*- and *p*-methylcinnamic acid, and ferulic acid to their corresponding benzoic acid derivatives. Oie et al. (2007) have studied the degradation of benzoate and salicylate by *Halomonas campisalis* isolated from an alkaline Soap Lake in the presence of 5–10% NaCl. This study showed that the organism degraded benzoate and salicylate to catechol and then to *cis*, *cis*-muconate thus indicating degradation *via* the *ortho*-cleavage pathway. Kim et al. (2008) have isolated a *Chromohalobacter* sp. strain HS-2 from salted fermented clams that degrades benzoate and *p*-hydroxybenzoate at 10% NaCl as the sole carbon and energy sources.

Studies have also documented aerobic degradation of benzoates by extremely halophilic archaea, often growing in near-saturated brines (>30% NaCl). For example, Emerson et al. (1994) isolated a *Haloferax* sp. D1227 from an oil-brine soil near Grand Rapids, Michigan and was shown to degrade benzoic acid, 3-hydroxybenzoic acid, 3-phenylpropionic acid, and cinnamic acid as the sole sources of carbon at salt concentration ranging from 5 to 30% NaCl. When grown on <sup>14</sup>C-benzoate, strain D1227 converted 70% of the substrate to <sup>14</sup>CO<sub>2</sub> and assimilated 19% of the <sup>14</sup>C-label into cell biomass. These compounds were degraded *via* a gentisate pathway (Fu and Oriel, 1998, 1999). Fairley et al. (2002) have isolated a novel halophilic archaeon, *Haloarcula* sp. D1 from a high salt enrichment culture and shown to degrade *p*-hydroxybenzoic acid as the sole source of carbon. Cuadros-Orellana et al. (2006) have isolated 44 archaeal strains from five geographically different saline environments including the Uyuni salt marsh in Bolivia, solar saltern in Chile, solar saltern in Puerto Rico, Dead Sea near Jordan, and sabkhas in Saudi Arabia. Analysis of lipid composition and restriction analysis of 16S rDNA-gene places all the strain in four groups in the Halobacteriaceae family. These strains degraded *p*-hydroxybenzoic acid as the sole carbon source in the presence of 20% NaCl. Similarly, Bonfá et al. (2011) have isolated 10 halophilic archaea, all belonging to the genus *Haloferax* from *p*-hydroxybenzoic acid -utilizing mixed cultures obtained from the above five hypersaline sites. These strains were also able to degrade a mixture of *p*-hydroxybenzoic acid, benzoic acid, and salicylic acid as growth substrates in a medium containing 20% NaCl. Recently, Cuadros-Orellana et al. (2012) have reported the isolation of 10 halophilic archaea from Dead Sea that degrade *p*-hydroxybenzoic acid as the sole carbon and energy source. In addition, strain L1, a member of the unclassified Halobacteriaceae family of the phylum, *Euryarchaeota* also degrades benzoic acid to gentisate. Erdogan et al. (2013) reported the ability of many archaeal strains belonging to *Halobacterium*, *Haloferax*, *Halorubrum*, and *Haloarcula* group to degrade *p*-hydroxybenzoic acid in a medium containing 20% NaCl. These studies clearly demonstrate that archaea that metabolize *p*-hydroxybenzoic are widespread in the environment. Among bacteria, *Halomonas* spp. have been frequently reported for their ability to degrade phenolics and benzoates and only few reports exist on their potential to degrade non-oxygenated hydrocarbons. Therefore, to fully realize their remediation potential, more studies are needed to determine their capacity to degrade BTEX and PAHs.

## MOLECULAR MECHANISM OF HYDROCARBON DEGRADATION IN HIGH SALINITY ENVIRONMENT

In the last two decades there has been impressive progress in the area of hydrocarbon degradation in hypersaline environments. Pure cultures of aerobic bacteria, archaea, and some eukaryotes have been isolated that degrade hydrocarbons over a broad range of salinities. However, similar progress on genetics and biochemistry of hydrocarbon degradation is severely lacking. Extensive information exists in the literature on the degradation pathways and enzymes involved in the aerobic metabolism of petroleum compounds for many non-halophiles (Reineke, 2001; Van Hamme et al., 2003; Cao et al., 2009). In non-halophiles, monooxygenases initiate degradation of aliphatic hydrocarbons by the addition of oxygen atom (s) to the terminal or subterminal carbon and converting them to corresponding fatty acids which are then assimilated *via* beta-oxidation (Patzelt, 2007). The integral-membrane non-heme di-iron monooxygenase (*AlkB*) and the cytochrome P450 CYP153 family alkane hydroxylases (van Beilen and Funhoff, 2007) catalyze the hydroxylation of medium-chain-length alkanes (C<sub>8</sub>–C<sub>16</sub>), while a flavin-binding monooxygenase (*AlmA*) and a long chain alkane monooxygenase (*LadA*) have shown to be responsible for the degradation of long chain alkanes with chain length >C<sub>18</sub> (Feng et al., 2007; Throne-Holst et al., 2007).

Similarly, a wide variety of aromatic hydrocarbons are degraded by monooxygenases or dioxygenases by the addition of oxygen atom (s) to the alkyl moiety or aromatic ring (Reineke, 2001; Van Hamme et al., 2003; Cao et al., 2009; Pérez-Pantoja et al., 2010) converting them to a few central intermediates such as catechols, protocatechuate, and gentisate through convergent pathways. These ring intermediates are cleaved by *ortho*- or *meta*-cleavage dioxygenases such as catechol 1, 2-dioxygenase (1,2-CAT), catechol 2,3-dioxygenase (2,3-CAT), protocatechuate 3,4-dioxygenases (3,4-PCA), and protocatechuate 4,5-dioxygenase (4,5-PCA), and gentisate 1,2-dioxygenase (1,2-GDO) enzymes (Lack, 1959; Harwood and Parales, 1996; Reineke, 2001) into intermediary metabolites such as acetyl Co-A, succinyl Co-A, and pyruvate that feed into the Krebs cycle (Fuchs et al., 2011). The genes encoding these enzymes have been characterized for a variety of aerobic microorganisms including several members of the genera *Pseudomonas*, *Rhodococcus*, *Ralstonia*, and *Mycobacterium*, *Acinetobacter* (Luz et al., 2004; Cao et al., 2009).

To date little information exists about the pathways and enzymes for hydrocarbon degradation in high salinity environments. A few recent studies have shown that the degradation of hydrocarbons at high salinity occurs using enzymes described for many non-halophiles. For example, detection of ring-oxidation and ring-cleavage intermediates such as catechol and *cis*, *cis*-muconate in benzoate and phenol degrading *Halomonas* spp. indicate the role of *ortho*-cleaving enzymes in the beta-ketoadipate pathway for aromatic metabolism (Hinteregger and Streischsberg, 1997; Alva and Peyton, 2003; Oie et al., 2007). Garcia et al. (2005b) have used PCR and degenerate primers for the detection of genes that code for 1,2-CAT and 3,4-PCA enzymes in several strains of phenol- and

benzoate degrading *Halomonas* spp. Furthermore, activity of these enzymes was measured in cell free extract of *Halomonas organivorans* cells grown on various aromatic compounds. Recently, Moreno et al. (2011) have further characterized the genes involved in the metabolism of phenol and benzoate in *Halomonas organivorans* in much detail. The gene cluster *catR-BCA* involved in the utilization of catechol was isolated from *H. organivorans*. The genes *catA*, *catB*, *catC*, and *catR* that encode for 1,2-CAT, *cis,cis*-muconate cycloisomerase, muconolactone delta-isomerase and a LysR-type transcriptional regulator, respectively, were detected. Downstream of these genes were flanked by the benzoate catabolic genes, *benA* and *benB* that code for large and small subunit of benzoate 1, 2 dioxygenase, respectively. This gene organization in *H. organivorans* was found to be similar to that of the catabolic genes identified in other non-halophilic eubacteria. Abdelkafi et al. (2006) studied the metabolism of *p*-coumaric acid by *Halomonas* strain IMPC under halophilic conditions. Strain IMPC degraded *p*-coumaric acid to *p*-hydroxybenzaldehyde, *p*-hydroxybenzoic acid and then to protocatechuic acid as the final aromatic product before ring fission. The identity of these intermediates was confirmed using a gas chromatography and mass-spectrometry (GC-MS). Kim et al. (2008) isolated a benzoate- and *p*-hydroxybenzoate metabolizing halophile, *Chromohalobacter* sp. strain HS-2. Using a combination of molecular and biochemical approaches, these researchers have elucidated the catabolic pathways for benzoate and *p*-hydroxybenzoate in HS-2. Their work showed that benzoate induces the expression of benzoate 1,2-dioxygenase, 1,2-CAT, *p*-hydroxybenzoate hydroxylase (*pobA*), and 3,4-PCA, while *p*-hydroxybenzoate only induced the expression of *pobA*. Interestingly, the role of *pobA* and 3,4-PCA genes in benzoate grown HS-2 cells is not clear because benzoate is usually degraded via catechol by 1,2-CAT or 2,3-CAT. Dastgheib et al. (2011) have obtained a mixed culture, Qphe-SubIV consisting of only two organisms, *Halomonas* sp and *Marinobacter* sp. These organisms degrade phenanthrene. Metabolite analysis showed that 2-hydroxy 1-naphthoic acid and 2-naphthol were among the major metabolites accumulated in the culture media, indicating that an initial dioxygenation step might have proceeded by a novel mechanism at C1 and C2 positions. Recently, Dalvi et al. (2012) have analyzed the draft genome sequence of the extremely halophilic benzene and toluene degrading *Arhodomonas* sp. strain Seminole. The analysis predicted 13 putative genes that encode upper and lower pathway enzymes for aromatic compound degradation. These proteins share 44–77% sequence identity with proteins previously described in non-halophilic organisms. The results indicate that benzene is converted to phenol and then to catechol in two steps by monooxygenase-like enzymes closely related to phenol hydroxylases. Thus, formed catechol undergoes ring cleavage via the *meta* pathway by 2, 3-CAT to form 2-hydroxymuconic semialdehyde, which subsequently enters the tricarboxylic acid cycle. To corroborate these predicted enzymes that benzene is converted to first phenol and then to catechol prior to ring cleavage by 2,3-CAT, the authors grew a closely related species *Arhodomonas* sp. strain Rozel on deuterated benzene and deuterated phenol was detected by GC-MS as the initial intermediate of benzene degradation. A 2-D

gel electrophoresis and Tandem mass-spectrometry has identified the phenol hydroxylase-like and 2,3-CAT in the cell extract of strain Rozel grown on benzene as the sole carbon source. More recently, Bonfá et al. (2013) showed the presence of 1,2-CTD and 3,4-PCD genes in three of phenol degrading bacteria, *Halomonas organivorans*, *Arhodomonas aquaeolei*, and *Modicisalibacter tunisie*.

A few recent studies have provided information about aliphatic hydrocarbon degradation in saline environments. Dastgheib et al. (2011) have used PCR and degenerate primers to amplify two putative *alkB* genes that code for alkane hydroxylases needed for the hydroxylation of aliphatic hydrocarbons in *Alcanivorax* sp. strain Qtet3. The strain Qtet3 degrades a wide range of *n*-alkanes in the presence of 0–15% NaCl. More recently, Nie et al. (2013) have analyzed the full genome of the alkane-metabolizing *Amycolicoccus subflavus* isolated from an oily sludge at Daqing Oilfield, China (Wang et al., 2010). The organism grew utilizing C<sub>10</sub>–C<sub>36</sub> alkanes as the sole sources of carbon in the presence of 1–12% NaCl. Four types of alkane hydroxylase coding genes were identified in the genome. A quantitative real-time reverse transcription PCR was used to determine the induction of various alkane-degrading genes. Homologs of *AlkB* alkane hydroxylases were induced by C<sub>10</sub>–C<sub>36</sub> alkanes with maximum expression in the presence of C<sub>16</sub>–C<sub>24</sub>. Similarly, cytochrome P450 CYP153 genes were upregulated by alkanes, C<sub>10</sub>–C<sub>20</sub> and C<sub>24</sub>. In addition, *LadA* and propane monooxygenase genes responsible for the oxidation of C<sub>16</sub>–C<sub>36</sub> and propane, respectively, were also detected. Interestingly, analysis showed that key genes necessary for the degradation of aromatic compounds were missing in the genome. These physiological, genomic, and transcriptional analyses clearly reveals the *Amycolicoccus subflavus*'s potential to utilize a range of *n*-alkanes typically found in crude oil.

A few reports also exist in the literature on the degradation mechanism of hydrocarbons by archaea in the presence of high salt. For example, an extremely halophilic archaeon, *Halofexax* sp. strain D1227 that degrades benzoate, cinnamate, and phenylpropanoate, was shown to possess 1,2-GDO (Emerson et al., 1994; Fu and Oriel, 1998). Fairley et al. (2002, 2006) also found a closely related gene encoding 1,2-GDO in 4-hydroxybenzoate-degrading *Haloarcula* sp. strain D1. A recent study reported the isolation of nine archaeal isolates belonging to various genera that degraded *p*-hydroxybenzoate, naphthalene, phenanthrene, and pyrene as the sole carbon and energy sources in the presence of 20% NaCl. This study showed that the isolates possessed genes that encode 1,2-CAT and 3,4-PCA and the expression of these genes was measured spectrophotometrically (Erdogmuş et al., 2013).

Overall, these recent few studies show that microorganisms in high-salinity environments degrade hydrocarbons using enzymes and steps similar to those found in non-halophiles. However, in-depth studies are needed to obtain greater insights into degradation pathways and steps leading to intermediates that enter central metabolism. In addition, molecular studies can help develop specific probes to identify and monitor specific degradative organisms in the environment and their *in-situ* activity.

## CONCLUSIONS

As summarized in this review, knowledge on the ability of microorganisms capable of degrading hydrocarbons in hypersaline environments has accumulated significantly in the past two decades. Studies show that much richer microbial diversity exists in the environment that can efficiently degrade hydrocarbons over a broad range of salinity. Among microbial taxa, members of the genus *Halomonas*, *Marinobacter*, and *Alcanivorax* are common inhabitants of high salinity environments with the potential to degrade variety hydrocarbons. Among archaea, *Haloferax*, *Haloarchula*, and *Haloarchaeum* seem to play important role in the degradation of hydrocarbons, especially in extremely high salinity conditions. The implication that pigment-mediated ATP synthesis help archaea better survive and degrade hydrocarbons in oxygen deficient hypersaline environments explains the dominance of these organisms in such environments. Martins and Peixoto (2012) have suggested that halophilic photoautotrophs can be a critical factor for the degradation of hydrocarbons since their activity could compensate the lack of oxygen imposed by hypersalinity. In any case, the synergistic interactions between photosynthetic organisms and hydrocarbon degrading bacteria or archaea can lead to effective biodegradation of hydrocarbons in hypersaline environments is noteworthy and needs further investigation. Studies have revealed that many organisms are capable of degrading a mixture of hydrocarbons in widely fluctuating salinities and some produce surfactants and also some fix nitrogen thus underscoring the importance of such microbes in the cleanup of contaminated sites. Though appreciable progress has been made recently in understanding diversity of microorganisms responsible for hydrocarbon degradation under aerobic conditions, similar information under anaerobic condition is lacking. Also, information on genes, enzymes and molecular mechanism of hydrocarbon degradation, in high salinity environments is not fully understood. A few recent studies have shown that the degradation of hydrocarbons in moderate and to high salinity occurs using enzymes described for many non-halophiles. Recent advances in high-throughput DNA sequencing are providing new tools and capabilities for discovering novel hydrocarbon-degrading microorganisms, especially with new dioxygenases. A better knowledge of the diversity of catabolic pathways would certainly bring valuable information for the development of robust bioremediation processes for hypersaline environments.

## ACKNOWLEDGMENTS

This work was supported by funding from the Oklahoma Transportation Center (grant DTRT06-G-0016) and by the National Science Foundation (grant OCE1049301).

## REFERENCES

- Abdelkafi, S., Labat, M., Casalot, L., Chamkha, M., and Sayadi, S. (2006). Isolation and characterization of *Halomonas* sp. strain IMPC, a p-coumaric acid-metabolizing bacterium that decarboxylates other cinnamic acids under hypersaline conditions. *FEMS Microbiol. Lett.* 255, 108–114. doi: 10.1111/j.1574-6968.2005.00061.x
- Abed, R. M. M., Al-Thukair, A., and De Beer, D. (2006). Bacterial diversity of a cyanobacterial mat degrading petroleum compounds at elevated salinities and temperatures. *FEMS Microbiol. Ecol.* 57, 290–301. doi: 10.1111/j.1574-6941.2006.00113.x
- Al-Mailem, D. M., Elias, M., and Radwan, S. S. (2012). Enhanced haloarchaeal oil removal in hypersaline environments via organic nitrogen fertilization and illumination. *Extremophiles* 16, 751–758. doi: 10.1007/s00792-012-0471-y
- Al-Mailem, D. M., Elias, M., and Radwan, S. S. (2013). Oil-bioremediation potential of two hydrocarbonoclastic, diazotrophic *Marinobacter* strains from hypersaline areas along the Arabian Gulf coasts. *Extremophiles* 17, 463–470. doi: 10.1007/s00792-013-0530-z
- Al-Mailem, D. M., Sorkhoh, N. A., Al-Awadhi, H., Elias, M., and Radwan, S. S. (2010). Biodegradation of crude oil and pure hydrocarbons by extreme halophilic archaea from hypersaline coasts of the Arabian Gulf. *Extremophiles* 14, 321–328. doi: 10.1007/s00792-010-0312-9
- Al-Mallah, M., Goutx, M., Mille, G., and Bertrand, J.-C. (1990). Production of emulsifying agents during Growth of a marine *Alteromonas* in sea water with eicosane as carbon source, a solid hydrocarbon. *Oil Chem. Pollut.* 6, 289–305. doi: 10.1016/S0269-8579(05)80005-X
- Al-Mueini, R., Al-Dalali, M., Al-Amri, I. S., and Patzelt, H. (2007). Hydrocarbon degradation at high salinity by a novel extremely halophilic actinomycete. *Environ. Chem.* 4, 5–7. doi: 10.1071/EN06019
- Alva, V., and Peyton, B. M. (2003). Phenol and catechol biodegradation by the haloalkaliphile *Halomonas campialis*: influence of pH and salinity. *Environ. Sci. Technol.* 37, 4397–4402. doi: 10.1021/es0341844
- Ashok, T., Saxena, S., and Musarrat, J. (1995). Isolation and characterization of four polycyclic aromatic hydrocarbon degrading bacteria from soil near an oil refinery. *Lett. Appl. Microbiol.* 21, 246–248. doi: 10.1111/j.1472-765X.1995.tb01052.x
- Azetsu, S., Nicholson, C., Najar, F., Roe, B., and Fathepure, B. (2009). “Physiological and genomic analysis of BTEX degradation in novel *Arhodomonas* strains, isolated from hypersaline environments,” in *109th American Society for Microbiology* (Philadelphia, PA).
- Bastos, A. E. R., Moon, D. H., Rossi, A., Trevors, J. T., and Tsai, S. M. (2000). Salt-tolerant phenol-degrading microorganisms isolated from Amazonian soil samples. *Arch. Microbiol.* 174, 346–352. doi: 10.1007/s002030000216
- Berlendis, S., Cayol, J.-L., Verhe, F., Laveau, S., Tholozan, J.-C., Ollivier, B., et al. (2010). First evidence of aerobic biodegradation of BTEX compounds by pure cultures of *Marinobacter*. *Appl. Biochem. Biotechnol.* 160, 1992–1999. doi: 10.1007/s12100-009-8746-1
- Bertrand, J.-C., Almallah, M., Acquaviva, M., and Mille, G. (1990). Biodegradation of hydrocarbons by an extremely halophilic archaeabacterium. *Lett. Appl. Microbiol.* 11, 260–263. doi: 10.1111/j.1472-765X.1990.tb00176.x
- Bonfá, M. R. L., Grossman, M. J., Mellado, E., and Durrant, L. R. (2011). Biodegradation of aromatic hydrocarbons by Haloarchaea and their use for the reduction of the chemical oxygen demand of hypersaline petroleum produced water. *Chemosphere* 84, 1671–1676. doi: 10.1016/j.chemosphere.2011.05.005
- Bonfá, M. R. L., Grossman, M. J., Piubeli, F., Mellado, E., and Durrant, L. R. (2013). Phenol degradation by halophilic bacteria isolated from hypersaline environments. *Biodegradation* 24, 699–709. doi: 10.1007/s10532-012-9617-y
- Borzenkov, I. A., Milekhina, E. I., Gotoeva, M. T., Rozanova, E. P., and Belyaev, S. S. (2006). The properties of hydrocarbon-oxidizing bacteria isolated from the oilfields of Tatarstan, Western Siberia, and Vietnam. *Microbiology* 75, 66–72. doi: 10.1134/S0026261706010127
- Cao, B., Nagarajan, K., and Loh, K.-C. (2009). Biodegradation of aromatic compounds: current status and opportunities for biomolecular approaches. *Appl. Microbiol. Biotechnol.* 85, 207–228. doi: 10.1007/s00253-009-2192-4
- Chamkha, M., Mnif, S., and Sayadi, S. (2008). Isolation of a thermophilic and halophilic tyrosol-degrading *Geobacillus* from a Tunisian high-temperature oil field. *FEMS Microbiol. Lett.* 283, 23–29. doi: 10.1111/j.1574-6968.2008.01136.x
- Cuadros-Orellana, S., Pohlschroder, M., Grossman, M. J., and Durrant, L. R. (2012). Biodegradation of aromatic compounds by a halophilic archaeon isolated from the Dead Sea. *Chem. Eng. Trans.* 27, 13–18. doi: 10.3303/cet1227003
- Cuadros-Orellana, S., Pohlschroder, M., and Durrant, L. R. (2006). Isolation and characterization of halophilic archaea able to grow in aromatic compounds. *Int. Biodeter. Biodegrad.* 57, 151–154. doi: 10.1016/j.ibiod.2005.04.005
- Dalvi, S., Azetsu, S., Patrauchan, M. A., Aktas, D. F., and Fathepure, B. Z. (2012). Proteogenomic elucidation of the initial steps in the benzene degradation pathway of a novel halophile, *Arhodomonas* sp. Strain Rozel, isolated

- from a hypersaline environment. *Appl. Environ. Microbiol.* 78, 7309–7316. doi: 10.1128/AEM.01327-12
- Dastgheib, S. M. M., Amoozegar, M. A., Khajeh, K., Shavandi, M., and Ventosa, A. (2012). Biodegradation of polycyclic aromatic hydrocarbons by a halophilic microbial consortium. *Appl. Microbiol. Biotechnol.* 95, 789–798. doi: 10.1007/s00253-011-3706-4
- Dastgheib, S. M. M., Amoozegar, M. A., Khajeh, K., and Ventosa, A. (2011). A halotolerant *Alcanivorax* sp. strain with potential application in saline soil remediation. *Appl. Microbiol. Biotechnol.* 90, 305–312. doi: 10.1007/s00253-010-3049-6
- Diaz, M. P., Boyd, K. G., Grigson, S. J. W., and Burgess, J. G. (2002). Biodegradation of crude oil across a wide range of salinities by an extremely halotolerant bacterial consortium MPD-M, immobilized onto polypropylene fibres. *Biotechnol. Bioeng.* 79, 145–153. doi: 10.1002/bit.10318
- Diaz, M. P., Grigson, S. J. W., Peppiatt, C. J., and Burgess, J. G. (2000). Isolation and characterization of novel hydrocarbon-degrading euryhaline consortia from crude oil and mangrove sediments. *Mar. Biotechnol.* 2, 522–532. doi: 10.1007/s101260000037
- Duran, R. (2010). “Marinobacter,” in *Handbook of Hydrocarbon and Lipid Microbiology*, ed K. N. Timmis (Berlin, Heidelberg: Springer-Verlag), 1726–1733. doi: 10.1007/978-3-540-77587-4\_122
- Emerson, D., Chauhan, S., Oriel, P., and Breznak, J. A. (1994). *Haloferax* sp. D1227, a halophilic archaeon capable of growth on aromatic compounds. *Arch. Microbiol.* 161, 445–452. doi: 10.1007/BF00307764
- Erdogmuş, S. F., Mutlu, B., Korcan, S. E., Guven, K., and Konuk, M. (2013). Aromatic hydrocarbon degradation by halophilic archaea isolated from Çamaltı Saltern, Turkey. *Water Air Soil Pollut.* 224, 1449. doi: 10.1007/s11270-013-1449-9
- Fairley, D. J., Boyd, D. R., Sharma, N. D., Allen, C. C., Morgan, P., and Larkin, M. J. (2002). Aerobic metabolism of 4-hydroxybenzoic acid in Archaea via an unusual pathway involving an intramolecular migration (NIH shift). *Appl. Environ. Microbiol.* 68, 6246–6255. doi: 10.1128/AEM.68.12.6246-6255.2002
- Fairley, D. J., Wang, G., Rensing, C., Pepper, I. L., and Larkin, M. J. (2006). Expression of gentisate 1,2-dioxygenase (gdoA) genes involved in aromatic degradation in two haloarchaeal genera. *Appl. Microbiol. Biotechnol.* 73, 691–695. doi: 10.1007/s00253-006-0509-0
- Feng, L., Wang, W., Cheng, J., Ren, Y., Zhao, G., Gao, C., et al. (2007). Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5602–5607. doi: 10.1073/pnas.0609650104
- Fernandez-Linares, L., Acuaviva, M., Bertrand, J.-C., and Gauthier, M. (1996). Effect of sodium chloride concentration on growth and degradation of eicosane by the marine halotolerant bacterium *Marinobacter hydrocarbonoclasticus*. *System Appl. Microbiol.* 19, 113–121. doi: 10.1016/S0723-2020(96)80018-X
- Fowler, S. W., Readman, J. W., Oregoni, B., Villeneuve, J. P., and McKay, K. (1993). Petroleum hydrocarbons and trace metals in nearshore gulf sediments and biota before and after the 1991 war - an assessment of temporal and spatial trends. *Mar. Pollut. Bull.* 27, 171–182. doi: 10.1016/0025-326X(93)90022-C
- Fu, W., and Oriel, P. (1998). Gentisate 1,2-dioxygenase from *Haloferax* sp. D1227. *Extremophiles* 2, 439–436. doi: 10.1007/s007920050090
- Fu, W., and Oriel, P. (1999). Degradation of 3-phenylpropionic acid by *Haloferax* sp. D1227. *Extremophiles*. 3, 45–53. doi: 10.1007/s007920050098
- Fuchs, G., Boll, M., and Heider, J. (2011). Microbial degradation of aromatic compounds—from one strategy to four. *Nat. Rev. Microbiol.* 9, 803–816. doi: 10.1038/nrmicro2652
- Gao, W., Cui, Z., Li, Q., Xu, G., Jia, X., and Zheng, L. (2013). *Marinobacter nanhaiiticus* sp. nov., polycyclic aromatic hydrocarbon-degrading bacterium isolated from the sediment of the South China Sea. *Antonie Van Leeuwenhoek* 103, 485–491. doi: 10.1007/s10482-012-9830-z
- Garcia, M. T., Gallego, V., Ventosa, A., and Mellado, E. (2005a). *Thalassobacillus devorans* gen. nov., sp. nov., a moderately halophilic, phenol-degrading, Gram-positive bacterium. *Int. J. Syst. Evol. Microbiol.* 55, 1789–1795. doi: 10.1099/ijss.0.63560-0
- Garcia, M. T., Mellado, E., Ostos, J. C., and Ventosa, A. (2004). *Halomonas organivorans* sp. nov., a moderate halophile able to degrade aromatic compounds. *Int. J. Syst. Evol. Microbiol.* 54, 1723–1728. doi: 10.1099/ijss.0.63114-0
- Garcia, M. T., Ventosa, A., and Mellado, E. (2005b). Catabolic versatility of aromatic compound-degrading halophilic bacteria. *FEMS Microbiol. Ecol.* 54, 97–109. doi: 10.1016/j.femsec.2005.03.009
- Gauthier, M. J., Lafay, B., Christen, R., Fernandez, L., Acuaviva, M., Bonin, P., et al. (1992). *Marinobacter hydrocarbonoclasticus* gen. nov., sp. nov., a new, extremely halotolerant, hydrocarbon- degrading marine bacterium. *Int. J. Syst. Bacteriol.* 42, 568–576. doi: 10.1099/00207713-42-4-568
- Gibbs, G. W. (1997). Estimating residential polycyclic aromatic hydrocarbon (PAH) related lung cancer risk using occupational data. *Ann. Occup. Hyg.* 41, 49–53. doi: 10.1093/annhyg/41.inhaled\_particles\_VIII.49
- Harayama, S., Kishira, H., Kasai, Y., and Shutsubo, K. (1999). Petroleum biodegradation in marine environments. *J. Mol. Microbiol. Biotechnol.* 1, 63–70.
- Harwood, C. S., and Parales, R. E. (1996). The beta-ketoadipate pathway and the biology of self identity. *Ann. Rev. Microbiol.* 50, 553–590. doi: 10.1146/annurev.micro.50.1.553
- Hassan, H. A., Rizk, N. M. H., Hefnawy, M. A., and Awad, A. M. (2012). Isolation and characterization of halophilic aromatic and chloroaromatic degrader from Wadi El-Natrun Soda lakes. *Life Sci. J.* 9, 1565–1570.
- Head, I. M., Jones, D. M., and Roling, W. F. (2006). Marine microorganisms make a meal of oil. *Nat. Rev. Microbiol.* 4, 173–182. doi: 10.1038/nrmicro1348
- Head, I. M., and Swannell, R. P. (1999). Bioremediation of petroleum hydrocarbon contaminants in marine habitats. *Curr. Opin. Biotechnol.* 10, 234–239. doi: 10.1016/S0958-1669(99)80041-X
- Hinteregger, C., and Streischberg, F. (1997). *Halomonas* sp., a moderately halophilic strain, for biotreatment of saline phenolic waste-water. *Biotechnol. Lett.* 19, 1099–1102. doi: 10.1023/A:1018488410102
- Holden, P. A., LaMontagne, M. G., Bruce, A. K., Miller, W. G., and Lindow, S. E. (2002). Assessing the role of *Pseudomonas aeruginosa* surface-active gene expression in hexadecane biodegradation in sand. *Appl. Environ. Microbiol.* 68, 2509–2518. doi: 10.1128/AEM.68.5.2509-2518.2002
- Huu, N. B., Denner, E. B. M., Ha, D. T. C., Wanner, G., and Stan-Lotter, H. (1999). *Marinobacter aquaeolei* sp. nov., a halophilic bacterium isolated from a Vietnamese oil-producing well. *Int. J. Syst. Bacteriol.* 49, 367–375. doi: 10.1099/00207713-49-2-367
- Kim, D., Kim, S. W., Choi, K. Y., Lee, J. S., and Kim, E. (2008). Molecular cloning and functional characterization of the genes encoding benzoate and p-hydroxybenzoate degradation by the halophilic *Chromohalobacter* sp. strain HS-2. *FEMS Microbiol. Lett.* 280, 235–241. doi: 10.1111/j.1574-6968.2008.01067.x
- Kleinsteuber, S., Müller, R. H., and Babel, W. (2001). Expression of the 2,4-D degradative pathway of pJP4 in an alkaliphilic, moderately halophilic soda lake isolate, *Halomonas* sp. EF43. *Extremophiles* 5, 375–384. doi: 10.1007/s007920100202
- Kulichevskaya, I. S., Milekhina, E. I., Borzenkov, I. A., Zvyagintseva, I. S., and Belyaev, S. S. (1992). Oxidation of petroleum hydrocarbons by extremely halophilic archaeabacteria. *Microbiology* 60, 596–601.
- Kumar, M., Vladimir, L., de Sistro Materano, A., and Ilzins, O. A. (2007). A halotolerant and thermotolerant *Bacillus* sp. degrades hydrocarbons and produces tension active emulsifying agent. *World J. Microbiol. Biotechnol.* 23, 211–220. doi: 10.1007/s11274-006-9215-4
- Kushner, D. J. (1978). “Life in high salt and solute concentrations: halophilic bacteria,” in *Microbial Life in Extreme Environments*, ed D. J. Kushner (London: Academic Press, Ltd.), 317–368.
- Kuznetsov, V. D., Zaitseva, T. A., Vakulenko, L. V., and Filippova, S. N. (1992). *Streptomyces albiaxialis* sp. nov.: a new petroleum hydrocarbon-degrading species of thermo- and halotolerant *Streptomyces*. *Microbiology* 61, 62–67.
- Lack, L. (1959). The enzymatic oxidation of gentisic acid. *Biochem. Biophys. Acta* 34, 117–123. doi: 10.1016/0006-3002(59)90239-2
- Leahy, J. G., and Colwell, R. R. (1990). Microbial degradation of hydrocarbons in the environment. *Microbiol. Rev.* 54, 305–315.
- Le Borgne, S., Paniagua, D., and Vazquez-Duhalt, R. (2008). Biodegradation of organic pollutants by halophilic bacteria and archaea. *J. Mol. Microbiol. Biotechnol.* 15, 74–92. doi: 10.1159/000121323
- Li, H., Liu, Y. H., Luo, N., Zhang, X. Y., Luan, T. G., Hu, J. M., et al. (2006). Biodegradation of benzene and its derivatives by a psychrotolerant and moderately haloalkaliphilic *Planococcus* sp. strain ZD22. *Res. Microbiol.* 157, 629–636. doi: 10.1016/j.resmic.2006.01.002

- Luz, A. P., Pellizari, V. H., Whyte, L. G., and Greer, C. W. (2004). A survey of indigenous microbial hydrocarbon degradation genes in soils from Antarctica and Brazil. *Can. J. Microbiol.* 50, 323–333. doi: 10.1139/w04-008
- Martins, L. F., and Peixoto, R. S. (2012). Biodegradation of petroleum hydrocarbons in hypersaline environments. *Braz. J. Microbiol.* 43, 865–872. doi: 10.1590/S1517-83822012000300003
- McGenity, T. J. (2010). “Halophilic hydrocarbon degraders,” in *Handbook of Hydrocarbon and Lipid Microbiology*, ed K. N. Timmis (Berlin; Heidelberg: Springer-Verlag), 1939–1948.
- McGenity, T. J., Folwell, B. D., McKew, B. A., and Sanni, G. O. (2012). Marine crude-oil biodegradation: a central role for interspecies interactions. *Aquat. Biosyst.* 8, 10. doi: 10.1186/2046-9063-8-10
- Menzie, C. A., Potocki, B. B., and Santodonato, J. (1992). Exposure to carcinogenic PAH in the environment. *Environ. Sci. Technol.* 26, 1278–1284. doi: 10.1021/es00031a002
- Mnif, S., Chamkha, M., Labat, M., and Sayadi, S. (2011). Simultaneous hydrocarbon biodegradation and biosurfactant production by oilfield-selected bacteria. *J. Appl. Microbiol.* 111, 525–536. doi: 10.1111/j.1365-2672.2011.05071.x
- Mnif, S., Chamkha, M., and Sayadi, S. (2009). Isolation and characterization of *Halomonas* sp. strain C2SS100, a hydrocarbon-degrading bacterium under hypersaline conditions. *J. Appl. Microbiol.* 107, 785–794. doi: 10.1111/j.1365-2672.2009.04251.x
- Moreno, M. D. L., Sanchez-Porro, C., Piubeli, F., Frias, L., Garcia, M. T., and Mellado, E. (2011). Cloning, characterization and analysis of *cat* and *ben* genes from the phenol degrading halophilic bacterium *Halomonas organivorans*. *PLoS ONE* 6:e21049. doi: 10.1371/journal.pone.0021049
- Neff, J., Lee, K., and DeBlois, E. M. (2011). “Produced water: overview of composition, fates, and effects,” in *Produced Water: Environmental Risks and Advances in Mitigation Technologies* eds K. Lee and J. Neff (New York, NY: Springer Science+Business Media, LLC), 3–54. doi: 10.1007/978-1-4614-0046-2\_1
- Nicholson, C. A., and Fathepure, B. Z. (2004). Biodegradation of benzene by halophilic and halotolerant bacteria under aerobic conditions. *Appl. Environ. Microbiol.* 70, 1222–1225. doi: 10.1128/AEM.70.2.1222-1225.2004
- Nicholson, C. A., and Fathepure, B. Z. (2005). Aerobic biodegradation of benzene and toluene under hypersaline conditions at the Great Salt Plains, Oklahoma. *FEMS Microbiol. Lett.* 245, 257–262. doi: 10.1016/j.femsle.2005.03.014
- Nicholson, C. A., and Fathepure, B. Z. (2006). “Isolation and characterization of a halophile that degrades benzene as the sole carbon and energy source,” in *106th American Society for Microbiology* (Orlando, FL).
- Nie, Y., Fang, H., Li, Y., Chi, C. Q., Tang, Y. Q., and Wu, X. L. (2013). The genome of the moderate halophile *Amycolicicoccus subflavus* DQS3-9A1T reveals four alkane hydroxylation systems and provides some clues on the genetic basis for its adaptation to a petroleum environment. *PLoS ONE* 8:e70986. doi: 10.1371/journal.pone.0070986
- Obuekwe, C. O., Badrudeen, A. M., Al-Saleh, E., and Mulder, J. L. (2005). Growth and hydrocarbon degradation by three desert fungi under conditions of simultaneous temperature and salt stress. *Intern. Biodeter. Biodegrad.* 56, 197–205. doi: 10.1016/j.ibiod.2005.05.005
- Oren, A. (2013). “Life at high salt concentrations,” in *The Prokaryotes – Prokaryotic Communities and Ecophysiology*, eds E. Rosenberg, E. F. Delong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin; Heidelberg: Springer-Verlag), 421–440. doi: 10.1007/978-3-642-30123-0\_57
- Oren, A., Gurevich, P., Azachi, M., and Hents, Y. (1992). Microbial degradation of pollutants at high salt concentrations. *Biodegradation* 3, 387–398. doi: 10.1007/BF00129095
- Oie, C. S., Albaugh, C. E., and Peyton, B. M. (2007). Benzoate and salicylate degradation by *Halomonas campialis*, an alkaliphilic and moderately halophilic microorganism. *Water Res.* 41, 1235–1242. doi: 10.1016/j.watres.2006.12.029
- Patzelt, H. (2005). “Hydrocarbon degradation under hypersaline conditions, some facts, some experiments and many open questions,” in *Adaptation to Life at High Salt Concentrations in Archaea, Bacteria, and Eukarya*, eds Gunde-Cimerman et al. (Netherlands: Springer publishers), 105–122. doi: 10.1007/1-4020-3633-7\_8
- Patzelt, H. (2007). “Hydrocarbon metabolism,” in *Encyclopedia of Life Sciences* (John Wiley and Sons, Ltd), 1–9. doi: 10.1002/9780470015902.a0000473
- Pérez-Pantoja, D., González, B., and Pieper, D. H. (2010). “Aerobic degradation of aromatic hydrocarbons,” in *Handbook of Hydrocarbon and Lipid Microbiology*, ed K. N. Timmis (Berlin: Springer-Verlag), 799–837. doi: 10.1007/978-3-540-77587-4\_60
- Pernetti, M., and Di Palma, L. (2005). Experimental evaluation of inhibition effects of saline wastewater on activated sludge. *Environ. Technol.* 26, 695–703. doi: 10.1080/09593330.2001.9619509
- Philip, J. C., Bamforth, S. M., Singleton, I., and Atlas, R. M. (2005). “Environmental pollution and restoration: a role for bioremediation,” in *Bioremediation: Applied Microbial Solutions for Real-World Environmental Cleanup* eds R. M. Atlas and J. Philip (Washington, DC: ASM Press), 1–48.
- Plotnikova, E. G., Alyntseva, O. V., Kosheleva, I. A., Puntus, I. F., Filonov, A. E., Gavrilov Elu, et al. (2001). Bacterial degraders of polycyclic aromatic hydrocarbons isolated from salt-contaminated soils and bottom sediments in salt mining areas. *Microbiology* 70, 51–58. doi: 10.1023/A:1004892804670
- Plotnikova, E. G., Yastrebova, O. V., Anan’ina, L. N., Dorofeeva, L. V., Lysanskaya, V. Ya., and Demakov, V. A. (2011). Halotolerant bacteria of the genus *Arthrobacter* degrading polycyclic aromatic hydrocarbons. *Russian J. Ecol.* 42, 502–509. doi: 10.1134/S1067413611060130
- Reineke, W. (2001). “Aerobic and anaerobic biodegradation potentials of microorganisms,” in *The Handbook of Environmental Chemistry*. Vol. 2 Part K Biodegradation and Persistence, ed B. Beek (Berlin; Heidelberg: Springer-Verlag), 1–140.
- Riis, V., Kleinsteuber, S., and Babel, W. (2003). Influence of high salinities on the degradation of diesel fuel by bacteria consortia. *Can. J. Microbiol.* 49, 713–721. doi: 10.1139/w03-083
- Rosenberg, A. (1983). *Pseudomonas halodurans* sp. nov., a halotolerant bacterium. *Arch. Microbiol.* 136, 117–123. doi: 10.1007/BF00404785
- Sass, A. M., McKew, B. A., Sass, H., Fichtel, J., Timmis, K. N., and McGinity, T. J. (2008). Diversity of *Bacillus*-like organisms isolated from deep-sea hypersaline anoxic sediments. *Saline Syst.* 4, 8. doi: 10.1186/1746-1448-4-8
- Sei, A., and Fathepure, B. Z. (2009). Biodegradation of BTEX at high salinity by an enrichment culture from hypersaline sediments of Rozel Point at Great Salt Lake. *J. Appl. Microbiol.* 107, 2001–2008. doi: 10.1111/j.1365-2672.2009.04385.x
- Tapilatu, Y. H., Grossi, V., Acquaviva, M., Milton, C., Bertrand, J.-C., and Cuny, P. (2010). Isolation of hydrocarbon-degrading extremely halophilic archaea from an uncontaminated hypersaline pond (Camargue, France). *Extremophiles* 14, 225–231. doi: 10.1007/s00792-010-0301-z
- Throne-Holst, M., Wentzel, A., Ellingsen, T. E., Kotlar, H. K., and Zotchev, S. B. (2007). Identification of novel genes involved in long-chain n-alkane degradation by *Acinetobacter* sp. strain DSM 17874. *Appl. Environ. Microbiol.* 73, 3327–3332. doi: 10.1128/AEM.00064-07
- van Beilen, J. B., and Funhoff, E. G. (2007). Alkane hydroxylases involved in microbial alkane degradation. *Appl. Microbiol. Biotechnol.* 74, 13–21. doi: 10.1007/s00253-006-0748-0
- Van Hamme, J. D., Singh, A. M., and Ward, O. P. (2003). Recent advances in petroleum microbiology. *Microbiol. Mol. Biol. Rev.* 67, 503–549. doi: 10.1128/MMBR.67.4.503-549.2003
- Veil, J. H., Puder, M. G., Elcock, D., and Redweik, R. J. (2004). “A white paper describing produced water from production of crude oil, natural gas, and coal bed methane,” in *U.S. Department of Energy, National Energy Technology Laboratory. Prepared by Argonne National Laboratory* (Washington, DC), 1–87.
- Ventosa, A., and Nieto, J. J. (1995). Biotechnological applications and potentialities of halophilic microorganisms. *World J. Microbiol. Biotechnol.* 11, 85–94. doi: 10.1007/BF00339138
- Wang, Y. N., Chi, C. Q., Cai, M., Lou, Z. Y., Tang, Y. Q., Zhi, X. Y., et al. (2010). *Amycolicicoccus subflavus* gen. nov., sp. nov., an actinomycete isolated from a saline soil contaminated by crude oil. *Int. J. Syst. Evol. Microbiol.* 60, 638–643. doi: 10.1099/ijss.0.010546-0
- Ward, D. M., and Brock, T. D. (1978). Hydrocarbon biodegradation in hypersaline environments. *Appl. Environ. Microbiol.* 35, 353–359.
- Woolard, C. R., and Irvine, R. L. (1995). Treatment of hypersaline wastewater in the sequencing batch reactor. *Water Res.* 29, 1159–1168. doi: 10.1016/0043-1354(94)00239-4
- Yemashova, N. A., Murygina, V. P., Zhukov, D. V., Zakharyantz, A. A., Gladchenko, M. A., Appanna, V., et al. (2007). Biodeterioration of crude oil and oil derived products: a review. *Rev. Environ. Sci. Biotechnol.* 6, 315–337. doi: 10.1007/s11157-006-9118-8
- Wang, Y. N., Cai, H., Chi, C. Q., Lu, A. H., Lin, X. G., Jian, Z. F., et al. (2007). *Halomonas shengliensis* sp. nov., a moderately halophilic, denitrifying, crude-oil-utilizing bacterium. *Int. J. Syst. Evol. Microbiol.* 57, 1222–1226. doi: 10.1099/ijss.0.64973-0

- Zhao, B., Wang, H., Mao, X., and Li, R. (2009). Biodegradation of phenanthrene by a halophilic bacterial consortium under aerobic conditions. *Curr. Microbiol.* 58, 205–210. doi: 10.1007/s00284-008-9309-3
- Zvyagintseva, I. S., Belyaev, S. S., Borzenkov, I. A., Kostrikina, N. A., Milekhina, E. I., and Ivanov, M. V. (1995). Halophilic archaeabacteria from the Kalamkass oil field. *Microbiology* 64, 67–71.
- Zvyagintseva, I. S., Poglasova, M. N., Gotoeva, M. T., and Belyaev, S. S. (2001). Effect of the medium salinity on oil degradation by Nocardioform bacteria. *Microbiology* 70, 652–656. doi: 10.1023/A:1013179513922

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 08 January 2014; accepted: 30 March 2014; published online: 23 April 2014.

Citation: Fathepure BZ (2014) Recent studies in microbial degradation of petroleum hydrocarbons in hypersaline environments. *Front. Microbiol.* 5:173. doi: 10.3389/fmicb.2014.00173

This article was submitted to Extreme Microbiology, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Fathepure. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Mining a database of single amplified genomes from Red Sea brine pool extremophiles—improving reliability of gene function prediction using a profile and pattern matching algorithm (PPMA)

**Stefan W. Grötzingler<sup>1</sup>, Intikhab Alam<sup>2</sup>, Wail Ba Alawi<sup>2</sup>, Vladimir B. Bajic<sup>2</sup>, Ulrich Stingl<sup>3</sup> and Jörg Eppinger<sup>1\*</sup>**

<sup>1</sup> Division of Physical Sciences and Engineering, KAUST Catalysis Center, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia

<sup>2</sup> Division of Biological Sciences and Engineering, Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia

<sup>3</sup> Division of Biological Sciences and Engineering, Red Sea Research Center, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia

**Edited by:**

R. Thane Papke, University of Connecticut, USA

**Reviewed by:**

William D. Orsi, Woods Hole Oceanographic Institution, USA

James A. Coker, University of Maryland, University College, USA

**\*Correspondence:**

Jörg Eppinger, Division of Physical Sciences, KAUST Catalysis Center, King Abdullah University of Science and Technology, PO Box 2011, Thuwal 23955-6900, Kingdom of Saudi Arabia

e-mail: jorg.eppinger@kaust.edu.sa

Reliable functional annotation of genomic data is the key-step in the discovery of novel enzymes. Intrinsic sequencing data quality problems of single amplified genomes (SAGs) and poor homology of novel extremophile's genomes pose significant challenges for the attribution of functions to the coding sequences identified. The anoxic deep-sea brine pools of the Red Sea are a promising source of novel enzymes with unique evolutionary adaptation. Sequencing data from Red Sea brine pool cultures and SAGs are annotated and stored in the Integrated Data Warehouse of Microbial Genomes (INDIGO) data warehouse. Low sequence homology of annotated genes (no similarity for 35% of these genes) may translate into false positives when searching for specific functions. The Profile and Pattern Matching (PPM) strategy described here was developed to eliminate false positive annotations of enzyme function before progressing to labor-intensive hyper-saline gene expression and characterization. It utilizes InterPro-derived Gene Ontology (GO)-terms (which represent enzyme function profiles) and annotated relevant PROSITE IDs (which are linked to an amino acid consensus pattern). The PPM algorithm was tested on 15 protein families, which were selected based on scientific and commercial potential. An initial list of 2577 enzyme commission (E.C.) numbers was translated into 171 GO-terms and 49 consensus patterns. A subset of INDIGO-sequences consisting of 58 SAGs from six different taxons of bacteria and archaea were selected from six different brine pool environments. Those SAGs code for 74,516 genes, which were independently scanned for the GO-terms (profile filter) and PROSITE IDs (pattern filter). Following stringent reliability filtering, the non-redundant hits (106 profile hits and 147 pattern hits) are classified as reliable, if at least two relevant descriptors (GO-terms and/or consensus patterns) are present. Scripts for annotation, as well as for the PPM algorithm, are available through the INDIGO website.

**Keywords:** bioinformatics, single amplified genomes, halophiles, extremophile, protein sequence consensus patterns, PROSITE IDs, GO-terms, functional genomics

## INTRODUCTION

Discovery of extremophilic enzymes has developed into a major driver for the biotech industry. Although many industrially relevant enzymes were isolated from organisms growing at high temperature, high salt concentration, or in environments contaminated with organic solvents, significant challenges and limitations exist for bio-prospecting of extremophilic enzymes (Liszka et al., 2012). It was estimated that only as few as 0.001–0.1% of microbes in the seawater are currently cultivatable (Amann et al., 1995) and until recently the bottleneck of cultivation not only biased the view of microbial diversity but limited the appreciation of the microbial world in general (Hugenholtz and Tyson, 2008). Novel culture-independent techniques allow

the identification of thousands of novel protein motifs, domains and families from different environments (Yooseph et al., 2007). Despite the vast expectations, metagenomic data have not yet lead to the expected boost of biotechnology (Chistoserdova, 2010), mostly because they suffer from short read length, a low probability to identify rare populations (below 1%) (Kunin et al., 2008), and difficulties in assembling larger contigs of genetic material for members of complex communities. Single-cell genomics (Lasken, 2007) circumvents this problem, and larger contigs from uncultured organisms can be analyzed. A major challenge in mining genomic data of uncultured organisms is a lack of homology to genes of established organisms resulting in limited reliability of gene annotation.

A promising source of novel organisms are the deep-sea anoxic brine pools in the northern part of the Red Sea, formed by tectonic shifts (Gurvich, 2006). Interstitial brine was expelled due to tectonic movements that allowed re-dissolution of evaporitic deposits, and/or phase separation due to temperature variations (Cita, 2006; Hovland et al., 2006). The salt-enriched waters drifted to the seafloor and accumulated in geographical depressions where the brine pools remain stable because of their high density (DasSarma and Arora, 2001). The combination of different extreme physicochemical parameters makes the deep-sea anoxic brine pools one of the most remote, challenging and extreme environments on Earth, while remaining one of the least studied (Antunes et al., 2011). The Red Sea brine pools are extreme in salinity and show a characteristic sharp brine-seawater interface with steep gradients of dissolved O<sub>2</sub>, density, pH, salinity, and temperature (Emery et al., 1969; Ross, 1972; Anschutz and Blanc, 1995). Except for the connected brine pools Atlantis II, Chain, and Discovery Deep (Backer and Schoell, 1972; Faber et al., 1998), environmental conditions vary drastically between the pools, e.g., temperatures range from 22.6°C (Oceanographer) to 68.2°C (Atlantis II) and the NaCl concentration vary from 2.6 M (Suakin) to 5.6 M (Discovery) (Antunes et al., 2011). While the brine pools were detected more than 65 years ago by the Swedish RV Albatross expedition (1947–1948) (Bruneau et al., 1953), microbiological analysis did not start until the late 1960's. The first sampling led to the assumption that under the harsh environmental conditions of the brines life is not possible (Watson and Waterbury, 1969). The search for life in those extreme habitats continuously intensified after the high scientific and economic potential of halophilic organisms became evident (Karan et al., 2012). Since 2010, several sampling expeditions to the Red Sea brine pools have provided a large amount of genomic data, which are collected and annotated at KAUST within the recently described Integrated Data Warehouse of Microbial Genomes (INDIGO) data ware house (Alam et al., 2013). Data stored in INDIGO will stepwise become publicly available.

Analysis and management of next generation whole genome sequencing (NGS) data utilizes comprehensive package of software applications for assembly of sequence reads, mapping to reference genome, variants/SNP calling and annotation, transcript assembly/quantification, and identification of rRNA (Horner et al., 2010; Garber et al., 2011; Pabinger et al., 2014), yet further improvements are required (Dolled-Filhart et al., 2013). Large-scale annotation of DNA sequences with a low homology to genes of experimentally verified function may be flawed and hence represents a major drawback for biomining. The homology-based annotation faces one intrinsic issue: annotation reliability and protein diversity are reciprocal. The situation is complicated by error propagation. The function of the encoded protein was validated experimentally only for a small and continuously diminishing fraction of the gene sequences available. Initially, functions of novel genes were annotated based on gene sequences with experimentally verified function. Based on these data more genes were annotated and so on. While in this chain two proteins are always highly similar, the last annotated

gene and the experimentally verified source may possess distinct sequences and functions. In comparison to genomic sequencing, experimental characterization of Single Amplified Genome (SAG) gene products requires gene synthesis, expression, purification as well as functional characterization and therefore is by several orders of magnitude more time consuming. Hence, false positive results from flawed annotation are much more problematic than false negative (due to non-complete annotation) when genomic data are searched for a desired function. This is particularly true for genes from extremophilic organisms, which require slow growing expression systems. Here we present a strategy to minimize false positive identification of the gene product's function. The Profile and Pattern Matching (PPM) algorithm describe below collates complementary information available from (a) InterPro-derived Gene Ontology (GO) terms (Ashburner et al., 2000), which connect an enzyme's function to amino acid sequence profiles and (b) annotated PROSITE IDs (Sigrist et al., 2013), which are linked to an amino acid consensus pattern. This PPM algorithm was tested on 15 protein families of scientific or commercial interest. The strict PPM algorithm initially extracted the most reliably annotated genes, which in this example represent about 1.5% of the genes in the database. Subsequent removal of incomplete genes followed by PPM selection lead to further condensation of gene hits (0.1% of genes in database). A final ranking extracted 11 genes as most likely candidates to code for one of the Protein of Interest (POI) functions.

## MATERIALS AND METHODS

### SAMPLE COLLECTION

All samples were collected during leg 2 of the RV *Aegaeo* WHOI, AUC—KAUST Red Sea Cruise in October/November 2011. Samples were taken at different depths and locations in the Red Sea, in and outside the brine pools as well as from sediments. For all brine pools, samples were taken in the brine itself, the sediment and at different depths of the brine seawater interphase (Eder et al., 2001). In total 46 casts were done containing 7030 L of water, as well as seven sediment samples. The collected liquid samples were immediately filtered using a TFF (tangential flow filtration) system, concentrated and immediately afterwards stored at -80°C. During the sampling, different chemical parameters including salinity (conductivity) and temperature were measured. The five brine pools sampled were Kebrat Deep, Nereus Deep, Atlantis II Deep, Discovery Deep, and Erba Deep (Backer and Schoell, 1972; Searle and Ross, 1975; Karbe, 1987; Hartmann et al., 1998).

### SINGLE AMPLIFIED GENOME GENERATION

For the production of SAG from single cells, the “SCGC SAG generation service” (cat. no. S-101) at the “BIGELOW Laboratory single cell genomics center,” which is part of the Bigelow Laboratory for Ocean Sciences in Boothbay Harbor, Lincoln County, Maine, United States, was used. The service includes initial sample evaluation for FACS suitability, individual cell separation into wells of a 384-well plate, cell lysis, and single cell multiple displacement amplification (MDA).

## WHOLE GENOME SEQUENCING AND ASSEMBLY

The whole genome sequencing was performed at the “BIGELOW Laboratory single cell genomics center” using the “Prokaryote SAG whole genome sequencing” service (cat. no. S-014). The service includes sequencing library preparation, genomic sequencing, *de novo* assembly, and assembly quality control. Service products include contig fasta files and assembly statistics. Assemblies of the single-cell amplified genomes (SAGs) were generated using a pipeline that employs a choice of assemblers designed for single-cell sequencing data including VelvetSC (Chitsaz et al., 2011), SPAdes (Bankevich et al., 2012), and IDBA-UD (Peng et al., 2012), along with several pre- and post-assembly data quality checks using Trimmomatic (Lohse et al., 2012). IDBA-UD was benchmarked as the overall best assembler for our SAGs as it did reconstruct longer contigs with higher accuracy to the reference genome of *Nitrosopumilus maritimus* SCM1 (Könneke et al., 2005).

## DATASET

The data used in this work consisted of 87 SAGs covering 16 different taxonomic groups, sampled in 11 different environments. A total of 26,626 contigs covering 111,269 ORFs and containing 79.8 Mbp genomic information (**Table 1**) were analyzed.

### Annotation of the dataset

The assembled contig sequences were integrated into the INDIGO data warehouse (Alam et al., 2013) for microbial genomes. INDIGO is a dynamic system using the InterMine framework (Smith et al., 2012), one of the highest benchmarked data warehouses (Triplet and Butler, 2013). INDIGO allows Automatic Annotation of Microbial Genomes (AAMG), extensive query building for annotation integration, creation of customized feature/attribute/entity lists and enrichment analysis for GO concepts, which are crucial steps of the following analysis. Using INDIGO the assembled contig sequences were (i) annotated, (ii) converted into an XML schema, and (iii) implemented into the data warehouse. **Figure 1** gives an overview of the workflow (Alam et al., 2013). Assignments of GO-terms are largely independent from PROSITE IDs. GO-terms emerge from domain associations provided by InterPro (Quevillon et al., 2005) (one of several domain resources may be PROSITE). PROSITE consensus patterns are predicted by the PS\_Scan (De Castro et al., 2006) tool.

### Automatic annotation of microbial genomes (AAMG) pipeline.

Functional annotation of archaeal or bacterial genomes is available via the INDIGO website interface (<http://www.cbrc.kaust.edu.sa>).

[edu.sa/indigo/mymine.do?subtab=aamg](http://edu.sa/indigo/mymine.do?subtab=aamg)). Completed genome annotations may be included into the INDIGO database. This enables application of the scripts presented in this work for any novel genetic data.

## PHYLOGENETIC ANALYSIS

The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). All illustrated trees are drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA6 (Tamura et al., 2013).

## PPM METHODOLOGY

The PPM algorithm was automated by including two new scripts into INDIGO, which are publicly available from the homepage.

### AutoTECNo: automated translation of E.C. numbers

The E.C. No. translator (AutoTECNo) automatically converts a list of given enzyme commission (E.C.) numbers into GO-terms (Kanehisa and Goto, 2000) as well as PROSITE IDs, using open source PROSITE files (Sigrist et al., 2002). Preliminary, transferred and deleted E.C. numbers are ignored. The AutoTECNo provides two XML scripts for the independent profile and pattern search via INDIGO. AutoTECNo is available at the following website: <http://www.cbrc.kaust.edu.sa/ppma/ec2gops.html>.

### PPM processor: automated extraction and ranking of the most reliable hits

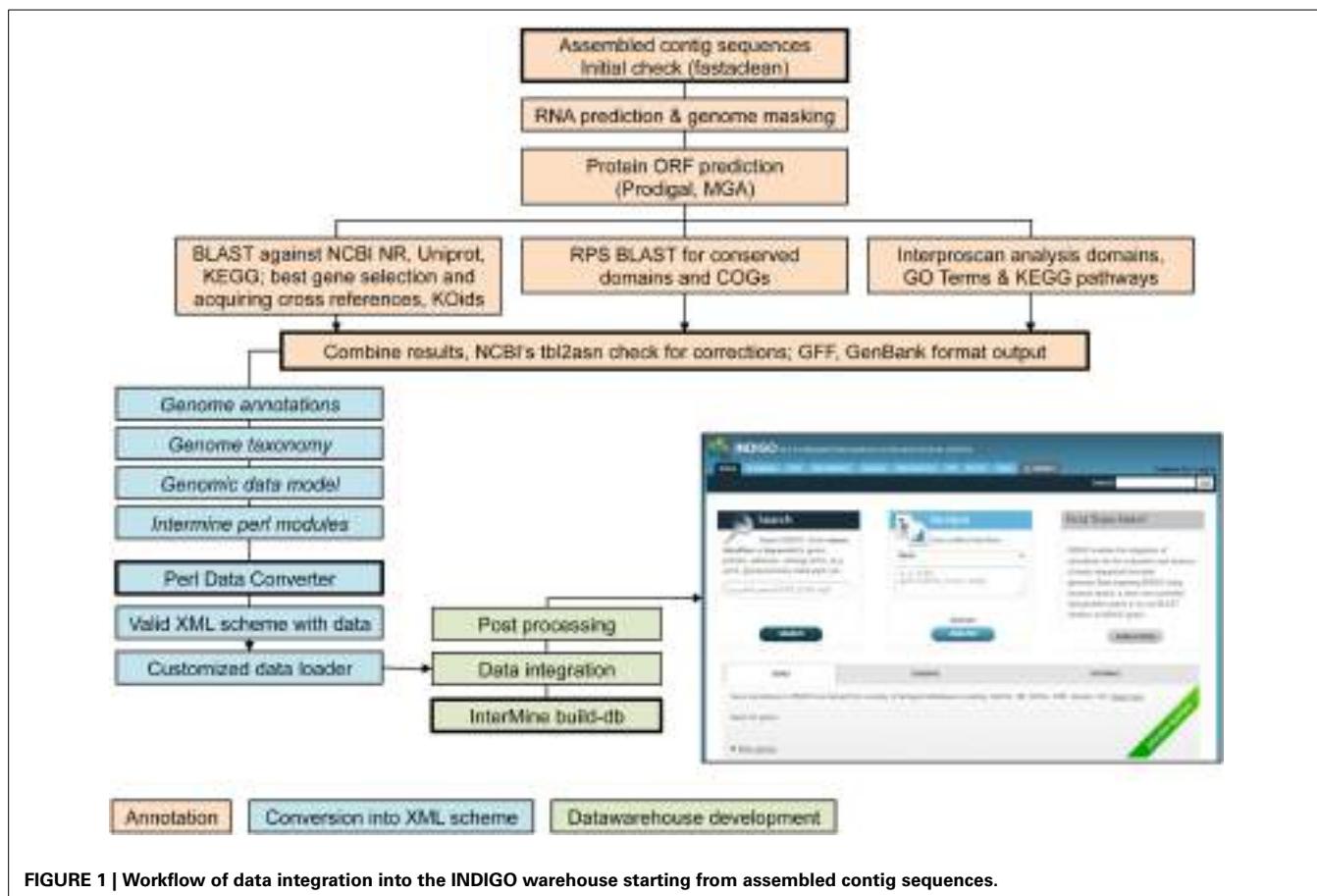
The PPM Processor requires one or more tab separated spreadsheets (.tsv) of the independent profile analysis (via GO-terms) and/or pattern analysis (via PROSITE IDs) as input file. The processor generates sets of genes according to their profile and pattern distribution. The resulting list is ranked regarding to the amount of profile and pattern combinations. The PPM processor is available at the following website: <http://www.cbrc.kaust.edu.sa/ppma/indigoTbl2PSgoSets.html>.

### The PPM workflow, starting from a non annotated genome

First, an assembled genome is annotated using the AAMG pipeline as part of the INDIGO data warehouse. Second the E.C. number based list of POI (list) is translated into profile and pattern values (GO-terms and PROSITE IDs) by using AutoTECNo. The resulting XML lists (of pattern and profile values) are separately imported into the INDIGO

**Table 1 | Two example and summary (*italic*) of the SAG data in INDIGO used for this work.**

Organism	Habitat (°C; % salt)	Contigs (number)	Contigs (min. bp)	Genome (Mbp)	N50 (kbp)	N90 (kbp)	ORFs
SAR86 clade	25.0; 4.0	86	312	0.74	83	6.9	853
MSBL1	54.0; 15.2	849	200	2.20	15	0.7	3293
<i>87 SAGs, 16 taxonomic groups</i>	<i>11 types 23.4 – 63.0; 4 – 26</i>	<i>26,626 (16 – 849)</i>	<i>200 – 317</i>	<i>79.8 (0.1 – 2.2)</i>	<i>1,4 – 1,200</i>	<i>0.4 – 13.9</i>	<i>111,269 (242 – 3,293)</i>



data warehouse to analyze any listed genome at the following URL: <http://www.cbrc.kaust.edu.sa/indigo/importQueries.do?querybuilder=yes>. The two resulting tab separated spreadsheets can be uploaded into the PPM processor to generate three PPM sets of genes: (i) profile set, (ii) pattern set, and (iii) profile and pattern set.

## RESULTS

### PPM: PROFILE AND PATTERN MATCHING FOR FUNCTION IDENTIFICATION

Analysis of the huge amount of data resulting from next generation whole genome sequencing (NGS) requires modern bioinformatic tools. Comparisons of annotation pipelines reveal a surprising level of uncertainty in gene annotation. Annotations of the same genome (strain TY2482) of the enterohemorrhagic diarrhea causing shiga-toxin-producing *E. coli* O104 (Rohde et al., 2011) by several groups allowed a comparison of the three main annotation pipelines: Broad, BG7, and RAST. Compared 5164 coding sequences (CDS) of to the Broad annotation the BG7 annotation resulted in 5210 CDS with 163 (3.1%) false negatives and 271 (5.2%) false positives, and RAST annotation gave in 5446 CDS with 116 (2.1%) false negatives and 321 (5.9%) false positives (Alam et al., 2013). The AAMG based annotation stored in INDIGO, which is used for this article, gave results similar to those of the Broad institute. Annotation of the *E. coli* K12 strain W3110 by INDIGO resulted in 4340 CDS (NCBI 4337), with

236 (5.4%) false positives and 235 (5.4%) false negatives in comparison to the NCBI annotation. These examples illustrate, that state-of-the-art annotation still yields about 5.5% false positives for strains of the standard organisms *E. coli* and a significantly higher rate of false positives may be expected for novel genomes. While this might not impact *in silico* analysis e.g., for identification of pathways, a substantial amount of false positives can lead to costly failures in experimental bioprospecting campaigns.

Among the descriptors INDIGO annotation associates with genes, two are particularly suited to evaluate the correct assignment of an enzymatic function to a gene product: (i) the GO-term and (ii) the PROSITE ID. The GO project describes genes (gene products) using terms from three structured vocabularies: biological process, cellular component and molecular function. Correspondingly, a list of GO-terms associated with a gene can be seen as the gene's profile. A PROSITE ID relates to a single consensus pattern as "amino acid sequence signature" to characterize protein function. Genes from INDIGO with matching function description of GO-term and PROSITE ID(s) should represent a subset of genes with highly reliable annotation. To extract such genes based on an input list of E.C. numbers of interest, we developed a protein PPM algorithm.

### From proteins of interest to bioinformatics descriptors

Initially, we established a set of proteins, which potentially are of scientific and/or commercial interest. Protein classes selected

include a variety of hydrolases, ene reductases, dehydrogenases, and carbonic anhydrases (CAs) as well as a range of metalloproteins, porines and potentially new aminoacyl tRNA synthetases. The selected 15 protein families of interest (POI families) are summarized in **Table 2**. Bioinformatic matching of the POIs vs. the INDIGO database requires a translation of the POI list into terms of the selected descriptors (GO-terms and PROSITE ID). For enzymes, E.C. numbers can be associated with the enzyme family name as well as GO-terms and PROSITE ID and therefore can be used to interconvert these terms. The POI list was translated into the E.C. numbers using BRENDA (Braunschweig Enzyme Database) (Schomburg et al., 2013). Of the resulting 2577 E.C. numbers (Table S1) 434 were non-redundant. Removal of preliminary/transferred and deleted E.C. numbers provided a final list of 265 E.C. numbers (Table S2). The list of E.C. numbers was converted into profiles (GO-terms) and pattern (PROSITE IDs). For gene expression products without enzymatic function like aquaporins and pyltRNA, the respective GO-terms and PROSITE IDs were added manually. The resulting protein profile filter consist of 171 non-redundant GO-terms (BRENDA) (Table S3). The independent pattern filter consisted of 52 non-redundant PROSITE consensus pattern (Sigrist et al., 2013). Three consensus patterns (PS00198, PS00455, PS00143) were removed because of their low specificity (consensus pattern specificity can be derived from the information available at PROSITE web page: <http://prosite.expasy.org>), resulting in a final pattern list of 49 consensus pattern (Table S4).

**AutoTECNo: automated translation of E.C. numbers.** The web-based AutoTECNo script simplifies conversion of POI classes into the two bioinformatic PPM descriptors described above. A user may enter one or more distinct or flexible E.C. numbers, which are automatically converted into GO-terms and PROSITE IDs. A numeric value is required for the first three digits of flexible E.C. numbers (e.g., 1.1.1.\*). AutoTECNo automatically ignores preliminary, transferred and deleted E.C. numbers. The

AutoTECNo output provides two XML scripts, one for each of the independent profile and pattern search, which can be imported directly into the INDIGO data warehouse by using the direct links on the output page.

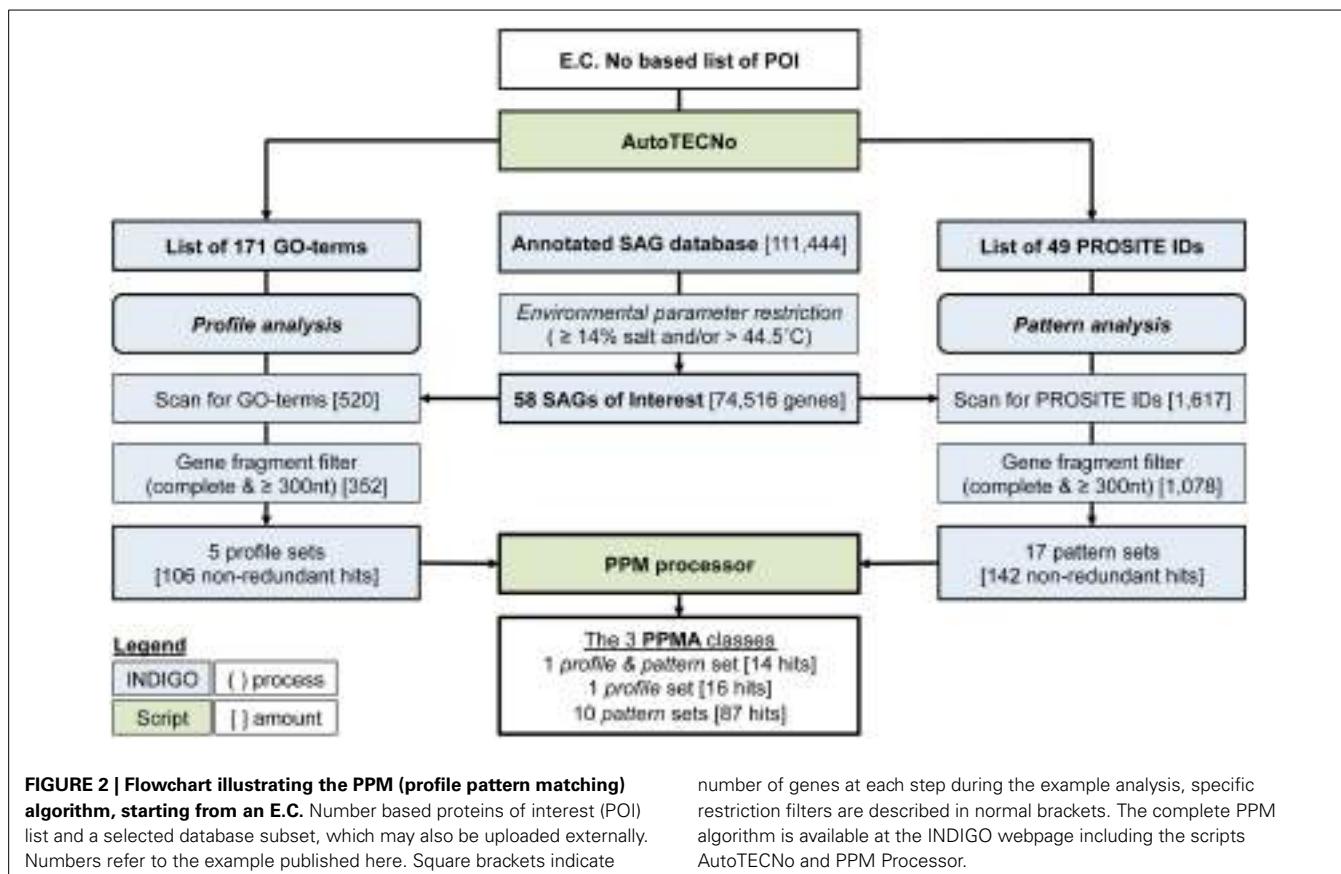
#### The PPM (Profile and Pattern Matching) algorithm

The PPM algorithm retrieves those POIs from a database, which are most likely to be annotated correctly. Initially, the GO-term list (profile) and the consensus pattern list (coded by the PROSITE IDs) are matched independently onto the dataset of interest. From each of the resulting subset of genomic data, gene fragments commonly present in SAGs or metagenomic data a gene fragment filter eliminates (i) genes with less than 300 nucleotides (to sustain a minimal length required for functionality) and (ii) genes that are not annotated as complete (indicating that a 3' or 5' part of the gene is missing). In a last step, both filtered lists are transferred to the PPM processor (see below), which arranges all hits into sets of genes having the same combination of identifiers (GO-terms and/or Prosite IDs). Three classes of sets are listed: (i) the profile sets, containing genes with one or more GO-term describing the respective POI, (ii) the pattern sets, containing genes with one or more PROSITE ID of the respective POI and (iii) the profile and pattern set, consisting of genes with at least one GO-term and PROSIT ID of the POI. The annotation of genes is ranked as more reliable with increasing numbers of associated identifiers. The complete PPM algorithm is illustrated in **Figure 2**.

Identification of the most reliably annotated genes in INDIGO that match our POI served as test-case for the PPM. The genetic database search was restricted to certain brine pool SAGs based on environmental parameters of the sampling locations (salinity  $\geq 14\%$  and/or a temperature  $>44.5^{\circ}\text{C}$ ). The habitats selected were set to reflect the upper part of moderate halophilic conditions (5–20% salt) as well as extreme halophilic conditions (20–30% salt) (Ollivier et al., 1994) and/or thermophilic conditions [ $45$ – $80^{\circ}\text{C}$  (Madigan et al., 2003)]. The sample subset comprises 58 SAGs from three different brine pools (Atlantis II

**Table 2 | List of proteins of interest (POIs), which were selected for this study.**

No	POI group	Description	Interest
1	Alcohol DH	Interconversion of aldehydes/ketones and alcohols	Biocatalytic synthesis of chiral intermediates
2	Formate DH	Conversion of $\text{CO}_2$ into formate	Biological carbon capture
3	Formaldehyde DH	Interconversion of formaldehyde and formate	Biological carbon capture, methanol conversion
4	Carbon monoxide DH	Interconversion of CO and $\text{CO}_2$	Biological carbon capture, metalloenzyme structures
5	Ene reductase	Stereoselective reduction of alkenes	Biocatalytic synthesis of chiral intermediates
6	Protease	Hydrolysis of peptide bonds	Detergents, food, and leather processing
7	Terpene synthase	Synthesis of basic, (mulit-)cyclic terpene structures	Biocatalytic synthesis of complex intermediates
8	Nitrogenase	Fixation of nitrogen from air	metalloenzyme structure and function
9	Lipase	Hydrolysis of triglyceride esters	Detergents, biodiesel synthesis
10	Carbonic anhydrase	Interconversion of $\text{CO}_2$ and Bicarbonate	Biological carbon capture, metalloenzyme structures
11	Acetylene hydratase	Synthesis of aldehydes from acetylene	Biocatalytic synthesis intermediates
12	Acetyl-CoA synthetase	Activation of acetate for further conversion	Biological carbon capture metabolism
13	pylRS	Aminoacyl tRNA synthetase, acting on pyrrolysine	Synthetic biology, expanding the genetic code
14	pyltRNA	tRNA coding for pyrrolysine	Synthetic biology, expanding the genetic code
15	Aquaporin	Integral membrane proteins controlling osmotic pressure	Water desalination membranes



**FIGURE 2 | Flowchart illustrating the PPM (profile pattern matching) algorithm, starting from an E.C.** Number based proteins of interest (POI) list and a selected database subset, which may also be uploaded externally. Numbers refer to the example published here. Square brackets indicate

number of genes at each step during the example analysis, specific restriction filters are described in normal brackets. The complete PPM algorithm is available at the INDIGO webpage including the scripts AutoTECNo and PPM Processor.

deep, Discovery, and Kebrit), covering six different environmental conditions. These SAGs contain a total of 73,688 ORFs coding for 74,516 genes. The ORFs were assembled out of 21,519 contigs into genomes of a combined size of 48.2 mega base pairs (**Table 3**).

As described above, the POI list was transformed into a protein profile filter consisting of 171 non-redundant GO-terms (BRENDA) and an independent pattern filter of 49 PROSITE IDs (Sigrist et al., 2013). Profile matching of the 74,516 pre-selected genes with the 171 GO-terms resulted in 520 hits, which were further reduced by the gene fragment filter to 352 (**Table 4**). Elimination of duplicates (genes associated to multiple GO-term or PROSITE ID occur multiple times in the output) yielded 106 non-redundant hits, which could then be grouped into five different profile sets, based on the gene-associated GO-terms. The five profile sets contain six different GO-terms, four profiles with only one GO-term and one profile with two GO-terms (**Table 5**). Categorizing the 106 genes into five profile sets clarifies what functions and functional diversity can be expected from the hits.

The independent pattern filter was applied according to the same scheme. Screening all 58 SAGs against the 49 PROSITE IDs resulted in 1617 hits. Applying the gene reliability filter reduced this number to 1078 hits, which could be further condensed to 142 non-redundant hits. These 142 genes fall into 17 pattern sets containing 25 different PROSITE IDs.

Since the presence of several GO terms, PROSITE IDs or a combination of both indicates a more reliable gene annotation, we used the PPM processor to identify genes which are

associated with multiple descriptors. The list (**Table 5**) contains three sub-sets: (i) the profile sets (one set of 16 hits), (ii) the pattern sets (10 sets containing 87 hits), and (iii) the profile and pattern sets (one set of 14 hits). Only the profile and pattern set contains genes, which were found independently by both, PPM. In other words, when the INDIGO subset of 74,516 genes is screened for the 434 non-redundant E.C. numbers, only 14 genes have a matching GO-term and PROSITE ID. All 14 hits belong to the same E.C. number (1.3.1.26, dihydrolipocinolate reductase, DHPR). Since some profile or pattern sets stand for the same enzyme type the total amount of 117 most reliably annotated genes that were identified by the PPM algorithm fall under only nine different enzyme families: prephenate DH (1.3.1.13), iron containing ADH (1.1.1.), dkgA (1.1.1.274), glyoxylate reductase (1.1.1.26), Clp protease (3.4.21.92), molybdo-pterin oxidoreductase (e.g., 1.2.2.1), nitrogenase (1.18.6.1), subtilisin (3.4.21.\*), and DHPR (1.3.1.26). The relatively small number of highly reliable hits is helpful for an experimental scientist, who is aiming to characterize novel gene expression products. A reduction of 111,444 potential expression targets to only 117 provides the necessary experimental focus (see below).

**Semi-automatic, XML based PPM algorithm.** The PPM algorithm was integrated into the INDIGO web page via a XML script. The semi-automated work flow requires three steps: (i) conversion of the POI list into GO-terms and PROSITE IDS with

**Table 3 | Bacterial (*italic*) and archeal SAGs from thermophilic and hypersalinic sampling regions selected for this study.**

SAGs	Taxonomy	Genome (Mbp)	Contigs	ORFs	Salinity [% (w/v)]	T (°C)	Location
1	<i>Desulfo-bacterales</i>	1.2	401	1658	9.4	47	Atlantis II
13	MSBL1	8.8	4262	14,159	16.8	63	
10	MBGE	9.5	4809	14,809			
3	MSBL1	2.8	1019	4114	15.2	54	
3	SA2 cluster	0.9	386	1367			
1	<i>Candidate division</i>	0.4	176	598			
2	MSBL1	1.2	321	1716	14	32	Discovery
17	MSBL1	17.9	7462	27,110	26.2	44.8	
5	MSBL1	3.2	1647	4989	26	23.4	Kebrit
3	<i>UMSBL6</i>	2.3	1036	3168			
58	6 taxon. groups	48.2	21,519	73,688	6 habitats		3 pools

**Table 4 | Stepwise overview of the conversion the 15 selected enzyme groups into non-redundant GO terms and PROSITE ID.**

No	POI-group	E.C.-No.			GO-terms			PROSITE IDs		
		T	NR	S	T	NR	Hits	T	NR	Hits
1	Alcohol DH	101	32	25	20	20	0	12	8	2–6
2	Formate DH	29	6	6	4	4	1	7	6	6
3	Formaldehyde DH	23	9	4	3	2	0	0	0	0
4	Carbon monoxide DH	19	4	4	4	4	1	1	0	0
5	Ene reductase	1162	107	65	61	61	4	9	1	1–5
6	Protease	741	217	111	39	39	0	45	20	8
7	Terpene synthase	35	023	17	9	9	0	0	0	0
8	Nitrogenase	18	4	2	2	2	0	4	4	4
9	Lipase	380	26	25	24	24	0	9	6	0
10	Carbonic anhydrase	58	1	1	1	1	0	3	3	0
11	Acetylene hydratase	2	1	1	1	1	0	0	0	0
12	Acetyl-CoA synthetase	8	3	3	3	2	0	1	0	0
13	Aquaporin	0	0	0	2	2	0	1	1	0
Total		2576	433	264	173	171	6	92	49	25

T, Total in class; NR, non-redundant ones; S, selected for this study.

AutoTECNo, (ii) individual profile as well as pattern matching via a query in INDIGO and (iii) extraction and ranking of the most reliable result in pattern, profile and profile and pattern by the PPM processor. This process requires two input files: (i) an assembled genome, which can be annotated using the AAMG pipeline and (ii) an E.C. number based POI list. The POI list can directly be copied into the AutoTECNo input mask. After submitting the E.C. number list, AutoTECNo will generate a list of all E.C. numbers and the associated GO-terms and PROSITE IDs. At the bottom of the output mask, three links are provided: “GO xml,” “Prosite xml,” and “INDIGO datawarehouse.” Clicking either of the first two links will open a window, which provides XML-formatted files (for either GO-terms or PROSITE IDs). These files can be edited and used separately to build INDIGO queries. In such a query, INDIGO is used to match each of the two XML lists against the selected genomes. Clicking on the “INDIGO datawarehouse” link opens the INDIGO XML

input mask, which can be used to initiate a query by pasting the XML script from AutoTECNo. A graphical overview of the query will be shown and further customization can be done (pre-set columns should not be deleted). At this stage, both, profile (GO-term) and pattern (Prosite ID) filters can be applied individually in connection with the optional gene fragment filter. Hits will be organized in a table summarizing all information available in INDIGO. The table still may contain duplicates, since one gene can be found under several GO-terms and/or PROSITE IDs. The results-table can be downloaded as “Spreadsheet (tab separated values)” (.tsv file) for import into the PPM processor. The PPM processor output provides a list of non-redundant genes, grouped into subsets of the three classes of hits (profile sets, pattern sets and profile and pattern sets) as well as ranked based on the amount of associated patterns and profiles. A link back to INDIGO allows listing of the obtained hits for a detailed analysis.

**Table 5 | Enzyme hits identified using the PPM algorithm and the respective PPM descriptors.**

PPM class	Profile (GO-term) or pattern (PROSITE IDs)		Hits	PPM set	Enzyme hit
Profile	GO:0008839		3	–	
	GO:0009326		25	–	
	GO:0018492		17	–	
	GO:0043115		31	–	
	GO:0004665	GO:0008977	16	Pro 1	Prephenate DH [1.3.1.13]
Pattern	PS00059		15	–	
	PS00061		3	–	
	PS00136		3	–	
	PS00137		1	–	
	PS00138		2	–	
	PS00141		11	–	
	PS00501		6	–	
	PS00060	PS00913	4	Pat 1	Fe—ADH [1.1.1.1]
	PS00062	PS00063	PS00798	Pat 2	dkgA [1.1.1.274]
	PS00065	PS00670	PS00671	Pat 3	Glyoxylate red. [1.1.1.26]
	PS00381	PS00382		Pat 4	Clp protease [3.4.21.92]
	PS00490	PS00551	PS00932	Pat 5	Molybdopt. OR [e.g., 1.2.2.1]
	PS00090	PS00699		Pat 6	Nitrogenase [1.18.6.1]
	PS00692	PS00746		Pat 7	
Profile and Pattern	PS00136	PS00137	2	Pat 8	Subtilisin [3.4.21.*]
	PS00136	PS00138	4	Pat 9	
	PS00137	PS00138	1	Pat 10	
Profile and Pattern	GO:0008839	PS01298	14	PP 1	DHPR [1.3.1.26]

#### MANUAL HIT SELECTION FROM THE PPM PROCESSOR OUTPUT

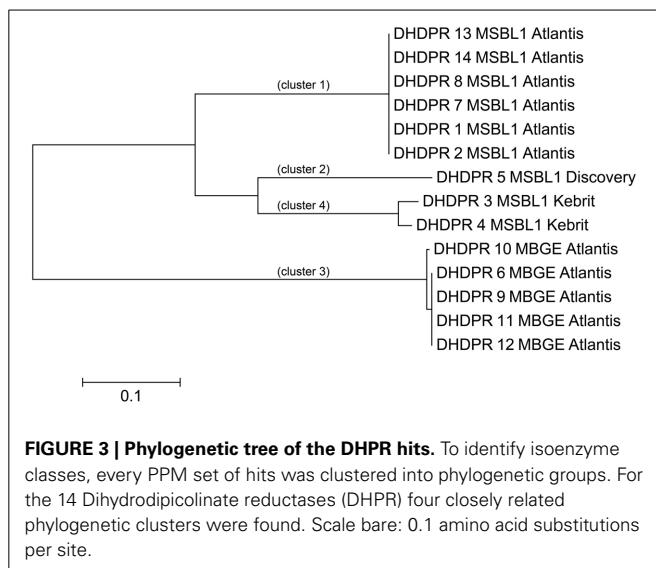
Grouping of genes into PPM classes and sets immediately highlights expected functional similarities of gene expression products. PPM sets of patterns and/or profiles, which are characteristic for the same protein, can be condensed further into one meta-set. For example pattern sets with combinations of PS00136 and PS00137, PS00136, and PS00138 or PS00137 and PS00138 are all indicative of subtilase type serine proteases and these pattern sets were condensed into one meta set. In total nine functionally distinct PPM sets remained after manual condensing (**Table 5**).

For experimental characterization, synthesis and expression of 117 genes from halophilic extremophiles still represent an enormous challenge, which mandates identification of those extremophilic proteins as expression targets, which are most typical for each functionality-set. For five of the nine functionally different PPM sets, we were able to pinpoint nine genes representing all three PPM classes (profile, pattern, profile, and pattern) (**Table 5**). Amino acid based phylogenetic analysis within each PPM set revealed phylogenetic relations and sequence clusters. The sequence representing most of the set-members was selected, e.g., the PP1 DHPR PPM set contains 14 different hit sequences (isoenzymes). Phylogenetic analysis resulted in four clusters of phylogenetically closely related groups (**Figure 3**). For each of those four clusters the sequence representing most of the members was selected. This was straightforward for three DHPR clusters, since one sequence contained all elements of the others. In the fourth

case as well as for cluster of other sets the selection was more complicated, because phylogenetic sequence clusters showed either an equal distribution of mutations in one cluster or an unequal length of sequences. To address this problem an additional protein BLAST (BLASTp) (Johnson et al., 2008) was performed and the sequence with the highest similarity was chosen for the fourth DHPR and the halolysin cluster. In case of no difference in similarity according to BLASTp, the gene product providing more functional side chains was chosen (e.g., for subtilisin) since additional chemical functionality may indicate more diverse enzyme characteristics (e.g., hydrogen bonding, allosteric pockets, metal complexation etc.). Amino acid sequences typically differed in less than 10 positions [amino acid sequence length: 401 (ADH), 348 (2-hydroxyacid DH), 498–565 (halolysin; the 565 amino acid sequence contains all shorter ones), 528 (subtilisin), 435–440 (prephenate DH), 272–285 (four subgroups of DHRPs)].

#### FUNCTION IDENTIFICATION OF PROTEINS WITHOUT EXISTING GO-TERMS OR PROSITE IDs

The initial search for CAs was not successful. While distinct GO-terms and consensus patterns exist for  $\alpha$ - and  $\beta$ -CAs (**Table 5**), none are available for the other three CA families ( $\gamma$ ,  $\delta$  and  $\zeta$ ). According to Ferry the CS chain A from *Methanoscincus thermophilic* (Cam) can be considered the archetype of the  $\gamma$ -CA family, and a distinct, 180 amino acid sequence (no 34–214) is indicative for a  $\gamma$ -CA protein (Smith and Ferry, 2000). An



**FIGURE 3 | Phylogenetic tree of the DHPR hits.** To identify isoenzyme classes, every PPM set of hits was clustered into phylogenetic groups. For the 14 Dihydrodipicolinate reductases (DHPR) four closely related phylogenetic clusters were found. Scale bar: 0.1 amino acid substitutions per site.

INDIGO internal BLAST of this 180 amino acid motif against all genes yielded 17 potential  $\gamma$ -CAs. Applying the gene fragment filter reduced the candidate pool to six.

As discussed above, an additional pattern matching should increase the reliability of the profile-based protein identification. The analysis of the only two  $\gamma$ -CA class crystal structures reported (Cam from *M. thermophila* Kisker et al., 1996, see also pdb 3OW5 and a CamH homolog from *P. horikoshii* Jeyakanthan et al., 2008) revealed nine amino acids in two peptide sequences of 26 and six amino acids as most relevant for enzyme function (Smith and Ferry, 2000). The resulting two initial consensus patterns are shown below [color code: yellow, metal binding motifs ( $H_{81}$ ,  $H_{117}$ ,  $H_{122}$ ); green, residues directly involved in catalysis ( $E_{62}$ ,  $N_{73}$ ,  $Q_{75}$ ,  $E_{84}$ ); blue, structurally important residues ( $R_{59}$ ,  $D_{61}$ ); not highlighted, residues of no specific function as they appear in the  $\gamma$ -CA sequence].

59 - R<sub>59</sub>S<sub>D61</sub>E<sub>62</sub>GMPIFVGDRSN<sub>73</sub>VQ<sub>75</sub>DGVVLH<sub>81</sub>ALE<sub>84</sub> - 84 and

117 - H<sub>117</sub>QSQVH<sub>122</sub> - 122

No hit was found for a strict pattern matching of the six potential  $\gamma$ -CAs. This is not surprising, since it is common for consensus patterns that some functionally important amino acids can be altered within a certain threshold. Alignment of the initial  $\gamma$ -CA consensus pattern with the six  $\gamma$ -CA candidate sequences revealed that the 10 amino acid long stretch from  $E_{62}$  to  $N_{73}$  was shortened by one amino acid in all six candidates. The resulting structural alteration is unlikely to affect function. Further, the two structurally important residues  $R_{59}$  and  $D_{61}$  were conserved as well as two out of the three metal binding histidines ( $H_{81}$  and  $H_{117}$ ) (Table 6). The third metal binding amino acid  $H_{122}$  was replaced by an N in hit number 6, a mutation, which potentially affects function. Further sequence variations involve the replacement of catalytic  $E_{84}$  by either D (four cases, potentially not influencing function), or K (two cases, potentially affecting function). The remaining catalytically important residues  $E_{62}$ ,  $N_{73}$ , and  $E_{75}$ , which are involved in a hydrogen-bonding network

in the *M. thermophile* protein, are highly variable among the six candidates sequences. Assuming that some of these candidates are CAs because of profile and pattern similarity to the *M. thermophile* archetype enzyme we concluded that  $E_{62}$  is not generally important for the function of this enzyme type and that  $N_{73}$  and  $E_{75}$  can be replaced by the hydrogen bonding amino acids C or K, respectively. Correspondingly, we suggest the following two consensus patterns for  $\gamma$ -CAs:

R-x-D-x(10,11)-[NC]-x-[QK]-x(5)-H-x(2)-[ED] and  
H-x(3)-H

Application of the PPM algorithm using the 180 amino acid profile stretch identified from pdb 3OW5 and the new consensus patterns delivered three  $\gamma$ -CAs candidates. Because of high sequence similarity in two out of the three sequences, the sequences of gene 2 (annotated as ferripyochelin binding protein 01) from Atlantis II deep and gene 3 (annotated as predicted acetyltransferase) from discovery deep (Table 6) were selected as best candidates for experimental studies of  $\gamma$ -CAs [CA\_A (Atlantis II deep) and CA\_D (Discovery deep) in Table 7].

## DISCUSSION

Proteins, which are suitable for the harsh conditions of many biotechnological applications can be obtained through protein engineering, discovery and mining of novel extremophilic genomes or a combination of both. The major challenge in mining genomic data from extreme environments is, that, with increasing extremeness of the habitat, the possibility of culturing the organism thriving under these conditions shrinks substantially (Alain and Querellou, 2009). However, SAGs can provide genomic data from uncultured organism. We believe that improving the quality of SAGs assemblies (higher sequence coverage, longer contigs, and advanced annotation programs) should enable us to utilize SAGs as a rich source for discovery of extremophilic enzymes of scientific interest and commercial value. However, annotation reliability is lowered for both, extremophilic genomes (for which commonly no close relative is known) and SAGs (which may suffer from gaps, incomplete genes, or generally sequencing data of lower quality) and therefore a highly reliable algorithm for identification of genes of interest from extremophilic SAG databases is mandatory before entering labor-intensive expression and characterization of these genes.

## PROBLEMS OF SINGLE PROFILE OR PATTERN ANALYSIS AND THE PPM ALGORITHM

Consensus patterns show a good reliability, yet a considerable amount of hits identified via PROSITE ID are false positives (has the motif but not the function), false negatives (has the function but not the motif), unknown (has the motif but no verified function), or partial hits (has the function but only parts of the motif) (Sigrist et al., 2002). Table 8 combines examples illustrating the reliability for consensus pattern based annotation of enzyme function. Reliability may be as low as 55% false positives (PS00136) or 90% false negatives (PS00065). A further problem of pattern-based annotation is the low flexibility because of the short pattern lengths (about 10–20 amino acids Sigrist et al., 2002), typically covering only 1.9–7.9% of the total protein length. Due to the short length of the consensus pattern, a

**Table 6 | Alignment of functional important residues the  $\gamma$ -CA chain A from *Methanoscincus thermophiles* with  $\gamma$ -CA candidates.**

No.	Annotation	Organism and habitat	R <sub>59</sub>	D <sub>61</sub>	E <sub>62</sub>	N <sub>73</sub>	O <sub>75</sub>	H <sub>81</sub>	E <sub>84</sub>	H <sub>117</sub>	H <sub>122</sub>
1	Carbonic anhydrase acetyltransferase	MSBL1, Atlantis II (63°C, 16.8% salt)	Conserved	F	C	Q		D			
2	Ferripyochelin binding protein 01	MSBL1, Atlantis II (63°C, 16.8% salt)		F	C	Q		D			
3	Predicted acetyltransferase	MSBL1, Discovery (44.8°C, 26.2% salt)		I	C	K		D			
4	Ferripyochelin binding protein 02	MSBL1, Kebrit (23.4°C, 26.0% salt)		F	C	G		D			
5	Ferripyochelin binding protein 03	MBGE, Atlantis II (63°C, 16.8% salt)		K	N	Q		K			
6	Ferripyochelin binding protein 04	MSBL1, Discovery (44.8°C, 26.2% salt)		F	V	E		K			N

**Table 7 | Hits identified as reliable using the PPM algorithm.**

Gen	Enzyme [organism]	PPM ident.	Closest related protein (sequence identity) [organism]
AF_D	ADH [MSBL1]	Pat 1	ADH, iron-containing (60%) [ <i>Thermotoga neapolitana</i> DSM 4359]
HD_K	2-Hydroxyacid DH [USMBL6]	Pat 3	NAD-binding 2-hydroxyacid DH (63%) [ <i>Petrotoga mobilis</i> SJ95]
HP_D	Halolysin [MSBL1]	Pat 8-10	Peptidase S8 (53%) [ <i>Haladaptatus paucihalophilus</i> ]
SP_A	Subtilisin [MBGE]		Peptidase S8 (50%) [ <i>Bacillus</i> sp. SG-1]
PD_A	Prephenate DH [MBGE]	Pro 1	Prephenate dehydrogenase (48%) [ <i>Methanobacterium formicicum</i> ]
DR_A1	DHPR [MSBL1]	PP 1	DHPR (56%) [ <i>Methanobacterium</i> sp. SWAN-1]
DR_A2	DHPR [MBGE]		DHPR (57%) [ <i>Methanocaldococcus vulcanius</i> M7]
DR_D	DHPR [MSBL1]		DHPR (54%) [ <i>Methanobacterium</i> sp. SWAN-1]
DR_K	DHPR [MSBL1]		DHPR (53%) [ <i>Methanobacterium</i> sp. SWAN-1]
CA_A	$\gamma$ -CA [MSBL1]	Manual PPM	Ferripyochelin binding protein (fbp, 43%) [ <i>Thermosiliminibacter oceanii</i> DSM 16646]
CA_D	$\gamma$ -CA [MSBL1]		Hypothetical protein (53%) [ <i>Corynebacterium pyruviciroducens</i> ]

Last letter of gene name indicates habitat: D, Discovery; A, Atlantis II; K, Kebrit.

higher reliability requires reducing the permissible flexibility. In the CAs example above, three consensus patterns were available with high reliability (**Table 8**). Hence we expected to identify several CAs through pattern matching. Yet, no CA was found in the entire database since the rigidity of these consensus patterns prevented identification of novel enzymes with the same function. Finally, a consensus pattern may not be specific for a specific function, e.g., NADH or ATP binding motifs typically are associated with consensus patterns, which occur in several enzyme families. **Table 7** illustrates this issue. Four PROSITE IDs are related to both, either alcohol dehydrogenase or ene reductase function. Identifying combinations of patterns can circumvent these problems and increase reliability. According to the PROSITE web page, one of the strongest pattern combinations is PS00136–PS00138. If a protein includes at least two of the three active site signatures, the probability of it showing a protease activity is assumed to be 100%.

Ontologies are widely used for functional annotation (Radivojac et al., 2013). Gene ontologies are commonly expressed by GO-terms. The source for GO-terms in the UniProt Gene Ontology Annotation database falls into three categories: (i) the smallest but most reliable category, experimental annotations, (ii) curated non-experimental annotations and last electronic annotations, (iii) with less reliability. Over 98% of the repository of the UniProt Gene Ontology Annotation database is inferred in silico without curator oversight (Škunca et al., 2012). GO-terms

are highly flexible, which is reflected in the gene's sequence length associated with it, e.g., annotation of GO-terms in this study covered 1.9–100% of the total gene. The particular sources used for GO-term identification leads to this large range. GO-terms based on consensus pattern naturally are reflected by a short associated sequence length (e.g., the 1.9% lower limit in this study). GO-terms determined by different methods (e.g., Hamap, TIGRFam, PIRSF) can take up to 100% of the sequence into consideration. In this analysis GO-terms association to ORFs was in average based on about 65% of the total sequence length. Recent studies could show that electronic annotations are more reliable than generally believed and that the overall reliability of electronically determined GO-annotations is increasing, but still very low. The mean value of reliability was ≈30% in 2006 and increase to 50% in 2011 (Škunca et al., 2012). The variations are significant among different inference methods, types of annotations, and organisms. Further, functional annotation, which is only based on GO-terms can result in a considerable bias (Schnoes et al., 2013). INDIGO utilizes all InterProScan derived GO-terms whether they are emerging from longer domains such as PFAM, TIGRFam, or PROSITE short consensus patterns. It is common that PROSITE IDs do not relate to any GO term, yet a longer domain in the vicinity or around a PROSITE pattern yields a GO-term associated to a POI. Currently, 11,910 ORFs (10.6%) annotated in INDIGO are associated with a GO-term and a PROSITE ID, which both describe the same function. The INDIGO data warehouse

**Table 8 | Reliability of consensus patterns found in chosen hits as well as for carbonic anhydrases.**

PROSITE ID	Description	Hits	TP (%)	FP (%)	FN	FN (%)
PS00059	Zinc-containing AD signature	491	97.4	2.6	40	8.1
PS00061	Short-chain DHR family signature	720	82.5	17.5	192	26.7
PS00065	NAD-binding 2-hydroxyacid DH signature	235	77.4	22.6	210	89.4
PS00136	Subtilase family, aspartic acid active site	328	45.1	54.9	90	27.4
PS00137	Subtilase family, histidine active site	200	92.5	7.5	56	28.0
PS00138	Subtilase family, serine active site	261	88.1	11.9	29	11.1
PS00671	NAD-binding 2-hydroxyacid DH signature 3	319	99.7	0.3	66	20.7
PS00913	Iron-containing ADH signature 1	42	81.0	19.0	26	61.9
PS01298	DHPR signature	541	100.0	0.0	19	3.5
PS00162	A-CA signature	64	100.0	0.0	32	50.0
PS00704	Prokaryotic-type CA signature 1	22	95.5	4.5	10	45.5
PS00705	Prokaryotic-type CA signature 2	25	100.0	0.0	5	20.0

TP, True positive; FP, False positive; FN, False negative (Sigrist et al., 2002).

based annotation (AAMG) combines various annotation methods. Unlike other data warehouses, INDIGO keeps and organizes all annotation meta data even if these are not in agreement with the final annotation (Alam et al., 2013). All GO-term and PROSITE IDs, which are available from these meta data are used by the PPM algorithm. In two cases, the PPM algorithm based function predictions differ from the INDIGO annotation.  $\gamma$ -CA identified by the PPM algorithm were previously annotated as “predicted acetyltransferase isoleucine patch superfamily” or “Ferricyochelin binding protein.” Other PPM algorithm based functions narrowed the INDIGO annotation down to only one function. The prephenate DHs were originally annotated as both, Chorismate and Prephenate DH.

In summary both, consensus patterns and GO-terms are standard tools to identify the function of a gene, yet they have weaknesses. The key to increase reliability is combination of descriptors. Since GO-terms (profiles) and PROSITE IDS (patterns) provide orthogonal information of protein function (with the exception of GO-terms based on consensus patterns) selecting combination of both descriptors is a powerful tool to identify the function of a gene product with higher reliability, particularly for novel and distantly related organisms. The PPM algorithm combines those advantages and is able to select for all three combinations of descriptors: the profile sets, the pattern sets and the profile and pattern sets. The strict PPM algorithm extracts and ranks in our case the top 0.1% of most reliably annotated genes. Since genomic data are growing at a much faster pace than experimental verification can proceed, a focus on quality rather than quantity is required. The PPM algorithm guides experimentalists to relevant starting points for successful expression, characterization, and verification of gene products.

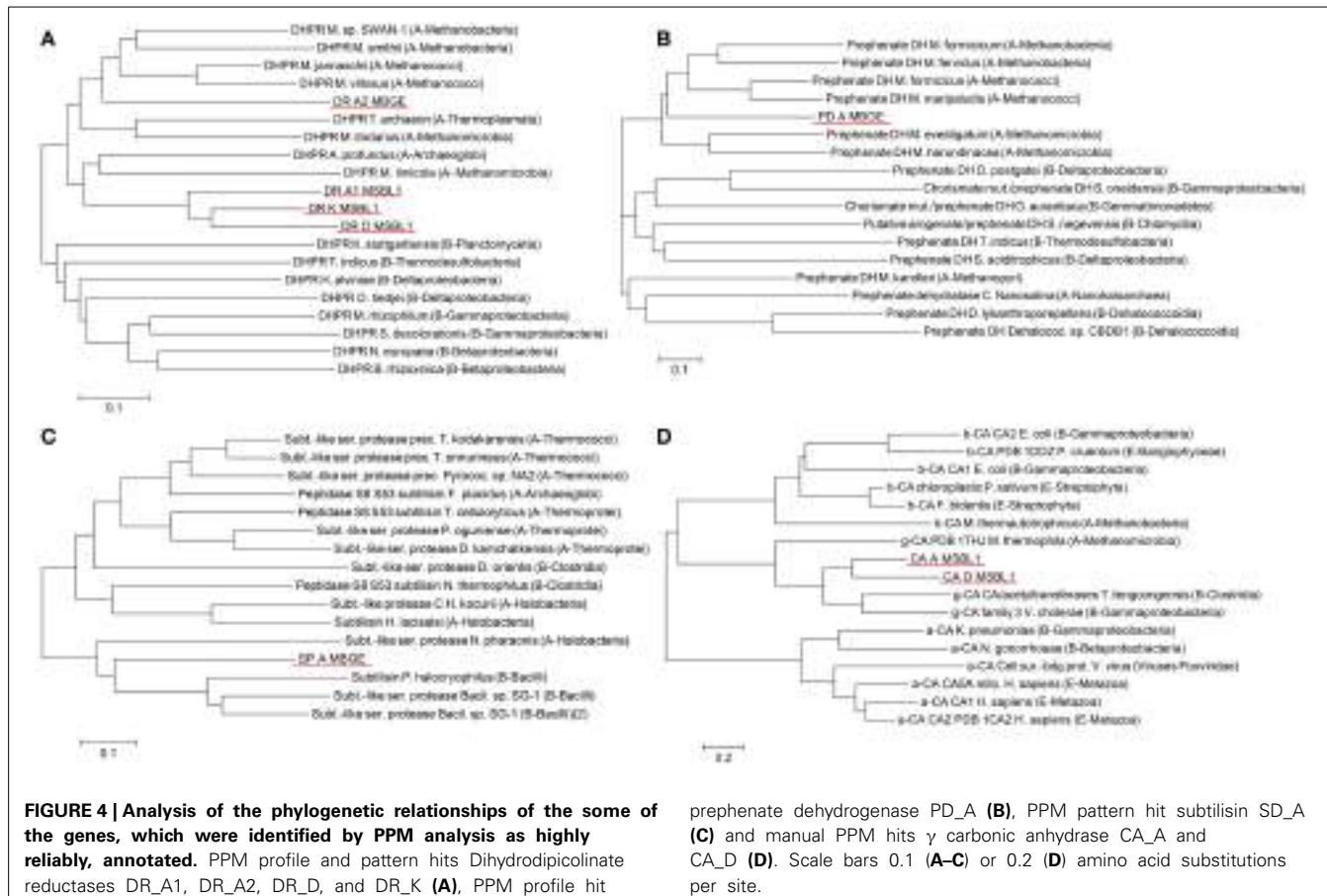
#### DISTANTLY RELATED SEQUENCES FROM NOVEL ORGANISMS

Phylogenetic analysis of gene sequences identified as candidates for expression tests revealed a high evolutionary distance to any known sequence (Figure 4). In case of the PPM profile and pattern set hits, which all are DHPRs, the phylogenetic tree with the closest related organisms includes both, the archeal and bacterial domains of life (Figure 4A). The four identified hits are all

in the archeal branch. The three hits from the organism MSBL1 (DR\_A1, DR\_D, and DR\_K) are clustering together in a separate branch, connected to *Acheoglobales* and *Methanomicroba*. The hit from the organism MBGE (DR\_A2) is in a separate branch and closer related to *Methanobacteria* and *Methanococci*. As indicated by the long branches the junction to the closest previously known sequences occurs at 0.3–0.35 amino acid substitutions per site. The PPM multi-profile hit prephenate dehydrogenase from MBGE (Figure 4B) shows phylogenetic relations similar to DHPR. The closest related enzymes found are from archaea and the closest related sequences are from *Methanococci* and *Methanobacteria*. The junction to the closest previously known sequences occurs at 0.33 amino acid substitutions per site. The subtilase type sequence from the PPM multi-pattern hit has a different phylogenetic footprint (Figure 4C). Based on the amino acid sequence the novel subtilisin shows equal evolutionary relations to archaea and bacteria, which indicates comparatively low sequence mutations in the two different domains compared to their common ancestor. For the  $\gamma$ -CA hits, which are based on a combination of a new profile and pattern, the phylogenetic tree includes all three known classes of CA (Figure 4D). The tree reveals clearly, that the identified sequences fall into the  $\gamma$  class of CAs with very distant relations to the  $\alpha$  and  $\beta$  class. Distant phylogenetic relationships are also found for all other hits, underlining the novelty of the SAGs analyzed (Figures S1–S3).

#### CURRENT LIMITATIONS OF THE PPM APPROACH

The PPM approach intrinsically leads to a high number of false negatives, because not all protein of interest groups can be translated into GO-terms and PROSITE IDs. During conversion from E.C. numbers to profiles (GO-terms) or pattern (PROSITE ID) about 35 or 81% of the POIs are lost, respectively. This limitation will be overcome through the exponential growth of biological data, which will increase the number and precision of GO-terms and PROSITE IDS. The combination of self-derived profiles and pattern can also enhance/enable PPM analysis, even with comparatively flexible sequences that show individually low reliability, as shown for the  $\gamma$ -CA example. Reducing the rigidity of consensus pattern with a high false negative rate may further help to increase



**FIGURE 4 | Analysis of the phylogenetic relationships of the some of the genes, which were identified by PPM analysis as highly reliably, annotated.** PPM profile and pattern hits Dihydrodipicolinate reductases DR\_A1, DR\_A2, DR\_D, and DR\_K (**A**), PPM hit

prephenate dehydrogenase PD\_A (**B**), PPM pattern hit subtilisin SD\_A (**C**) and manual PPM hits  $\gamma$  carbonic anhydrase CA\_A and CA\_D (**D**). Scale bars 0.1 (**A–C**) or 0.2 (**D**) amino acid substitutions per site.

hit rates. However, as discussed above, from an experimentalist point of view false positives are of much higher concern and these can be eliminated very effectively by the PPM approach.

#### OUTLOOK AND CONCLUSION—THE RED SEA EXTREMOPHILES AS SOURCE FOR NOVEL ENZYMES WITH HIGH SCIENTIFIC AND INDUSTRIAL POTENTIAL

For the first time SAGs were used to identify proteins for biotechnological applications. The eleven different genes, which were extracted from the INDIGO database during this study as candidates for expression just give a glimpse of the potential the Red Sea brine pools extremophiles have for discovery of novel enzymes. Not only the great phylogenetic distance to any described organism but also the extreme anoxic, high temperature, and hypersaline environment makes the enzymes of those organisms highly valuable. Enzymatic activity at high temperature and with low water activity can enable biocatalysis to be a tool for complex chemical reactions giving high yield and enantiomeric excess and under conditions that were so far out of reach for biological applications. Investigation of the enzymes, for which genes were identified here, will help understanding the limitations and adaptation of life at such extreme places.

The PPM algorithm is not intended to be a competitor for standard annotation. However, it is a powerful tool to analyze functions of proteins of extremophilic organisms that are only

distantly related to organisms described so far. The PPM algorithm helps experimentalists to extract proteins and particularly enzymes with high confidence from databases with only limited annotation reliability, e.g., when SAGs of extremophiles are used.

The combination of orthogonal descriptors may also facilitate screening of other genomic data for proteins of interest, e.g., those resulting from metagenomic or metatranscriptomic sampling as well as from shotgun sequencing. For metagenomic sequences the most reliable functional annotations are achieved using homology-based approaches against publicly available reference sequence databases including GO. Recently, it was recommended for metagenomic data to run a motif-based analysis (e.g., using PROSITE-IDs) in parallel to the homology-based functional prediction (Prakash and Taylor, 2012). The PPM algorithm provides an example using this approach. However, since the PPM algorithm was developed to minimize the number of false positive hits when experimentalists search genomic databases for proteins of interest and we expect also for metagenomic data that the increased reliability of genes identified by this algorithm will be its main advantage.

The publicly available scripts used in this study (i) AutoTECNo, (ii) PPM processor in combination with (iii) the INDIGO data warehouse are powerful tools, with a minimalist character to keep handling of extreme large datasets simple. The PPM algorithm will facilitate experimental characterization of extremophilic proteins and therefore help to increase the general

understanding of life at extreme conditions and exploiting its biotechnological potential. The enzymes identified in this study will be the first of many proteins on this path.

## ACKNOWLEDGMENTS

We gratefully acknowledge financial support by the Saudi Economic and Development Company (SEDCO) Research Excellence Award “The Deep-Sea Brine Pools of the Red Sea: From Novel Extreme Organisms to Commercial Applications” and the King Abdullah University of Science and Technology (KAUST) faculty baseline funding (Jörg Eppinger, Ulrich Stingl, and Vladimir Bajic). We thank Prof. James Gregory Ferry from the Department of Biochemistry and Molecular Biology, Pennsylvania State University, for helpful discussion leading to identification of the  $\gamma$  CA candidate genes described in this article. We are grateful for the valuable help of the scientists and crew on board RV Aegeo (3rd KAUST/WHOI Red Sea Expedition). We thank the Coastal and Marine Resources Core Lab (CMRC) of the King Abdullah University of Science and Technology for support and technical assistance and the Computational Bioscience Research Center (CBRC) computer cluster administration team, especially Allan Anthony Kamau for their help in making computational resources available for this work.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00134/abstract>

## REFERENCES

- Alain, K., and Querellou, J. (2009). Cultivating the uncultured: limits, advances and future challenges. *Extremophiles* 13, 583–594. doi: 10.1007/s00792-009-0261-3
- Alam, I., Antunes, A., Kamau, A. A., Kalkatawi, M., Stingl, U., and Bajic, V. B. (2013). INDIGO-INtegrated Data Warehouse of Microbial GenOmes with examples from the red sea extremophiles. *PLoS ONE* 8:e82210. doi: 10.1371/journal.pone.0082210
- Amann, R. I., Ludwig, W., and Schleifer, K.-H. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59, 143–169.
- Anschutz, P., and Blanc, G. (1995). Geochemical dynamics of the Atlantis II Deep (Red Sea): silica behavior. *Mar. Geol.* 128, 25–36. doi: 10.1016/0025-3227(95)00085-D
- Antunes, A., Ngugi, D. K., and Stingl, U. (2011). Microbiology of the Red Sea (and other) deep-sea anoxic brine lakes. *Environ. Microbiol. Rep.* 3, 416–433. doi: 10.1111/j.1758-2229.2011.00264.x
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Backer, H., and Schoell, M. (1972). New deeps with brines and metalliferous sediments in the Red Sea. *Nat. Phys. Sci.* 240, 153–158. doi: 10.1038/physci240153a0
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Bruneau, L., Jerlov, N. G., and Koczy, F. F. (1953). “Physical and chemical methods,” in *Reports of the Swedish Deep Sea Expedition*, Vol. 3. (Elanders: Göteborg), 99–112.
- Chistoserdova, L. (2010). Recent progress and new challenges in metagenomics for biotechnology. *Biotechnol. Lett.* 32, 1351–1359. doi: 10.1007/s10529-010-0306-9
- Chitsaz, H., Yee-Greenbaum, J. L., Tesler, G., Lombardo, M.-J., Dupont, C. L., Badger, J. H., et al. (2011). Efficient *de novo* assembly of single-cell bacterial genomes from short-read data sets. *Nat. Biotechnol.* 29, 915–921. doi: 10.1038/nbt.1966
- Cita, M. B. (2006). Exhumation of Messinian evaporites in the deep-sea and creation of deep anoxic brine-filled collapsed basins. *Sediment. Geol.* 188, 357–378. doi: 10.1016/j.sedgeo.2006.03.013
- DasSarma, S., and Arora, P. (2001). “Halophiles,” in *Encyclopedia of Life Sciences*, (London: John Wiley & Sons, Ltd), 1–9. doi: 10.1038/npg.els.0000394
- De Castro, E., Sigrist, C. J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., et al. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* 34, W362–W365. doi: 10.1093/nar/gkl124
- Dolled-Filhart, M. P., Lee, M. Jr., Ou-Yang, C.-W., Haraksingh, R. R., and Lin, J. C.-H. (2013). Computational and bioinformatics frameworks for next-generation whole exome and genome sequencing. *Sci. World J.* 2013:730210. doi: 10.1155/2013/730210
- Eder, W., Jahnke, L. L., Schmidt, M., and Huber, R. (2001). Microbial diversity of the brine-seawater interface of the Kebrit Deep, Red Sea, studied via 16S rRNA gene sequences and cultivation methods. *Appl. Environ. Microbiol.* 67, 3077–3085. doi: 10.1128/AEM.67.7.3077-3085.2001
- Emery, K., Hunt, J., and Hays, E. (1969). “Summary of hot brines and heavy metal deposits in the Red Sea,” in *Hot Brines and Heavy Metal Deposits in the Red Sea*, eds E. T. Degens, and D. A. Ross (Berlin: Springer Verlag), 557–571. doi: 10.1007/978-3-662-28603-6\_50
- Faber, E., Botz, R., Poggensburg, J., Schmidt, M., Stoffers, P., and Hartmann, M. (1998). Methane in Red Sea brines. *Org. Geochem.* 29, 363–379. doi: 10.1016/S0146-6380(98)00155-7
- Garber, M., Grabherr, M. G., Guttman, M., and Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods* 8, 469–477. doi: 10.1038/nmeth.1613
- Gurvich, E. (2006). Metalliferous sediments of the world ocean: fundamental theory of deep-sea hydrothermal sedimentation. Heidelberg: Springer Berlin Heidelberg.
- Hartmann, M., Scholten, J. C., Stoffers, P., and Wehner, F. (1998). Hydrographic structure of brine-filled deeps in the Red Sea—new results from the Shaban, Kebrit, Atlantis II, and Discovery Deep. *Mar. Geol.* 144, 311–330. doi: 10.1016/S0025-3227(97)00055-8
- Horner, D. S., Pavesi, G., Castrignanò, T., De Meo, P. D. O., Liumi, S., Sammethyl, M., et al. (2010). Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Brief. Bioinformatics* 11, 181–197. doi: 10.1093/bib/bbp046
- Hovland, M., Kuznetsova, T., Rueslätten, H., Kvamme, B., Johnsen, H. K., Fladmark, G. E., et al. (2006). Sub-surface precipitation of salts in supercritical seawater. *Basin Res.* 18, 221–230. doi: 10.1111/j.1365-2117.2006.00290.x
- Hugenholtz, P., and Tyson, G. W. (2008). Microbiology: metagenomics. *Nature* 455, 481–483. doi: 10.1038/455481a
- Jeyakanthan, J., Rangarajan, S., Mridula, P., Kanaujia, S. P., Shiro, Y., Kuramitsu, S., et al. (2008). Observation of a calcium-binding site in the-class carbonic anhydrase from Pyrococcus horikoshii. *Acta Crystallogr. D Biol. Crystallogr.* 64, 1012–1019. doi: 10.1107/S0907444908024323
- Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36, W5–W9. doi: 10.1093/nar/gkn201
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Karan, R., Kumar, S., Sinha, R., and Khare, S. (2012). “Halophilic microorganisms as sources of novel enzymes,” in *Microorganisms in Sustainable Agriculture and Biotechnology*, eds T. Satyanarayana, B. Narain Johri, and A. Prakash (New Delhi: Springer), 555–579.
- Karbe, L. (1987). “Hot brines and the deep sea environment,” in *Red Sea. Key Environments Series*, eds A. J. Edwards, and S. M. Head (Oxford: Pergamon Press), 70–89.
- Kisker, C., Schindelin, H., Alber, B. E., Ferry, J. G., and Rees, D. C. (1996). A left-hand beta-helix revealed by the crystal structure of a carbonic anhydrase from the archaeon Methanoscincus thermophila. *EMBO J.* 15, 2323.
- Könneke, M., Bernhard, A. E., José, R., Walker, C. B., Waterbury, J. B., and Stahl, D. A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 437, 543–546. doi: 10.1038/nature03911

- Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., and Hugenholtz, P. (2008). A bioinformatician's guide to metagenomics. *Microbiol. Mol. Biol. Rev.* 72, 557–578. doi: 10.1128/MMBR.00009-08
- Lasken, R. S. (2007). Single-cell genomic sequencing using multiple displacement amplification. *Curr. Opin. Microbiol.* 10, 510–516. doi: 10.1016/j.mib.2007.08.005
- Liszka, M. J., Clark, M. E., Schneider, E., and Clark, D. S. (2012). Nature versus nurture: developing enzymes that function under extreme conditions. *Annu. Rev. Chem. Biomol. Eng.* 3, 77–102. doi: 10.1146/annurev-chembioeng-061010-114239
- Lohse, M., Bolger, A. M., Nagel, A., Fernie, A. R., Lunn, J. E., Stitt, M., et al. (2012). RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* 40, W622–W627. doi: 10.1093/nar/gks540
- Madigan, M., Martinko, J. M., and Parker, J. (2003). *Brock Biology of Microorganisms*. New Jersey: Pearson Education International.
- Ollivier, B., Caumette, P., García, J.-L., and Mah, R. A. (1994). Anaerobic bacteria from hypersaline environments. *Microbiol. Rev.* 58, 27–38.
- Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efremova, M., et al. (2014). A survey of tools for variant analysis of next-generation genome sequencing data. *Brief. Bioinform.* 15, 256–278. doi: 10.1093/bib/bbs086
- Peng, Y., Leung, H. C., Yiu, S. M., and Chin, F. Y. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi: 10.1093/bioinformatics/bts174
- Prakash, T., and Taylor, T. D. (2012). Functional assignment of metagenomic data: challenges and applications. *Brief. Bioinformatics* 13, 711–727. doi: 10.1093/bib/bbs033
- Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., et al. (2005). InterProScan: protein domains identifier. *Nucleic Acids Res.* 33, W116–W120. doi: 10.1093/nar/gki442
- Radivojac, P., Clark, W.T., Oron, T.R., Schnoes, A.M., Wittkop, T., Sokolov, A., et al., (2013). A large-scale evaluation of computational protein function prediction. *Nat. Methods* 10, 221–227. doi: 10.1038/nmeth.2340
- Rohde, H., Qin, J., Cui, Y., Li, D., Loman, N. J., Hentschke, M., et al. (2011). Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104: H4. *N. Engl. J. Med.* 365, 718–724. doi: 10.1056/NEJMoa1107643
- Ross, D. A. (1972). Red Sea hot brine area: revisited. *Science* 175, 1455–1457. doi: 10.1126/science.175.4029.1455
- Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Schnoes, A. M., Ream, D. C., Thorman, A. W., Babbitt, P. C., and Friedberg, I. (2013). Biases in the experimental annotations of protein function and their effect on our understanding of protein function space. *PLoS Comput. Biol.* 9:e1003063. doi: 10.1371/journal.pcbi.1003063
- Schomburg, I., Chang, A., Placzek, S., Söhngen, C., Rother, M., Lang, M., et al. (2013). BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res.* 41, D764–D772. doi: 10.1093/nar/gks1049
- Searle, R. C., and Ross, D. A. (1975). A geophysical study of the Red Sea axial trough between 20° 5' and 22° N. *Geophys. J. Int.* 43, 555–572. doi: 10.1111/j.1365-246X.1975.tb00647.x
- Sigrist, C. J., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., et al. (2002). PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief. Bioinformatics* 3, 265–274. doi: 10.1093/bib/3.3.265
- Sigrist, C. J. A., De Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., et al. (2013). New and continuing developments at PROSITE. *Nucleic Acids Res.* 41, D344–D347. doi: 10.1093/nar/gks1067
- Škunca, N., Altenhoff, A., and Dessimoz, C. (2012). Quality of computationally inferred gene ontology annotations. *PLoS Comput. Biol.* 8:e1002533. doi: 10.1371/journal.pcbi.1002533
- Smith, K. S., and Ferry, J. G. (2000). Prokaryotic carbonic anhydrases. *FEMS Microbiol. Rev.* 24, 335–366. doi: 10.1111/j.1574-6976.2000.tb00546.x
- Smith, R. N., Aleksić, J., Butano, D., Carr, A., Contrino, S., Hu, F., et al. (2012). InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data. *Bioinformatics* 28, 3163–3165. doi: 10.1093/bioinformatics/bts577
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Triplet, T., and Butler, G. (2013). A review of genomic data warehousing systems. *Brief. Bioinform.* doi: 10.1093/bib/bbt031. [Epub ahead of print].
- Watson, S. W., and Waterbury, J. B. (1969). "The sterile hot brines of the Red Sea," in *Hot Brines and Recent Heavy Metal Deposits in the Red Sea*, eds E. T. Degens, and D. A. Ross (New York, NY: Springer Berlin Heidelberg), 272–281. doi: 10.1007/978-3-662-28603-6\_27
- Yooseph, S., Sutton, G., Rusch, D. B., Halpern, A. L., Williamson, S. J., Remington, K., et al. (2007). The sorcerer II global ocean sampling expedition: expanding the universe of protein families. *PLoS Biol.* 5:e16. doi: 10.1371/journal.pbio.0050016
- Zuckerkandl, E., and Pauling, L. (1965). Evolutionary divergence and convergence in proteins. *Evolving Genes Proteins* 97, 97–166.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received:** 10 January 2014; **accepted:** 16 March 2014; **published online:** 07 April 2014.  
**Citation:** Grötzing SW, Alam I, Ba Alawi W, Bajic VB, Stingl U and Eppinger J (2014) Mining a database of single amplified genomes from Red Sea brine pool extremophiles—improving reliability of gene function prediction using a profile and pattern matching algorithm (PPMA). *Front. Microbiol.* 5:134. doi: 10.3389/fmicb.2014.00134  
*This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.*  
*Copyright © 2014 Grötzing, Alam, Ba Alawi, Bajic, Stingl and Eppinger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# Population and genomic analysis of the genus *Halorubrum*

Matthew S. Fullmer<sup>1</sup>, Shannon M. Soucy<sup>1</sup>, Kristen S. Swithers<sup>1,2</sup>, Andrea M. Makkay<sup>1</sup>, Ryan Wheeler<sup>1</sup>, Antonio Ventosa<sup>3</sup>, J. Peter Gogarten<sup>1</sup> and R. Thane Papke<sup>1\*</sup>

<sup>1</sup> Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT, USA

<sup>2</sup> Department of Cell Biology, Yale School of Medicine, Yale University, New Haven, CT, USA

<sup>3</sup> Department of Microbiology and Parasitology, University of Seville, Seville, Spain

**Edited by:**

Jesse Dillon, California State University, Long Beach, USA

**Reviewed by:**

Jesse Dillon, California State University, Long Beach, USA  
Federico Lauro, University of New South Wales, Australia

**\*Correspondence:**

R. Thane Papke, Microbiology Program, Department of Molecular and Cell Biology, University of Connecticut, 91 N. Eagleville Rd., Storrs, CT 06269-3125, USA  
e-mail: thane@uconn.edu

The Halobacteria are known to engage in frequent gene transfer and homologous recombination. For stably diverged lineages to persist some checks on the rate of between lineage recombination must exist. We surveyed a group of isolates from the Aran-Bidgol endorheic lake in Iran and sequenced a selection of them. Multilocus Sequence Analysis (MLSA) and Average Nucleotide Identity (ANI) revealed multiple clusters (phylogroups) of organisms present in the lake. Patterns of intein and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) presence/absence and their sequence similarity, GC usage along with the ANI and the identities of the genes used in the MLSA revealed that two of these clusters share an exchange bias toward others in their phylogroup while showing reduced rates of exchange with other organisms in the environment. However, a third cluster, composed in part of named species from other areas of central Asia, displayed many indications of variability in exchange partners, from within the lake as well as outside the lake. We conclude that barriers to gene exchange exist between the two purely Aran-Bidgol phylogroups, and that the third cluster with members from other regions is not a single population and likely reflects an amalgamation of several populations.

**Keywords:** Halobacteria, Multilocus Sequence Analysis (MLSA), Average Nucleotide Identity (ANI), intein, CRISPR

## INTRODUCTION

Besides an obligate requirement for high concentrations of NaCl, a unifying trait of Halobacteria (often referred to colloquially as the haloarchaea)—a class within the archaeal phylum Euryarchaeota, is their propensity for horizontal gene transfer (HGT) (Legault et al., 2006; Rhodes et al., 2011; Nelson-Sathi et al., 2012; Williams et al., 2012). Although HGT occurs continuously, events that provide an adaptive advantage and are maintained in modern lineages can be detected. For instance, HGTs from bacterial lineages into the Halobacteria occurred before their last common ancestor and brought respiration and nutrient transport genes that transformed them from a methanogen to their current aerobic heterotrophic state (Nelson-Sathi et al., 2012). Other examples including rhodopsins (Sharma et al., 2006), tRNA synthetases (Andam et al., 2012), 16S rRNA genes (Boucher et al., 2004), membrane proteins (Cuadros-Orellana et al., 2007), and genes allowing the assembly of novel pathways (Khomyakova et al., 2011) have been reported for this group and reflect the adaptive benefit of acquiring these genes.

HGT into the Halobacteria has profoundly impacted their evolution; however, understanding this contribution is only part of their evolutionary picture. The study of recombination frequency among this class has been utilized to address population genetics questions that address whether they are clonal (i.e., linked alleles at different loci) or “sexual” in the sense that alleles at different loci are randomly associated. Several studies have addressed those questions by assessing the impact of

frequent HGT on Halobacteria. Homologous replacement of loci was inferred within and between phylogenetic clusters (phylogroups) using Multilocus Sequence Analysis (MLSA) on closely related strains (Papke et al., 2004) and comparative analyses of genomes (Williams et al., 2012). Within phylogroups where genetic diversity was less than one percent divergent for protein coding genes, alleles at different loci were randomly associated whereas between phylogroups they were not (Papke et al., 2007) indicating haloarchaea are highly sexual. Measurements of frequency across the breadth of halobacterial diversity indicates no absolute barrier to homologous recombination; rather between relatives, there is a log-linear decay in recombination frequency relative to phylogenetic distance (Williams et al., 2012).

Laboratory experiments also support these results. Mating experiments measuring the rate of recombination using *Halofexax (Hfx) volcanii* and *Hfx. mediterranei* auxotrophs demonstrated the degree of genetic isolation between species was much lower than expected. The observed rate of exchange between species suggested that given an opportunity over time these species would homogenize, indicating strong barriers to recombination would have to exist for speciation to occur, and for lineages to be maintained (Naor et al., 2012). Further, mating experiments demonstrated that enormous genomic fragments (i.e., 300–500 kb, ~18% of the chromosome size) could be exchanged in a single event (Naor et al., 2012). Similar large fragment exchange events were recently observed in natural isolates from Deep Lake (Antarctic hypersaline lake): Distantly related strains

(<75% average nucleotide identity) shared up to 35 kb with nearly 100% sequence identity (DeMaere et al., 2013).

The Halobacteria have clearly been shaped by gene transfer and are actively engaged in substantial genetic exchange. However, little is known about genomic diversity within populations, and the impact of gene flow is unknown at these scales. In this study we report the intra and inter population sequence diversity of *Halorubrum* spp. strains cultivated from the same location and compare them to the genomic diversity of type strains from the same genus. Our results lead to insights on the genomic diversity that comprises haloarchaeal species.

## METHODS

### GROWTH CONDITIONS AND DNA EXTRACTION

*Halorubrum* spp. cultures were grown in Hv-YPC medium (Allers et al., 2004) at 37°C with agitation. DNA from Halobacteria was isolated as described in the Halohandbook (Dyall-Smith, 2009). Briefly, stationary-phase cells were pelleted at 10,000 × g, supernatant was removed and the cells were lysed in distilled water. An equal volume of phenol was added, and the mixture was incubated at 65°C for 1 h prior to centrifugation to separate the phases. The aqueous phase was reserved and phenol extraction was repeated without incubation, and followed with a phenol/chloroform/iso-amyl alcohol (25:24:1) extraction. The DNA was precipitated with ethanol, washed, and re-suspended in TE (10 mM tris, pH 8.0, 1 mM EDTA).

### MULTILOCUS SEQUENCE ANALYSIS (MLSA)

Five housekeeping genes were amplified using PCR. The loci were *atpB*, *ef-2*, *glnA*, *ppsA*, and *rpoB* and the primers used for each locus are listed in Table 1. To more efficiently sequence PCR products, an 18 bp M13 sequencing primer was added to the 5' end of each degenerate primer (Table 1). Each PCR reaction was 20 μl in volume. The PCR reaction was run on a Mastercycler Ep Thermocycler (Eppendorf) using the following PCR cycle protocol: 30 s initial denaturation at 98°C, followed by 40 cycles of 30 s at 98°C, 5 s at the annealing temperature for each set of primers and 15 s at 72°C. Final elongation occurred at 72°C for 1 min. Table 2 provides a detailed list of reagents and the PCR mixtures for each amplified locus. The PCR products were separated by gel electrophoresis with agarose (1%). Gels were stained with ethidium bromide. An exACTGene mid-range plus DNA ladder (Fisher Scientific International Inc.) was used to estimate the size of the amplicons, which were purified using Wizard SV gel and PCR cleanup system (Promega). The purified amplicons were sequenced by Genewiz Inc. using Sanger sequencing technology.

### GENOME SEQUENCING

DNA purity was analyzed with a Nanodrop spectrophotometer, was quantified using a Qubit fluorometer (Invitrogen) and then prepared for sequencing using the Illumina Nextera XT sample preparation kit as described by the manufacturer. Fragmented and amplified libraries were either normalized using the normalization beads and protocol supplied with the kit, or manually as described in protocols for the Illumina Nextera kit. Libraries were loaded onto 500 cycle MiSeq reagent kits with a 5% spike-in PhiX control, and sequenced using an Illumina MiSeq benchtop sequencer. The genomes to be sequenced were selected based

**Table 1 | Degenerate primers used to PCR amplify and sequence the genes for MLSA.**

MLSA primer sequence 5'-3'				
Locus	Forward	Reverse		
atpB	tgt aaa acg acg gcc agt aac ggg gag scv ats aac cc	cag gaa aca gct atg act tca ggt cvg trt aca tgt a		
ef-2	tgt aaa acg acg gcc agt atc ccg gct bta yaa stg g	cag gaa aca gct atg act ggt cga tgg wyt cga ahg g		
glnA	tgt aaa acg acg gcc agt cag gta cgg gtt aca sga cgg	cag gaa aca gct atg acc ctc gcs ccg aar gac ctc gc		
ppsA	tgt aaa acg acg gcc agt ccg ccg tar ccv agc atc gg	cag gaa aca gct atg aca tcg tca ccg acg arg gyg g		
rpoB	tgt aaa acg acg gcc agt tcg aag agc cgg acg aca tgg	cag gaa aca gct atg acc ggt cag cac ctg bac cgg ncc		

**Table 2 | PCR conditions for each locus.**

	atpB	ef-2	glnA	ppsA	rpoB
Water (μl)	11.6	8.2	11.8	7.9	11.9
5× phire reaction buffer (μl)	4.0	4.0	4.0	4.0	4.0
DMSO (μl)	0.6	0	0.4	0.6	0.6
Acetamide (25%, μl)	0	4.0	0	4.0	0
dNTP mix (10 mM, μl)	0.4	0.4	0.4	0.4	0.4
Forward primer (10 mM, μl)	1.0	1.0	1.0	1.0	1.0
Reverse primer (10 mM, μl)	1.0	1.0	1.0	1.0	1.0
Phire II DNA polymerase (μl)	0.4	0.4	0.4	0.4	0.4
Template DNA (20 ng/μl, μl)	1.0	1.0	1.0	0.7	0.7
Annealing temperature (°C)	60.0	61.0	69.6	66.0	63.7

upon the results of the initial PCR MLSA data analysis (see Results).

### GENOME ASSEMBLY

Type strain genomes were obtained from the NCBI ftp repository. *Halorubrum lacusprofundi* and the non-*Halorubrum* genomes (*Haloarcula marismortui* ATCC 43049 and *Har. hispanica* ATCC 33960 as well as *Haloferax volcanii* DS2 and *Hfx. mediterranei* ATCC 33500) are completed projects. The other *Halorubrum* genomes are drafts, also obtained from the NCBI ftp repository. New draft genomes were sequenced using an Illumina MiSeq platform. Assembly on strain Ga2p was carried out using the nogo pipeline (Tritt et al., 2012) while all others were assembled via the CLC Genomics Workbench 6.0.5 suite with a trim and merge workflow with scaffolding enabled.

To ensure equal gene calling across the genomes all genomes, including the 19 draft and completed *Halorubrum*, *Haloferax*, and *Haloarcula* genomes available on the NCBI ftp site as of June 2013, were reannotated using the rapid annotation using subsystem technology (RAST) server (Aziz et al., 2008). Assembled contigs were reconstructed from the RAST-generated genbank files for all genomes using the seqret application of the emboss package (Rice et al., 2000).

## PHYLOGENETIC METHODOLOGY

Top scoring BLASTn hits for each MLSA target gene (*atpB*, *ef-2*, *glnA*, *ppsA*, and *rpoB*) in each genome were identified. Multiple-sequence alignments (MSAs) were generated by translating the genes to protein sequences in SeaView (Gouy et al., 2010), aligning the proteins using MUSCLE (v.3.8.31) (Edgar, 2004) and then reverting back to the nucleotide sequences. In-house scripts created a concatenated alignment of all five genes. The best model of evolution was determined by calculating the Akaike Information Criterion with correction for small sample size (AICc) in jModelTest 2.1.4 (Guindon et al., 2010; Darriba et al., 2012). The best-fitting model was GTR + Gamma estimation + Invariable site estimation. A maximum likelihood (ML) phylogeny was generated from the concatenated MSA and individual gene phylogenies from the individual gene MSAs using PhyML (v3.0\_360-500M)(Guindon et al., 2010). PhyML parameters consisted of GTR model, estimated p-invar, 4 substitution rate categories, estimated gamma distribution, subtree pruning, and regrafting enabled with 100 bootstrap replicates.

## PAIRWISE SEQUENCE IDENTITY CALCULATION

Calculation of pairwise identities was carried out using Clustal Omega on the EMBL-EBI webserver (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). The alignments were uploaded and percent identity matrices calculated (Sievers et al., 2011).

## INTEIN METHODOLOGY

To retrieve haloarchaeal intein sequences Position-Specific Scoring Matrices (PSSMs) were created using the collection of all inteins from InBase, the Intein database, and registry (Perler, 2002). A custom database was created with all inteins, and each intein was used as a seed to create a PSSM using the custom database. These PSSMs were then used as a seed for PSI-BLAST (Altschul et al., 1997) against each of the halobacterial genomes available from NCBI. A size exclusion step was then performed to remove false positives. Inteins were then aligned using MUSCLE (Edgar, 2004) with default parameters in the SeaView version 4.0 software package (Gouy et al., 2010). Insertions, which passed the size exclusion step but did not contain splicing domains, were filtered out and the previous steps were repeated using the resulting dataset on this study's dataset. Once the collection of haloarchaeal inteins was complete, sequences were re-aligned using SATé v2.2.2 (Liu et al., 2012) to generate a final alignment.

## INTEIN PHYLOGENETIC METHODOLOGY

Intein protein sequences were retrieved using in house scripts. Each intein allele was aligned separately using MUSCLE (v.3.8.31) (Edgar, 2004). In-house scripts created a concatenated alignment from the allele alignments. ProtTest v3.4 (Darriba et al., 2011) evaluated the protein sequences for an optimal model using the AICc and returned WAG\_I+G+F. A presence-absence matrix of zeros and ones was amended to each taxon's alignment data. The presence-absence data allows for grouping of taxa by sharing or lacking an allele. This complements the protein data, and allows the resolution of taxa with few inteins from those lacking them entirely or possessing many. To accommodate the two different formats of data simultaneously MrBayes v3.2.2 (Ronquist and

Huelsenbeck, 2003; Ronquist et al., 2012) was employed for the phylogenetic reconstruction.

## AVERAGE NUCLEOTIDE IDENTITY/TETRAMER ANALYSIS

JSpecies1.2.1 (Richter and Rosselló-Móra, 2009) was used to analyze the genomes for Average Nucleotide Identity (ANI) and tetramer frequency patterns. As the relationships of interest for this study are within the same genus only the nucmer and tetra algorithms were used. The BLAST-based ANI was not used as we were primarily interested in understanding the degree of relatedness between closely related organisms, which the nucmer method is equally capable of (Richter and Rosselló-Móra, 2009). Additionally, the increased rate of drop-off between moderately divergent sequences (<90%) the nucmer method yields relative to the BLAST method (Richter and Rosselló-Móra, 2009) was useful in highlighting when organisms were dissimilar. The default settings for both algorithms were used (Richter and Rosselló-Móra, 2009).

## CODON POSITION GC CONTENT

Complete sets of nucleotide sequences for all called ORFs were downloaded from RAST. In house scripts confirmed that all ORF calls were divisible by three and thus could be taken as in-frame. In house scripts were used to calculate the GC percentages for each codon position in each genome. Two-tailed *t*-tests were calculated using the StatsPlus software package (AnalystSoft, 2009).

## CRISPRs

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) presence/absence patterns were determined using the CRISPR Recognition Tool (CRT) v1.2 (Bland et al., 2007) with minimum repeat and minimum spacer parameters set to 30 nucleotides. All other parameters were the CRT defaults.

## RESULTS

### ASSEMBLED GENOMES

The assembled genomes ranged in size from 2.3 to 4.2 Mb. The median assembled genome size is 3.6 Mb. The median N50 (the size of the contig where 50% of the basepairs in the assembly are part of a contig that size or larger. N75 and N90 are similar but use 75 and 90% cutoffs) was 47.5 kb with a range from 1.86 to 80.3 kb (see Table 3, for statistics on the assembled genomes). Plasmids were not identified during assembly. As such, if some isolates possess differing numbers or types of plasmids then some of the genome-to-genome size variability may be attributable to this. A list of genomes used in this study can be found in Table 4.

### PHYLOGENETIC ASSIGNMENT OF PHYLOGROUPS

Initial MLSA analysis (5-genes: *atpD*, *ef-2*, *glnA*, *radA*, *rpoB*) revealed the presence of three well-supported clusters [hereafter referred to as phylogroups *in sensu* (Papke et al., 2007)] within the canonical *Halorubrum* population of Aran-Bidgol (Figures 1, 2). A phylogroup was initially defined as a cluster of isolates with very low sequence divergence across the sequenced (MLSA) loci (<~1%). Seventeen of these isolates were then selected for genome sequencing for a higher resolution assessment. Selection criteria were biased toward the two larger phylogroups (A and B) to facilitate comparison between clusters. Only a single genome

**Table 3 | Assembly statistics for the genomes sequenced in this study.**

	C191	C3	C49	Cb34	E3	E8	Ea1	Ea8	Eb13	Ec15	Fb21	G37	Ga2p	Ga36	Hd13	Ib24	LD3	LG1
N75 (kb)	18.9	2.3	23.2	24.7	1.1	30.0	25.1	25.4	42.7	25.3	27.2	41.1	23.8	32.1	23.2	21.4	8.4	
N50 (kb)	54.9	4.4	56.3	42.9	1.9	2.3	43.8	51.6	80.3	42.7	68.1	74.9	51.2	64.4	43.4	39.6	32.1	
N25 (kb)	97.3	7.8	99.8	73.4	3.5	4.0	77.5	95.4	95.7	131.8	90.3	118.4	118.9	91.9	83.0	68.2	76.0	67.9
Minimum (kb)	0.5	0.4	0.5	0.5	0.4	0.4	0.5	0.5	0.5	0.6	0.5	0.5	0.3	0.5	0.5	0.5	0.4	
Maximum (kb)	180.2	40.5	183.6	123.4	26.7	25.0	203.3	169.6	268.1	412.4	174.7	230.0	246.3	145.6	122.0	190.3	145.8	153.4
Average (kb)	16.6	2.9	22.5	23.1	1.5	1.8	24.7	22.6	23.3	44.3	20.6	25.7	40.3	21.0	27.9	19.6	17.5	4.4
Contig count	233	1165	159	145	2784	1278	159	166	156	74	176	138	83	160	137	189	213	1090
Length (Mb)	3.87	3.33	3.58	3.35	4.21	2.26	3.93	3.75	3.63	3.28	3.63	3.55	3.35	3.36	3.82	3.70	3.73	4.79
Base composition (GC%)	66.0	65.8	65.8	67.6	65.5	66.3	67.0	67.6	67.5	67.6	66.6	67.1	67.8	67.6	67.6	66.2	66.0	
Number of coding sequences	3908	3379	3529	3323	4147	2187	3977	3672	3544	3245	3860	3617	3400	3382	3718	3612	3724	4615
Number of RNAs	57	37	49	54	51	31	50	49	48	47	65	48	49	47	51	48	56	69

from phylogroup C was sequenced. Once genomic data were available, the PCR amplicons were replaced with the full-length genes from the assemblies. Further analysis made use of only these genomic sequences. The addition of the 19 NCBI genomes was made to provide context to the placement of the phylogroups within the genus and to determine their relationship with each other. The phylogenetic reconstruction including the type strains sequences revealed the presence of a fourth phylogroup (designated D) composed of three isolates from Aran-Bidgol and five type strains isolated from Central Asia and China (Figure 2).

#### PHYLOGROUPS A AND B ARE WELL-SUPPORTED AS DISCRETE AND COHESIVE ENTITIES

The bootstrap values provided by the phylogenetic reconstruction strongly supported both phylogroups A and B. Individual gene trees and the concatenated gene tree returned support values of 99% or higher for all of the clusters (Figures 1, 2) and the trees showed no paraphyly with other taxa. Both phylogroups also displayed sequence divergence below 1% across the five loci (Table 5). Further, genome-level analysis (ANI) demonstrated similar results to the MLSA data (Figure 3). Additional support for these phylogroups came from the tetramer frequency analysis, which found no discordance amongst the members of either group, and each phylogroup displayed an intra-group ANI  $\geq 98\%$ . An analysis of G+C composition in the protein coding ORFs found that the strains within phylogroups A and B had a statistically different content in overall coding G+C and at the third codon position ( $P < 0.05$  for both, Figure 4). Analyses of the inter-phylogroup differences showed the two phylogroups were quite different from each other and all other examined taxa. Both clusters were less than 97% similar in their pairwise MLSA distance to any other taxon in this study. Additionally phylogroups A and B were different from each other in tetramer frequency (below the 0.9900 correlation of Richter and Rosselló-Móra, 2009), ANI (only  $\sim 87\%$  identity), and G+C content in the third codon position ( $P < 0.05$ ; two-tailed  $t$ -test, Figure 4). Taken together these data support the notion that these phylogroups are discrete entities within a single environment, and that the individual phylogroups are cohesive.

To further evaluate the cohesion of the phylogroups a survey of inteins was performed. Inteins are molecular parasites that invade new hosts through horizontal transmission (Okuda et al., 2003; Swithers et al., 2013). Their patterns of presence and absence have been used as a barometer for horizontal transfer between closely and distantly related lineages (Swithers et al., 2013). Analysis of intein distributions supported earlier findings of cohesion within phylogroups and major distinctions between the phylogroups (Figure 5). Phylogroup A contains three non-fixed intein alleles that are present in more than half of the isolates, *cdc21a*, *cdc21b*, and *pol-IIa*. Phylogroup B contains four non-fixed intein alleles also present in half or more of its isolates, *rir1-b*, *rfc-a*, *polBa*, and *polBb* but are absent from phylogroup A. Closer examination of the two shared alleles reveals that these inteins are not the same between the phylogroups. The *pol-IIa* inteins in phylogroup B are 515aa long while those in phylogroup A are 494aa long, indicating an insertion or deletion event occurred in one of the phylogroups before the intein spread through the population. The preservation

**Table 4 | List of genomes used in this study.**

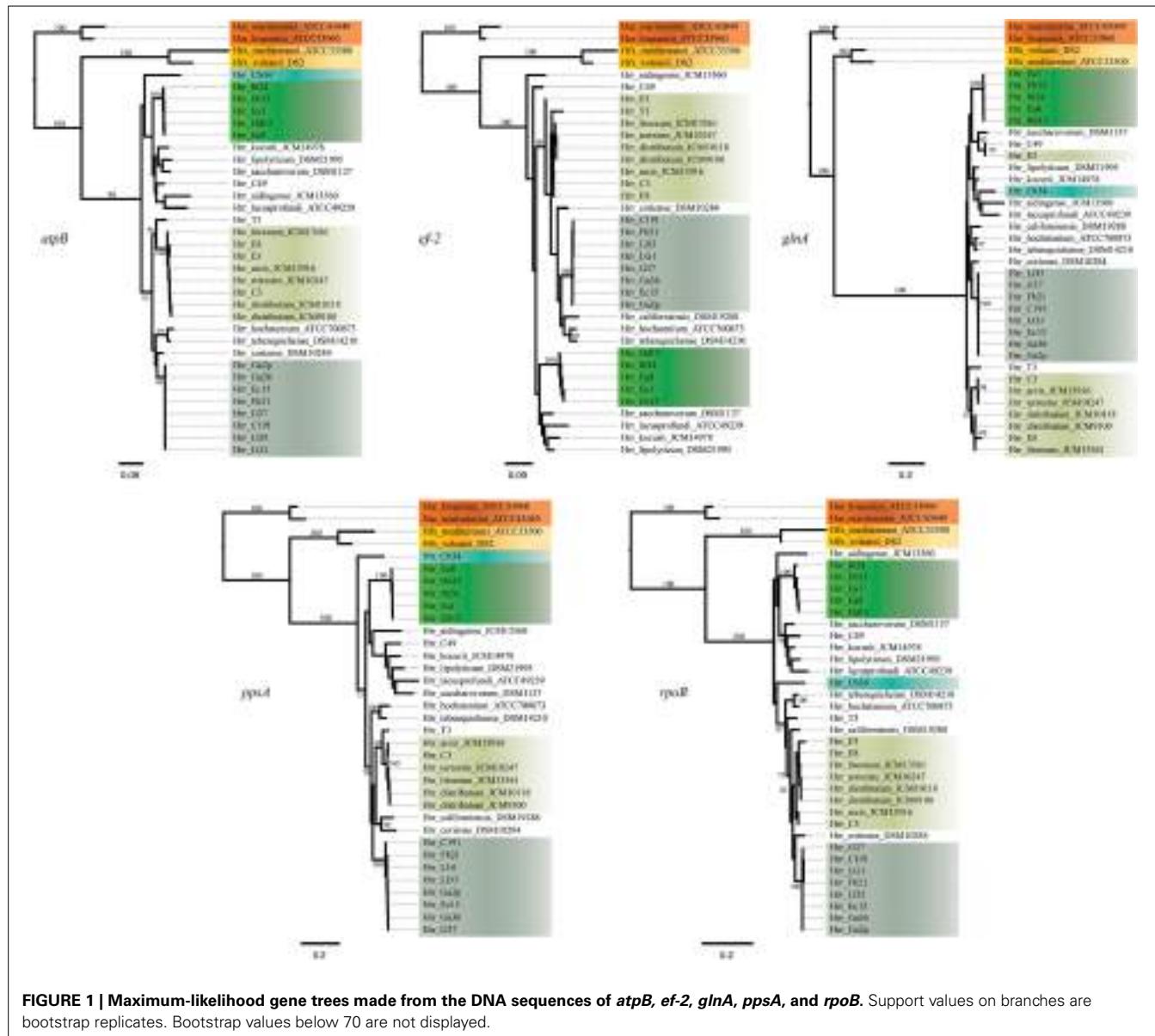
Organism name	NCBI identifier	Sequence source	Isolation site	Environment	Status
<i>Haloarcula hispanica</i> ATCC 33960	PRJNA72475	NCBI	Alicante, Spain	Solar saltern	Complete
<i>Haloarcula marismortui</i> ATCC 43049	PRJNA57719	NCBI	Dead Sea, Israel	Saline lake/sea	Complete
<i>Haloferax mediterranei</i> ATCC 33500	PRJNA167315	NCBI	Alicante, Spain	Solar saltern	Complete
<i>Haloferax volcanii</i> DS2	PRJNA46845	NCBI	Dead Sea, Israel	Saline lake/sea	Complete
<i>Halorubrum</i> sp. T3	PRJNA199598	NCBI	Yunnan, China	Solar saltern	Draft
<i>Halorubrum aidingense</i> JCM 13560	PRJNA188616	NCBI	Xin-Jiang, China	Saline lake	Draft
<i>Halorubrum arcis</i> JCM 13916	PRJNA188617	NCBI	Xin-Jiang, China	Saline lake	Draft
<i>Halorubrum californiensis</i> DSM 19288	PRJNA188618	NCBI	California, United States	Solar saltern	Draft
<i>Halorubrum coriense</i> DSM 10284	PRJNA188619	NCBI	Geelong, Australia	Solar saltern	Draft
<i>Halorubrum distributum</i> JCM 10118	PRJNA188621	NCBI	Turkmenistan	Saline soils	Draft
<i>Halorubrum distributum</i> JCM 9100	PRJNA188620	NCBI	Turkmenistan	Saline soils	Draft
<i>Halorubrum hochstetnium</i> ATCC 700873	PRJNA188622	NCBI	California, United States	Solar saltern	Draft
<i>Halorubrum kocurii</i> JCM 14978	PRJNA188615	NCBI	Inner Mongolia, China	Saline lake	Draft
<i>Halorubrum lacusprofundi</i> ATCC 49239	PRJNA58807	NCBI	Deep Lake, Antarctica	Saline lake	Complete
<i>Halorubrum lipolyticum</i> DSM 21995	PRJNA188614	NCBI	Xin-Jiang, China	Saline lake	Draft
<i>Halorubrum litoreum</i> JCM 13561	PRJNA188613	NCBI	Fujian, China	Solar saltern	Draft
<i>Halorubrum saccharovorum</i> DSM 1137	PRJNA188612	NCBI	California, United States	Solar saltern	Draft
<i>Halorubrum tebenquichense</i> DSM 14210	PRJNA188611	NCBI	Atacama, Chile	Solar saltern	Draft
<i>Halorubrum terrestre</i> JCM 10247	PRJNA188610	NCBI	Turkmenistan	Saline soils	Draft
Hrr. Cb34	PRJNA232799 (in submission)	This study	Aran-Bidgol, Iran	Saline lake	Draft
Hrr. C49	PRJNA232799 (in submission)	This study	Aran-Bidgol, Iran	Saline lake	Draft
Hrr. Ea1	PRJNA232799 (in submission)	This study	Aran-Bidgol, Iran	Saline lake	Draft
Hrr. Eb13	PRJNA232799 (in submission)	This study	Aran-Bidgol, Iran	Saline lake	Draft
Hrr. Ib24	PRJNA232799 (in submission)	This study	Aran-Bidgol, Iran	Saline lake	Draft
Hrr. Ea8	PRJNA232799 (in submission)	This study	Aran-Bidgol, Iran	Saline lake	Draft
Hrr. Hd13	PRJNA232799 (in submission)	This study	Aran-Bidgol, Iran	Saline lake	Draft
Hrr. C3	PRJNA232799 (in submission)	This study	Aran-Bidgol, Iran	Saline lake	Draft
Hrr. E8	PRJNA232799 (in submission)	This study	Aran-Bidgol, Iran	Saline lake	Draft
Hrr. E3	PRJNA232799 (in submission)	This study	Aran-Bidgol, Iran	Saline lake	Draft
Hrr. LG1	PRJNA232799 (in submission)	This study	Aran-Bidgol, Iran	Saline lake	Draft
Hrr. Fb21	PRJNA232799 (in submission)	This study	Aran-Bidgol, Iran	Saline lake	Draft
Hrr. Ga2p	PRJNA232799 (in submission)	This study	Aran-Bidgol, Iran	Saline lake	Draft
Hrr. G37	PRJNA232799 (in submission)	This study	Aran-Bidgol, Iran	Saline lake	Draft
Hrr. LD3	PRJNA232799 (in submission)	This study	Aran-Bidgol, Iran	Saline lake	Draft
Hrr. Ec15	PRJNA232799 (in submission)	This study	Aran-Bidgol, Iran	Saline lake	Draft
Hrr. Ga36	PRJNA232799 (in submission)	This study	Aran-Bidgol, Iran	Saline lake	Draft

of the insertion or deletion within the phylogroups indicates that gene flow is occurring more readily within phylogroups than between, even when the same intein allele is shared. In accordance with earlier evidence, within phylogroups the intein sequence similarity is much higher than between phylogroups. It is unlikely that intein lengths are the result of sequencing or assembly artifacts, as they are constant within phylogroups.

The phylogenetic reconstruction derived from the combined presence-absence data and intein sequence data (**Figure 6**) shows clustering among phylogroup A and B of their constituent taxa. None of the taxa placed anywhere else but with the other members of its phylogroups and the posterior probabilities for these placements are high (0.991 for A and 0.923 for B). These results indicate that inteins are diverging mainly along cluster boundaries, as phylogroups A and B are distinct and separate,

which further suggests that it is more challenging for the inteins to migrate outside compared to inside their phylogroups.

Another genetic element that serves to distinguish phylogroups A from B is the relative presence of CRISPRs. CRISPRs are a type of microbial innate immunity that provides a record of MGEs previously encountered by the lineage that carries them. This record serves the organism by recognizing and destroying sequences that resemble previously encountered MGEs. CRISPRs have been reported in 90% of surveyed archaeal genomes (Kunin et al., 2007), thus the presence and similarity of CRISPR loci provides a means for comparing the phylogroups. The distribution of CRISPRs was surprisingly patchy in phylogroup A and the genus as a whole; however, even more surprisingly was that putative CRISPRs were absent in phylogroup B indicating its members may be devoid of them entirely (**Figure 5**). To assess if the absence

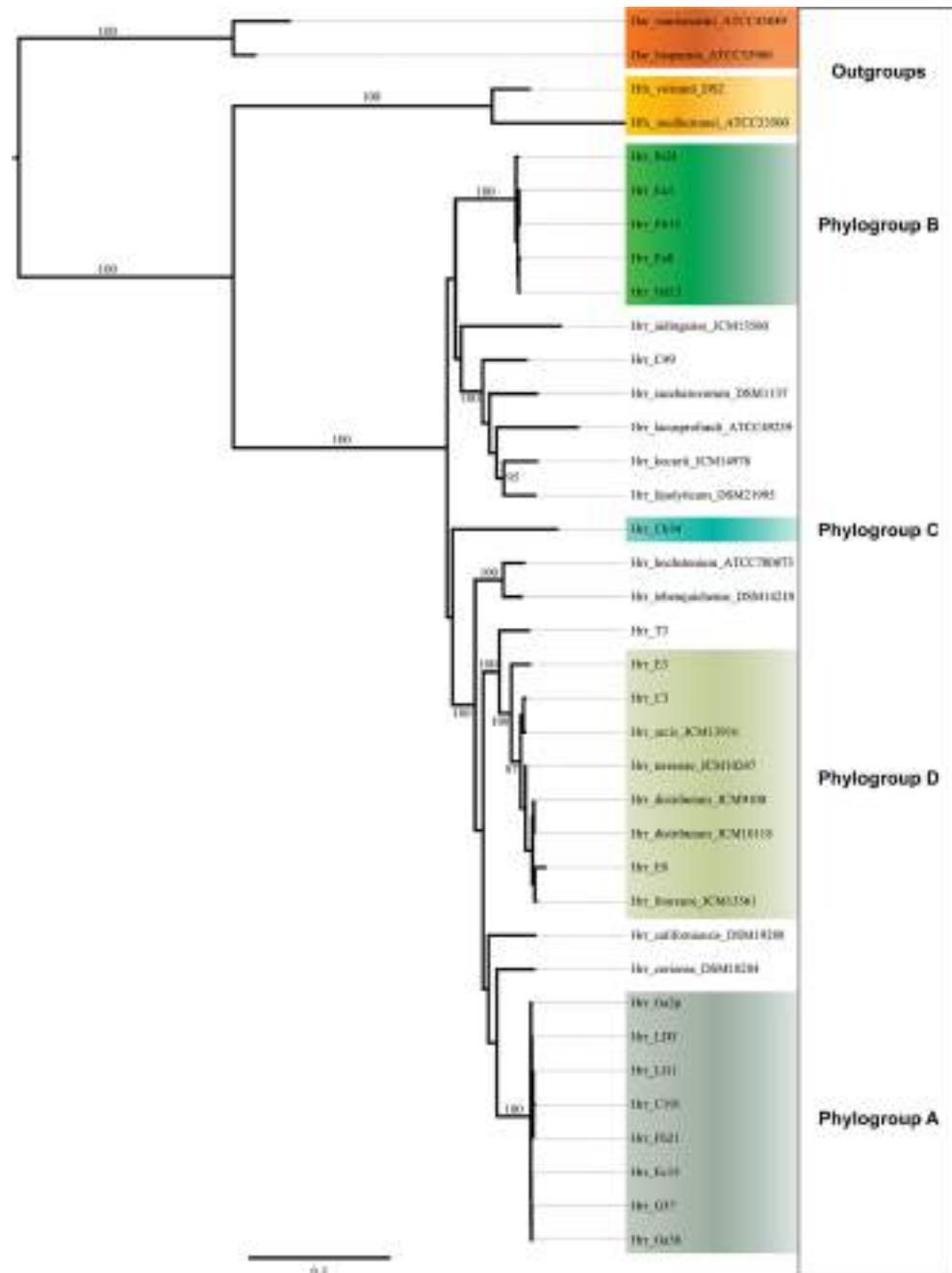


of CRISPRs was an artifact of using draft genome assemblies, we tested for a correlation by relating N50 to CRISPR counts per genome and found there to be no correlation ( $R^2 = 0.105$ ,  $P > 0.05$ ). Therefore, the CRISPR absences do not appear to be a result of genome assembly.

#### PHYLOGROUP D IS NOT A COHESIVE AND DISCRETE ENTITY

Phylogroup D appeared in the phylogenetic reconstructions of MLSA genes after the inclusion of the NCBI *Halorubrum* genomes. It includes five genomes representing four previously described *Halorubrum* species (*Hrr. arcis*, *Hrr. terrestre*, *Hrr. Distributum*, and *Hrr. litoreum*). It was surprising that multiple named species formed such a unit, but evidence suggests it is not discrete and cohesive like phylogroups A and B: much of the data conflict leading to an ambiguous demarcation of its boundary (see below).

The phylogenetic reconstruction of this cluster is supported by the bootstrap values, with exceptions. The concatenated phylogeny has a bootstrap value of 100 at its base and the individual gene trees each support the cluster with bootstrap value of greater than 80 (Figures 1, 2). Pairwise identity between the MLSA genes shows phylogroup D meets the initial criterion of <1% sequence divergence (Table 5). While high, the intra-cluster sequence identity is statistically lower than both phylogroup A and B values ( $P < 0.05$ , two-tailed  $t$ -test). ANI gives similar results to the pairwise identity (Figure 3): the intra-cluster value is ~97%. However some members of the group do not meet the 96% threshold identity, such as E3. Tetramer analysis shows good cohesion within the group, as all but one genome (E3) passed the cutoff. Both E3 and *Hrr. litoreum*'s tetramer frequency patterns are poorly correlated and are below the 0.99 coefficient cutoff advocated by the JSpecies 1.2.1 (Richter and Rosselló-Móra, 2009) package.



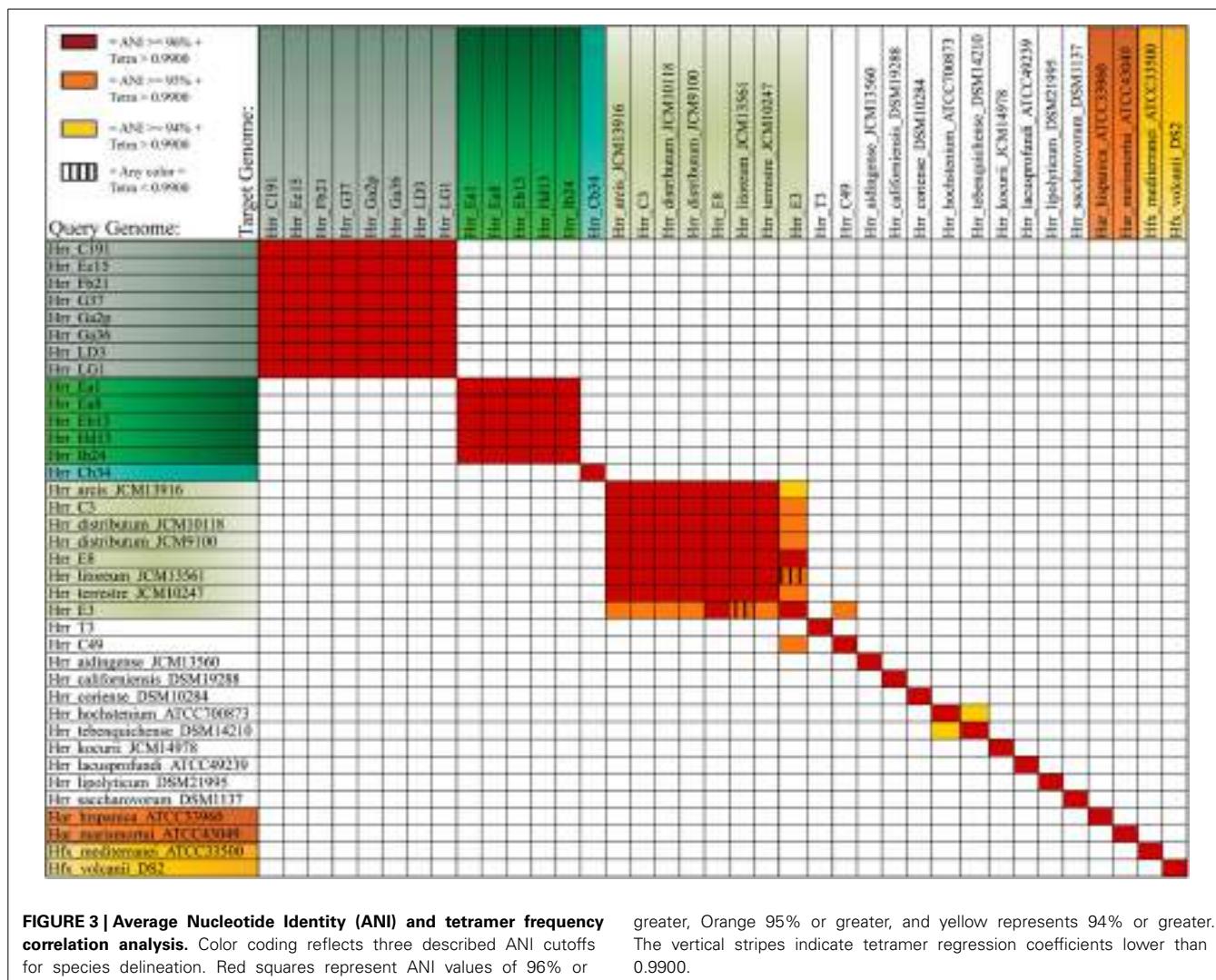
**FIGURE 2 | Maximum-likelihood tree made from the concatenated DNA sequences of five housekeeping genes (*atpB*, *ef-2*, *glnA*, *ppsA*, and *rpoB*).** Support values on branches are bootstrap replicates. Bootstraps values below 70 are not displayed.

As tetramer patterning is largely a granular filter, it strongly suggests that E3 and *Hrr. litoreum* may be distantly related, which is further supported by the ANI analysis.

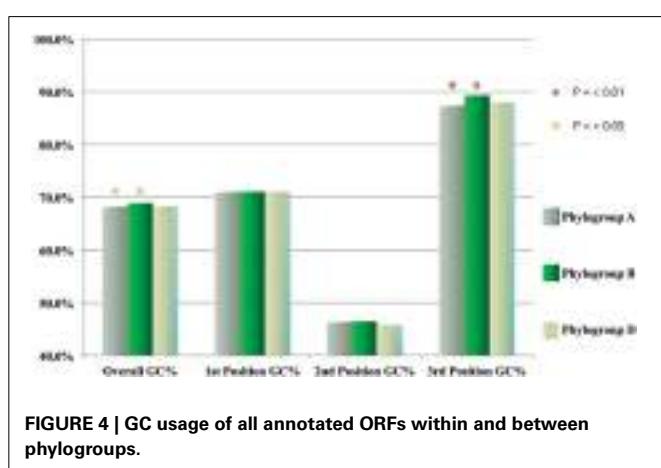
The phylogroup D intein distribution patterns and sequences identities are dissimilar to phylogroup A and B (**Figure 5**). The intra-phylogroup identity of *pol-IIa* is quite low in D compared to phylogroups A and B (~78 vs. ~99% and ~89%, respectively). The inter-group identities are much higher between B and D

than in any other phylogroup relationship (~71%). These relationships are partly explained by *Hrr. terrestris*, which features an intein of much greater length and sequence divergence than the other alleles. This intein shares no more than 55% identity with any other phylogroup D *pol-IIa* allele. If it is removed from consideration, the phylogroup D intra-cluster identity increases to ~99%. The relatedness to phylogroup A rises to ~53% while the value to phylogroup B is 76%. Intra-phylogroup D *cdc21b*

**Table 5 |** Pairwise distances of the concatenated alignment of the five MLSA genes.



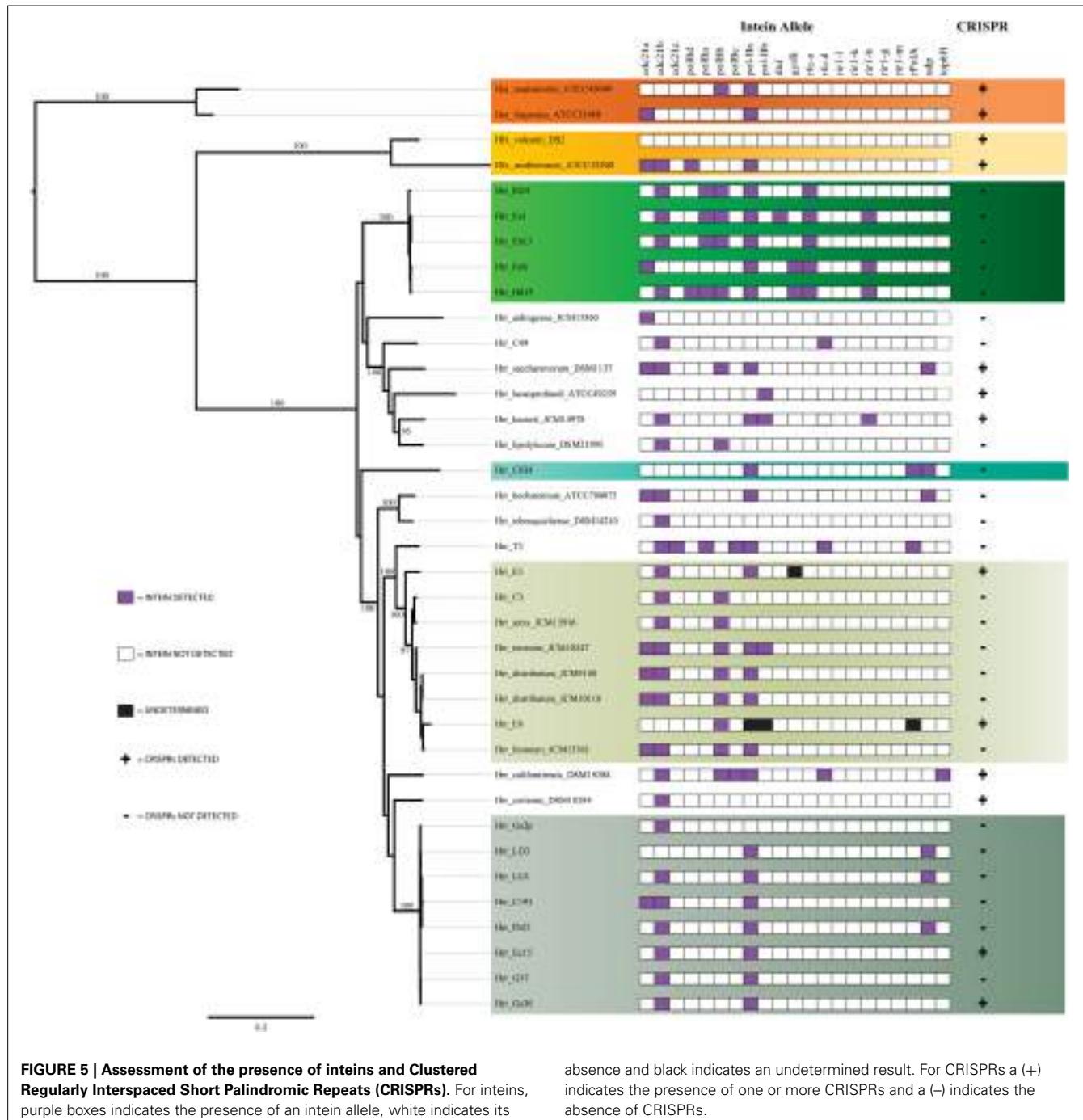
greater, Orange 95% or greater, and yellow represents 94% or greater. The vertical stripes indicate tetramer regression coefficients lower than 0.9900.



diversity is nearly the same as its inter-phylogroup D diversity, which further indicates phylogroup D is a fuzzy entity. The intra-phylogroup identity for the *cdc21b* intein is ~91% (as compared to ~100% for A and ~99% for B) and its inter-phylogroup values

are not much lower with D vs. B at ~83% and D vs. A at ~87%. However, the remaining taxa (*Hrr. arcis*, *Hrr. litoreum*, *Hrr. distributum*, *Hrr. terrestre*, E8, and C3), including the named species appear to form a stable phylogroup. These data suggest that phylogroup D as constructed in our analysis is an amalgamation of populations that resembles other analyzed phylogroups but is not a cohesive unit upon additional investigation. The phylogenetic reconstruction derived from the combined presence-absence data and intein sequence data (Figure 6) shows that phylogroup D does not retain monophyly. Members place at four locations in the tree. The phylogroup displays high identities for core members, but “fringe” members are at the edge of inclusion.

*Hrr. T3* and E3 presented significant challenges to defining the boundary of phylogroup D. As mentioned above, *Hrr. T3* placed directly sister to the phylogroup in three of five gene phylogenies and inside the group in a fourth (Figure 1). In the fifth phylogeny it placed several nodes away from the cluster. The concatenation also places it sister to the cluster with maximum bootstrap support. However, its branch is long relative to the phylogroup. As noted, the pairwise identities and ANI values (Figure 3) both

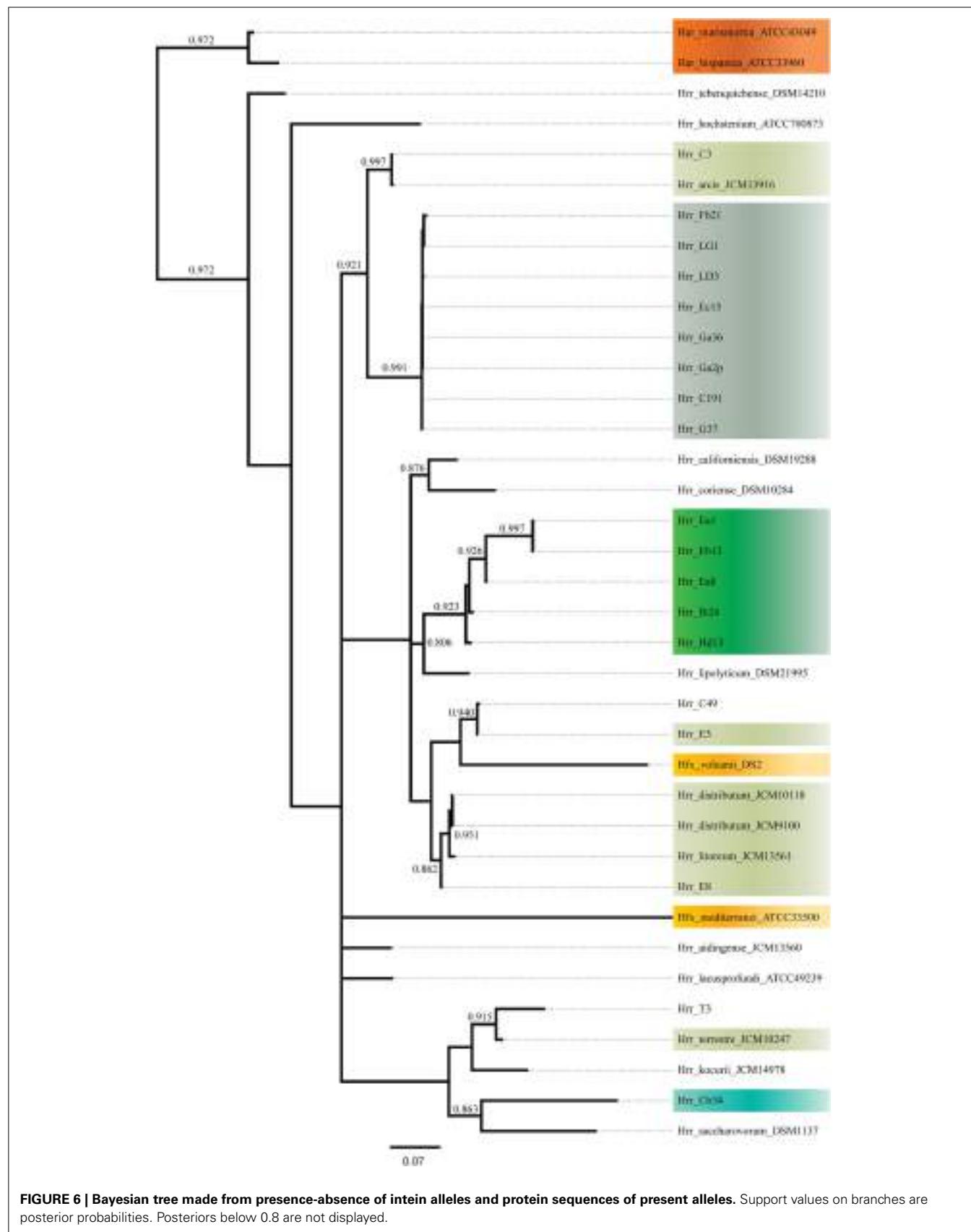


**FIGURE 5 | Assessment of the presence of inteins and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs).** For inteins, purple boxes indicate the presence of an intein allele, white indicates its

absence and black indicates an undetermined result. For CRISPRs a (+) indicates the presence of one or more CRISPRs and a (-) indicates the absence of CRISPRs.

place it below the values seen inside the cluster. These notably lower values were used to exclude this taxon from the phylogroup. *Hrr. E3* is less of a clean-cut case. Its *glnA* gene is outside of the phylogroup. It also falls on a branch by itself at the base of the cluster with rest of the phylogroup supported by an 87% bootstrap score. However, its intra-cluster pairwise and ANI values are several percent higher than *Hrr. T3* and only a percent or two below most of the other members of the phylogroup. Overall, the ANI support was on the edge of current cutoffs for species delineation (95% or 96%) (Konstantinidis et al., 2006; Richter

and Rosselló-Móra, 2009). Its genome had ANIs ~95% to most of the others in the phylogroup and was only 94% to *Hrr. arcis*. Further, E3's tetramer frequency was also substantially different from *Hrr. litoreum*. A possible explanation for some of these differences is that C49 and E3 show a high degree of sequence identity (95% ANI). It is also C49 with which E3's *glnA* gene associates. Finally, the combined presence-absence and intein phylogeny places these taxa together (Figure 6). These data suggest that the two lineages may have engaged in a recent round of genetic exchange, which might explain why E3 is on the periphery



of phylogroup D. Ultimately, it was concluded to include E3 as a member of the phylogroup with the acceptance that it was probably an arbitrary distinction in either direction. It was this difficulty in defining the border that resulted in closer examination of phylogroup D and the ultimate rejection of it representing the same sort of entity that phylogroups A and B are.

## DISCUSSION

### ARE PHYLOGROUPS SPECIES?

The data presented here raise the question: are phylogroups species? We use the term “phylogroup” because a polyphasic analysis (currently defined for the Halobacteria by Oren and Ventosa, 2013) for species description has yet been published on any of the clusters. Still, an evaluation of the data strongly suggests that at least some phylogroups will be eventually described as new species. From the phylogenetic data the perspective provided by the type strain sequences would indicate that phylogroups A and B are unique species. The ANI data support the idea of phylogroups A and B belonging to separate, novel species as several studies advocate cutoffs for species delineation (Konstantinidis and Tiedje, 2005; Konstantinidis et al., 2006; Richter and Rosselló-Móra, 2009) and phylogroups A and B meet all of them. Additionally, both phylogroups form a cohesive cluster with no particular affinity for other clusters, as evidenced by the strong bootstrap support at the base of each cluster. Also, phylogroups A and B are separated from the others by multiple type strains that place between them. Despite many of these branches being poorly supported, their placement and the strong cohesion within the phylogroups argue that the clusters indicate meaningful phylogenetic splits. These splits likely represent barriers that affect the frequency of gene flow between phylogroups, but not within.

Despite the phylogroups’ seemingly species-like attributes, each gene analyzed demonstrates a different topological relationship for them, which means species cannot be viewed as a group of individuals that have a common ancestor, as would be expected from eukaryotic species. While the individual organisms in a prokaryotic species do not share a common ancestor, some of their genes will. For instance, analysis of marine *Vibrio* strains showed that ~1% of the genes within populations shared a common heritage (Shapiro et al., 2012), thus the term species in prokaryotes reflects a process of homogenization, but not heritage, the assumption of Darwinian tree-like speciation. A model that could explain the data is that genes are recombined frequently within *Halorubrum* populations and less so between them. Within the high frequency recombination background new genes that confer selective advantage constantly enter phylogroups from outside the population. These advantageous genes/alleles rise rapidly in frequency throughout the recombining population causing them to diverge in comparison to other phylogroups, yet remaining homogenized within. Like continental drift gives the appearance of discreet units yet are comprised of parts derived from other continents, so too are these two *Halorubrum* phylogroups.

Phylogroup D demonstrates further the model above, as recombination from outside the group is causing divergence, and

disallowing a clean species prediction compared to phylogroups A or B. Therefore, phylogroups D is unlikely to be a single species because it is less cohesive in other measurements, which reflects that it contains several previously described species and also that it has engaged in numerous gene exchanges with not-to-distantly-related organisms. Alternatively, since species assignment is a pragmatic endeavor it could be argued from our data and analyses that phylogroup D is a single species with more genetic diversity than found in A and B. The ambiguous relationships of *Hrr*. T3 and E3 suggest there are different recombination partners available to the cluster members. Such differential exchange partners are key elements in microbial speciation (Papke and Gogarten, 2012) and it could be that T3 and E3 are in the process of speciation from the other members of D, but is incomplete. Tetramer frequency data, which has been demonstrated to convey phylogenetic information (Bohlin et al., 2008a,b) casts doubt on the phylogroup representing a single species. It is less stringent than ANI, being more inclusive with the clusters it forms at typical cutoff values (Richter and Rosselló-Móra, 2009). For this reason, when tetramer frequencies are in disagreement it is likely that the two sequences being compared are not closely related. Thus, the tetramer frequency difference between E3 and *Hrr. litoreum* is also strong evidence for those two taxa not belonging to the same species. Interestingly, if T3 and E3 belong to different species and are removed from consideration, the remaining members of phylogroup D would be a single species by all measurements and cutoffs, and yet are still comprised of four named species. However, these strains were isolated from three different geographic regions of Asia at three different time points (Zvyagintseva and Tarasov, 1987; Ventosa et al., 2004; Cui et al., 2007; Xu et al., 2007), from Chinese solar salterns to Turkmenistani saline soils. While the role of geography and ecology in haloarchaeal speciation is unsettled (Oh et al., 2010; DeMaere et al., 2013; Dillon et al., 2013; Zhaxybayeva et al., 2013) all four of the named species have undergone polyphasic characterization, including DNA-DNA hybridization (Ventosa et al., 2004; Cui et al., 2007; Xu et al., 2007). Presumably, if these taxa lived in the same environments and exchanged genes with each other in a positively biased manner like phylogroups A and B, they would be homogenized and indistinguishable by current polyphasic description processes. What sets phylogroup D apart in our analysis is that we do not have population data on members from the same site, and cannot compare equivalently: if we had more data from natural populations like we do for phylogroups A and B, it might be possible to detect reliable differences that separate the named species into different MLSA phylogroups. For example, dozens of *Sulfolobus* strains isolated from geographically distant sites were less than 1% divergent across multiple loci, yet population data analysis demonstrated they fall into discreet clusters associated with geography (Whitaker et al., 2003) While the taxonomy of the Halobacteria is in flux (for example: McGenity and Grant, 1995; Oren and Ventosa, 1996) it seems unlikely that these four separate species will be merged into one. Recent work has served to split *Hrr. terrestris* from *Hrr. distributum* (Ventosa et al., 2004). Thus, it is challenging to conceive of phylogroup D as a single species, which serves as a strong example of the limits

to MLSA and ANI in regards to being the defining measurements of species.

### CRISPR DISTRIBUTION MAY BE THE RESULT OF SELECTION

It is important to acknowledge that the patchy CRISPR distribution may be in part an artifact of genome assembly. Repeats can prove a challenge to assembly of short read data (Miller et al., 2010; Magoc et al., 2013) and CRISPRs are repeat heavy. However, false negatives that may exist are unlikely to be directly correlated with assembly quality, and no significant correlation is found between N50 score and the number of CRISPR arrays detected ( $P > 0.05$ ). Additionally, the use of a different CRISPR detector, Crass v0.3.6 (Skennerton et al., 2013), which analyzes raw sequencing reads, rather than finding them in assemblies, supported the CRISPRs reported and found only slight evidence for three additional taxa possessing CRISPRs (data not shown). This would only represent individual CRISPR repeats no larger than about three spacers. While CRISPRs this size have been reported (Kunin et al., 2007) the evidence is inconclusive and if these three taxa do possess CRISPRs their distribution would remain sparse. Only seven of the 18 genomes sequenced in this study would possess them.

CRISPRs have been reported to be very common in the archaea (Jansen et al., 2002; Godde and Bickerton, 2006; Kunin et al., 2007; Held et al., 2010) with reported incidence as high as 90% (Koonin and Makarova, 2009). The incidence in bacteria is closer to 50%. The higher incidence in the archaea may be due to the underrepresentation of archaeal genomes in databases. With viruses and other MGEs so common (for discussion of haloviruses see Dyall-Smith et al., 2003; Porter et al., 2007) and horizontal transfer of CRISPRs a frequent occurrence (Kunin et al., 2007; Sorek et al., 2008), why does selection ever conjure a no-CRISPR lineage? One possibility is that the benefit provided is not strong enough to outweigh the costs, as CRISPR systems require precise matches with their target, and a “proto-spacer” with one or two mismatches can eliminate functionality (Deveau et al., 2008). The loss of cassettes in CRISPR arrays is not uncommon (Deveau et al., 2008; Díez-Villaseñor et al., 2010; Touchon and Rocha, 2010), while loss of an entire array is less so (Held et al., 2010; Touchon and Rocha, 2010). Possession of large CRISPR arrays may not offer extra protection against the viruses in an environment (Díez-Villaseñor et al., 2010). It might be that if predation level by MGEs rise and fall then the value of the CRISPR system might follow those trends. *Escherichia* and *Salmonella* CRISPR arrays do not appear to deteriorate rapidly enough to be lost entirely and they show a high rate of transfer and loss of the *cas* proteins that form the machinery of the functional system (Touchon and Rocha, 2010). This might suggest that the need for the system may not be constant. Another reason for degradation of the system could be related to it behaving in an auto-immune fashion. When challenged by artificial constructs including a proto-spacer and a gene complementing an autotrophic defect in the strain, *Sulfolobus* cells developed a surprisingly large number of deletion mutants in the spacer providing immunity to the construct (Gudbergsdottir et al., 2011). The authors speculated that there might be some small degree of feedback where the system attacks the host's spacer in addition to

that of the MGE. The cellular repair systems may then easily delete the spacer during the repair process. Feedback against self and similar to self DNA, such as targeting closely related housekeeping genes (Gophna and Brodt, 2012) could also impact mating proficiency if the CRISPR system degrades the DNA of exchange partners before it can experience recombination events. It is also important to consider that mechanisms other than CRISPRs have major roles in developing resistance to MGEs (Wilson and Murray, 1991; Bickle and Krüger, 1993; Díez-Villaseñor et al., 2010). For instance, there could be a balance between CRISPRs and restriction/modification systems where one system is lost and another replaces, or complements it such that any one anti-MGE mechanism at any moment in time is in flux.

### THE ABSENCE OF INTEINS SUGGESTS BARRIERS TO RECOMBINATION BETWEEN PHYLOGROUPS

Inteins are found pervasively among the archaea (Perler, 2002). They insert into genes and once translated their splicing domains use an auto-catalytic mechanism to self-excise from the protein and re-join the two halves of the polypeptide to generate a functional protein. Inteins associate with homing endonucleases (HEN), found between the splicing domains, to allow their transmission into new hosts. HENs target highly conserved sites in highly conserved genes (Swithers et al., 2009). These HENs appear to be extremely specific in their target sequences as inteins are only found inserted among the most conserved residues of highly conserved protein coding genes (Swithers et al., 2009). Their means of dissemination from host to host is, as yet, unknown although it is clear that it relies on established methods of gene flow within a population (Goddard and Burt, 1999; Gogarten and Hilario, 2006). This suggests that if two hosts have no method of transmitting genes between themselves then the resident inteins will not cross hosts, either. Thus, the patchy distribution of inteins can be interpreted as evidence for a barrier to transfer. This is particularly relevant for the alleles that are not shared between phylogroups A and B. The presence of multiple alleles not seen in the other group argues that the allele has been unable to spread. This is not implying that members of phylogroups A and B do not exchange genes, rather, the sequence divergence and lack of intein spread implies that the recombination process is hindered relative to within group genetic exchange. Indeed, if the mating observed between different *Haloflexax* species (see Naor et al., 2012) is possible then almost any sequence divergence between *Halorubrum* phylogroups is akin to a speed bump rather than a mountain in slowing the rate of genetic exchange. Additionally, studies of homologous recombination have found transfers across class-level phylogenetic distance, only at increasingly lower rates as the genetic distance increases (Vulić et al., 1997; Williams et al., 2012).

### AUTHOR CONTRIBUTIONS

Matthew S. Fullmer, J. Peter Gogarten, Antonio Ventosa, and R. Thane Papke participated in the design of this study and helped to draft the manuscript. Shannon M. Soucy generated the intein data and performed the majority of the intein analysis and helped to draft the manuscript. Kristen S. Swithers performed the CRT analysis and helped to draft the manuscript. Andrea M.

Makkay and Ryan Wheeler performed the MLSA PCR. Andrea M. Makkay performed the genome sequencing. All authors read and approved the final manuscript.

## ACKNOWLEDGMENTS

The authors would like to thank Dr. Mohammad A. Amoozegar (University of Tehran, Iran) for allowing us to analyze the Aran-Bidgol strains, and the UConn Bioinformatics Facility for providing computing resources. This research was supported by the National Science Foundation (award numbers, DEB0919290 and DEB0830024) and NASA Astrobiology: Exobiology and Evolutionary Biology Program Element (Grant Number NNX12AD70G).

## REFERENCES

- Allers, T., Ngo, H.-P., Mevarech, M., and Lloyd, R. G. (2004). Development of additional selectable markers for the halophilic archaeon *Haloflexax volcanii* based on the *leuB* and *trpA* genes. *Appl. Environ. Microbiol.* 70, 943–953. doi: 10.1128/AEM.70.2.943-953.2004
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- AnalystSoft. (2009). *Statistical Analysis Program for Mac OS*. Alexandria, VA: AnalystSoft Inc.
- Andam, C. P., Harlow, T. J., Papke, R. T., and Gogarten, J. P. (2012). Ancient origin of the divergent forms of leucyl-tRNA synthetases in the Halobacteriales. *BMC Evolutionary Biology* 12:85. doi: 10.1186/1471-2148-12-85
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75
- Bickle, T. A., and Krüger, D. H. (1993). Biology of DNA restriction. *Microbiol. Rev.* 57, 434–450.
- Bland, C., Ramsey, T. L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N. C., et al. (2007). CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8:209. doi: 10.1186/1471-2105-8-209
- Bohlin, J., Skjerve, E., and Ussery, D. W. (2008a). Investigations of oligonucleotide usage variance within and between prokaryotes. *PLoS Comput. Biol.* 4:e1000057. doi: 10.1371/journal.pcbi.1000057
- Bohlin, J., Skjerve, E., and Ussery, D. W. (2008b). Reliability and applications of statistical methods based on oligonucleotide frequencies in bacterial and archaeal genomes. *BMC Genomics* 9:104. doi: 10.1186/1471-2164-9-104
- Boucher, Y., Douady, C. J., Sharma, A. K., Kamekura, M., and Doolittle, W. F. (2004). Intrageneric heterogeneity and intergeneric recombination among haloarchaeal rRNA genes. *J. Bacteriol.* 186, 3980–3990. doi: 10.1128/JB.186.12.3980-3990.2004
- Cuadros-Orellana, S., Martin-Cuadrado, A.-B., Legault, B., D'Auria, G., Zhaxybayeva, O., Papke, R. T., et al. (2007). Genomic plasticity in prokaryotes: the case of the square haloarchaeon. *ISME J.* 1, 235–245. doi: 10.1038/ismej.2007.35
- Cui, H.-L., Lin, Z.-Y., Dong, Y., Zhou, P.-J., and Liu, S.-J. (2007). *Halorubrum litoreum* sp. nov., an extremely halophilic archaeon from a solar saltern. *Int. J. Syst. Evol. Microbiol.* 57, 2204–2206. doi: 10.1099/ijs.0.65268-0
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27, 1164–1165. doi: 10.1093/bioinformatics/btr088
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772–772. doi: 10.1038/nmeth.2109
- DeMaere, M. Z., Williams, T. J., Allen, M. A., Brown, M. V., Gibson, J. A. E., Rich, J., et al. (2013). High level of intergenera gene exchange shapes the evolution of haloarchaea in an isolated Antarctic lake. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16939–16944. doi: 10.1073/pnas.1307090110
- Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., et al. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* 190, 1390–1400. doi: 10.1128/JB.01412-07
- Díez-Villaseñor, C., Almendros, C., García-Martínez, J., and Mojica, F. J. M. (2010). Diversity of CRISPR loci in *Escherichia coli*. *Microbiology* 156, 1351–1361. doi: 10.1099/mic.0.036046-0
- Dillon, J. G., Carlin, M., Gutierrez, A., Nguyen, V., and McLain, N. (2013). Patterns of microbial diversity along a salinity gradient in the Guerrero Negro solar saltern, Baja CA Sur, Mexico. *Front. Microbiol.* 4:399. doi: 10.3389/fmicb.2013.00399
- Dyall-Smith, M. (2009). *The Halohandbook - Protocols for Haloarchaeal Genetics*. Available online at: <http://www.haloarchaea.com/resources/halohandbook/index.html>
- Dyall-Smith, M., Tang, S.-L., and Bath, C. (2003). Haloarchaeal viruses: how diverse are they? *Res. Microbiol.* 154, 309–313. doi: 10.1016/S0923-2508(03)00076-7
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Goddard, M. R., and Burt, A. (1999). Recurrent invasion and extinction of a selfish gene. *Proc. Natl. Acad. Sci. U.S.A.* 96, 13880–13885. doi: 10.1073/pnas.96.24.13880
- Godde, J. S., and Bickerton, A. (2006). The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J. Mol. Evol.* 62, 718–729. doi: 10.1007/s00239-005-0223-z
- Gogarten, J. P., and Hilario, E. (2006). Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evolutionary Biology* 6:94. doi: 10.1186/1471-2148-6-94
- Gophna, U., and Brodt, A. (2012). CRISPR/Cas systems in archaea. *Mob. Genet. Elements* 2, 63–64. doi: 10.4161/mge.19907
- Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224. doi: 10.1093/molbev/msp259
- Gudbergsdóttir, S., Deng, L., Chen, Z., Jensen, J. V. K., Jensen, L. R., She, Q., et al. (2011). Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* 79, 35–49. doi: 10.1111/j.1365-2958.2010.07452.x
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Held, N. L., Herrera, A., Cadillo-Quiroz, H., and Whitaker, R. J. (2010). CRISPR associated diversity within a population of *Sulfolobus islandicus*. *PLoS ONE* 5:e12988. doi: 10.1371/journal.pone.0012988
- Jansen, R., van Embden, J. D. A., Gaastra, W., and Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* 43, 1565–1575. doi: 10.1046/j.1365-2958.2002.02839.x
- Khomyakova, M., Bükmmez, Ö., Thomas, L. K., Erb, T. J., and Berg, I. A. (2011). A Methylation cycle in haloarchaea. *Science* 331, 334–337. doi: 10.1126/science.1196544
- Konstantinidis, K. T., Ramette, A., and Tiedje, J. M. (2006). The bacterial species definition in the genomic era. *Philos. Trans. R. Soc. B Biol. Sci.* 361, 1929–1940. doi: 10.1098/rstb.2006.1920
- Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102
- Koonin, E. V., and Makarova, K. S. (2009). CRISPR-Cas: an adaptive immunity system in prokaryotes. *F1000 Biol. Rep.* 1:95. doi: 10.3410/B1-95
- Kunin, V., Sorek, R., and Hugenholtz, P. (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* 8:R61. doi: 10.1186/gb-2007-8-4-r61
- Legault, B. A., Lopez-Lopez, A., Alba-Casado, J. C., Doolittle, W. F., Bolhuis, H., Rodriguez-Valera, F., et al. (2006). Environmental genomics of “*Haloquadratum walsbyi*” in a saltern crystallizer indicates a large pool of accessory genes in an otherwise coherent species. *BMC Genomics* 7:171. doi: 10.1186/1471-2164-7-171
- Liu, K., Warnow, T. J., Holder, M. T., Nelesen, S. M., Yu, J., Stamatakis, A. P., et al. (2012). SATé-II: very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. *Syst. Biol.* 61, 90–106. doi: 10.1093/sysbio/syr095
- Magoc, T., Pabinger, S., Canzar, S., Liu, X., Su, Q., Puiu, D., et al. (2013). GAGE-B: an evaluation of genome assemblers for bacterial organisms. *Bioinformatics* 29, 1718–1725. doi: 10.1093/bioinformatics/btt273

- McGenity, T. J., and Grant, W. D. (1995). Transfer of *Halobacterium saccharorum*, *Halobacterium sodomense*, *Halobacterium trapanicum* NRC 34021 and *Halobacterium lacusprofundi* to the Genus *Halorubrum* gen. nov., as *Halorubrum saccharorum* comb. nov., *Halorubrum sodomense* comb. nov., *Halorubrum trapanicum* comb. nov., and *Halorubrum lacusprofundi* comb. nov. *Syst. Appl. Microbiol.* 18, 237–243. doi: 10.1016/S0723-2020(11)80394-2
- Miller, J. R., Koren, S., and Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics* 95, 315–327. doi: 10.1016/j.ygeno.2010.03.001
- Naor, A., Lapierre, P., Mevarech, M., Papke, R. T., and Gophna, U. (2012). Low species barriers in halophilic Archaea and the formation of recombinant hybrids. *Curr. Biol.* 22, 1444–1448. doi: 10.1016/j.cub.2012.05.056
- Nelson-Sathi, S., Dagan, T., Landan, G., Janssen, A., Steel, M., McInerney, J. O., et al. (2012). Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. *Proc. Natl. Acad. Sci. U.S.A.* 109, 20537–20542. doi: 10.1073/pnas.1209119109
- Oh, D., Porter, K., Russ, B., Burns, D., and Dyall-Smith, M. (2010). Diversity of *Haloquadratum* and other haloarchaea in three, geographically distant, Australian saltern crystallizer ponds. *Extremophiles* 14, 161–169. doi: 10.1007/s00792-009-0295-6
- Okuda, Y., Sasaki, D., Nogami, S., Kaneko, Y., Ohya, Y., and Anraku, Y. (2003). Occurrence, horizontal transfer and degeneration of VDE intein family in Saccharomyces yeasts. *Yeast* 20, 563–573. doi: 10.1002/yea.984
- Oren, A., and Ventosa, A. (1996). A proposal for the transfer of *Halorubrobacterium distributum* and *Halorubrobacterium coriense* to the genus *Halorubrum* as *Halorubrum distributum* comb. nov. and *Halorubrum coriense* comb. nov., respectively. *Int. J. Syst. Bacteriol.* 46, 1180–1180. doi: 10.1099/00207713-46-4-1180
- Oren, A., and Ventosa, A. (2013). Subcommittee on the taxonomy of Halobacteriaceae and Subcommittee on the taxonomy of Halomonadaceae: minutes of the joint open meeting, 24 June 2013, Storrs, Connecticut, USA. *Int. J. Syst. Evol. Microbiol.* 63, 3540–3544. doi: 10.1099/ijts.0.055988-0
- Papke, R. T., and Gogarten, J. P. (2012). How bacterial lineages emerge. *Science* 336, 45–46. doi: 10.1126/science.1219241
- Papke, R. T., Koenig, J. E., Rodriguez-Valera, F., and Doolittle, W. F. (2004). Frequent recombination in a saltern population of *Halorubrum*. *Science* 306, 1928–1929. doi: 10.1126/science.1103289
- Papke, R. T., Zhaxybayeva, O., Feil, E. J., Sommerfeld, K., Muise, D., and Doolittle, W. F. (2007). Searching for species in haloarchaea. *Proc. Natl. Acad. Sci. U.S.A.* 104, 14092–14097. doi: 10.1073/pnas.0706358104
- Perler, F. B. (2002). InBase: the Intein Database. *Nucleic Acids Res.* 30, 383–384. doi: 10.1093/nar/30.1.383
- Porter, K., Russ, B. E., and Dyall-Smith, M. L. (2007). Virus-host interactions in salt lakes. *Curr. Opin. Microbiol.* 10, 418–424. doi: 10.1016/j.mib.2007.05.017
- Rhodes, M. E., Spear, J. R., Oren, A., and House, C. H. (2011). Differences in lateral gene transfer in hypersaline versus thermal environments. *BMC Evolutionary Biology* 11:199. doi: 10.1186/1471-2148-11-199
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277. doi: 10.1016/S0168-9525(00)02024-2
- Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106
- Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Shapiro, B. J., Friedman, J., Cordero, O. X., Preheim, S. P., Timberlake, S. C., Szabó, G., et al. (2012). Population genomics of early events in the ecological differentiation of Bacteria. *Science* 336, 48–51. doi: 10.1126/science.1218198
- Sharma, A. K., Spudich, J. L., and Doolittle, W. F. (2006). Microbial rhodopsins: functional versatility and genetic mobility. *Trends Microbiol.* 14, 463–469. doi: 10.1016/j.tim.2006.09.006
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7. doi: 10.1038/msb.2011.75
- Skennerton, C. T., Imelfort, M., and Tyson, G. W. (2013). Crass: identification and reconstruction of CRISPR from unassembled metagenomic data. *Nucleic Acids Res.* 41, e105. doi: 10.1093/nar/gkt183
- Sorek, R., Kunin, V., and Hugenholtz, P. (2008). CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.* 6, 181–186. doi: 10.1038/nrmicro1793
- Swithers, K. S., Senejani, A. G., Fournier, G. P., and Gogarten, J. P. (2009). Conservation of intron and intein insertion sites: implications for life histories of parasitic genetic elements. *BMC Evolutionary Biology* 9:303. doi: 10.1186/1471-2148-9-303
- Swithers, K. S., Soucy, S. M., Lasek-Nesselquist, E., Lapierre, P., and Gogarten, J. P. (2013). Distribution and evolution of the mobile *vma-1b* intein. *Mol. Biol. Evol.* 30, 2676–2687. doi: 10.1093/molbev/mst164
- Touchon, M., and Rocha, E. P. C. (2010). The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. *PLoS ONE* 5:e11126. doi: 10.1371/journal.pone.0011126
- Tritt, A., Eisen, J. A., Facciotti, M. T., and Darling, A. E. (2012). An integrated pipeline for *de novo* assembly of microbial genomes. *PLoS ONE* 7:e42304. doi: 10.1371/journal.pone.0042304
- Ventosa, A., Gutiérrez, M. C., Kamekura, M., Zvyagintseva, I. S., and Oren, A. (2004). Taxonomic study of *Halorubrum distributum* and proposal of *Halorubrum terrestre* sp. nov. *Int. J. Syst. Evol. Microbiol.* 54, 389–392. doi: 10.1099/ijts.0.02621-0
- Vulić, M., Dionisio, F., Taddei, F., and Radman, M. (1997). Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. *Proc. Natl. Acad. Sci. U.S.A.* 94, 9763–9767. doi: 10.1073/pnas.94.18.9763
- Whitaker, R. J., Grogan, D. W., and Taylor, J. W. (2003). Geographic barriers isolate endemic populations of hyperthermophilic Archaea. *Science* 301, 976–978. doi: 10.1126/science.1086909
- Williams, D., Gogarten, J. P., and Papke, R. T. (2012). Quantifying homologous replacement of loci between haloarchaeal species. *Genome Biol. Evol.* 4, 1223–1244. doi: 10.1093/gbe/evs098
- Wilson, G. G., and Murray, N. E. (1991). Restriction and modification systems. *Annu. Rev. Genet.* 25, 585–627. doi: 10.1146/annurev.ge.25.120191.003101
- Xu, X.-W., Wu, Y.-H., Zhang, H., and Wu, M. (2007). *Halorubrum arcis* sp. nov., an extremely halophilic archaeon isolated from a saline lake on the Qinghai-Tibet Plateau, China. *Int. J. Syst. Evol. Microbiol.* 57, 1069–1072. doi: 10.1099/ijts.0.64921-0
- Zhaxybayeva, O., Stepanauskas, R., Mohan, N. R., and Papke, R. T. (2013). Cell sorting analysis of geographically separated hypersaline environments. *Extremophiles* 17, 265–275. doi: 10.1007/s00792-013-0514-z
- Zvyagintseva, I. S., and Tarasov, A. L. (1987). Extreme halophilic bacteria from saline soils. *Mikrobiologiya* 56, 839–844.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received:** 03 January 2014; **accepted:** 18 March 2014; **published online:** 11 April 2014. **Citation:** Fullmer MS, Soucy SM, Swithers KS, Makkay AM, Wheeler R, Ventosa A, Gogarten JP and Papke RT (2014) Population and genomic analysis of the genus *Halorubrum*. *Front. Microbiol.* 5:140. doi: 10.3389/fmicb.2014.00140
- This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.
- Copyright © 2014 Fullmer, Soucy, Swithers, Makkay, Wheeler, Ventosa, Gogarten and Papke. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Patterns of microbial diversity along a salinity gradient in the Guerrero Negro solar saltern, Baja CA Sur, Mexico

Jesse G. Dillon\*, Mark Carlin, Abraham Gutierrez, Vivian Nguyen and Nathan McLain

Department of Biological Sciences, California State University, Long Beach, CA, USA

**Edited by:**

R. Thane Papke, University of Connecticut, USA

**Reviewed by:**

R. Thane Papke, University of Connecticut, USA

Lejla Pašić, University of Ljubljana, Slovenia

**\*Correspondence:**

Jesse G. Dillon, Department of Biological Sciences, California State University, 1250 Bellflower Blvd., Long Beach, CA 90840, USA  
e-mail: jesse.dillon@csulb.edu

The goal of this study was to use environmental sequencing of 16S rRNA and *bop* genes to compare the diversity of planktonic bacteria and archaea across ponds with increasing salinity in the Exportadora de Sal (ESSA) evaporative saltern in Guerrero Negro, Baja CA S., Mexico. We hypothesized that diverse communities of heterotrophic bacteria and archaea would be found in the ESSA ponds, but that bacterial diversity would decrease relative to archaea at the highest salinities. Archaeal 16S rRNA diversity was higher in Ponds 11 and 12 (370 and 380 g l<sup>-1</sup> total salts, respectively) compared to Pond 9 (180 g l<sup>-1</sup> total salts). Both Pond 11 and 12 communities had high representation (47 and 45% of clones, respectively) by *Haloquadratum walsbyi*-like (99% similarity) lineages. The archaeal community in Pond 9 was dominated (79%) by a single uncultured phylotype with 99% similarity to sequences recovered from the Sfax saltern in Tunisia. This pattern was mirrored in *bop* gene diversity with greater numbers of highly supported phylotypes including many *Haloquadratum*-like sequences from the two highest salinity ponds. In Pond 9, most *bop* sequences, were not closely related to sequences in databases. Bacterial 16S rRNA diversity was higher than archaeal in both Pond 9 and Pond 12 samples, but not Pond 11, where a non-*Salinibacter* lineage within the Bacteroidetes >98% similar to environmental clones recovered from Lake Tuz in Turkey and a saltern in Chula Vista, CA was most abundant (69% of community). This OTU was also the most abundant in Pond 12, but only represented 14% of clones in the more diverse pond. The most abundant OTU in Pond 9 (33% of community) was 99% similar to an uncultured gammaproteobacterial clone from the Salton Sea. Results suggest that the communities of saltern bacteria and archaea vary even in ponds with similar salinity and further investigation into the ecology of diverse, uncultured halophile communities is warranted.

**Keywords:** halophile, gradient, saltern, 16S rRNA gene, *bop* gene, haloarchaea

## INTRODUCTION

Some of the best examples of chemical gradients are found in solar salterns around the world (Anton et al., 2000; Litchfield et al., 2001; Baati et al., 2008; Oren, 2008; Manikandan et al., 2009; Oh et al., 2010). Like most hypersaline habitats, these evaporation ponds typically contain abundant microbial populations including members of all three domains of life (Javor, 1989). Traditional cultivation studies as well as molecular sequencing and FISH studies in salterns have revealed diverse communities dominated by phototrophs like *Dunaliella* as well as aerobic heterotrophic prokaryotes (Anton et al., 1999; Benloch et al., 2001; Pašić et al., 2005; Maturrano et al., 2006; Papke et al., 2007; Baati et al., 2008). However, biogeographic differences in the specific communities found in salterns have been observed, perhaps due to dispersal limitation (Oren and Rodriguez-Valera, 2001; Zhaxybayeva et al., 2013).

Variation in prokaryotic communities along salinity gradients has also been reported as studies have found that microbial species richness decreases with increasing salinity, often resulting in a few dominant phylotypes found in highest salinity ponds (Casamayor et al., 2000; Benloch et al., 2002; Baati et al., 2008).

In many salterns the highest salinities are dominated by haloarchaea such as *Haloquadratum* and *Halorubrum* (Anton et al., 1999; Burns et al., 2004; Maturrano et al., 2006; Oh et al., 2010) as well as having significant proportions (up to 15–27%) of extremely halophilic members of *Salinibacter* (Anton et al., 2000; Oren, 2002; Øvreås et al., 2003; Anton et al., 2008). However, in other salterns, *Haloquadratum* (Pašić et al., 2005) or *Salinibacter* (Maturrano et al., 2006) are absent or rare.

The saltern Exportadora de Sal (ESSA) in Guerrero Negro, Baja CA S., Mexico covers an area over 300 km<sup>2</sup> and is the world's largest producer of evaporative salt. Salt water is pumped year-round from the modestly hypersaline Ojo de Liebre lagoon and then slowly pumped through a series of large (many >1 km<sup>2</sup>), shallow (~1 m deep), interconnected ponds. The ponds display a gradient of chemical make-up as salts precipitate (e.g., gypsum) as the water moves up the evaporation scale for over a year (Javor, 1983). As with many salterns, environmental conditions are quite stable over time. Nutrient levels are relatively low at the ESSA saltern and it has been classified as oligotrophic (Javor, 1983, 1989). The ESSA saltern has been intensively studied. However, most of the research on the diversity of gene sequences of

microbes in this system has focused either on the well-developed, benthic microbial mats found at moderate salinities ( $\sim 70\text{--}100 \text{ g l}^{-1}$ ) (Nübel et al., 2001; Ley et al., 2006; Feazel et al., 2008; Kunin et al., 2008; Dillon et al., 2009b; Robertson et al., 2009) or within evaporites (Sahl et al., 2008). Few studies have been performed in the ESSA ponds at higher salinities ( $>150\text{--}160 \text{ g l}^{-1}$ ), where planktonic communities dominate (Javor and Castenholz, 1981) and these were primarily cultivation-based (Javor, 1984; Sabet et al., 2009). This study represents the first use of culture-independent, molecular techniques to examine the diversity of planktonic microbes in the ESSA saltern at the highest salinities where benthic microbial mats are not found. Based on the prior cultivation work, we hypothesized that diverse communities of both halophilic bacteria and archaea would be found in the ESSA ponds, but that bacterial community diversity would decrease in comparison with archaea at the highest salinities.

## METHODS AND MATERIALS

### WATER SAMPLE COLLECTION AND ANALYSIS

In conjunction with a broader study aimed at cultivation of halophiles (Sabet et al., 2009), water samples were collected in February, 2006, from three ponds along a salinity gradient at the ESSA saltworks, Guerrero Negro, Baja California Sur, Mexico. Replicate 50 ml water samples ( $n = 3\text{--}5$ ) were collected via near-shore surface grabs in evaporative ponds (Ponds 9 and 11) and in a crystallizer pond (Pond 12). Samples were frozen in liquid nitrogen for transport back to CSU Long Beach and stored at  $-80^{\circ}\text{C}$  prior to analysis. Pond 9 bottom was covered with gypsum precipitate, while Pond 12 contained halite with Pond 11 showing a soft sediment bottom with evidence of both gypsum and halite precipitates. Salinity (total salts) of each water source was measured using a refractometer and when salinities were off scale (i.e.,  $>280 \text{ g l}^{-1}$ ) by dilution prior to reading. Water temperature was measured for each pond using a handheld probe (Russell RL060P, Thermo Electron Corp., Beverly, MA, USA). Water samples ( $n = 1$ ) were filtered through a  $0.2 \mu\text{m}$  filter, diluted, and analyzed for major cation content using inductively-coupled plasma mass spectrometry (ICP-MS) at the Institute for Integrated Research in Materials, Environments, and Society (IIRMES) on the CSULB campus and anions via ion chromatography (882 Compact IC plus, Metrohm, Riverview, FL, USA) using the EPA 300.0 method at Physis Environmental Laboratories (Anaheim, CA).

### DNA EXTRACTION

The 50 ml water samples were thawed and pre-filtered through a  $10 \mu\text{m}$  dia. pore size nylon membrane filters (GE Osmonics, Minnetonka, MN) to remove large particles and algae. The water was re-filtered through a  $0.22 \mu\text{m}$  polysulfone membrane (GE Osmonics) to collect the bacterial and archaeal cell fraction. Nucleic acids were extracted from the filter using a modified protocol of Benloch et al. (2001). Filters were cut into pieces, washed with 2 ml of sterile nanopure water, vortexed, and the supernatant treated with SDS (1% w/v) and proteinase K (0.5 mg  $\text{ml}^{-1}$ ) and the samples incubated at  $55^{\circ}\text{C}$  for 2 h, then boiled for 2 min. Nucleic acids were extracted twice with 1 volume of

phenol/chloroform/isoamyl alcohol (IAA) (50:49:1), centrifuged at  $3300 \times g$  for 20 min, and extracted again with an equal volume of chloroform:IAA (49:1) and centrifuged for an additional 5 min. The aqueous supernatants were precipitated in 2 volumes of 100% ethanol and centrifuged at  $3220 \times g$  for 20 min at  $4^{\circ}\text{C}$ . The supernatant was decanted and the tube was allowed to dry. The pellet was resuspended in 100  $\mu\text{L}$  sterile nanopure water, incubated at  $55^{\circ}\text{C}$  for 30 min and stored at  $-20^{\circ}\text{C}$ .

### PCR, CLONING, AND SEQUENCING

PCR amplifications of 16S rRNA were performed using purified nucleic acid and bacterial 16S rRNA primers GM3f (5'-AGAGTTGATCMTGGC) and GM4r (5'-TACCTTGTTACGAC TT) (Muyzer et al., 1995) and archaeal 16S rRNA primers arch21f (TTCCGGTTGATCCYGCCGGA) (Delong, 1992) with either the archaea-specific 958r (YCCGGCGTTGAMTCCAATT) (Delong, 1992) or the universal 1392r reverse primer (Stahl et al., 1988). Amplifications of bacteriorhodopsin (*bop*) genes used the *bop*401F (GACTGGTTGTTYACVACGCC) and *bop*795R (AAG CCGAAGCCGAYCTTBGC) primers (Papke et al., 2007). The 20  $\mu\text{L}$  reaction mixtures contained 1 $\times$  PCR buffer (Invitrogen, Carlsbad, CA), 0.2 mM each dNTP (Promega, Madison, WI), 10 pmol each primer (Operon, Huntsville, AL), 1 U Platinum *Taq* polymerase (Invitrogen), and 50–100 ng of purified nucleic acids. For most samples, 1  $\mu\text{l}$  of bovine serum albumin (0.4% w/v) was added to reaction mixtures to facilitate amplification. Reaction conditions were as follows: initial denaturation ( $94^{\circ}\text{C}$  for 5 min) followed by 30 cycles of denaturation ( $94^{\circ}\text{C}$  for 30 s), annealing ( $53^{\circ}\text{C}$  for 30 s), and extension ( $72^{\circ}\text{C}$  for 90 s) and a final extension ( $72^{\circ}\text{C}$  for 10 min) using a mastercycler (Eppendorf, Hauppauge, NY). The resulting amplicons were ligated into pCR4 TOPO vector with the TOPO® Cloning PCR Cloning Kit (Invitrogen, Carlsbad, CA) and transformed into One Shot® TOP10 chemically competent *E. coli* cells according to the manufacturer's instructions. Transformants were plated on LB plates with  $100 \mu\text{g L}^{-1}$  ampicillin. For each sample, colonies were picked with a sterile toothpick and grown up in 75  $\mu\text{L}$  of LB + ampicillin broth in 96-well plates, diluted to a final concentration of 15% (w/v) with sterile glycerol, and stored at  $-80^{\circ}\text{C}$ . For all clones, cells were grown in 2 ml of LB + ampicillin broth at  $37^{\circ}\text{C}$  overnight and plasmid minipreps performed using the GenCatch plasmid DNA purification kit (Epoch Biolabs, Sugar Land, TX). Plasmids were sequenced with M13forward and reverse primers by a commercial sequencing facility (University of WA High Throughput Genomic Center, Seattle, WA).

### PHYLOGENETIC AND STATISTICAL ANALYSES

For 16S rRNA a total of 296 archaeal sequences and 254 bacterial sequences were obtained from 1–2 clone libraries for each pond separately for each gene. Chimera detection of sequences was performed using Mallard software analysis (Ashelford et al., 2006) and Pintail (Ashelford et al., 2005). Non-chimeric 16S rRNA sequences were aligned using the SINA aligner on the SILVA website (Pruesse et al., 2007) and imported into ARB software and manually refined with reference to nearest neighbor taxa from the v. 102 database (Ludwig et al., 2004). A total of 167 high quality

bop sequences were derived from a total of 288 clones generated via a single 96-clone library from each pond. These sequences were initially aligned to all available gene sequences in the NCBI website using Clustal X (Larkin et al., 2007), and then imported into a custom-created ARB database. Custom lane masks of aligned sequences were created excluding hypervariable regions (16S rRNA genes) and ambiguous nucleotide positions. This resulted in the export of 874 nt (archaeal 16S rRNA) and 1033 nt (bacterial 16S rRNA) and 312 nt (*bop*). Maximum Likelihood trees were constructed using the Blackbox RaxML tool on CIPRES Science Gateway v. 7.2 (Miller et al., 2010) with 750 bootstrap pseudoreplications or fewer if stopped using the automated MRE bootstrapping criterion (e.g., 250, bacterial 16S rRNA). Where full-length sequences were not obtained, partial sequences were added to the 16S rRNA trees using the parsimony tool in ARB.

Additional statistical analyses of the three ponds' clone library sequences were performed using distance matrices generated in ARB (Ludwig et al., 2004). These datasets differed from those used for phylogenetic analyses in that redundant sequences were included, and new custom filters were made for archaeal (304 nt), bacterial (412 nt), and *bop* (356 nt) sequences. Exported similarity matrices were used to cluster sequences into OTU by pair-wise sequence identity with the average neighbor algorithm at a evolutionary distance of 0.00 (actually represents <0.005), 0.01, 0.03, and 0.05 (Schloss et al., 2009) corresponding to the 100, 99, 97, and 95% similarity cut-offs, respectively. Rarefaction curves were generated using a resampling without replacement approach. Estimations of alpha diversity metrics (Chao1, ACE Richness, Shannon's Index, Simpson's Index) were also performed using MOTHUR. Richness and diversity estimates were calculated on random subsamples set to the size of the smallest library to alleviate biases by sample size (Youssef and Elshahed, 2008; Gihring et al., 2012). Percent coverage of the libraries, which determines the probability that all genotypes present in a given set of samples were recovered at least once, was calculated as follows:  $[1 - (n_i/N)] * 100$  where  $n_i$  is the number of unique OTU and N is the total number of clones sampled in the library (Good, 1953). The statistical comparison of community structure among the three ponds' clone libraries was tested using  $\chi^2$ -Libshuff and analysis of molecular variance (AMOVA) as implemented in Mothur software (Schloss et al., 2009). AMOVA tests whether the genetic diversity within communities is significantly different from their pooled genetic diversity (Schloss, 2008).  $\chi^2$ -Libshuff uses the integral form of Cramér-von Mises-type statistic as described in Schloss et al. (2009). To account for multiple comparisons among the three libraries (i.e., Ponds 9, 11, 12), Bonferroni corrections for  $P$ -values were used to determine significance.

Physicochemical data were analyzed using a Principal Components Analysis (PCA) using Primer software v. 6.1.11 (Primer-E Ltd., Plymouth, UK).

#### NUCLEOTIDE SEQUENCE ACCESSION NUMBERS

Sequences were deposited in GenBank and were assigned the accession numbers KF234269-KF234397, KF814118-KF814651, and KF870833-KF870836.

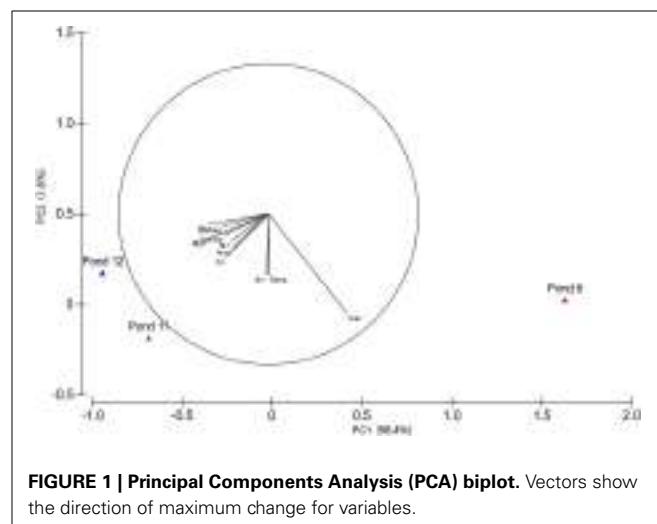
## RESULTS

### PHYSICOCHEMICAL CHARACTERIZATION OF PONDS

Total salinity increased across the three ponds with Pond 9 at 180 g l<sup>-1</sup> and the other two ponds being more similar at 370 and 380 g l<sup>-1</sup> salinity. PCA based on variation in physicochemical variables confirmed that Pond 11 and 12 were more similar to each other than to Pond 9 with the majority of the variation (98.4%) observed along the primary axis, which can be explained by variation in chemical species, but not temperature (Figure 1). Most major cations and anions increased in concentrations with the exceptions of Ca<sup>2+</sup> (Table 1). This is reflected by the difference in directionality of the eigenvector for calcium compared to the other ions. The drop in calcium concentration in Pond 11 and 12 is likely due to precipitation of gypsum (CaSO<sub>4</sub>), which begins in Pond 9. Somewhat surprisingly, sulfate concentrations did not decline in parallel with calcium concentration. The slight drop in sodium and chloride ions in Pond 12 water compared to Pond 11 is likely due to the precipitation of halite (NaCl) at these elevated salinities; halite deposition was apparent in Pond 12.

### ARCHAEAL 16S rRNA SEQUENCE DIVERSITY ACROSS ESSA SALTERN PONDS

Archaeal 16S rRNA sequence diversity included members of genera previously cultured from the ESSA saltern including



**FIGURE 1 | Principal Components Analysis (PCA) biplot.** Vectors show the direction of maximum change for variables.

**Table 1 | Physicochemical parameters in evaporation ponds.**

Ponds	9	11	12
Temp. (°C)	17.3	18.9	16.2
Salinity (g l <sup>-1</sup> )	180	370	380
Cl <sup>-</sup> (mg l <sup>-1</sup> )	85,460	174,160	168,880
SO <sub>4</sub> <sup>2-</sup> (mg l <sup>-1</sup> )	11,256	23,234	24,434
Na <sup>+</sup> (mg l <sup>-1</sup> )	46,460	86,250	84,650
Mg <sup>2+</sup> (mg l <sup>-1</sup> )	6130	16,390	17,160
K <sup>+</sup> (mg l <sup>-1</sup> )	1982	5155	5585
Ca <sup>2+</sup> (mg l <sup>-1</sup> )	1204	407	280
B <sup>+</sup> (mg l <sup>-1</sup> )	14	27	27
Sr <sup>2+</sup> (mg l <sup>-1</sup> )	26	29	25

numerous *Halorubrum*-like sequences (OTU 14–19, 90–99% similarity to cultured species) obtained from all three ponds at clonal abundances of 1–11% of the community (**Figure 2**). We also recovered one *Haloarcula*-like clone (98% similarity) from Pond 12. Pond 11 and 12 clone libraries were dominated (62 and 70%, respectively) by a diverse assemblage of closely related *Halococcus*-like sequences (OTU 1–6). This was especially true of a single lineage 99% similar to the type species *H. walsbyi* (OTU 1, 47 and 45% of Ponds 11 and 12 community, respectively). The other five, less abundant OTU in this group ranged from 1–12% of community and were 92–99% similar to *H. walsbyi*. Some of these lineages were highly similar (98 ≥ 99%) to environmental sequences obtained in Australian crystallizer ponds (Oh et al., 2010) and the Santa Pola saltern in Spain (Zhaxybayeva et al., 2013) (**Figure 2**). Sequences related to *Halorhabdus utahense* (OTU 10) represented ~2–3% of clones in Pond 11 (98% similar). Additionally, lineages related to other uncultured haloarchaeal lineages with no known cultured representative within the family Halobacteriaceae were observed at all sites.

In Pond 9, a *Halorubrum*-like lineage (OTU 18, 1% of community) and singletons (~6% total) were nearly identical (>99.5% similar) to environmental sequences from the Santa Pola saltern (Zhaxybayeva et al., 2013). Aside for these, Pond 9 clone sequences were exclusively related to uncultured haloarchaeal lineages. One lineage (OTU 12, 12% of the community) was >97% similar to an environmental clone from a Chinese saltern. However, the preponderance (79% of community) was a single, highly redundant lineage most closely related (99% similarity) to sequences recovered from the Sfax saltern in Tunisia (Trigui et al., 2011) (OTU 7, **Figure 2**). This redundancy of phylotypes in Pond 9 was reflected in much lower species richness and diversity metrics in this community compared with the communities in the two salt-saturated ponds (**Table 2**) and a flatter rarefaction curve in the Pond 9 sample (**Figure 3A**) at most similarity levels. We analyzed the alpha diversity results at 4 different OTU cut-offs: 0, 1, 3, and 5% evolutionary distance levels among sequences. Regardless of cut-off used, the number of OTUs calculated and the Shannon's and Simpson's diversity (shown as  $1/D = \text{Dominance}$ ) were always lower for Pond 9 than the other two ponds, and the Pond 9 rarefaction curves were always lower than Pond 11 and 12 at each cut-off (**Figure 3A**). However, the richness estimates (Chao1 and ACE) were much higher at the 0% cut-off than at other similarity levels within Pond 9, and both the estimates were even higher than Pond 11 at this level. These results suggest that there were a number of unique, but highly similar (<1%) OTUs in Pond 9, many in the abundant OTU 7 (data not shown). Significant community overlap between the archaeal 16S rRNA communities in Ponds 11 and 12 was confirmed with the  $\beta$ -LIBSHUFF and AMOVA comparisons both of which showed no significant difference between the two libraries (**Table 3**), but highly significant differences when those libraries were compared with the Pond 9 library. This was also reflected in the Venn diagram showing overlap of OTUs among pond communities at the 4 levels of similarity (**Figure 3D**). High degrees of overlap were observed between Ponds 11 and 12 at all similarity levels, but less overlap (2 OTUs) were observed between Ponds 9 and 11 and no overlap was observed between Ponds 9 and 12 until the 5%

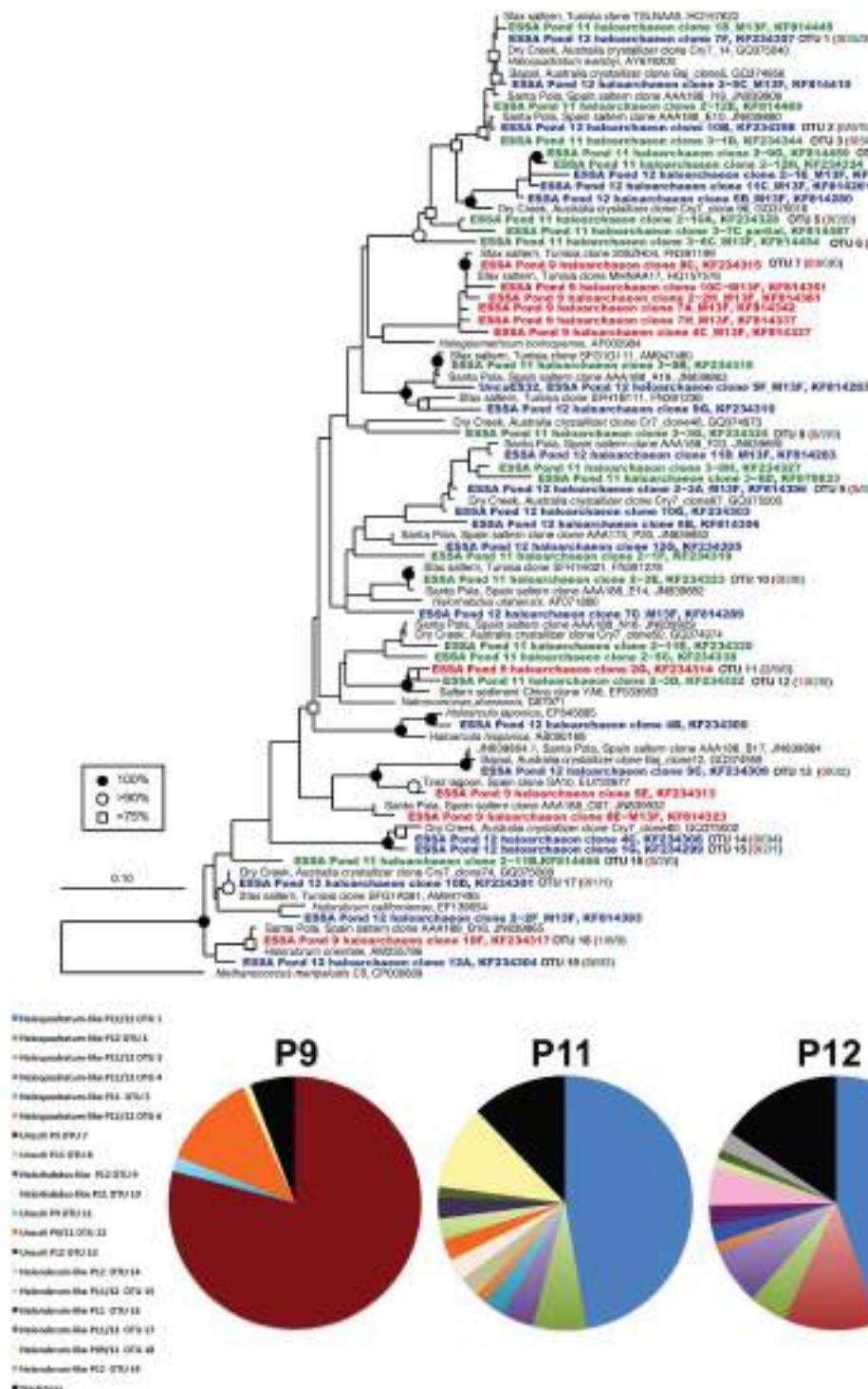
evolutionary distance threshold was employed. The overlapping OTUs between Ponds 9 and 11 were the *Halorubrum*-like lineage (OTU 18, 99% similarity) and the uncultured lineage recovered from a Chinese saltern (OTU12, 97% similarity).

## BACTERIORHODOPSIN GENE DIVERSITY

We successfully created an alignment and ARB database of *bop* sequences from this study and those downloaded from the Genbank database and created a phylogenetic tree (**Figure 4**). Overall, similar patterns of community shifts across the ponds to the archaeal rRNA library were observed when *bop* nucleotide sequences were analyzed. Many environmental clones were recovered from Ponds 11 and 12 (58 and 77% of community, respectively) that had >90% bootstrap support for clustering with *H. walsbyi* on the tree (**Figure 4**). These varied in similarity from 98% (OTU 4), 95% (OTUs 5–6), 90% (OTU 3) to 75% (OTUs 1–2) to the *H. walsbyi bop* gene sequence. Sequences in the latter group were 96–100% similar to environmental clones reported in one of the few published environmental *bop* surveys from the Santa Pola saltern (Papke et al., 2003). Only a single sequence, 88% similar to a cultured *Halorubrum* from the same study, was obtained from Pond 11 and no *Haloarcula* were found in the environmental *bop* library. This was unexpected since we detected both these groups with 16S rRNA genes in Pond 12 and we have successfully cultivated members of these genera from the ESSA ponds (see bolded culture *bop* sequences in **Figure 4**). Additionally, one Pond 11 lineage (OTU 10, **Figure 4**) was closely related (~96% similar) to environmental clones from both Santa Pola and a saltern in Chiku, Taiwan (Lin et al., unpublished). However, none of the Pond 9 clones were closely related to any sequences in Genbank (**Figure 4**). One cluster of closely related OTUs (7–9, 95–98% similar to each other, 71% similar to *H. walsbyi*) represented 44% of clones in Pond 9 and was only found in that location. Only one well-supported cluster (100% bootstrap) of sequences was found in both Ponds 9 and 11 (OTUs 12–13 plus singletons). This phylogroup comprised 48% of the community in Pond 9 and 25% in Pond 11, and again had no close cultured relative (70–71% similar to *Natronococcus* and *Halobiforma* *bop* sequences). As with the archaeal 16S rRNA data, higher species richness and diversity was found in Ponds 11 and 12 compared to Pond 9 (**Table 2**) at most OTU cutoffs, reflected in higher OTU redundancy in the Pond 9 rarefaction curves (**Figure 3B**). Once again the exception was the 0% evolutionary distance cut-off, where very high richness estimates were observed for Pond 9 resulting in the steepest rarefaction curve of all (**Figure 3B**). For *bop* gene analyses, the OTU number and diversity metrics in Pond 9 were even higher than the other ponds at the 0% level, while they were the lowest for all measures with this sample at the 1 and 3% OTU distance level (**Table 2**). As with archaeal 16S rRNA genes, more overlap was found between Ponds 11 and 12 at all similarity levels with only 2 OTUs (uncultured OTU 12–13) shared between Pond 9 and 11 and none between Ponds 9 and 12 (**Figure 3E**).

## BACTERIAL DIVERSITY PATTERNS

Compared with the archaeal 16S rRNA libraries, even greater diversity of bacterial 16S rRNA phylotypes was recovered



**FIGURE 2 | Unrooted maximum likelihood dendrogram depicting relationships among representative archaeal 16 rRNA gene sequences at the 1% evolutionary difference level for clones obtained from the Ponds 9 (red), 11 (green), and 12 (blue) and closely related sequences. The tree was constructed using an alignment of 874 nucleotide positions (gaps and ambiguous residues were excluded using a custom filter in ARB). Symbols at branches represent nodes with bootstrap support  $\geq 75\%$  (□), 90% (○), and 100% (●) for maximum**

likelihood trees (750 replicates). Partial ESSA sequences labeled with M13F/R and short Santa Pola saltern sequences (accession = JN839XXX) were added using the parsimony addition tool in ARB. For OTU ( $\geq 99\%$  similarity) containing more than 1 sequence, the number of duplicates from each pond (9/11/12) is shown in parentheses. *Methanococcus maripaludis* was used as the outgroup. Pie charts show the relative representation of each of the identified OTUs for each pond, with those found only once (singletons) grouped. Legend is for all three charts.

**Table 2 |** 16S rRNA and *bop* nucleotide diversity analyses among ESSA source ponds.

Sequence	Source pond	Sequences (n)	OTU	Coverage (%)	H	1/D	Ace richness	Chao richness
Archaeal 16S rRNA	9	105	27,10,06 <sup>1</sup>	74,91,94	2.0,0.8,0.6	3.6,1.6,1.4	159,38,12	104,21,08
	11	74	37,21,17	50,72,77	3.1,2.1,1.8	14.5,4.2,3.1	103,52,26	96,29,20
	12	83	42,24,16	49,71,81	3.2,2.2,1.6	17.1,4.6,2.5	356,142,66	218,54,31
<i>Bop</i> (nt)	9	66	47,11,05	29,84,93	3.7,1.8,1.0	65.0,5.0,2.4	140,33,06	107,14,05
	11	48	31,14,11	35,71,77	3.2,2.2,2.0	31.3,8.2,6.6	134,28,24	64,19,21
	12	47	32,16,14	32,66,70	3.3,2.0,1.8	30.0,5.1,4.0	157,66,80	132,49,42
Bacterial 16S rRNA	9	81	39,18,16	52,78,80	3.2,2.2,2.0	17.5,6.3,5.5	222,50,43	141,30,25
	11	69	17,07,07	75,90,90	1.8,1.1,1.1	3.4,2.0,2.0	45,07,07	32,07,07
	12	77	47,31,24	39,60,69	3.6,3.0,2.5	39.6,16.6,9.1	386,184,370	399,108,69

H, Shannon-Weaver Diversity Index; D, Simpson's diversity index; shown as 1/D = Dominance.

<sup>1</sup>Commas separate mean estimates calculated using 0,1, and 3% OTU dissimilarity cut-offs respectively for all analyses.

from the ESSA ponds. This included members of the Alpha-, Delta-, and Gammaproteobacteria, Bacteroidetes, Firmicutes, Verrucomicrobia, and algal plastid sequences (Figure 5). Pond 9 communities had relatively high representation of Verrucomicrobial *Puniceicoccus*-like sequences (OTU 10–11, 22% of clone), but the largest proportion (47%) was found in two Gammaproteobacterial phylogroups (Figure 5). The first of these (OTU 12, 14% of community) was 97% similar to *Spiribacter salinus*, a recently isolated photoheterotroph that was found to be abundant (~16% of community) at ~19‰ salinity in the Santa Pola saltern in Spain (Ghai et al., 2011; Leon et al., 2013). The second (33% of community) was 99% similar to an environmental clone from the Salton Sea, a modestly hypersaline (~40 g l<sup>-1</sup>) lake in southern California.

Pond 11 diversity was quite limited, not only in diversity metrics, but also taxonomically since only members of the Phylum Bacteroidetes were detected, including relatives of *Psychroflexus*, *Sediminibacterium*, *Owenweeksia*, and *Salinibacter* (Figure 5). *Psychroflexus*-like (~90% similarity to cultured species) sequences comprised ~16% of the Pond 11 library, *Sediminibacterium*-like species (94% similarity to *S. salmoneum*) made up over 7% of the community. Only three sequences (2 from Pond 11, 1 from Pond 12) were closely related (99% similar) to a *Salinibacter* culture isolated from ESSA (Sabet et al., 2009), which is >98% similar to the type species *S. ruber*. However, 69% of the Pond 11 library was comprised of a single uncultured Bacteroidetes lineage (OTU 5, <80% similar to nearest cultured representative *Saprosira grandis*). This lineage was closely related (>98% similar) to environmental clones recovered from Lake Tuz in Turkey and a saltern in Chula Vista, CA (Zhaxybayeva et al., 2013). This OTU also made up ~14% of the community in Pond 12. Despite overlap in this phylogroup, Pond 12 differed from Pond 11 in having a greater number of Deltaproteobacterial sequences (40% of community). This included one abundant group (13% of community) that was <85% similar to cultured sulfate-reducing species such as *Desulfobacca acetoxidans*, but was 96–99% similar to metagenomic sequences recently obtained in Lake Tyrell in Australia (Podell et al., 2013). Chlorophyte plastid sequences comprised 10% of the Pond 12 community.

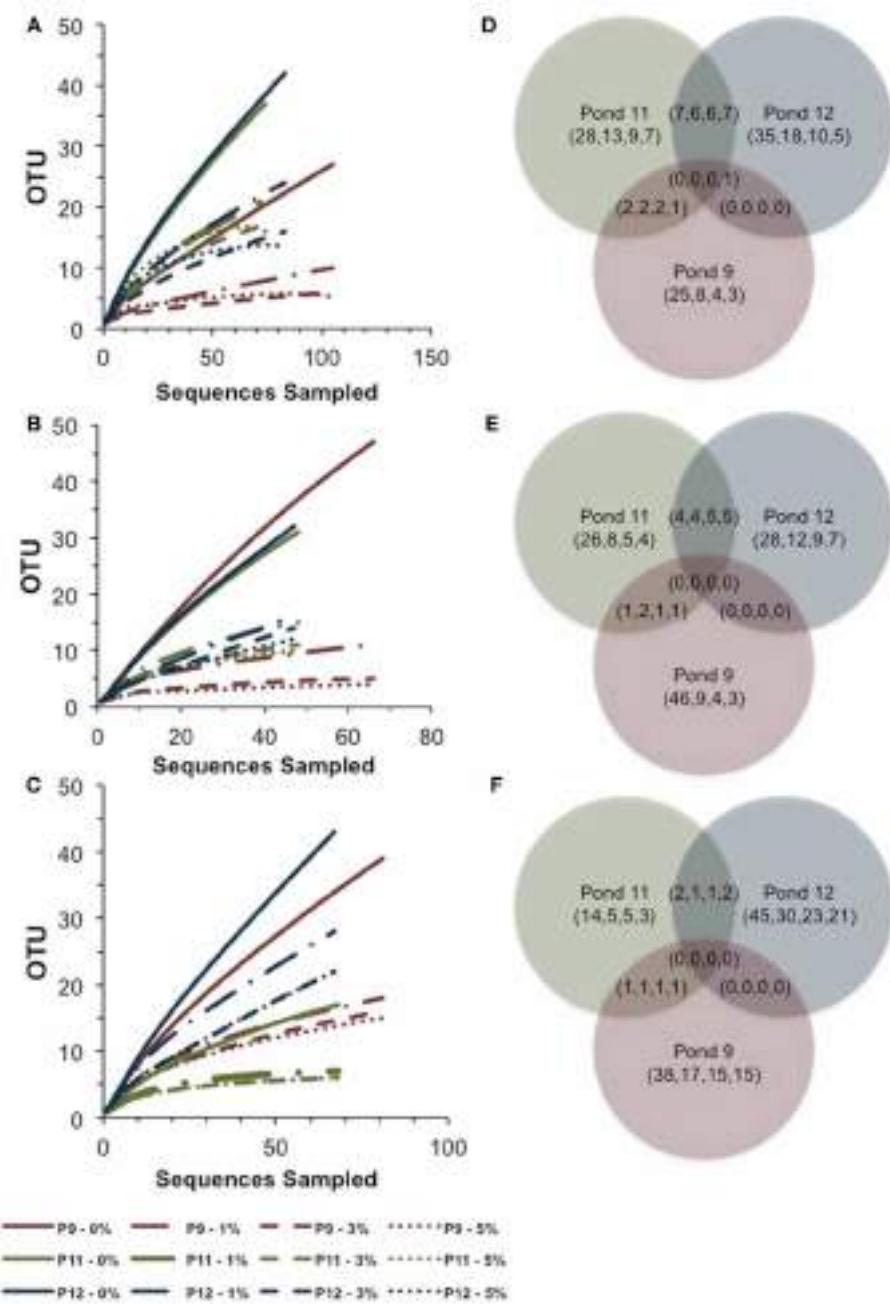
This included one sequence related to the halophilic green alga, *Dunaliella* (~97% similar), but most were only distantly related to cultured algae (<80% similar to *Monomastix* sequences). Nearly one third of sequences in Pond 12 were singlettons, found only once in the library.

The low number of phylotypes in Pond 11 was reflected in low richness and diversity estimates (Table 2) and a nearly flat rarefaction curves for the Pond 11 community (Figure 3C). By contrast, the diversity and richness values in Ponds 9 and 12 were much higher. In Pond 9, there was less OTU redundancy for bacteria than the archaeal community. Somewhat surprisingly, at all OTU levels, richness and diversity metrics were highest in Pond 12 compared to the other ponds, exceeding Shannon index values of 3.0 at the 0% evolutionary distance. In contrast to the archaeal 16S rRNA results, there was almost no overlap among bacterial sequences obtained from the three ponds with only 2 shared OTUs (OTU 1,5) between communities (Figure 3F). This was confirmed by pairwise *f*-LIBSHUFF and AMOVA comparisons that revealed highly significant differences between all ponds including 11 and 12 (Table 3).

## DISCUSSION

### ARCHAEAL DIVERSITY

Archaeal 16S rRNA sequence diversity estimates in Ponds 11 and 12 were comparable to those reported for the Santa Pola saltern and somewhat lower than in the Sfax ponds (Baati et al., 2008). The increase in archaeal diversity metrics at higher salinities was driven primarily by increased representation of culturable haloarchaeal groups including *Halorubrum* and *Haloarcula*, groups that have been previously isolated from the ESSA saltern (Sabet et al., 2009). Over 60% of sequences recovered in these two ponds were *Haloquadratum*-like. Among these, the most abundant lineage recovered from Ponds 11 and 12 (>40% in each) was only 1% divergent from the type species *Haloquadratum walsbyi*. We also identified sequences with lower similarity (93–97%) to *H. walsbyi*, that clustered in the phylogeny with a clone recovered in a more recent study in Australian crystallizers (Oh et al., 2010). We suggest that these are likely members of another species of *Haloquadratum* that has yet to be cultivated. Variation among



**FIGURE 3 |** Rarefaction curves and Venn diagrams showing the shared and unique OTUs across Ponds for archaeal 16S rRNA genes (A,D), *bop* genes (B,E), and bacterial 16S rRNA genes (C,F). Pond 9 results are shown

in red, Pond 11 in green, and Pond 12 in blue. Analyses were performed using a 0, 1, 3, 5% evolutionary distance cut-offs shown in legend for rarefaction and in parentheses (0,1,3,5) for Venn.

*Haloquadratum*-like genotypes assayed by DGGE in different ponds in the Sfax saltern has also been reported (Boujelben et al., 2012). It has been speculated that the seemingly cosmopolitan occurrence of *Haloquadratum* in the highest salinity ponds in salterns may be due to their tolerance of extreme salinity fluctuations and dispersal mechanisms (Oh et al., 2010), although at this point little data is available especially regarding dispersal.

The haloarchaeal community was much less diverse in Pond 9, the main site of gypsum precipitation ( $\text{CaSO}_4$ ) in the ESSA saltern. The Pond 9 archaeal library was dominated by uncultured lineages, especially a single phylotype >99% identical to an environmental sequence recovered from the Sfax multipond saltern in Tunisia (Baati et al., 2010). Interestingly, the phylotype from that study was also recovered from a pond with similar salinity (~18%), suggesting this group may be selected by this salinity or

**Table 3 | Community comparisons among ponds for 16S rRNA sequence libraries.**

Library	Statistic	Comparison	P-value
Archaea	AMOVA <sup>a</sup>	<b>9 vs. 11</b>	<0.001
		<b>9 vs. 12</b>	<0.001
		11 vs. 12	0.22
	$\beta$ -Libshuff <sup>b</sup>	<b>9 vs. 11</b>	<0.001 (XY), <0.001 (YX)
		<b>9 vs. 12</b>	<0.001 (XY), <0.001 (YX)
		11 vs. 12	0.0002 (XY), 0.1776 (YX)
Bacteria	AMOVA	<b>9 vs. 11</b>	<0.001
		<b>9 vs. 12</b>	<0.001
		<b>11 vs. 12</b>	<0.001
	$\beta$ -Libshuff	<b>9 vs. 11</b>	<0.001 (XY), <0.001 (YX)
		<b>9 vs. 12</b>	<0.001 (XY), <0.001 (YX)
		<b>11 vs. 12</b>	<0.001 (XY), <0.001 (YX)

<sup>a</sup>For AMOVA analyses,  $P < 0.017$  was considered significant based on Bonferroni corrections for multiple comparisons (significant comparisons bolded).

<sup>b</sup>For Libshuff analyses, pairwise comparisons were made in both directions, so  $P < 0.0085$  in both directions was considered significant (in bold) based on the Bonferroni correction (Singleton et al., 2001).

some other physicochemical factors across these geographically isolated coastal salterns.

### Bop GENE DIVERSITY

Our findings from archaeal ribosomal gene sequences were largely confirmed in our analyses of the functional gene (*bop*) coding for bacteriorhodopsins. Overall, we found a diverse assemblage of sequences that clustered within the Halobacterales with comparable rhodopsin gene diversity ( $H' = 1.0\text{--}2.2$  at the 1 and 3% OTU level) compared with past studies of *bop* gene diversity ( $H' = \sim 1.4$ ) (Papke et al., 2003; Pašić et al., 2005). There was greater overlap in the sequences recovered in ESSA with those from the Santa Pola saltern (Papke et al., 2003) than the Sečovlje saltern, which was not found to have *Halocladratum*-like sequences (Pašić et al., 2005). However, in contrast with Papke et al. (2003), we found more evidence of potential pond-specific communities (i.e., *bop* phylotypes only found in one of the three ponds). A single *bop* gene lineage was found to have similarity with a database sequence from a saltern in Chiku, Taiwan (Lin et al., unpublished). No overlap was observed with bacteriorhodopsin sequences recently reported from the Dead sea, an athalassohaline habitat (Bodaker et al., 2012).

Many of the sequences we recovered, especially in Pond 9, were from uncultured lineages, neither closely related to any in the NCBI database, nor closely related to *bop* sequences from cultures recovered in the ESSA saltern. The sequence novelty may be due in part to unique ESSA-specific phylotypes, but is also likely due to global undersampling and the relative paucity of available environmental *bop* gene sequence data in databases.

Since this is one of the only studies with directly parallel 16S rRNA and *bop* gene sequencing from the environment, we can attempt to overcome some of the limitations of the limited database for identification. For example, in the archaeal 16S rRNA

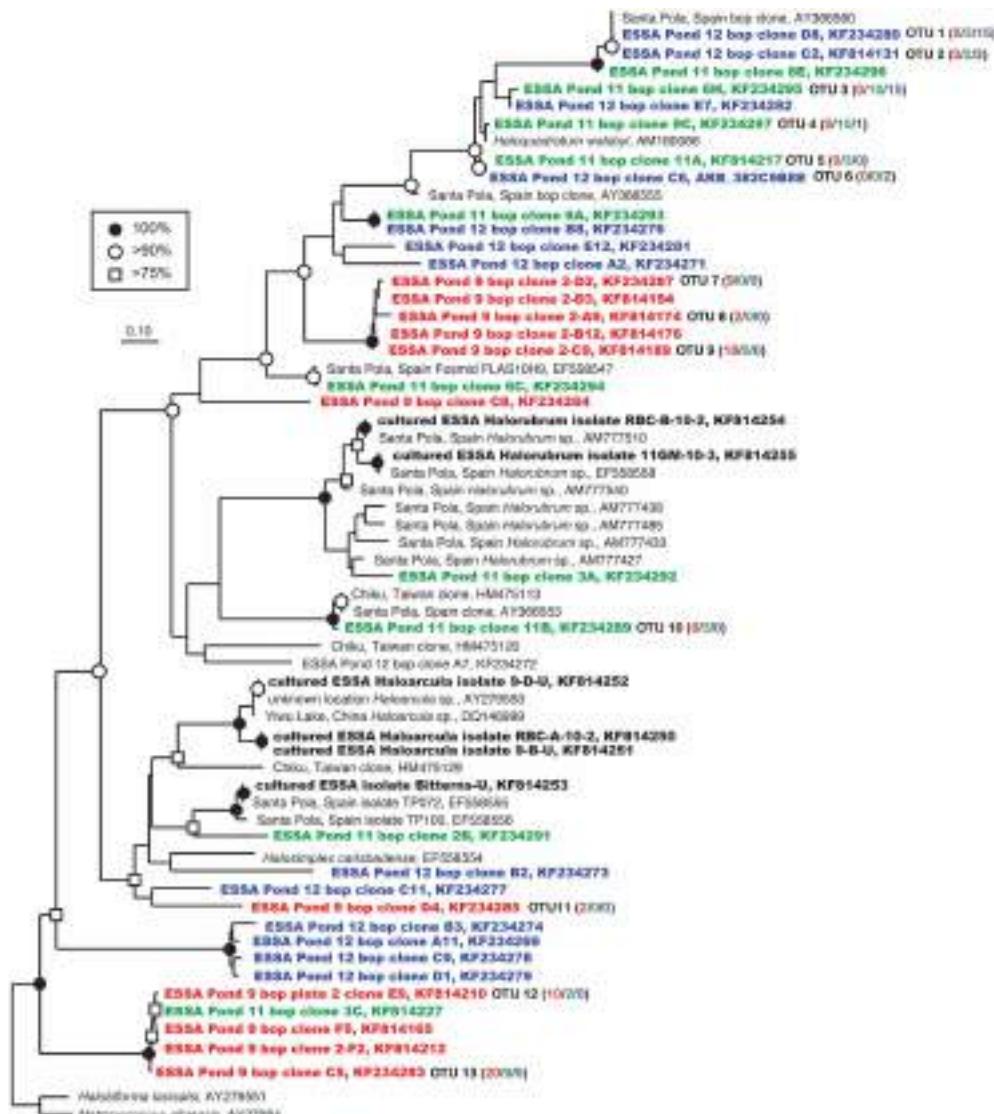
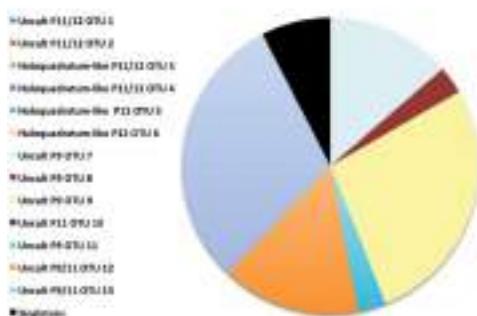
library in Pond 9, a single phylotype was highly abundant (79% of library) and exclusively found in this pond. In our *bop* gene library, an abundant cluster of closely related OTUs (44% of library) was also exclusively found in Pond 9. These sequences may all represent variants of bacteriorhodopsin within the same species, as functional genes typically display more genetic variation than highly conserved 16S rRNA gene.

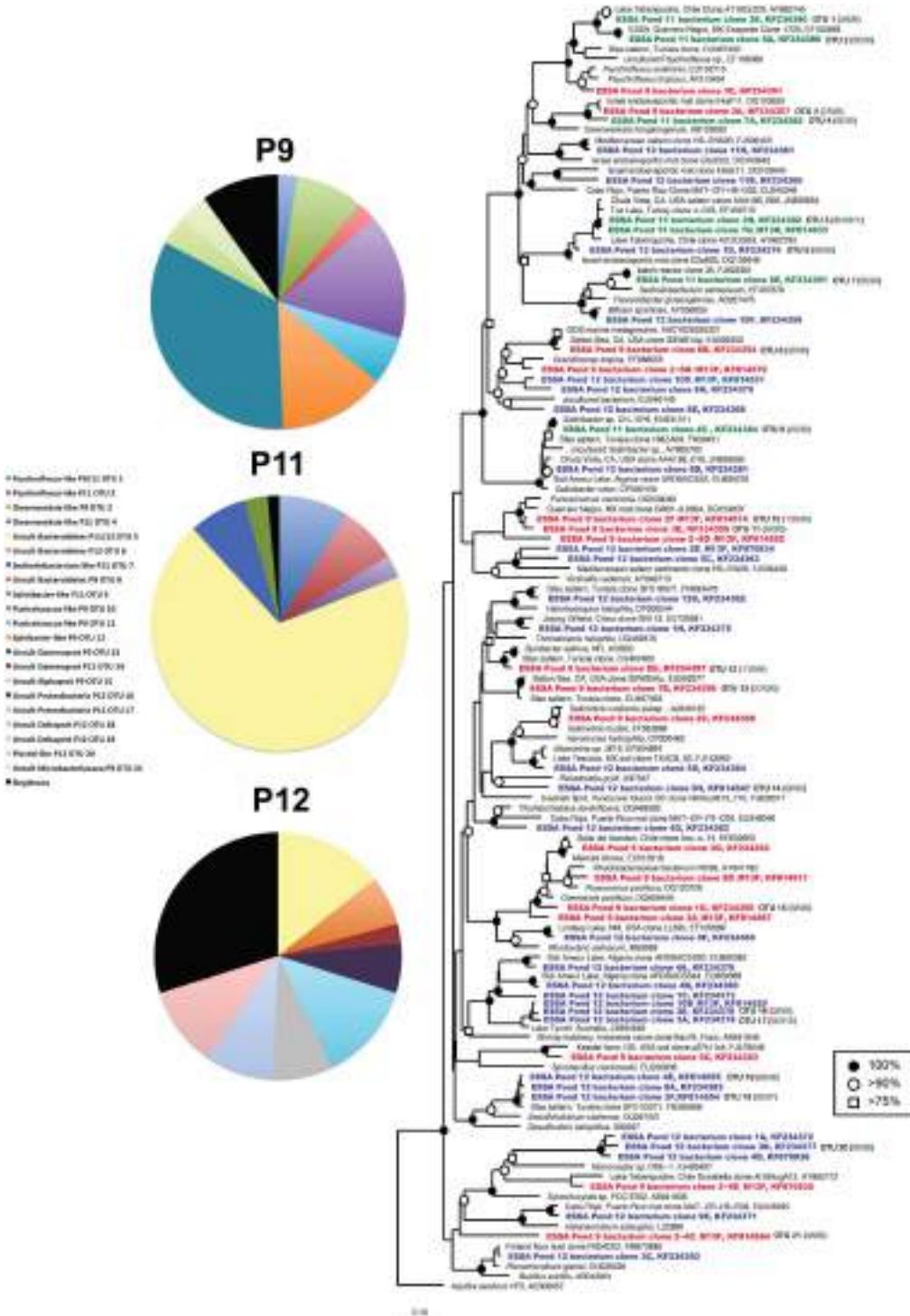
### BACTERIAL DIVERSITY

Among our ESSA Pond 9 bacterial sequences, the largest group recovered was 99% similar to an uncultured gammaproteobacterial sequence from the Salton Sea, a moderately saline ( $\sim 40\text{ g l}^{-1}$ ) endorheic lake in southern California (Dillon et al., 2009a). This suggests that this lineage may live within the lower range of hypersaline conditions and may explain why it was not recovered in Ponds 11 and 12. Interestingly, a number of bacterial sequences from Pond 9 were closely related (~95% similar) to sequences recovered in a previous study of the photosynthetic microbial mats found in Pond 4 of the ESSA saltern (Ley et al., 2006) as well as evaporitic mats from Eilat, Israel (>99% similarity) (Sørensen et al., 2005). This suggests that in interconnected saltern systems, the microbial mats found at lower salinities may serve as a source of bacteria resident in the plankton further up the salinity gradient, although recovery of genes via PCR-based methods does not confirm that they were actively growing at these higher salinities.

Algal plastid sequences were obtained in Pond 12. We recovered a single clone of the halophilic alga *Dunaliella*, which was somewhat unexpected since a past report noted the absence of this group from the ESSA ponds despite its prevalence in other more nutrient-rich salterns (Javor, 1983). However, the most abundant group of plastids sequences (10% of clones in Pond 12) showed only modest (~94%) similarity with an environmental clone found in Lake Tebenquiche in Chile (Demergasso et al., 2008), and among cultured relatives was distantly related (<80% similar) to the freshwater Prymnesiophyte *Monomastix* (Turmel et al., 2009), suggesting this may represent a previously unknown halophilic algal lineage.

In contrast to the other two ponds, which had relatively high numbers of taxonomic groups, only members of the Phylum Bacteroidetes were recovered in Pond 11. We found high clonal abundance (~27% of community) of sequences related to *Psychroflexus*, *Owenweeksia*, and *Sediminibacterium*. *Psychroflexus* strains have been previously cultured from hypersaline habitats (Donachie et al., 2004; Zhang et al., 2010) and members of this genus have been observed in salterns and high altitude athalassohaline lakes in Tibet and Chile (Benloch et al., 2002; Wu et al., 2006; Dorador et al., 2009). No known halophilic members have been cultured from the *Owenweeksia* and *Sediminibacterium* genera, although members of these lineages have been isolated from marine habitats (Lau et al., 2005; Khan et al., 2007). These findings contrast with past studies of hypersaline lakes and salterns where members of the *Salinibacter* genus were abundant (Anton et al., 2000; Benloch et al., 2002; Demergasso et al., 2004; Baati et al., 2008). We only rarely detected *Salinibacter*-like lineages in ESSA, with two clones from Pond 11 closely related (>99% similar) to a cultured ESSA *Salinibacter* (Sabet et al., 2009) and one

**P9**



**FIGURE 5 | Unrooted maximum likelihood dendrogram depicting relationships among representative bacterial 16 rRNA gene sequences at the 1% evolutionary difference level for clones obtained from the Ponds 9 (red), 11 (green), and 12 (blue) and closely related sequences. The tree was constructed using an alignment of 1033 nucleotide positions (gaps and ambiguous residues were excluded using a custom filter in ARB). Symbols at branches represent nodes with bootstrap support  $\geq 75\%$  (□), 90% (○), and 100% (●) for**

maximum likelihood trees (250 replicates). Partial sequences labeled with M13F/R and short Chula Vista saltern sequences (accession = JN839XXX) were added using the parsimony addition tool in ARB. For OTU ( $\geq 99\%$  similarity) containing more than 1 sequence, the number of duplicates from each pond (9/11/12) is shown in parentheses. *Aquifex aeolicus* was used as the outgroup. Pie charts show the relative representation of each of the identified OTUs for each pond, with those found only once (singletons) grouped. Legend is for all three charts.

clone 98% similar to an environmental clone from the saltern in Chula Vista, CA, USA (Zhaxybayeva et al., 2013).

Additionally, an uncultured Bacteroidetes lineage >98% similar to clones from Lake Tuz in Turkey (Mutlu et al., 2008) and the Chula Vista, CA, USA saltern (Zhaxybayeva et al., 2013) comprised 69 and 14% of all bacterial 16S rRNA clones recovered from Ponds 11 and 12, respectively. This group was also the most commonly identified (29/58 bacteria) using single cell genomics approaches from high salinity ponds ( $320\text{--}350\text{ g l}^{-1}$ ) in the Chula Vista saltern, which is near the US-Mexico border. Both the Chula Vista and ESSA salterns are derived from evaporation of pacific coastal waters and are less than 500 miles apart. The Zhaxybayeva et al. (2013) study found little overlap in bacterial communities between the Chula Vista saltern and the Santa Pola saltern in Spain, suggesting geographic or other unidentified environmental differences may be responsible. The abundance of this Bacteroidetes phylotype in Pond 11 of ESSA and Chula Vista at high salinity suggests there may be environmental factors in common or dispersal mechanisms between the two salterns that explains this. However, not all groups showed the same pattern. Gammaproteobacteria from commonly cultivated genera such as *Salicola* and *Halomonas* sequences were abundant in the Chula Vista saltern. Members of these bacterial genera were not detected in this study, despite being isolated from Ponds 9 and 11 in a cultivation study performed using samples collected in parallel to this one (Sabet et al., 2009). The difference in abundance of these culturable groups between the ESSA and Chula Vista salterns may be due to differences in nutrient levels as the latter saltern has been reported to be eutrophic (Javor, 1989). Elevated nutrients in the Chula Vista ponds may more closely resemble culture conditions that favor those lineages.

The novel bacterial community members in Pond 12 were primarily Deltaproteobacteria. Deltaproteobacteria, especially sulfate-reducing lineages, have been commonly identified in saltern sediments (Baati et al., 2010; Lopez-Lopez et al., 2010) and benthic mats (Caumette et al., 1994; Fourçans et al., 2004; Dillon et al., 2009b), but less commonly reported in saltern waters. We would not expect the lineages we recovered in these well-mixed, aerobic ponds to be sulfate reducers. One abundant cluster (17% of community) was not closely related to cultured members of this subphylum (<15% similar to cultured Deltaproteobacterial species), but was 96–99% similar to metagenomic sequences recently obtained in Lake Tyrell in Australia (Podell et al., 2013) suggesting it may be globally distributed. Overall, these findings combined with other recent studies (Jiang et al., 2007; Pagaling et al., 2009; Baati et al., 2010) indicate that *Salinibacter* may not always be the most abundant bacterial type in saturated brines and that other Bacteroidetes as well as Proteobacteria may be similarly well-adapted to such extreme salinities and should be targeted for cultivation and further study.

## DIVERSITY ALONG THE SALINITY GRADIENT

Our findings seemed to follow our hypothesized pattern of increasing archaeal and declining bacterial diversity along the salinity gradient from Ponds 9 and 11, but the increase in bacterial diversity, even higher than archaeal diversity, in Pond 12 runs counter to this. This was surprising, since most studies in

salterns have found that bacterial diversity declines with increasing salinity. For example, studies comparing the relative abundance of bacterial and archaeal 16S rRNA clones in the Santa Pola and Sfax multipond salterns found very low bacterial diversity (i.e.,  $H' < 1.0$ ) above  $300\text{ g l}^{-1}$  salinity (Benloch et al., 2002; Casamayor et al., 2002; Baati et al., 2008). Bacterial phylotypes have been found to outnumber archaea in an athalassohaline lake in the Atacama desert (Demergasso et al., 2004) and an alkaline, hypersaline depression in the Sahara (Mesbah et al., 2007), but the dramatic differences in bacterial populations we observed between Ponds 11 and 12 were unexpected given the similar chemical nature of these two ponds. The only obvious difference between the two sites was the higher degree of precipitation of halite in Pond 12, although this does not rule out some unmeasured physicochemical difference between the two ponds.

Of course, it must be noted that what we were measuring in this study (like the majority of similar studies) is the relative abundance of phylotypes recovered using PCR on a limited sample set, not direct environmental abundances. We used different primer sets for the bacterial and archaeal clone libraries and did not use quantitative PCR, so we cannot directly compare the clonal abundance of bacteria and archaea in these ponds and we cannot assume that the relative clonal abundance represents the actual abundance of cells in the ponds. These patterns are intriguing, but must be confirmed with more quantitative methods (e.g., FISH).

What is clear is that even though the evaporation ponds in the ESSA multipond salterns are interconnected, with evaporating seawater being pumped between the ponds, there are unique, diverse communities of both bacteria and archaea, including diverse bacteriorhodopsin-containing lineages, found in each. Future studies using metagenomic approaches to connect functional genes with taxonomic genes and targeted cultivation of abundant lineages identified in this study are warranted.

## ACKNOWLEDGMENTS

Funding for this project was provided by NIH Minority Bridges Summer Program grant (5R25GM50089-12) and a CSULB SCAC mini-grant to Jesse G. Dillon. We thank Dr. David DesMarais and the NASA AMES EMERG group for logistical support and Dr. Shereen Sabet, Emmerleen Basiana, and Carlos Martinez for their assistance. We thank Dr. R. Thane Papke and the organizing committee for organizing the Halophiles 2013 conferences in Storrs, CT, USA and the two peer-reviewers for helpful suggestions for improving the manuscript.

## REFERENCES

- Anton, J., Llobet-Brossa, E., Rodriguez-Valera, F., and Amann, R. (1999). Fluorescence *in situ* hybridization analysis of the prokaryotic community inhabiting crystallizer ponds. *Environ. Microbiol.* 1, 517–523. doi: 10.1046/j.1462-2920.1999.00065.x
- Anton, J., Pena, A., Santos, F., Martinez-Garcia, M., Schmitt-Kopplin, P., and Rossello-Mora, R. (2008). Distribution, abundance and diversity of the extremely halophilic bacterium *Salinibacter ruber*. *Saline Syst.* 4, 15. doi: 10.1186/1748-1448-4-15
- Anton, J., Rossello-Mora, R., Rodriguez-Valera, F., and Amann, R. (2000). Extremely halophilic bacteria in crystallizer ponds from solar salterns. *Appl. Environ. Microbiol.* 66, 3052–3057. doi: 10.1128/AEM.66.7.3052-3057.2000
- Ashelford, K. E., Chuzhanova, N. A., Fry, J. C., Jones, A. J., and Weightman, A. J. (2005). At Least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies.

- Appl. Environ. Microbiol.* 71, 7724–7736. doi: 10.1128/AEM.71.12.7724-7736.2005
- Ashelford, K. E., Chuzhanova, N. A., Fry, J. C., Jones, A. J., and Weightman, A. J. (2006). New screening software shows that most recent large 16S rRNA gene clone libraries contain Chimeras. *Appl. Environ. Microbiol.* 72, 5734–5741. doi: 10.1128/AEM.00556-06
- Baati, H., Guermazi, S., Amdouni, R., Gharsallah, N., Sghir, A., and Ammar, E. (2008). Prokaryotic diversity of a Tunisian multipond solar saltern. *Extremophiles* 12, 505–518. doi: 10.1007/s00792-008-0154-x
- Baati, H., Guermazi, S., Gharsallah, N., Sghir, A., and Ammar, E. (2010). Novel prokaryotic diversity in sediments of Tunisian multipond solar saltern. *Res. Microbiol.* 161, 573–582. doi: 10.1016/j.resmic.2010.05.009
- Benloch, S., Acinas, S. G., Anton, J., Lopez-Lopez, S. P., and Rodriguez-Valera, F. (2001). Archaeal biodiversity in crystallizer ponds from a solar saltern: culture versus PCR. *Microb. Ecol.* 41, 12–19. doi: 10.1007/s002480000069
- Benloch, S., Lopez-Lopez, A., Casamayor, E. O., Øvreås, L., Goddard, V., Daae, F. L., et al. (2002). Prokaryotic genetic diversity throughout the salinity gradient of a coastal solar saltern. *Environ. Microbiol.* 4, 349–360. doi: 10.1046/j.1462-2920.2002.00306.x
- Bodaker, I., Suzuki, M. T., Oren, A., and Béjà, O. (2012). Dead Sea rhodopsins revisited. *Environ. Microbiol. Rep.* 4, 617–621. doi: 10.1111/j.1758-2229.2012.00377.x
- Boujelben, I., Gomariz, M., Martinez-Garcia, M., Santos, F., Pena, A., Lopez, C., et al. (2012). Spatial and seasonal prokaryotic community dynamics in ponds of increasing salinity of Sfax solar saltern in Tunisia. *Antonie Van Leeuwenhoek* 101, 845–857. doi: 10.1007/s10482-012-9701-7
- Burns, D. G., Camakaris, H. M., Janssen, P. H., and Dyall-Smith, M. L. (2004). Combined use of cultivation-dependent and cultivation-independent methods indicates that members of most haloarchaeal groups in an Australian crystallizer pond are cultivable. *Appl. Environ. Microbiol.* 70, 5258–5265. doi: 10.1128/AEM.70.9.5258-5265.2004
- Casamayor, E. O., Calderon-Paz, J. I., and Pedros-Alio, C. (2000). 5S rRNA fingerprints of marine bacteria, halophilic archaea and natural prokaryotic assemblages along a salinity gradient. *FEMS Microbiol. Ecol.* 34, 113–119. doi: 10.1111/j.1574-6941.2000.tb00760.x
- Casamayor, E. O., Massana, R., Benloch, S., Øvreås, L., Diez, B., Goddard, V. J., et al. (2002). Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern. *Environ. Microbiol.* 4, 338–348. doi: 10.1046/j.1462-2920.2002.00297.x
- Caumette, P., Matheron, R., Raymond, N., and Relexans, J. C. (1994). Microbial mats in the hypersaline ponds of Mediterranean salterns (Salins-De-Giraud, France). *FEMS Microbiol. Ecol.* 13, 273–286. doi: 10.1111/j.1574-6941.1994.tb00074.x
- Delong, E. F. (1992). Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. U.S.A.* 89, 5685–5689. doi: 10.1073/pnas.89.12.5685
- Demergasso, C., Casamayor, E. O., Chong, G., Galleguillos, P., Escudero, L., and Pedros-Alio, C. (2004). Distribution of prokaryotic genetic diversity in athalassohaline lakes of the Atacama Desert, Northern Chile. *FEMS Microbiol. Ecol.* 48, 57–69. doi: 10.1016/j.femsec.2003.12.013
- Demergasso, C., Escudero, L., Casamayor, E., Chong, G., Balagué, V., and Pedrós-Alio, C. (2008). Novelty and spatio-temporal heterogeneity in the bacterial diversity of hypersaline Lake Tebenquiche (Salar de Atacama). *Extremophiles* 12, 491–504. doi: 10.1007/s00792-008-0153-y
- Dillon, J. G., McMath, L. M., and Trout, A. L. (2009a). Seasonal changes in bacterial diversity in the Salton Sea. *Hydrobiologia* 632, 49–64. doi: 10.1007/s10750-009-9827-4
- Dillon, J. G., Miller, S., Bebout, B., Hullar, M., Pinel, N., and Stahl, D. A. (2009b). Spatial and temporal variability in a stratified hypersaline microbial mat community. *FEMS Microbiol. Ecol.* 68, 46–58. doi: 10.1111/j.1574-6941.2009.00647.x
- Donachie, S. P., Bowman, J. P., and Alam, M. (2004). *Psychroflexus tropicus* sp. nov., an obligately halophilic Cytophaga-Flavobacterium-Bacteroides group bacterium from an Hawaiian hypersaline lake. *Int. J. Syst. Evol. Microbiol.* 54, 935–940. doi: 10.1099/ijts.0.02733-0
- Dorador, C., Meneses, D., Urtuvia, V., Demergasso, C., Vila, I., Witzel, K. P., et al. (2009). Diversity of bacteroidetes in high-altitude saline evaporitic basins in northern chile. *J. Geophys. Res. Biogeosci.* 114, 11. doi: 10.1029/2008JG00837
- Feazel, L. M., Spear, J. R., Berger, A. B., Harris, J. K., Frank, D. N., Ley, R. E., et al. (2008). Eucaryotic diversity in a hypersaline microbial mat. *Appl. Environ. Microbiol.* 74, 329–332. doi: 10.1128/AEM.01448-07
- Fourçans, A., De Oteyza, T. G., Wieland, A., Solé, A., Diestra, E., Van Bleijswijk, J., et al. (2004). Characterization of functional bacterial groups in a hypersaline microbial mat community (Salins-de-Giraud, Camargue, France). *FEMS Microbiol. Ecol.* 51, 55–70. doi: 10.1016/j.femsec.2004.07.012
- Ghai, R., Pašić, L., Fernandez, A. B., Martin-Cuadrado, A.-B., Mizuno, C. M., McMahon, K. D., et al. (2011). New abundant microbial groups in aquatic hypersaline environments. *Sci. Rep.* 1. doi: 10.1038/srep00135. Available online at: <http://www.nature.com/srep/2011/111031/srep00135/full/srep00135.html>
- Gihring, T. M., Green, S. J., and Schadt, C. W. (2012). Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. *Environ. Microbiol.* 14, 285–290. doi: 10.1111/j.1462-2920.2011.02550.x
- Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* 40, 237–264. doi: 10.1093/biomet/40.3-4.237
- Javor, B. J. (1983). Planktonic standing crop and nutrients in a saltern ecosystem. *Limnol. Oceanogr.* 28, 153–159. doi: 10.4319/lo.1983.28.1.0153
- Javor, B. J. (1984). Growth potential of halophilic bacteria isolated from solar salt environments: carbon sources and salt requirements. *Appl. Environ. Microbiol.* 48, 352–360.
- Javor, B. J. (1989). *Hypersaline Environments: Microbiology and Biogeochemistry*, New York, NY: Springer-Verlag. doi: 10.1007/978-3-642-74370-2
- Javor, B. J., and Castenholz, R. W. (1981). Laminated microbial mats, Laguna Guerrero Negro, Mexico. *Geomicrobiol. J.* 3, 237–273. doi: 10.1080/01490458109377766
- Jiang, H., Dong, H., Yu, B., Liu, X., Li, Y., Ji, S., et al. (2007). Microbial response to salinity change in Lake Chaka, a hypersaline lake on Tibetan plateau. *Environ. Microbiol.* 9, 2603–2621. doi: 10.1111/j.1462-2920.2007.01377.x
- Khan, S. T., Nakagawa, Y., and Harayama, S. (2007). *Sediminibacter furfurosus* gen. nov., sp. nov. and *Gilvibacter sediminis* gen. nov., sp. nov., novel members of the family Flavobacteriaceae. *Int. J. Syst. Evol. Microbiol.* 57, 265–269. doi: 10.1099/ijts.0.64628-0
- Kunin, V., Raes, J., Harris, J. K., Spear, J. R., Walker, J. J., Ivanova, N., et al. (2008). Millimeter-scale genetic gradients and community-level molecular convergence in a hypersaline microbial mat. *Mol. Syst. Biol.* 4, 198. doi: 10.1038/msb.2008.35
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGgettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics(Oxford)* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404
- Lau, K. W. K., Ng, C. Y. M., Ren, J., Lau, S. C. L., Qian, P.-Y., Wong, P.-K., et al. (2005). *Owenweeksia hongkongensis* gen. nov., sp. nov., a novel marine bacterium of the phylum 'Bacteroidetes'. *Int. J. Syst. Evol. Microbiol.* 55, 1051–1057. doi: 10.1099/ijts.0.63155-0
- Leon, M. J., Ghai, R., Fernandez, A. B., Sanchez-Porro, C., Rodriguez-Valera, F., and Ventosa, A. (2013). Draft genome of *Spiribacter salinus* M19-40, an abundant gammaproteobacterium in aquatic hypersaline environments. *Genome Announce.* 1:e00179. doi: 10.1128/genomeA.00179-12
- Ley, R. E., Harris, J. K., Wilcox, J., Spear, J. R., Miller, S. R., Bebout, B. M., et al. (2006). Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. *Appl. Environ. Microbiol.* 72, 3685–3695. doi: 10.1128/AEM.72.5.3685-3695.2006
- Litchfield, C. D., Irby, A., Kis-Papo, T., and Oren, A. (2001). Comparative metabolic diversity in two solar salterns. *Hydrobiologia* 466, 73–80. doi: 10.1023/A:1014513317746
- Lopez-Lopez, A., Yarza, P., Richter, M., Suarez-Suarez, A., Anton, J., Niemann, H., et al. (2010). Extremely halophilic microbial communities in anaerobic sediments from a solar saltern. *Environ. Microbiol. Rep.* 2, 258–271. doi: 10.1111/j.1758-2229.2009.00108.x
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Kumar, Y., et al. (2004). ARB: a software environment for sequence data. *Nucleic Acids Res.* 32, 1363–1371. doi: 10.1093/nar/gkh293
- Manikandan, M., Kannan, V., and Pašić, L. (2009). Diversity of microorganisms in solar salterns of Tamil Nadu, India. *World J. Microbiol. Biotechnol.* 25, 1007–1017. doi: 10.1007/s11274-009-9980-y
- Maturrano, L., Santos, F., Rossello-Mora, R., and Anton, J. (2006). Microbial diversity in maras salterns, a hypersaline environment in the peruvian andes. *Appl. Environ. Microbiol.* 72, 3887–3895. doi: 10.1128/AEM.02214-05

- Mesbah, N. M., Abou-El-Ela, S. H., and Wiegel, J. (2007). Novel and unexpected prokaryotic diversity in water and sediments of the alkaline, hypersaline lakes of the Wadi an Natrun, Egypt. *Microb. Ecol.* 54, 598–617. doi: 10.1007/s00248-006-9193-y
- Miller, M., Pfeiffer, W., and Schwartz, T. (2010). “Creating the CIPRES science gateway for inference of large phylogenetic trees,” in *Proceedings of the Gateway Computing Environments Workshop*, (New Orleans, LA), 1–8.
- Mutlu, M. B., Martinez-Garcia, M., Santos, F., Pena, A., Guven, K., and Anton, J. (2008). Prokaryotic diversity in Tuz Lake, a hypersaline environment in Inland Turkey. *FEMS Microbiol. Ecol.* 65, 474–483. doi: 10.1111/j.1574-6941.2008.00510.x
- Muyzer, G., Teske, A., Wirsén, C. O., and Jannasch, H. W. (1995). Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel-electrophoresis of 16S rDNA fragments. *Arch. Microbiol.* 164, 165–172. doi: 10.1007/BF02529967
- Nübel, U., Bateson, M. M., Madigan, M. T., Kuhl, M., and Ward, D. M. (2001). Diversity and distribution in hypersaline microbial mats of bacteria related to *Chloroflexus* spp. *Appl. Environ. Microbiol.* 67, 4365–4371. doi: 10.1128/AEM.67.9.4365-4371.2001
- Oh, D., Porter, K., Russ, B., Burns, D., and Dyall-Smith, M. (2010). Diversity of *Halococcus* and other haloarchaea in three, geographically distant, Australian saltern crystallizer ponds. *Extremophiles* 14, 161–169. doi: 10.1007/s00792-009-0295-6
- Oren, A. (2002). Molecular ecology of extremely halophilic Archaea and Bacteria. *FEMS Microbiol. Ecol.* 39, 1–7. doi: 10.1111/j.1574-6941.2002.tb00900.x
- Oren, A. (2008). Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Syst.* 4, 2. doi: 10.1186/1746-1448-4-2
- Oren, A., and Rodriguez-Valera, F. (2001). The contribution of halophilic Bacteria to the red coloration of saltern crystallizer ponds. *FEMS Microbiol. Ecol.* 36, 123–130. doi: 10.1111/j.1574-6941.2001.tb00832.x
- Øvreås, L., Daee, F. L., Torsvik, V., and Rodriguez-Valera, F. (2003). Characterization of microbial diversity in hypersaline environments by melting profiles and reassociation kinetics in combination with terminal restriction fragment length polymorphism (T-RFLP). *Microb. Ecol.* 46, 291–301. doi: 10.1007/s00248-003-3006-3
- Pagaling, E., Wang, H., Venables, M., Wallace, A., Grant, W. D., Cowan, D. A., et al. (2009). Microbial biogeography of six salt lakes in Inner Mongolia, China, and a salt lake in Argentina. *Appl. Environ. Microbiol.* 75, 5750–5760. doi: 10.1128/AEM.00040-09
- Papke, R. T., Douady, C. J., Doolittle, W. F., and Rodriguez-Valera, F. (2003). Diversity of bacteriorhodopsins in different hypersaline waters from a single Spanish saltern. *Environ. Microbiol.* 5, 1039–1045. doi: 10.1046/j.1462-2920.2003.00501.x
- Papke, R. T., Zhaxybayeva, O., Feil, E. J., Sommerfeld, K., Muise, D., and Doolittle, W. F. (2007). Searching for species in haloarchaea. *Proc. Natl. Acad. Sci. U.S.A.* 104, 14092–14097. doi: 10.1073/pnas.0706358104
- Pašić, L., Bartual, S. G., Ulrich, N. P., Grabnar, M., and Velikonja, B. H. (2005). Diversity of halophilic archaea in the crystallizers of an Adriatic solar saltern. *FEMS Microbiol. Ecol.* 54, 491–498. doi: 10.1016/j.femsec.2005.06.004
- Podell, S., Ugalde, J. A., Narasingarao, P., Banfield, J. F., Heidelberg, K. B., and Allen, E. E. (2013). Assembly-driven community genomics of a hypersaline microbial ecosystem. *PLoS ONE* 8:e61692. doi: 10.1371/journal.pone.0061692
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., et al. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196. doi: 10.1093/nar/gkm864
- Robertson, C. E., Spear, J. R., Harris, J. K., and Pace, N. R. (2009). Diversity and stratification of archaea in a hypersaline microbial mat. *Appl. Environ. Microbiol.* 75, 1801–1810. doi: 10.1128/AEM.01811-08
- Sabet, S., Diallo, L., Hays, L., Jung, W., and Dillon, J. (2009). Characterization of halophiles isolated from solar salterns in Baja California, Mexico. *Extremophiles* 13, 643–656. doi: 10.1007/s00792-009-0247-1
- Sahl, J. W., Pace, N. R., and Spear, J. R. (2008). Comparative molecular analysis of endoevaporitic microbial communities. *Appl. Environ. Microbiol.* 74, 6444–6446. doi: 10.1128/AEM.00879-08
- Schloss, P. D. (2008). Evaluating different approaches that test whether microbial communities have the same structure. *ISME J.* 2, 265–275. doi: 10.1038/ismej.2008.5
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09
- Singleton, D. R., Furlong, M. A., Rathbun, S. L., and Whitman, W. B. (2001). Quantitative comparisons of 16S rRNA gene sequence libraries from environmental samples. *Appl. Environ. Microbiol.* 67, 4374–4376. doi: 10.1128/AEM.67.9.4374-4376.2001
- Sørensen, K. B., Canfield, D. E., Teske, A. P., and Oren, A. (2005). Community composition of a hypersaline endoevaporitic microbial mat. *Appl. Environ. Microbiol.* 71, 7352–7365. doi: 10.1128/AEM.71.11.7352-7365.2005
- Stahl, D. A., Flesher, B., Mansfield, H. R., and Montgomery, L. (1988). Use of phylogenetically based hybridization probes for studies of ruminal microbial ecology. *Appl. Environ. Microbiol.* 54, 1079–1084.
- Trigui, H., Masmoudi, S., Brochier-Armanet, C., Barani, A., Grégoire, G., Denis, M., et al. (2011). Characterization of heterotrophic prokaryote subgroups in the Sfax coastal solar salterns by combining flow cytometry cell sorting and phylogenetic analysis. *Extremophiles* 15, 347–358. doi: 10.1007/s00792-011-0364-5
- Turmel, M., Gagnon, M. C., O’Kelly, C. J., Otis, C., and Lemieux, C. (2009). The chloroplast genomes of the green algae *Pyramimonas*, *Monostatrix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. *Mol. Biol. Evol.* 26, 631–648. doi: 10.1093/molbev/msn285
- Wu, Q. L., Zwart, G., Schauer, M., Kamst-Van Agterveld, M. P., and Hahn, M. W. (2006). Bacterioplankton community composition along a salinity gradient of sixteen high-mountain lakes located on the Tibetan Plateau, China. *Appl. Environ. Microbiol.* 72, 5478–5485. doi: 10.1128/AEM.00767-06
- Youssef, N. H., and Elshahed, M. S. (2008). Species richness in soil bacterial communities: a proposed approach to overcome sample size bias. *J. Microbiol. Methods* 75, 86–91. doi: 10.1016/j.mimet.2008.05.009
- Zhang, H., Hosoi-Tanabe, S., Nagata, S., Ban, S., and Imura, S. (2010). *Psychroflexus lacialis* sp. nov., a moderate halophilic bacterium isolated from a hypersaline lake (Hunazoko-Ike) in Antarctica. *J. Microbiol.* 48, 160–164. doi: 10.1007/s12275-010-0018-9
- Zhaxybayeva, O., Stepanauskas, R., Mohan, N., and Papke, R. T. (2013). Cell sorting analysis of geographically separated hypersaline environments. *Extremophiles* 17, 265–275. doi: 10.1007/s00792-013-0514-z

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Received: 03 July 2013; paper pending published: 31 August 2013; accepted: 04 December 2013; published online: 20 December 2013.*

*Citation: Dillon JG, Carlin M, Gutierrez A, Nguyen V and McLain N (2013) Patterns of microbial diversity along a salinity gradient in the Guerrero Negro solar saltern, Baja CA Sur, Mexico. Front. Microbiol. 4:399. doi: 10.3389/fmicb.2013.00399*

*This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.*

*Copyright © 2013 Dillon, Carlin, Gutierrez, Nguyen and McLain. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# Evidence from phylogenetic and genome fingerprinting analyses suggests rapidly changing variation in *Halorubrum* and *Haloarcula* populations

Nikhil Ram Mohan<sup>1</sup>, Matthew S. Fullmer<sup>1</sup>, Andrea M. Makkay<sup>1</sup>, Ryan Wheeler<sup>1</sup>, Antonio Ventosa<sup>2</sup>, Adit Naor<sup>3</sup>, J. Peter Gogarten<sup>1</sup> and R. Thane Papke<sup>1\*</sup>

<sup>1</sup> Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT, USA

<sup>2</sup> Department of Microbiology and Parasitology, University of Seville, Seville, Spain

<sup>3</sup> Molecular Microbiology and Biotechnology, Tel Aviv University, Tel Aviv, Israel

## Edited by:

Jesse Dillon, California State University, Long Beach, USA

## Reviewed by:

Jocelyne DiRuggiero, The Johns Hopkins University, USA

James A. Coker, University of Maryland, University College, USA

## \*Correspondence:

R. Thane Papke, Department of Molecular and Cell Biology, University of Connecticut, 91 N. Eagleville Rd., Storrs, CT 06269, USA

e-mail: thane@uconn.edu

Halobacteria require high NaCl concentrations for growth and are the dominant inhabitants of hypersaline environments above 15% NaCl. They are well-documented to be highly recombinogenic, both in frequency and in the range of exchange partners. In this study, we examine the genetic and genomic variation of cultured, naturally co-occurring environmental populations of Halobacteria. Sequence data from multiple loci (~2500 bp) identified many closely and more distantly related strains belonging to the genera *Halorubrum* and *Haloarcula*. Genome fingerprinting using a random priming PCR amplification method to analyze these isolates revealed diverse banding patterns across each of the genera and surprisingly even for isolates that are identical at the nucleotide level for five protein coding sequenced loci. This variance in genome structure even between identical multilocus sequence analysis (MLSA) haplotypes indicates that accumulation of genomic variation is rapid: faster than the rate of third codon substitutions.

**Keywords:** Halobacteria, MLSA, genome fingerprinting, Aran-Bidgol lake, environmental population

## INTRODUCTION

Members of the class Halobacteria (Domain: Archaea; Phylum: Euryarchaeota) are the dominant inhabitants of hypersaline environments (Anton et al., 1999; Ghai et al., 2011). These hypersaline environments provide extreme growth conditions in the form of high salinity and ionic concentrations with variations in pH, and temperature (Oren, 2002). Such extreme conditions are necessary for Halobacteria, also called haloarchaea, to live. The environment is also subject to low solubility of gases, low diffusion rates, and very low water activity (Litchfield, 1998). To overcome many of these obstacles, haloarchaea can generate ATP from light energy (Lozier et al., 1975) and have gas vesicles to buoyantly lift themselves to the surface (Jones et al., 1991). Osmotic survival in these brines is managed by maintaining a cytosolic salinity in equilibrium with that of the environment, a feat that requires solubilized proteins under those conditions, and solved with a proteome enriched in acidic and depleted of basic amino acids (Oren, 2002).

Haloarchaea have a well-documented capacity for generating enormous amounts of genetic variation through horizontal gene transfer (HGT) (Papke et al., 2004, 2007; Cuadros-Orellana et al., 2007; Lynch et al., 2012; Naor et al., 2012; Williams et al., 2012; Demaere et al., 2013; Podell et al., 2013). From the very first genome sequence analysis of *Halobacterium* strain NRC-1, evidence was provided for the acquisition of aerobic respiration genes via HGT from Bacteria (Ng et al., 2000). Since then, several studies on specific genes of interest [e.g., rhodopsins (Sharma et al., 2007), ribosomal RNAs (Boucher et al., 2004),

and tRNA synthetases (Andam et al., 2012)] have further demonstrated gene transfer into and among the haloarchaea. A recent report suggested that this process of generating diversity has been ongoing since before the group's last universal common ancestor and that HGT played a huge role in changing their physiology from an autotrophic anaerobe to a heterotrophic aerobe (Nelson-Sathi et al., 2012). Population genetics analysis on strains from the genus *Halorubrum* using multilocus sequence analysis (MLSA) demonstrated that alleles at different loci are unlinked indicating that homologous recombination (HR) is frequent enough within phylogenetically defined groups to randomize traits among individuals (Papke et al., 2004, 2007), an observation once considered unique to sexually reproducing eukaryotes. Analysis of 20 haloarchaeal genomes showed that there are no absolute barriers to HR, which occurs regularly and proportionally to genetic distance throughout the haloarchaea (Williams et al., 2012). Community analyses using metagenomics revealed that genes are coming and going quickly within *Halorubrum* populations, suggesting there may be very few identical genomes within the species (Legault et al., 2006; Cuadros-Orellana et al., 2007). Perhaps most striking is their ability to exchange large swaths of genetic information. Mating experiments between *Haloferax volcanii* and *Haloferax mediterranei* demonstrated between ~10 and 18% (~300–500 kb) of their chromosome could be transferred in a single fragment (Naor et al., 2012). Also, genomes of highly divergent strains (e.g., <75% average nucleotide identity) isolated from Deep Lake, Antarctica were shown to share many ~100% identical DNA

sequences in fragments up to 35 Kb in length (Demaere et al., 2013).

MLSA has often been used as a technique for classifying microorganisms (Maiden et al., 1998), including halophiles (Papke et al., 2011; De la Haba et al., 2012), but it is also used to estimate population variation and gene flow (Feil et al., 2000). Assumptions using MLSA regarding how representative multiple genes are for capturing individual variation, and thus the appearance of clonality, can lead to erroneous conclusions. For instance, two strains may have identical sequences across multiple loci, but unexamined genomic variation might be high and belie the interpretation of little or no recombination. Indeed, studies are demonstrating that there are vast amounts of variation within bacterial species/populations. Environmental isolates with identical HSP-60 genes from a natural coastal *Vibrio* sp. population demonstrated that the overwhelming majority of individual strains were unique as determined by chromosome pulse field gel electrophoresis, with some strains differing by up to a megabase in genome size (Thompson et al., 2005). This variation in genome size and the existence of “open” (i.e., infinite) pan-genomes like that of *Prochlorococcus marinus* and others (Tettelin et al., 2008; Lapierre and Gogarten, 2009) suggest that HGT is so frequent that for at least some species every cell may be genetically distinct.

To get a better understanding for the genomic variation within closely related haloarchaeal strains we examined naturally co-occurring environmental strains from the genera *Halorubrum* and *Haloarcula* isolated from the Aran-Bidgol salt lake in Iran. We used MLSA to identify closely related strains, and a PCR genome fingerprinting technique that randomly primed amplification sites along the chromosome to generate a gel electrophoresis pattern that enabled us to inexpensively compare genomic variation of the isolates.

## MATERIALS AND METHODS

### GROWTH CONDITIONS AND DNA EXTRACTION

Aran-Bidgol *Halorubrum* and *Haloarcula* spp. cultures were grown in Hv-YPG medium (Allers et al., 2004) at 37°C with agitation. DNA from haloarchaea was isolated as described in the Halohandbook (<http://www.haloarchaea.com/resources/halohandbook/>). Briefly, stationary-phase cells were pelleted at 10,000 × g, supernatant was removed and the cells were lysed in distilled water. An equal volume of phenol was added, and the mixture was incubated at 65°C for 1 h prior to centrifugation to separate the phases. The aqueous phase was reserved and phenol extraction was repeated without incubation, and followed with a phenol/chloroform/iso-amyl alcohol (25:24:1) extraction. The DNA was precipitated with ethanol, washed, and resuspended in TE (10 mM tris, pH 8.0, 1 mM EDTA). Type strains were grown, and DNA was purified as described by Papke et al. (2011).

### SEQUENCE ACQUISITION FOR MLSA

Five housekeeping genes were amplified using PCR. The loci were *atpB*, *ef-2*, *glnA*, *ppsA*, and *rpoB* and the primers used for each locus are listed in Table 1. To more efficiently sequence PCR products, an 18 bp M13 sequencing primer was added to the 5' end of each degenerate primer (Table 1). Each PCR reaction was 20 µl in volume. Phire Hot Start II DNA polymerase (Thermo

Scientific) was used in the amplification reactions. The PCR reaction was run on a Mastercycler Ep Thermocycler (Eppendorf) using the following PCR cycle protocol: 30 s initial denaturation at 98°C, followed by 40 cycles of 30 s at 98°C, 5 s at the annealing temperature for each set of primers, and 15 s at 72°C. Final elongation occurred at 72°C for 1 min. Table 2 provides a detailed list of reagents and the PCR mixtures for each amplified locus. The PCR products were separated by gel electrophoresis with agarose (1%). Gels were stained with ethidium bromide. An exACTGene mid-range plus DNA ladder (Fisher Scientific International Inc.) was used to estimate the size of the amplicons, which were purified using Wizard SV gel and PCR cleanup system (Promega). The purified amplicons were sequenced by Genewiz Inc. The sequences obtained for the five genes in this study were submitted to Genbank under the following accession numbers: KJ152221–KJ152260, KJ152261–KJ152298, KJ152362–KJ152397, KJ152398–KJ152433, and KJ152323–KJ152361.

### PHYLOGENETIC ANALYSIS

Type strain genomes were obtained from the NCBI ftp repository. Blast searches identified DNA top hits for each MLSA target gene (*atpB*, *ef-2*, *glnA*, *ppsA*, and *rpoB*) in each genome. Multiple-sequence alignments (MSAs) were created from the DNA genome hits as well as the PCR amplicons using MUSCLE (Edgar, 2004) (alignments available upon request) with its refine function. The

**Table 1 | Degenerate primers used to PCR amplify and sequence the *atpB*, *ef-2*, *glnA*, *ppsA*, and *rpoB* genes for MLSA.**

MLSA primer sequence 5'-3'		
Locus	Forward	Reverse
atpB	tgt aaa acg acg gcc agt aac ggg gag scv ats aac cc	cag gaa aca gct atg act tca ggt cvg trt aca tgt a
ef-2	tgt aaa acg acg gcc agt atc ccg gct bta yaa stg g	cag gaa aca gct atg act ggt cga tgg wyt cga ahg g
glnA	tgt aaa acg acg gcc agt cag gta cgg gtt aca sga cgg	cag gaa aca gct atg acc ctc gcs ccg aar gac ctc gc
ppsA	tgt aaa acg acg gcc agt ccg ccg tar ccv agc atc gg	cag gaa aca gct atg aca tcg tca ccg acg arg gyg g
rpoB	tgt aaa acg acg gcc agt tcg aag agc cgg acg aca tgg	cag gaa aca gct atg acc ggt cag cac ctg bac cgg ncc

**Table 2 | PCR conditions for each locus.**

	atpB	ef-2	glnA	ppsA	rpoB
Water (µl)	11.6	8.2	11.8	7.9	11.9
5× phire reaction buffer (µl)	4.0	4.0	4.0	4.0	4.0
DMSO (µl)	0.6	0	0.4	0.6	0.6
Acetamide (25%)	0	4.0	0	4.0	0
dNTP mix (10 mM)	0.4	0.4	0.4	0.4	0.4
Forward primer (10 mM)	1.0	1.0	1.0	1.0	1.0
Reverse primer (10 mM)	1.0	1.0	1.0	1.0	1.0
Phire hot start II DNA polymerase (µl)	0.4	0.4	0.4	0.4	0.4
Template DNA (20 ng µl <sup>-1</sup> )	1.0	1.0	1.0	0.7	0.7
Annealing temperature (°C)	60.0	61.0	69.6	66.0	63.7

MSA length was manually trimmed down to the lengths of the PCR amplicons. In-house scripts created a concatenated alignment of all five genes. A model of evolution was determined using the Akaike Information Criterion with correction for small sample size (AICc). The jModelTest 2.1.4 (Darriba et al., 2012) program was used to compute likelihoods from the nucleotide alignment and to perform the AICc test (Akaike, 1974). The AICc reported the best-fitting model to be GTR + Gamma estimation + Invariable site estimation. A maximum likelihood (ML) phylogeny was generated from the concatenated MSA using the PhyML v3.0\_360–500 (Guindon et al., 2010). The model used in PhyML corresponded to the one favored by jModeltest: GTR model, estimated p-invar, 4 substitution rate categories, estimated gamma distribution with 100 bootstrap replicates. The number of nucleotide differences in pairwise comparisons were determined using MEGA 5 (Tamura et al., 2011).

### GENOMIC FINGERPRINTING

In total, DNA from 81 haloarchaeal type strains and 43 isolates from the Aran-Bidgol lake were tested. Each primer selected has successfully been used in genome fingerprinting in previous studies. Primers P1 and P2 were used to fingerprint *Vibrio harveyi* bacteriophages (Shivu et al., 2007), primers OPA-9 and OPA-13 were used to assess marine viral richness (Winget and Wommack, 2008). The last primer, FALL-A was adapted from the primer used (Barrangou et al., 2002; Winget and Wommack, 2008) to study bacteriophages isolated from an industrial sauerkraut fermentation. Amplification conditions for each strain were equal to enable accurate comparison between banding patterns obtained. Each sample was diluted to  $20\text{ ng }\mu\text{l}^{-1}$  and amplified within the following reaction mixture:  $12.5\text{ }\mu\text{l SYBR Universal Faststart Mastermix (Roche)}$ ,  $4.5\text{ }\mu\text{l dH}_2\text{O}$ ,  $1.5\text{ }\mu\text{l}$  for each of five primers at  $10\text{ ng }\mu\text{l}^{-1}$  (see Table 3), and  $0.5\text{ }\mu\text{l}$  of template DNA. Two thermocycler programs were used in succession. The first included an initial 10 min denaturation at  $94^\circ\text{C}$ , followed by 4 cycles of a 45 s denaturation also at  $94^\circ\text{C}$ , annealing at  $30^\circ\text{C}$  for 2 min, and extension at  $72^\circ\text{C}$  for 50 s. This was followed by another 35 cycle program:  $94^\circ\text{C}$  for 17 s,  $36^\circ\text{C}$  for 30 s, and  $72^\circ\text{C}$  for 45 s, and a final extension for 10 min at  $72^\circ\text{C}$ . The aim of these repeated programs with low annealing temperatures and long annealing times is to produce as many non-specific bands as possible for each sample, increasing the resolving power of the method. Strains were amplified in triplicate to ensure that a repeatable banding pattern could be obtained.

**Table 3 | Random primers for genomic fingerprinting.**

Primers	
Primer name	Sequence
P1	5'-CCGCAGCCAA-3'
P2	5'-ACGGGCAGC-3'
OPA-9	5'-GGGTAACGCC-3'
OPA-13	5'-CAGCAGCCAC-3'
FALL-A	5'-ACGCCGCCCTG-3'

### GEL ELECTROPHORESIS

Reactions mixtures from PCR experiments were held at  $4^\circ\text{C}$  prior to electrophoresis. Standard DNA electrophoresis was carried out with replicates from each strain. Gels were 1.5% agarose and run at 12 v for 16 h at  $4^\circ\text{C}$  with the goal of producing crisp bands easily distinguishable by the analysis software. Gels were stained with ethidium bromide prior to imaging.

### IMAGING AND ANALYSIS

A digital image of each gel was created using a GelDoc (UVP). Images were then analyzed using the Phoretix 1D Pro program from the TotalLab Inc. ([www.totallab.com](http://www.totallab.com)). Banding patterns were standardized for cross gel comparisons by calibrating Rf lines on individual gels. Phoretix 1D Pro converts banding patterns into a format that can be used to produce a dendrogram comparing the differences and similarities between the patterns of amplicons. The final dendrogram was created within Phoretix 1D Pro using UPGMA statistical analysis on Dice coefficients (Dice, 1945) for each of the lanes. A measure of the correlation between the matrix similarities and the dendrogram derived similarities, the cophenetic correlation coefficients (Sokal and Rohlf, 1962) were determined for each sub-cluster of the dendrogram and displayed on the nodes of the constructed dendrograms to estimate the robustness of each cluster.

## RESULTS

### GENOMIC FINGERPRINTING

The repeatability of banding patterns, and thus the success of the fingerprinting technique was tested on 81 haloarchaeal type strains. The PCR on each of the 81 strains was run in triplicate and the products were run on adjacent wells. Figure 1 demonstrates results of the banding pattern for 18 out of the 81 type strains, 15 from the genus *Halorubrum*, and one each from the genus *Halosarcina*, *Halosimplex*, and *Halostagnicola*. Repeatability for the other 63 was examined and they were consistent, as in Figure 1 (data are not shown). Repeatability of the technique indicated robustness of the conditions and primers used and provide confidence for estimating variation between strains.

We were interested to know if the random primers can be used as a screening technique. If banding patterns could reliably demonstrate similarity within genera for instance, newly cultured yet unidentified strains could be easily screened and a general taxonomic decision could be made. Therefore, the banding patterns for the 81 total haloarchaeal type strains were assessed using software that produced a dendrogram of the genomic fingerprints. Figure 2 is the UPGMA dendrogram determined for the above type strains. Compared to other studies (e.g., Shivu et al., 2007; Winget and Wommack, 2008), our genome fingerprinting technique offers very little banding pattern complexity. There are two possible reasons—the primers were designed for systems other than the haloarchaea and adopted for our purposes, and PCR bias, though if it occurs is reproducible (see Figure 1). Yet, species specific banding patterns observed earlier in haloarchaea (Martinez-Murcia and Rodriguez-Valera, 1994) are also observed here; each species appears to have a unique banding pattern. However, there is very little clustering at the genus level. For instance, some species within the same genus



**FIGURE 1 | Repeatability of the fingerprinting technique.** Each number represents a type strain analyzed in triplicate. (1) *Halorubrum arcis* JCM 13916 (2) *Halorubrum coriense* DSM 10284 (3) *Halorubrum distributum* JCM 9100 (4) *Halorubrum ejinorensense* JCM 14265 (5) *Halorubrum lacusprofundi* ATCC 49239 (6) *Halorubrum lipolyticum* DSM 21995 (7) *Halorubrum litoreum* JCM 13561 (8) *Halorubrum saccharovorum* DSM 1137 (9) *Halorubrum*

sodomense JCM 8880 (10) *Halorubrum tebenquichense* DSM 14210 (11) *Halorubrum terrestre* JCM 10247 (12) *Halorubrum tibetense* JCM 11889 (13) *Halorubrum trapanicum* JCM 10477 (14) *Halorubrum vacuolatum* JCM 9060 (15) *Halorubrum xinjiangense* JCM 12388 (16) *Halosarcina pallida* JCM 14848 (17) *Halosimplex carlsbadense* JCM 11222 (18) *Halostagnicola larsenii* JCM 13463.

have similar banding patterns according to the dendrogram analysis (e.g., *Natrinema ejinorensense* and *Natrinema altunense*) but other species from the same genus are found elsewhere (e.g., *Natrinema pelliruberum* and *Natrinema versiforme*). This pattern is observed for all the genera for which several species were analyzed (e.g., *Halorubrum*, *Haloferax*). Thus, this DNA fingerprinting should not be used to classify isolates to a genus level. The observed amount of variation displayed among species within the same genus, led to the hypothesis that this technique might also detect genomic variation among strains within the same species. Therefore, we tested this fingerprinting technique on several populations of naturally co-occurring closely and distantly related strains.

#### MLSA ON ENVIRONMENTAL STRAINS

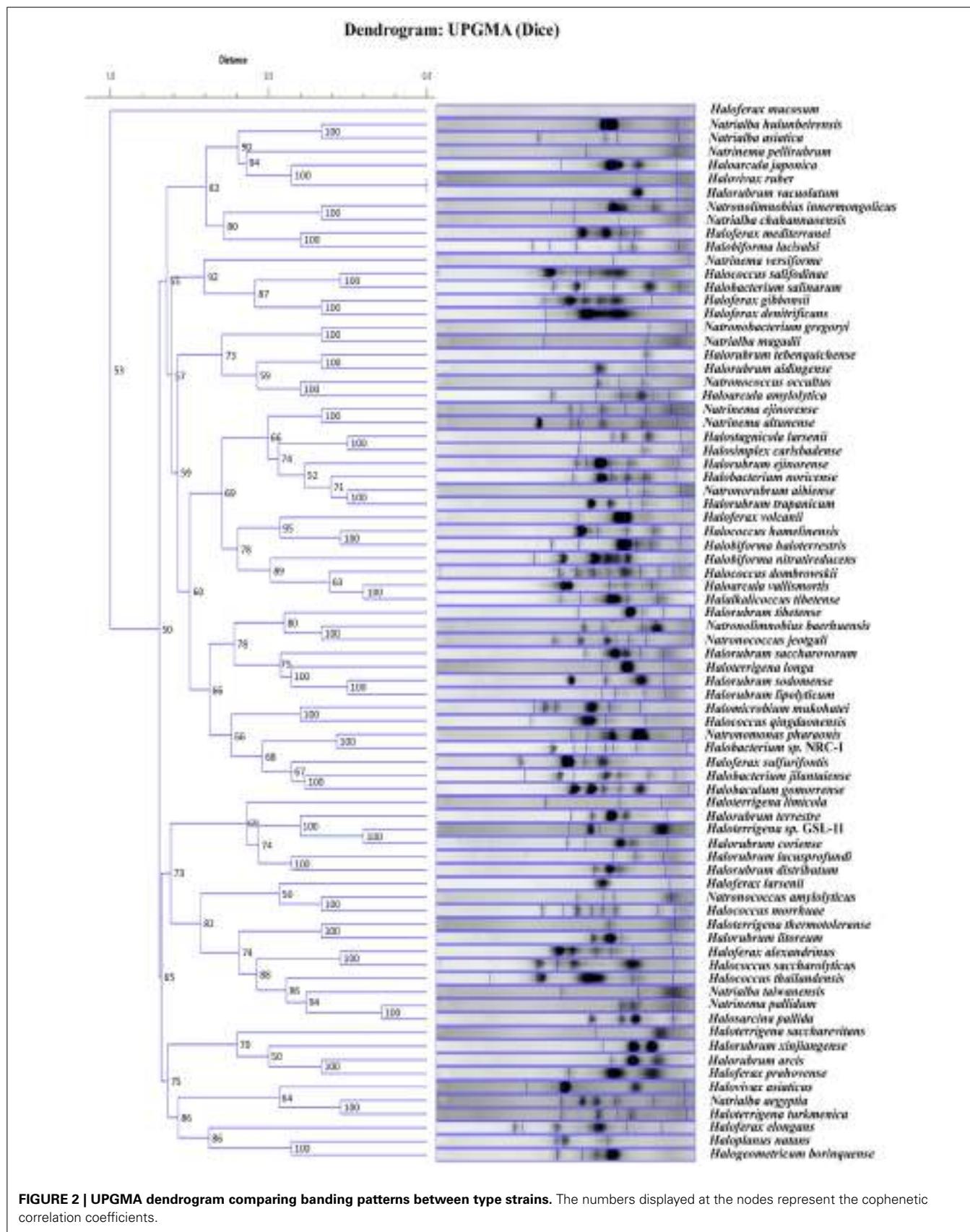
MLSA was performed in order to determine the genetic variation, and the evolutionary relationships of the isolates from Aran-Bidgol lake. Multiple sequence alignments were constructed from individual locus data from the new isolates and from genome data deposited in the NCBI database of type strains. Concatenated alignments were made from these and then a phylogenetic tree was constructed. The Aran-Bidgol isolates clustered into two main genera; *Halorubrum* and *Haloarcula* (Figure 3). Two polytomous groups, A and B, were observed within the genus *Halorubrum* and depicts evidence for distinct phylogroups with low sequence diversity as first seen for Spanish and Algerian isolates (Papke et al., 2007). Pairwise comparison of the number of nucleotides different within each of these phylogroups was carried out using MEGA 5 (Tamura et al., 2011). In both groups A and B, no two isolates had more than 10 nucleotide differences from one another across the concatenation of ~2500 bp (i.e., <1% sequence divergence; Table 4). This also holds true for group C (Table 5) within the *Haloarcula* cluster.

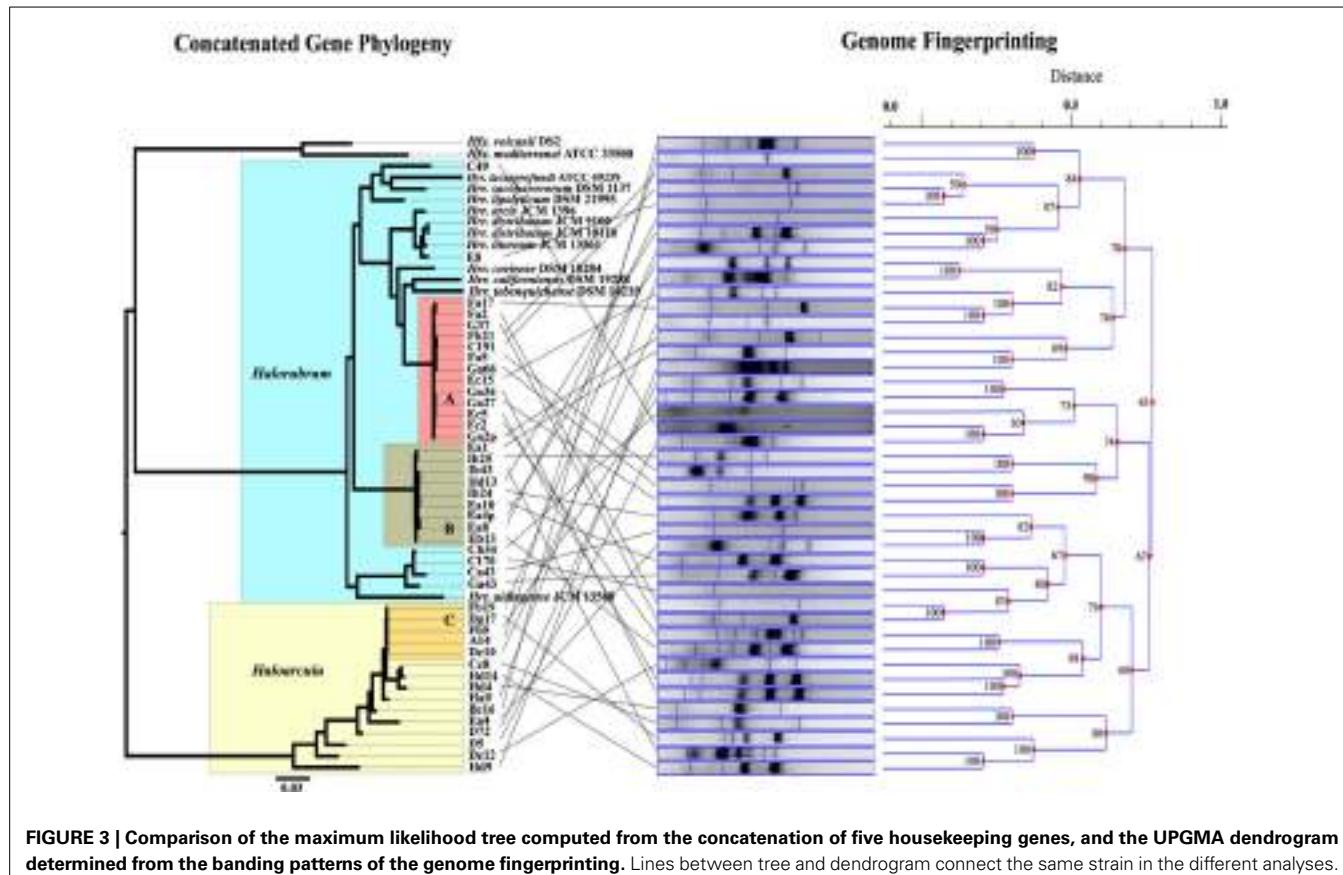
#### FINGERPRINTING THE ARAN-BIDGOL STRAINS

Genomic fingerprint analysis was run on each of the Aran-Bidgol lake environmental isolates. Banding patterns for each individual were generated and compared for similarity by dendrogram construction. The fingerprints and resulting dendrogram were then compared to the ML tree constructed from the MLSA data (Figure 3) for relating genetic and genomic variation within populations. It is noteworthy that despite limited numbers of bands produced for fingerprinting analysis, closely related strains from a single phylogroup displayed numerous variations in banding patterns, many of which were dissimilar to each other as determined by the dendrogram analysis. These widely different banding patterns reflect the variation in individual genomes. Comparison between sequence and banding pattern similarity demonstrates a lot of variation and no discernable patterns of relatedness even between strains that have zero differences across ~2500 nucleotides. Banding patterns of isolates within the genus *Halorubrum* seem as different as the banding patterns of isolates between the genera *Halorubrum* and *Haloarcula*. In some cases identical MLSA haplotypes have identical fingerprint patterns. We believe this can be attributed to the relatively low complexity of fingerprint bands produced, rather than two strains having identical genomes, and in such cases other methods of comparison like genome sequencing might reveal additional differences.

#### DISCUSSION

Our study employed DNA sequencing of multiple protein coding loci and random genomic amplification to test for variation in haloarchaeal isolates cultivated from the same location under the same conditions. The concatenated ML tree in Figure 3, and the number of pairwise nucleotide polymorphisms in Tables 4, 5, show that many isolates are closely related to one another across





**FIGURE 3 | Comparison of the maximum likelihood tree computed from the concatenation of five housekeeping genes, and the UPGMA dendrogram determined from the banding patterns of the genome fingerprinting.** Lines between tree and dendrogram connect the same strain in the different analyses.

**Table 4 | Pairwise comparison of number of nucleotide differences within polytomy Groups A and B defined on the maximum likelihood tree.**

GROUP A	Ga27		0	7	8	5	7	10	9	6	Ea8	GROUP B
	Ec5	5			7	8	5	7	10	9	6	Ea4p
	Ec15	8	5			5	2	8	7	6	5	Ea10
	Ga66	7	8	7			5	9	8	7	4	Hd13
	Fc2	5	4	5	6			6	9	8	3	Ib24
	Fa2	1	1	1	0	1			5	4	7	Eb13
	Fa5	7	8	7	4	6	0			1	8	Ib25
	Fa17	2	2	2	0	2	0	1			7	Ea1
	C191	8	9	8	5	7	0	1	1			Ib43
	Fb21	8	9	8	5	7	0	1	1	0		
	Ga36	4	3	6	7	3	1	7	3	8	8	
	G37	8	3	4	7	5	0	5	1	6	6	
	Ga2p	6	5	4	5	3	0	5	1	6	4	

the five loci and are more or less indistinguishable from each other by these methods. However, the DNA fingerprinting analysis on these same isolates revealed additional variation not captured by MLSA, indicating genomic changes occur faster than the rate of substitution in redundant codon positions. Unfortunately, the deeper branches of the UPGMA hierarchical clustering dendrogram are unreliable for determining relationships and do not provide a good description of the measured Dice coefficients. Yet, shallower branches in the clustering diagram that are a good

representation of the banding pattern differences show conflict with the MLSA phylogeny (Figure 3). Though the fingerprinting technique did not yield patterns of relatedness at the species level or genus level, it did demonstrate the high probability that the genomes of each isolates are unique. Whether that uniqueness is based on gene content or in genomic arrangements is undeterminable from this analysis.

However, given the known propensity for HGT in Halobacteria (Papke et al., 2004, 2007; Cuadros-Orellana

**Table 5 | Pairwise comparison of number of nucleotide differences within polytymous Group C defined on the maximum likelihood tree.**

G	Dc10						
R	A14	0					
O	Fb5	0	0				
P	Dg17	0	0	0			
C	Fb19	2	2	2	2		
	Hd14	30	30	30	30	32	
	Cc8	29	29	29	29	31	7
	Hd4	38	38	38	38	39	26
							19

Cells in blue represent members of Group C and cells in black represent the neighboring cluster on the ML tree.

et al., 2007; Lynch et al., 2012; Naor et al., 2012; Williams et al., 2012; Demaere et al., 2013; Podell et al., 2013), we surmise the fingerprint banding-pattern differences are largely due to gene transfer events. Discovery of recombinant hybrids (Naor et al., 2012) and the identification of enormous identical segments shared among the genomes of phylogenetically distant genera (Demaere et al., 2013) indicates the haloarchaea are subject to immense genomic variability from single gene transfer events. In another study, an influx of 303 transferred genes into *Haloferax mucosum* and *Haloferax mediterranei* were mostly of unknown function with some known transporters (Lynch et al., 2012), which is similar to the types of genes observed in the highly recombinogenic genomic islands of *Haloquadratum walsbyi* (Cuadros-Orellana et al., 2007). The *H. walsbyi* genome is 47.9% GC, but its genomic islands are GC rich by comparison, and enriched in transposable and repeat elements (Bolhuis et al., 2006) indicating a role for viruses in generating genomic diversity (Cuadros-Orellana et al., 2007). Similar to *H. walsbyi*, the genome of *Halobacterium NRC-1* was interspersed with 91 insertion sequence elements of diverse GC compositions (Ng et al., 2000; Kennedy et al., 2001). Apart from HR, IS elements have been attributed to inactivating the bacterio-opsin gene in *Halobacterium halobium* (Dassarma et al., 1983) and causing genomic rearrangements at AT-rich regions in *Halobacterium NRC-1* (Kennedy et al., 2001). Moreover, recent analysis indicates these Aran-Bidgol lake isolates display enormous variation in whole genome content with differences in group A ranging from 0.01 up to 0.51 Mb and from 0.07 up to 0.30 Mb in group B (Fullmer et al., 2014). Therefore, we hypothesize the drastic differences in fingerprints observed for the closest relatives (e.g., strains from groups A, B, and C) are more likely due to HGT, possibly mediated by insertion sequence elements (Dassarma et al., 1983; Ng et al., 2000; Kennedy et al., 2001), tRNAs (Naor et al., 2012), or other factors, rather than genome rearrangements.

We further suggest that the fingerprint banding patterns, especially for those within groups A, B, and C, were unlikely due to mutational events. Haloarchaea have low rates of spontaneous mutation, having been measured at  $1.90 \times 10^{-8}$  mutational events per cell division (Mackwan et al., 2007). Furthermore, haloarchaea are considered to have a high capacity for repairing DNA, as they have demonstrated the ability to survive radiation and desiccation damaged DNA (McCready, 1996; Kottemann et al., 2005), which is probably due to the prevalence of polyploidy

through the process of gene conversion (Lange et al., 2011). Preliminary *in silico* analysis to determine the binding sites for each of the five primers in *Haloquadratum walsbyi* DSM 16790 and *Halorubrum lacusprofundi* revealed priming mostly in conserved loci, although a few phage related loci were also detected. Because many of the compared strains are very closely related, having only a few (or zero) nucleotide polymorphisms in the ~2500 sequenced base pairs, yet display enormous differences in fingerprint banding patterns, it would be unlikely that a few, or even one of the PCR binding sites in every strain within groups A, B, or C, would be mutated. Therefore, substitutions in PCR primer binding sites seem unlikely to have played a role in generating all the observed differences in banding patterns, especially those from closely related strains.

Analysis of five housekeeping genes demonstrates the isolates form genetically similar and distinct populations in a single environmental community and yet each genome is apparently different. This observation agrees well with expectations from the distributed genome hypothesis (Ehrlich et al., 2010). According to this, the non-core genes available in the pangenome pool are dispensed uniquely amongst the individual cells of a species. The differences in haloarchaeal genomic banding patterns suggests that in nature populations are made of highly varied individuals rather than clones of a single individual. The number of distinct genotypes observed, most likely due to gene flow, suggests that haloarchaeal cells are acquiring genomic variation within populations at a rate faster than redundant codon position substitutions, and possibly at every replication event. Distribution of the non-core genes within a highly recombining population (defined by MLSA phylogeny) theoretically enables the individual to quickly adapt to new environmental selection conditions, especially virus predation (Cuadros-Orellana et al., 2007) but may also result from random processes like neutral drift (Gogarten and Townsend, 2005).

## AUTHOR CONTRIBUTIONS

R. Thane Papke, J. Peter Gogarten, and Antonio Ventosa conceived the research. Nikhil Ram Mohan, Matthew S. Fullmer, Andrea M. Makkay, and Ryan Wheeler gathered data, and performed the analyses. Nikhil Ram Mohan, Matthew S. Fullmer, Andrea M. Makkay, Ryan Wheeler, Antonio Ventosa, J. Peter Gogarten, and R. Thane Papke wrote the manuscript.

## ACKNOWLEDGMENTS

We would like to thank Mohammad A. Amozegar from the University of Tehran for cultivating the Aran-Bidgol salt lake isolates. This research was supported by the National Science Foundation (award numbers, DEB0919290 and DEB0830024) and NASA Astrobiology: Exobiology and Evolutionary Biology Program Element (Grant Number NNX12AD70G).

## REFERENCES

- Akaike, H. (1974). A new look at the statistical model. *IEEE Trans. Autom. Control* 19, 716–723. doi: 10.1109/TAC.1974.1100705
- Allers, T., Ngo, H. P., Mevarech, M., and Lloyd, R. G. (2004). Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. *Appl. Environ. Microbiol.* 70, 943–953. doi: 10.1128/AEM.70.2.943-953.2004

- Andam, C. P., Harlow, T. J., Papke, R. T., and Gogarten, J. P. (2012). Ancient origin of the divergent forms of leucyl-tRNA synthetases in the *Halobacteriales*. *BMC Evol. Biol.* 12:85. doi: 10.1186/1471-2148-12-85
- Anton, J., Llobet-Brossa, E., Rodriguez-Valera, F., and Amann, R. (1999). Fluorescence *in situ* hybridization analysis of the prokaryotic community inhabiting crystallizer ponds. *Environ. Microbiol.* 1, 517–523. doi: 10.1046/j.1462-2920.1999.00065.x
- Barrangou, R., Yoon, S. S., Breidt Jr. F. Jr., Fleming, H. P., and Klaenhammer, T. R. (2002). Characterization of six *Leuconostoc fallax* bacteriophages isolated from an industrial sauerkraut fermentation. *Appl. Environ. Microbiol.* 68, 5452–5458. doi: 10.1128/AEM.68.11.5452-5458.2002
- Bolhuis, H., Palm, P., Wende, A., Falb, M., Rampp, M., Rodriguez-Valera, F., et al. (2006). The genome of the square archaeon *Halocladratum walsbyi*: life at the limits of water activity. *BMC Genomics* 7:169. doi: 10.1186/1471-2164-7-169
- Boucher, Y., Douady, C. J., Sharma, A. K., Kamekura, M., and Doolittle, W. F. (2004). Intrageneric heterogeneity and intergeneric recombination among haloarchaeal rRNA genes. *J. Bacteriol.* 186, 3980–3990. doi: 10.1128/JB.186.12.3980-3990.2004
- Cuadros-Orellana, S., Martin-Cuadrado, A. B., Legault, B., D'Auria, G., Zhaxybayeva, O., Papke, R. T., et al. (2007). Genomic plasticity in prokaryotes: the case of the square haloarchaeon. *ISME J.* 1, 235–245. doi: 10.1038/ismej.2007.35
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772. doi: 10.1038/nmeth.2109
- Dassarma, S., Rajbhandary, U. L., and Khorana, H. G. (1983). High-frequency spontaneous mutation in the bacterio-opsin gene in *Halobacterium halobium* is mediated by transposable elements. *Proc. Natl. Acad. Sci. U.S.A.* 80, 2201–2205. doi: 10.1073/pnas.80.8.2201
- De la Haba, R. R., Marquez, M. C., Papke, R. T., and Ventosa, A. (2012). Multilocus sequence analysis of the family *Halomonadaceae*. *Int. J. Syst. Evol. Microbiol.* 62, 520–538. doi: 10.1099/ijts.0.032938-0
- Demaaere, M. Z., Williams, T. J., Allen, M. A., Brown, M. V., Gibson, J. A., Rich, J., et al. (2013). High level of intergenera gene exchange shapes the evolution of haloarchaea in an isolated antarctic lake. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16939–16944. doi: 10.1073/pnas.1307090110
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology* 26, 6. doi: 10.2307/1932409
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Ehrlich, G. D., Ahmed, A., Earl, J., Hiller, N. L., Costerton, J. W., Stoodley, P., et al. (2010). The distributed genome hypothesis as a rubric for understanding evolution *in situ* during chronic bacterial biofilm infectious processes. *FEMS Immunol. Med. Microbiol.* 59, 269–279. doi: 10.1111/j.1574-695X.2010.00704.x
- Feil, E. J., Enright, M. C., and Spratt, B. G. (2000). Estimating the relative contributions of mutation and recombination to clonal diversification: a comparison between *Neisseria meningitidis* and *Streptococcus pneumoniae*. *Res. Microbiol.* 151, 465–469. doi: 10.1016/S0923-2508(00)00168-6
- Fullmer, M. S., Soucy, S. M., Swithers, K. S., Makkay, A. M., Wheeler, R., Ventosa, A., et al. (2014). Population and genomic analysis of the genus *Halorubrum*. *Front. Microbiol.* 5:140. doi: 10.3389/fmicb.2014.00140
- Ghai, R., Pasie, L., Fernandez, A. B., Martin-Cuadrado, A. B., Mizuno, C. M., McMahon, K. D., et al. (2011). New abundant microbial groups in aquatic hypersaline environments. *Sci. Rep.* 1:135. doi: 10.1038/srep00135
- Gogarten, J. P., and Townsend, J. P. (2005). Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* 3, 679–687. doi: 10.1038/nrmicro1204
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Jones, J. G., Young, D. C., and Dassarma, S. (1991). Structure and organization of the gas vesicle gene cluster on the *Halobacterium halobium* plasmid pNRC100. *Gene* 102, 117–122. doi: 10.1016/0378-1119(91)90549-Q
- Kennedy, S. P., Ng, W. V., Salzberg, S. L., Hood, L., and Dassarma, S. (2001). Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res.* 11, 1641–1650. doi: 10.1101/gr.190201
- Kottmann, M., Kish, A., Iloanusi, C., Bjork, S., and Diruggiero, J. (2005). Physiological responses of the halophilic archaeon *Halobacterium* sp. strain NRC1 to desiccation and gamma irradiation. *Extremophiles* 9, 219–227. doi: 10.1007/s00792-005-0437-4
- Lange, C., Zerulla, K., Breuert, S., and Soppa, J. (2011). Gene conversion results in the equalization of genome copies in the polyploid haloarchaeon *Haloflexax volcanii*. *Mol. Microbiol.* 80, 666–677. doi: 10.1111/j.1365-2958.2011.07600.x
- Lapierre, P., and Gogarten, J. P. (2009). Estimating the size of the bacterial pan-genome. *Trends Genet.* 25, 107–110. doi: 10.1016/j.tig.2008.12.004
- Legault, B. A., Lopez-Lopez, A., Alba-Casado, J. C., Doolittle, W. F., Bolhuis, H., Rodriguez-Valera, F., et al. (2006). Environmental genomics of “*Halocladratum walsbyi*” in a saltern crystallizer indicates a large pool of accessory genes in an otherwise coherent species. *BMC Genomics* 7:171. doi: 10.1186/1471-2164-7-171
- Litchfield, C. D. (1998). Survival strategies for microorganisms in hypersaline environments and their relevance to life on early Mars. *Meteorit. Planet. Sci.* 33, 813–819. doi: 10.1111/j.1945-5100.1998.tb01688.x
- Lozier, R. H., Bogomolni, R. A., and Stoeckenius, W. (1975). Bacteriorhodopsin: a light-driven proton pump in *Halobacterium halobium*. *Biophys. J.* 15, 955–962. doi: 10.1016/S0006-3495(75)85875-9
- Lynch, E. A., Langille, M. G., Darling, A., Wilbanks, E. G., Haltiner, C., Shao, K. S., et al. (2012). Sequencing of seven haloarchaeal genomes reveals patterns of genomic flux. *PLoS ONE* 7:e41389. doi: 10.1371/journal.pone.0041389
- Mackwan, R. R., Carver, G. T., Drake, J. W., and Grogan, D. W. (2007). An unusual pattern of spontaneous mutations recovered in the halophilic archaeon *Haloflexax volcanii*. *Genetics* 176, 697–702. doi: 10.1534/genetics.106.069666
- Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., et al. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* 95, 3140–3145. doi: 10.1073/pnas.95.6.3140
- Martinez-Murcia, A. J., and Rodriguez-Valera, F. (1994). The use of arbitrarily primed PCR (AP-PCR) to develop taxa specific DNA probes of known sequence. *FEMS Microbiol. Lett.* 124, 265–270. doi: 10.1016/0378-1097(94)00440-4
- McCready, S. (1996). The repair of ultraviolet light-induced DNA damage in the halophilic archaeabacteria, *Halobacterium cutirubrum*, *Halobacterium halobium* and *Haloflexax volcanii*. *Mutat. Res.* 364, 25–32. doi: 10.1016/0921-8777(96)00018-3
- Naor, A., Lapierre, P., Mevarech, M., Papke, R. T., and Gophna, U. (2012). Low species barriers in halophilic archaea and the formation of recombinant hybrids. *Curr. Biol.* 22, 1444–1448. doi: 10.1016/j.cub.2012.05.056
- Nelson-Sathi, S., Dagan, T., Landan, G., Janssen, A., Steel, M., McInerney, J. O., et al. (2012). Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. *Proc. Natl. Acad. Sci. U.S.A.* 109, 20537–20542. doi: 10.1073/pnas.1209119109
- Ng, W. V., Kennedy, S. P., Mahairas, G. G., Berquist, B., Pan, M., Shukla, H. D., et al. (2000). Genome sequence of *Halobacterium* species NRC-1. *Proc. Natl. Acad. Sci. U.S.A.* 97, 12176–12181. doi: 10.1073/pnas.190337797
- Oren, A. (2002). Molecular ecology of extremely halophilic archaea and bacteria. *FEMS Microbiol. Ecol.* 39, 1–7. doi: 10.1111/j.1574-6941.2002.tb00900.x
- Papke, R. T., Koenig, J. E., Rodriguez-Valera, F., and Doolittle, W. F. (2004). Frequent recombination in a saltern population of *Halorubrum*. *Science* 306, 1928–1929. doi: 10.1126/science.1103289
- Papke, R. T., Zhaxybayeva, O., Feil, E. J., Sommerfeld, K., Muise, D., and Doolittle, W. F. (2007). Searching for species in haloarchaea. *Proc. Natl. Acad. Sci. U.S.A.* 104, 14092–14097. doi: 10.1073/pnas.0706358104
- Podell, S., Ugalde, J. A., Narasingarao, P., Banfield, J. F., Heidelberg, K. B., and Allen, E. E. (2013). Assembly-driven community genomics of a hyper-saline microbial ecosystem. *PLoS ONE* 8:e61692. doi: 10.1371/journal.pone.0061692
- Sharma, A. K., Walsh, D. A., Baptiste, E., Rodriguez-Valera, F., Doolittle, W. F., and Papke, R. T. (2007). Evolution of rhodopsin ion pumps in haloarchaea. *BMC Evol. Biol.* 7:79. doi: 10.1186/1471-2148-7-79
- Shivu, M. M., Rajeeva, B. C., Girisha, S. K., Karunasagar, I., and Krohne, G. (2007). Molecular characterization of *Vibrio harveyi* bacteriophages isolated from aquaculture environments along the coast of India. *Environ. Microbiol.* 9, 322–331. doi: 10.1111/j.1462-2920.2006.01140.x

- Sokal, R. R., and Rohlf, F. J. (1962). The Comparison of dendograms by objective methods. *Taxon* 11, 8. doi: 10.2307/1217208
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Tettelin, H., Riley, D., Cattuto, C., and Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* 11, 472–477. doi: 10.1016/j.mib.2008.09.006
- Thompson, J. R., Pacocha, S., Pharino, C., Klepac-Ceraj, V., Hunt, D. E., Benoit, J., et al. (2005). Genotypic diversity within a natural coastal bacterioplankton population. *Science* 307, 1311–1313. doi: 10.1126/science.1106028
- Williams, D., Gogarten, J. P., and Papke, R. T. (2012). Quantifying homologous replacement of loci between haloarchaeal species. *Genome Biol. Evol.* 4, 1223–1244. doi: 10.1093/gbe/evs098
- Winget, D. M., and Wommack, K. E. (2008). Randomly amplified polymorphic DNA PCR as a tool for assessment of marine viral richness. *Appl. Environ. Microbiol.* 74, 2612–2618. doi: 10.1128/AEM.02829-07

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 07 February 2014; accepted: 19 March 2014; published online: 09 April 2014.

Citation: Ram Mohan N, Fullmer MS, Makay AM, Wheeler R, Ventosa A, Naor A, Gogarten JP and Papke RT (2014) Evidence from phylogenetic and genome fingerprinting analyses suggests rapidly changing variation in *Halorubrum* and *Haloarcula* populations. *Front. Microbiol.* 5:143. doi: 10.3389/fmicb.2014.00143

This article was submitted to Extreme Microbiology, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Ram Mohan, Fullmer, Makay, Wheeler, Ventosa, Naor, Gogarten and Papke. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Inteins as indicators of gene flow in the halobacteria

Shannon M. Soucy, Matthew S. Fullmer, R. Thane Papke and Johann Peter Gogarten \*

Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT, USA

**Edited by:**

Jesse Dillon, California State University, Long Beach, USA

**Reviewed by:**

Julie L. Meyer, University of Florida, USA

Kenneth Mills, College of the Holy Cross, USA

**\*Correspondence:**

Johann Peter Gogarten,  
Microbiology Program, Department of Molecular and Cell Biology,  
University of Connecticut, 91 N. Eagleville Rd., Storrs, CT 06269-3125, USA  
e-mail: gogarten@uconn.edu;  
jpggarten@gmail.com

This research uses inteins, a type of mobile genetic element, to infer patterns of gene transfer within the Halobacteria. We surveyed 118 genomes representing 26 genera of Halobacteria for intein sequences. We then used the presence-absence profile, sequence similarity and phylogenies from the inteins recovered to explore how intein distribution can provide insight on the dynamics of gene flow between closely related and divergent organisms. We identified 24 proteins in the Halobacteria that have been invaded by inteins at some point in their evolutionary history, including two proteins not previously reported to contain an intein. Furthermore, the size of an intein is used as a heuristic for the phase of the intein's life cycle. Larger size inteins are assumed to be the canonical two domain inteins, consisting of self-splicing and homing endonuclease domains (HEN); smaller sizes are assumed to have lost the HEN domain. For many halobacterial groups the consensus phylogenetic signal derived from intein sequences is compatible with vertical inheritance or with a strong gene transfer bias creating these clusters. Regardless, the coexistence of intein-free and intein-containing alleles reveal ongoing transfer and loss of inteins within these groups. Inteins were frequently shared with other Euryarchaeota and among the Bacteria, with members of the Cyanobacteria (*Cyanothecae*, *Anabaena*), Bacteriodetes (*Salinibacter*), Betaproteobacteria (*Delftia*, *Acidovorax*), Firmicutes (*Halanaerobium*), Actinobacteria (*Longispora*), and Deinococcus-Thermus-group.

**Keywords:** gene symbiosis, genome as an ecosystem, inteins, mobile genetic elements, gene flow, horizontal gene transfer, halobacteria

## INTRODUCTION

Inteins are self-splicing genetic parasites located in highly conserved sites of slowly evolving genes. They are found in all three domains of life and in viruses (Perler et al., 1997; Pietrovoski, 2001; Gogarten et al., 2002; Swithers et al., 2009). Similar to group I introns, inteins are often associated with a homing endonuclease (HEN). An important difference between inteins and introns is the timing of the splicing activity, which occurs immediately after transcription in introns and after translation in inteins (Hirata et al., 1990; Kane et al., 1990). The association with a HEN domain enables a cyclic invasion pattern, called the homing cycle (Goddard and Burt, 1999; Gogarten and Hilario, 2006). The homing cycle consists of three phases: intein invasion, intein fixation, and eventually loss of the intein enabling invasion to occur again. During invasion and fixation the intein splicing domains are associated with a HEN domain forming a canonical intein (hereafter referred to as a large intein); however, during the loss phase the function of the HEN is often disrupted and begins to degrade, generating a mini-intein. Simulations have shown that intein-containing and intein-free alleles can coexist in well mixed populations under some sets of parameters (Yahara et al., 2009; Barzel et al., 2011). Also, inteins with functioning HEN domains were inferred to have persisted in some eukaryotic lineages for several 100 million years (Butler et al., 2006; Gogarten and Hilario, 2006).

Inteins do not have an apparatus to penetrate the cell envelope. Therefore, they must rely on mechanisms in place within the population for insertion into the cell such as: conjugation, mating, generalized DNA uptake, and viruses or gene transfer agents (Lang et al., 2012). The faster-than-Mendelian inheritance of the large inteins (Gimble and Thorner, 1992), along with a nearly neutral fitness burden, enables these mobile elements to persist in organisms over evolutionary time as long as there are new populations to invade (Goddard and Burt, 1999; Gogarten and Hilario, 2006). Furthermore, the size of the intein (mini or large) provides information about the genomic mobility of the element as mini inteins are rarely integrated into the recipient's genome; whereas large inteins are more frequently integrated due to the activity of the HEN. The conservation of the recognition site provides an invasion target even in distantly related strains and species. Also, inteins have a higher substitution rate relative to their extein hosts (Swithers et al., 2013). This substitution rate gives rise to many evolutionarily informative sites when comparing a large collection of homologous inteins. In this work, we take advantage of these traits and survey the distribution of inteins in the Halobacteria, a highly recombinant class of halophilic Archaea (Williams et al., 2012) known to contain several intein alleles (Perler, 2002). We make use of 118 halobacterial genomes (Supplementary Table 1) and the previously reported and newly discovered intein alleles to survey networks of gene transfer within and outside the

Halobacteria based on the presence-absence profile of the inteins, their sequence similarity, and the phylogenies reconstructed from intein sequences.

## MATERIALS AND METHODS

### HALOBACTERIAL INTEIN SEQUENCE RETRIEVAL AND ALIGNMENT

Position specific scoring matrices (PSSMs) were created using the collection of all inteins from InBase, the Intein database and registry (Perler, 2002). A custom database was created with all inteins, and each intein was used as a seed to create a PSSM using the custom database. These PSSMs were then used as a seed for PSI-BLAST (Altschul, 1997) searches against each of the halobacterial genomes available from NCBI as of June 2013 as well as a private collection sequenced by our collaborators. To remove false positives, a size exclusion step was then performed on each protein sequence as an intein domain adds 100–700 aa to invaded protein sequences. Inteins were then aligned using Muscle (Edgar, 2004) with default parameters in the SeaView version 4.0 software package (Gouy et al., 2010). Insertions, which passed the size exclusion step, but did not contain splicing domains, were removed and the previous steps were repeated using the resulting dataset on a collection of private genomes from the Papke lab. Prottest 3.2 (Guindon et al., 2010; Darriba et al., 2011) was used to determine an appropriate substitution model for the intein sequences, the WAG model was favored and used for all subsequent trees for consistency. Once the collection of halobacterial inteins was complete, sequences were re-aligned using SATé (Liu et al., 2012) to generate a final alignment using MAFFT (Katoh and Standley, 2013) to align, Muscle (Edgar, 2004) to merge, RAXML (Stamatakis, 2014) for tree estimation, and a WAG model for each allele.

To determine the relationship among all halobacterial inteins, the inteins were aligned using Muscle (Edgar, 2004). Subsequently a tree was built using PhyML v3.0 (Guindon et al., 2010) using a WAG substitution model with a Gamma shape parameter and the proportion of invariant sites estimated from the data.

### INTEIN RETRIEVAL OUTSIDE THE HALOBACTERIA

Each halobacterial intein was used as a BLAST (Altschul et al., 1990) query against the non-redundant database on NCBI. Any match with an *e*-value better than 0.000001 was aligned to the dataset to which its query belonged. Sequences were then filtered based on the protein annotation and goodness of fit to the existing alignment. As an additional filtering step each match was used as a query against the non-redundant database and the majority BLAST hit annotations were used to verify the protein identity, as annotations are not always reliable. Remaining sequences were aligned using Clustal Omega 1.1.0 (Sievers et al., 2011) with the profile alignment option in SeaView 4.0 (Gouy et al., 2010). Maximum-likelihood trees were built using PhyML (Guindon et al., 2010) with the WAG model, and rates estimated from the data.

To assess the relative contribution of different genera represented in each intein allele sequence data set, a stacked column graph was created. Sequence density was calculated for each intein allele by dividing the number of intein sequences in each genus by the number of total intein sequences in that allele.

### SYMBIOTIC STATE ASSIGNMENT

Intein sequence length was used to determine symbiotic state. For each intein allele the length of the intein sequence was determined. A cutoff length for mini-intein assignment was based on the presence of a gap in intein lengths greater than 100 amino acids within an allele. The third intein state “no-intein” was assigned where the intein was clearly absent from the orthologous protein containing an intein in any of the halobacterial genomes examined. Additionally, once an intein was noted as a mini-intein the alignment was analyzed to ensure the gaps in these sequences correspond to the location of the HEN domain.

### RIBOSOMAL PROTEIN REFERENCE TREE

Alignments of 55 ribosomal protein for 21 Halobacteria (Williams et al., 2012) were used to find orthologous proteins in the genomes used in this work. In-house python scripts (data file 1) were used to concatenate the alignments, and PhyML v3.0 (Guindon et al., 2010) was used to build a tree. The tree used the WAG substitution model with the Gamma shape parameter and the proportion of invariant sites and base frequencies estimated from the data.

### BAYESIAN CLUSTERING WITH INTEIN SEQUENCES

A concatenation of an intein presence-absence matrix and alignments for each intein allele were generated using in-house python scripts (data file 1). MrBayes version 3.2.1 (Ronquist et al., 2012) was then used to perform a clustering analysis using a partition allowing for character states in the presence-absence matrix and sequence information for each intein allele. The prior for the character portion of the data matrix used a symmetrical Dirichlet distribution with an exponential (1.0), and variable rates so each column was considered independent of the others. The likelihood for the character portion of the alignment used variable coding and 5 beta categories. The prior for the protein sequences in the alignment used a fixed WAG substitution model, with state frequencies estimated from the data, and the likelihood settings used a Gamma shape parameter and the proportion of invariant sites estimated from the data.

## RESULTS

### HALOBACTERIAL INTEINS

The intein content of a collection of halobacterial genomes was analyzed using an intein-allele-specific PSSM. This survey revealed 13 genes in the Halobacteria invaded by inteins at 24 distinct positions (intein alleles) (Table 1). Seven of these intein alleles were not previously reported in the Halobacteria, and two of the seven have not previously been reported to harbor inteins: a DNA ligase gene involved in double strand break repair, and a deaminase gene involved in nucleotide metabolism (Table 1). To determine if vertical inheritance was accountable for the distribution of intein alleles, the presence-absence matrix of intein alleles was mapped onto a reference phylogeny (Figure 1). Clearly, intein presence-absence is not concordant with the ribosomal protein phylogeny, implicating abundant horizontal genetic transfer (HGT) in creating the observed distribution. The presence of multiple intein alleles in the majority of genomes (70%) might be interpreted

**Table 1 | Extein in the halobacteria.**

Intein allele	Extein annotation
<i>cdc21-a</i>	Cell division control protein 21
<i>cdc21-b</i>	
<i>cdc21-c</i>	
<i>polB-d</i>	DNA polymerase B1
<i>polB-a</i>	
<i>polB-b</i>	
<i>polB-c</i>	
<i>pol-IIa</i>	DNA polymerase II large subunit
<i>pol-IIb*</i>	
<i>dtd**</i>	Deoxycytidine triphosphate deaminase
<i>gyrB</i>	DNA gyrase subunit B
<i>helicase-b*</i>	ATP-dependent helicase
<i>ligase**</i>	ATP-dependent DNA ligase I
<i>rfc-a</i>	Replication factor C small subunit
<i>rfc-d*</i>	
<i>rir1-l*</i>	Ribonucleoside-diphosphate reductase
<i>rir1-k</i>	
<i>rir1-b</i>	
<i>rir1-g</i>	
<i>rir1-m*</i>	
<i>rpolA</i>	DNA-directed RNA polymerase subunit A
<i>udp</i>	UDP-glucose 6-dehydrogenase
<i>topA</i>	DNA topoisomerase I
<i>topB6</i>	DNA topoisomerase VI subunit B

\*Denotes intein alleles discovered in this work.

\*\*Denotes extein sequences not previously reported to be invaded by an intein.

to suggest that inteins could spread locally within a single genome.

### INTEIN PROPAGATION WITHIN THE HALOBACTERIA

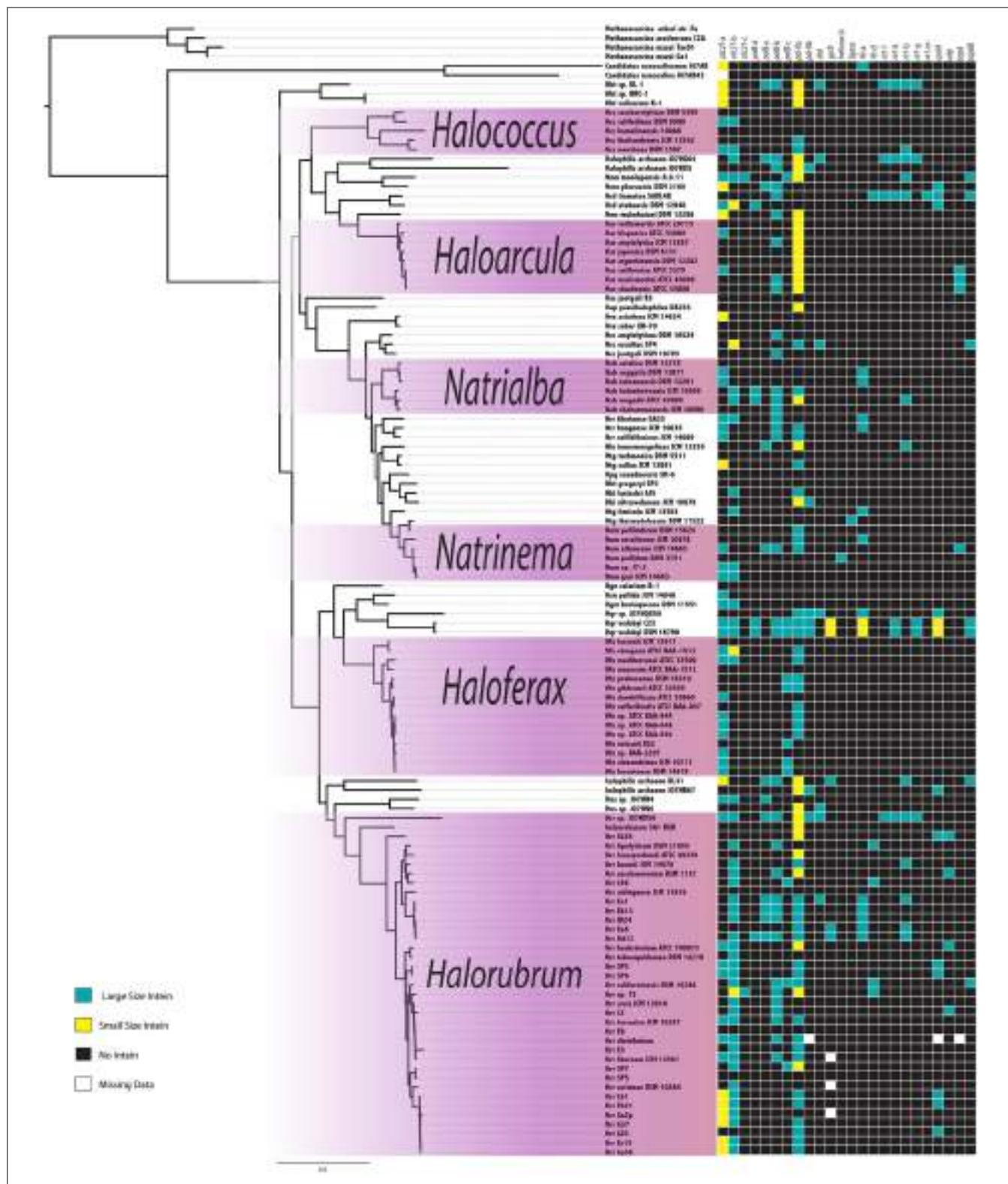
To address the possibility of inteins moving locally within a genome, the phylogenetic relationships among all halobacterial intein sequences were analyzed (Figure 2). All of the intein alleles form highly supported clusters with others of the same type, with the exception of two sequences: the *polB-c* inteins of *Haloferax larsenii* and *Haloferax elongans* group inside the *polB-b* intein allele cluster; however, this node is poorly supported (59/100 bootstraps) indicating this relationship could be an artifact produced by poor resolution of the relationships that connect various intein alleles. Furthermore, there is poor support linking all of the intein allele clusters together (less than 70% bootstrap support), indicating sequence conversion (an intein invading an ectopic or atypical locus) between intein alleles, even within the same host protein, is uncommon. Among the inteins analyzed here, at most one invasion of an ectopic site is supported by the data, confirming that this type of event is rare (Perler et al., 1997; Gogarten et al., 2002). These data indicate that HGT is the only plausible explanation for the large number of different intein alleles in this class of organisms. Incongruence between the presence of inteins and ribosomal phylogeny also support this conclusion.

### BAYESIAN PHYLOGENETIC ANALYSIS OF INTEINS

In an attempt to resolve the local events (transfers and vertical inheritance within the Halobacteria) that gave rise to the

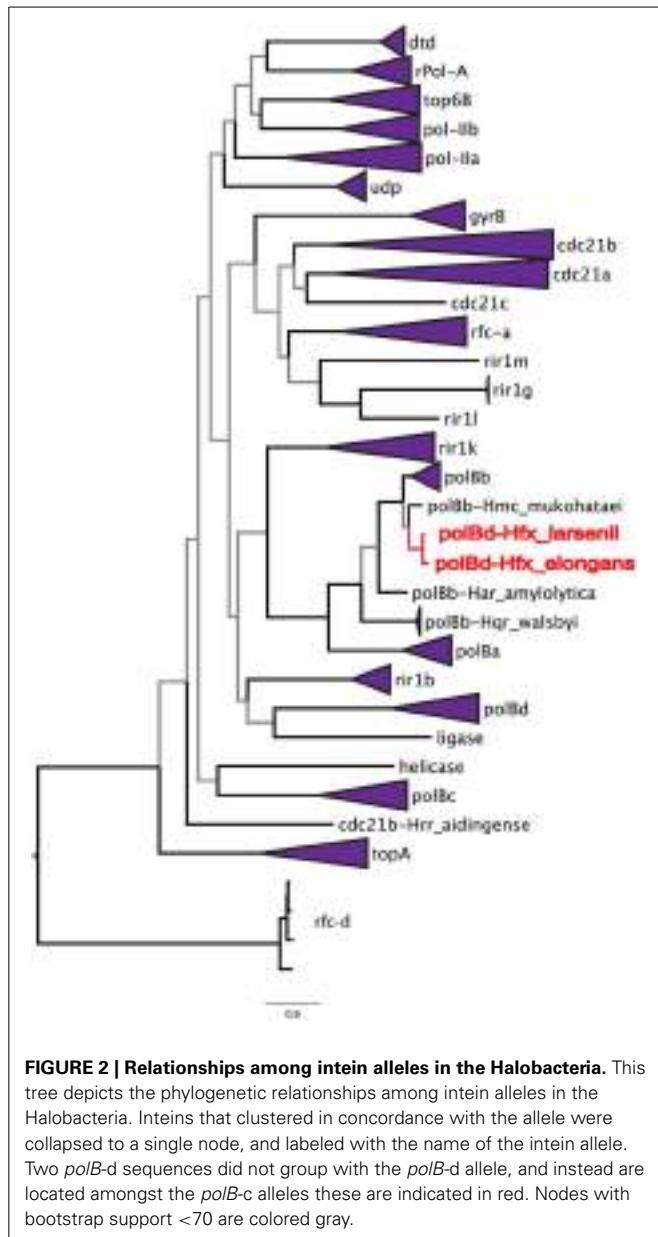
observed intein distribution in the Halobacteria, a Bayesian analysis based on the intein sequences for each allele and on the presence-absence pattern was performed (Figure 3). In this analysis two organisms may group together because they both inherited inteins from a common ancestor, or because an intein was recently transferred between them. The paucity of well-supported nodes (nodes with 0.95 or greater posterior probability were considered well-supported) in part reflects the extent to which our sample is biased toward very similar sequences (31% of halobacterial genomes in this study are from *Halorubrum*). Most of the well-supported clusters in the Bayesian tree also occur in the reference tree, suggesting these inteins may be the result of shared vertical inheritance. However, many of these clusters do not have identical intein profiles (clusters 1, 6, 8, and 10), thus HGT between close relatives is a better explanation than vertical inheritance for these clusters. Only three of the clusters, 2, 9, and 12, have branching orders that are different from those observed in the reference tree indicating HGT. Cluster 2 is made up of *Natrinema* spp. *pellirubrum* and *versiforme* which share only the *pol-IIa* intein. In the reference tree *Nnm. versiforme* groups with the rest of the *Natrinema*, and *Nnm. pellirubrum* groups with *Haloterrigena thermotolerans*. *Natrinema* sp. J7-2 is the only other member of the *Natrinema* that has an intein in the *pol-IIa* position, but the intein in this species is 14 aa shorter than the intein shared by *Nnm. pellirubrum* and *Nnm. versiforme*. *Htg. thermotolerans* shares no inteins with *Nnm. pellirubrum*. Cluster 9 is made up of *Halorubrum* spp. C49 and E3, which share only the *cdc21-b* intein. In the reference tree *Hrr. E3* groups with *Halorubrum litoreum* and the two share the *pol-IIa* intein allele, but no others. *Hrr. C49* groups with *Halorubrum saccharovorum* and they do not share any inteins. Cluster 12 is made up of *Haloferax* spp. *denitrificans*, *lucentense*, *alexandrinus*, and *Haloferax* sp. BAB2207, which all have an intein in the *cdc21-a* position. In the reference tree *Hfx. lucentense*, *Hfx. sp. BAB2207*, and *Hfx. alexandrinus* all group together, but *Hfx. denitrificans* groups with *Haloferax sulfurifontis*, and they do not share any inteins. The lack of shared inteins between clusters in the reference tree and differences among the inteins shared in these clusters cause these divergences in this tree as compared to the reference tree. This may indicate that the taxa in the Bayesian clusters are exchange partners, or that they share unsampled intermediate exchange partners. Additionally, the majority of clusters share 2 or fewer intein alleles between all members of the cluster (eight out of 12 clusters). The two clusters that share the most intein alleles between all members are Cluster 3, made up of *Halococcus walsbyi* strains DSM 16790 and C23 with 13 shared intein alleles, and cluster 7 made up of *Halorubrum* spp. strains SP3 and SP9 sharing 4 intein alleles. Both of these clusters have branching patterns identical to those on the reference tree, indicating that phylogenetic proximity plays a significant role in intein distribution.

Members of the *Halorubrum* genus, not surprisingly, were highly represented in the clusters (four of 12 total). All four of the clusters show a geographic bias. Clusters 6, 8, and 9 were all isolated from the Aran-Bidgol lake in Iran, and cluster 7 was isolated from the Sodom Ponds in Israel (Atanasova et al., 2012). Branch lengths in all of these clusters are very small, suggesting these populations are well mixed with respect to intein sequences. Geography does not seem to play a strong role in linking other



**FIGURE 1 | Intein Invasion Pattern in the Halobacteria.** Intein pattern of presence-absence is mapped onto the tips of a ribosomal reference tree, teal boxes indicate the presence of a full size intein, yellow boxes indicate the presence of a mini-intein, black boxes

indicate the absence of an intein, and white boxes indicate missing data. Purple shaded boxes indicate the genera with more than five species represented on the tree. Nodes with bootstrap support <70 are in gray.

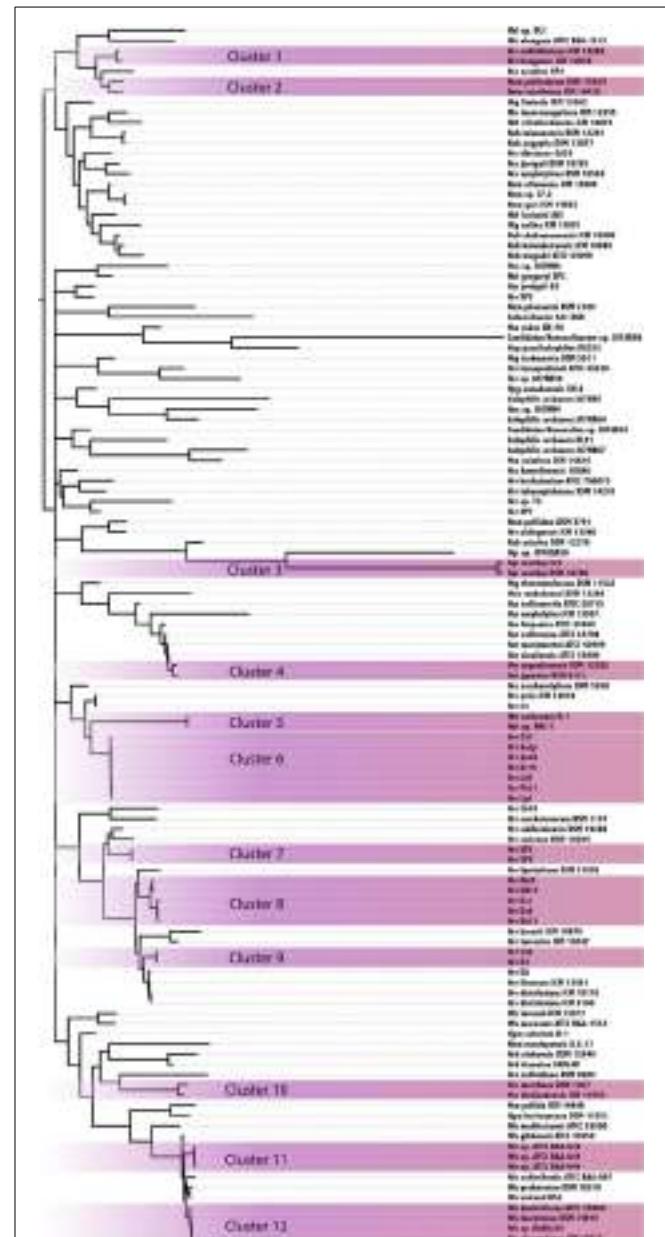


Halobacteria. Inteins that clustered in concordance with the allele were collapsed to a single node, and labeled with the name of the intein allele. Two *polB-d* sequences did not group with the *polB-d* allele, and instead are located amongst the *polB-c* alleles these are indicated in red. Nodes with bootstrap support <70 are colored gray.

well-supported clusters based on intein sequences. Furthermore, evidence of clustering based on geography in the *Halorubrum* is less interesting than the clear separation between groups isolated from the same location (cluster 6, 8, and 9). This separation of species of *Halorubrum* from the same location is echoed in the reference tree, and taken together with the short branch lengths in these clusters indicate that population structure plays a strong role in gene sharing at least for this location (see Fullmer et al., 2014 for in depth discussion). Increased geographical sampling could reveal similar trends in other locations.

## INTEIN HOMING IN THE HALOBACTERIA

The existence of a singleton in an intein allele in the genomes analyzed could represent intein invasion from outside the Halobacteria; but could also be due to incomplete sampling. To



**FIGURE 3 | Clustering of Halobacteria based on intein sequences and distribution.** Halobacteria were clustered based on intein sequences and the distribution in each genome. Clusters with posterior probability >95% are shaded purple.

investigate the phylogenetic distance of invasion events responsible for the observed distribution of inteins, the halobacterial inteins were used as queries to search for homologous sequences in the non-redundant database (Altschul et al., 1990). Intein sequences that matched the alleles in the Halobacteria were found in other Euryarchaeota (but not Crenarchaeota), and Bacteria (**Table 2**). To ascertain whether homing occurred between the Halobacteria and organisms outside the Halobacteria, a maximum likelihood tree was built for each intein allele. The

**Table 2 | Taxonomic distribution in each intein allele.**

Intein allele	Tree topology	Halobacteria	Bacteria	Other Euryarchaeota
<i>cdc21-a</i>	Monophyletic	55	4	16
<i>cdc21-c</i>	Monophyletic	1	0	0
<i>dtd**</i>	Monophyletic	6	0	0
<i>gyrB</i>	Monophyletic	6	19	1
<i>helicase-b*</i>	Monophyletic	1	2	1
<i>ligase**</i>	Monophyletic	1	0	0
<i>pol-IIb*</i>	Monophyletic	9	0	1
<i>polB-d</i>	Monophyletic	6	0	1
<i>rfc-a</i>	Monophyletic	16	0	13
<i>rfc-d*</i>	Monophyletic	5	0	0
<i>rir1-b</i>	Monophyletic	15	55	5
<i>rir1-g</i>	Monophyletic	4	15	0
<i>rir1-k</i>	Monophyletic	5	1	0
<i>rir1-l*</i>	Monophyletic	3	3	0
<i>rpolA</i>	Monophyletic	10	0	0
<i>top6B</i>	Monophyletic	8	0	0
<i>topA</i>	Monophyletic	4	0	1
<i>udp</i>	Monophyletic	7	2	6
<i>rir1-m*</i>	Monophyletic	1	4	0
<i>polB-c</i>	Monophyletic	20	1	1
<i>polB-a</i>	Polyphyletic-bacteria	16	2	1
<i>polB-b</i>	Polyphyletic-bacteria	38	3	0
<i>pol-IIa</i>	Polyphyletic-Euryarchaeota	75	0	16
<i>cdc21-b</i>	Polyphyletic-Euryarchaeota	51	1	3

\*Denotes intein alleles discovered in this work.

\*\*Denotes exteins discovered in this work.

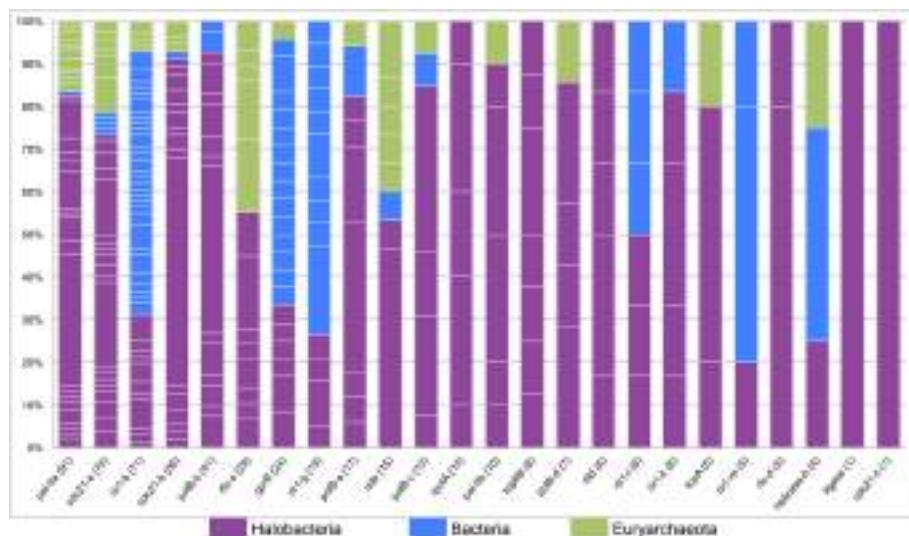
tree topologies were evaluated with respect to the halobacterial inteins. If the halobacterial inteins in the tree were monophyletic it was assumed that except for the initial invasion gene flow for that intein allele occurred within the Halobacteria exclusively. If the halobacterial inteins were polyphyletic, invasion events that generated the observed distribution likely involved organisms outside the Halobacteria either as donors or as recipients. The majority of intein trees, 83%, were monophyletic, reinforcing the idea that recombination is more successful between closely related organisms (Gogarten et al., 2002; Zhaxybayeva et al., 2006; Andam et al., 2010; Papke and Gogarten, 2012; Williams et al., 2012). Interestingly, for trees where the Halobacteria were polyphyletic, the organisms interrupting the clade were Bacteria for two out of the four polyphyletic intein alleles. The sample size restricts building strong claims about HGT between the Halobacteria and the Bacteria. However, this claim is supported by previous evidence of gene exchange between the Bacteria and the Halobacteria (Ng et al., 2000; Khomyakova et al., 2011).

The tight clustering of halobacterial intein sequences and short branches between closely related strains indicate that in the majority cases inteins are inherited vertically or are transferred

between closely related strains, and that successful invasion across large genetic distances is rare. Thus, intein alleles that are found in many different genera have been active for many generations, enabling invasion of many lineages, and accumulating examples of rare invasion events such as those that cross domain boundaries. Conversely, a lack of taxonomic diversity cannot be interpreted as a recent invasion as sampling limitations could be responsible for the paucity of samples in that intein allele. While many factors influence the success of intein transfer between divergent organisms, phylogenetic diversity of the organisms invaded by a particular intein allele also is a reflection of the time the intein allele has been present in a lineage. Furthermore, a high density of intein sequences in a particular domain or group of genera can be used to determine the most likely reservoir for the circulating intein allele. A stacked column chart was used to quantify the representation of each of the genera in each of the intein alleles (Figure 4). Five intein alleles, *cdc21b*, *pol-IIa*, *polBb*, *cdc21a*, and *rfc-d*, show polarity in intein density favoring the Halobacteria (specifically *Halorubrum*) as the reservoir for the intein population. This is not surprising as the data indicate that the majority of intein transfer in the Halobacteria is within the class. Additionally, the diversity in five of the intein alleles, *helicase-b*, *cdc21a*, *gyrB*, *rir1-b*, and *udp*, suggests these intein populations may be more ancient than the others in this study as they have had time to accumulate rare, long distance transfers such that the diversity within them spans both class and domain boundaries. Interestingly, the *helicase-b* intein was only recently discovered in this study, though the diversity in the allele gives the impression that this intein has been around for a long time.

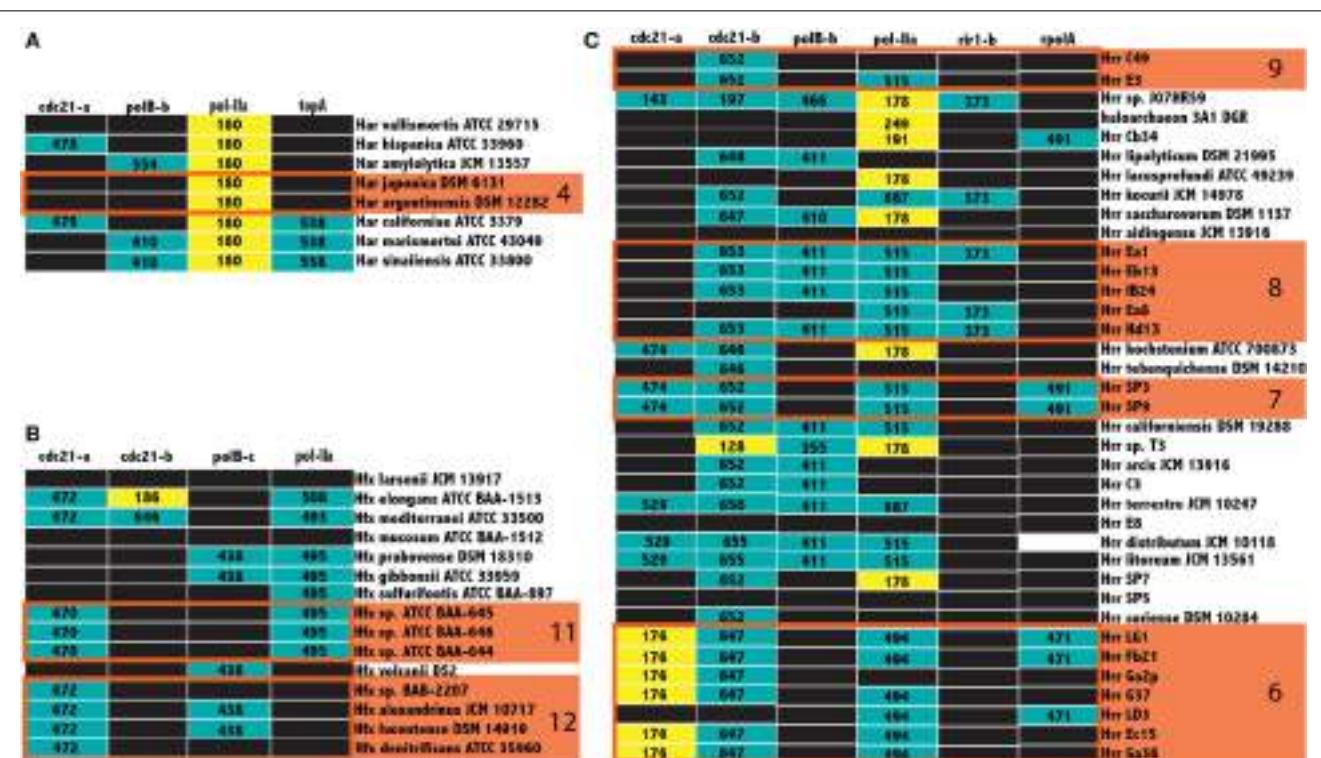
## TRANSFER OF INTEINS BETWEEN HALOBACTERIAL AND NON-HALOBACTERIAL LINEAGES

Not all inteins are transferred equally; the efficiency of intein invasion is affected largely by the state of the intein. The HEN domain in canonical inteins is required to induce a double strand break and the subsequent homologous repair that results in invasion (Pietrovovski, 2001). Thus, mini-inteins that have lost a functioning HEN domain are mainly transferred vertically (they may be transferred horizontally together with the host gene). If an intein containing allele has been fixed in a population, either a precise deletion of the mini intein encoding DNA could remove the intein from the population or homologous replacement by an intein-free allele transferred from outside the population. Thus, mini-inteins are maintained through strong purifying selection, because any mutation that decreases the self-splicing activity decreases the availability of the host protein (Barzel et al., 2011). The intein states were determined to infer patterns of homing in the Halobacteria. The size of inteins in each allele, along with the position of gaps in the alignment relative to the HEN domain were used as a heuristic for assigning mini-intein status. In most cases there was a clear separation in the distribution of intein lengths (at least 100 amino acids difference in length). The size of more populated intein alleles within the three genera of the Halobacteria with the largest number of available genomes, *Haloarcula*, *Halofervax*, and *Halorubrum*, were recorded in a matrix of intein alleles (Figure 5). Many intein alleles show



**FIGURE 4 | Phylogenetic diversity in halobacterial intein alleles.** A stacked column graph depicts the representation of the Halobacteria (in purple), the Bacteria (in blue), and other Euryarchaeota (in green). Intein alleles are ordered by the number of intein sequences recovered for each allele, which is reported in parenthesis after the

intein allele name on the x-axis. The number of genera for each intein allele is indicated by the number of breaks in the column (white lines) and the height of each of the fragments that make up a column indicate the proportion of sequences in that allele found in a particular genus.



**FIGURE 5 | Intein size distributions in the *Haloferax*, *Haloarcula*, and *Halorubrum*.** The size of inteins in the *Haloarcula* (A), *Haloferax* (B), and *Halorubrum* (C) are indicated in the column corresponding to the intein allele. Mini-inteins are colored yellow, large inteins are colored teal, black boxes indicate no intein, and white boxes indicate

missing data, clusters from **Figure 3** are indicated by numbered orange boxes. The *cdc21a*, and *b* sequences for *Halorubrum* sp. J07HR59, though smaller than the rest, cannot be considered mini-inteins, as the intein sequences in these positions are not complete.

a considerable size variation. This variability can be attributed to the accumulation of insertions and deletions in various lineages over time, which in some lineages leads to loss of the HEN domain. Notably, there is no variability in the size of intein sequences shared by the clusters recovered in the Bayesian analysis (orange boxes **Figure 5**) reinforcing the claim of ongoing gene exchange in these clusters.

Invasion from outside the Halobacteria is one explanation for the polyphyletic topology observed in some halobacterial intein alleles. To determine when these homing events could have occurred, the state of each intein was determined and mapped onto polyphyletic intein allele trees: the results of that analysis are summarized in **Table 3**, with mini-inteins indicated with a star (\*), and inteins that group within the Halobacteria indicated by a tilde (~) next to the name of the organism. Many of the intein sequences (5 out of 11) from taxa outside the Halobacteria that interrupt the clade are large-inteins, indicating that interactions between these taxa and the Halobacteria, though rare are ongoing (**Table 3**). Though the assignment of direction of transfers is extremely preliminary as limited sampling can affect the assignment of direction of transfer, there are some cases with an overwhelming signal where the majority of sequences originate from the Halobacteria, or the Bacteria in the case of *rir1-m*. The mixture of mini and large inteins represented in all of the intein alleles imply most of these inteins are active in the Halobacteria, and notably involve a wide distribution of taxonomic exchange partners.

## DISCUSSION

The importance of HGT throughout the tree of life demands the development of a system to monitor gene-flow within and between populations. This research provides fundamental evidence that mobile elements such as inteins can be used to uncover gene flow networks. Inteins have a unique combination of traits that make them ideal tools to study evolution in microbial populations. They have a naturally wide phylogenetic distribution, enabling detection of HGT between distantly related taxa. This is demonstrated in this work by the intein trees where the Halobacteria were polyphyletic (*pol-IIa*, *polB-a*, *polB-b*, and *cdc21b*) indicating intein transfer between the Halobacteria and the taxa that interrupt them, as well as by data from other studies where intein transfer has been detected across phyla and domains (Butler et al., 2006; Swithers et al., 2013). Inteins also have a high substitution rate relative to their extein hosts, and a propensity for accumulating insertions and deletions, which makes detection of transfers between close relatives (generally a difficult task) possible; for example, transfer within the *Halorubrum* clusters shown in **Figure 3**. Inteins can be associated with a HEN domain. If they are, they possess the ability to invade intein-free alleles following transfer; if they are not, they rely mainly on vertical inheritance together with the host gene, and the occasional transfer of the host gene. One intein allele, *pol-IIa*, is widely distributed in the Halobacteria and there are many examples of mini-intein sequences in this allele. These data suggest that invasion of this allele occurred early in the evolution of the Halobacteria, and that the intein may have been lost in some lineages, but retained as a mini intein in most of the genomes surveyed here. This could

also be true for the *cdc21-a* intein; however, the distribution is not as diverse, and considerably fewer mini-inteins were detected. This is more suggestive of an intein that has been active in the Halobacteria for a long period of time, with the different intein states (empty target site, target site invaded by an intein with active HEN, target site occupied by an intein without functioning HEN; Yahara et al., 2009; Barzel et al., 2011) existing and co-existing in different halobacterial lineages.

The genomes analyzed in this work were cultured from salty water and soil samples around the world. The diverse background of the genomes may contribute to the spotty distribution of intein alleles (**Figure 1**). However, genomes isolated from the same location show variation as well (**Figure 3**) (Fullmer et al., 2014), reinforcing the notion that inteins are currently actively propagating in and being eliminated from halobacterial populations. Additionally, previous data have shown recombination occurs at a higher rate than mutation within the Halobacteria, and very little linkage between genes is detected in these genomes (Papke et al., 2004, 2007). These observations indicated gene flow as an important method for niche adaptation in these organisms. In Deep Lake, Antarctica the freezing temperatures limit the rate of replication to approximately 6 times per year and evolution in the halobacterial populations there mainly occurs through gene flow (Demaere et al., 2013). Recent whole genome comparisons revealed frequent gene transfer followed by homologous replacement of the transferred gene within the Halobacteria, hampering attempts to resolve the phylogeny within this group (Williams et al., 2012). Gene flow and recombination between populations and species make it difficult to resolve the species phylogeny among the different genera of Halobacteria (Papke et al., 2004). The use of gene concatenation in building reference trees, as exemplified by the ribosomal protein reference tree used in this work, has been pivotal in determining a branching order for the major clades of organisms, such as the Halobacteria, that participate in a large amount of recombination with close relatives. However, because genetic transfer and homologous recombination occur frequently between close relatives, the resulting phylogeny reflects both, shared ancestry and frequency of gene transfer. Therefore, determining the network of gene flow that overlays the vertical signal is important to the understanding of the evolution of these organisms. Inteins cannot penetrate the cell wall, and thus capitalize on existing gene flow in populations to efficiently invade when the opportunity presents itself. This trait can be exploited to keep track of successful homing events revealed by sequence similarity of inteins in distinct strains.

*Halorubrum* was the only genus in this study that had a large enough sample size to begin to uncover a signal reflecting population structure. Many of the *Halorubrum* genomes in this study were isolated from the same location, and this collection of genomes showed a clear signal for a structured population. Sixteen genomes from Aran-Bidgol were separated into four well-supported clusters. Three of the four clusters have branching orders identical to those in the reference tree, and the support values for those clusters could be attributed to both transfer within the group and a background phylogenetic signal or ancestral inheritance of similar intein alleles. However, only cluster 7 in the *Halorubrum* shares all intein alleles between all members of

**Table 3 | Protein sequence identifiers for intein sequences.**

Intein allele	Species name	Accession number	Phylum
<i>cdc21-a</i>	<i>Archaeoglobus profundus</i> DSM 5631	YP_004340760.1	Euryarchaeota
	* <i>Archaeoglobus veneficus</i> SNP6	YP_003400528.1	Euryarchaeota
	* <i>Candidatus Methanomassiliicoccus intestinalis</i> Issoire Mx1	YP_008072558.1	Euryarchaeota
	* <i>Crocosphaera watsonii</i>	WP_021836378.1	Cyanobacteria
	* <i>Ferroglobus placidus</i> DSM 10642	YP_003435419.1	Euryarchaeota
	* <i>Halarchaeum acidiphilum</i>	WP_020220725.1	Halobacteria
	* <i>Lamprocystis purpurea</i>	WP_020504136.1	Gammaproteobacteria
	* <i>Methanomassiliicoccus luminyensis</i>	WP_019178416.1	Euryarchaeota
	* <i>Methanothermococcus okinawensis</i> IH1	YP_004576471.1	Euryarchaeota
	<i>Nocardia asteroides</i> NBRC 15531	GAD83132.1	Actinobacteria
	<i>Nocardiopsis potens</i>	WP_020380316.1	Actinobacteria
	<i>Pyrococcus abyssi</i> GE5	NP_127115.1	Euryarchaeota
	* <i>Pyrococcus furiosus</i> DSM 3638	NP_578211.1	Euryarchaeota
	* <i>Pyrococcus horikoshii</i> OT3	NP_142122.1	Euryarchaeota
	* <i>Pyrococcus</i> sp. NA2	YP_004424138.1	Euryarchaeota
	<i>Thermococcus litoralis</i> DSM 5473	YP_008429717.1	Euryarchaeota
	* <i>Thermococcus onnurineus</i> NA1	YP_002306424.1	Euryarchaeota
	* <i>Thermococcus sibiricus</i> MM 739	YP_002994932.1	Euryarchaeota
	* <i>Thermococcus</i> sp. AM4	YP_002582218.1	Euryarchaeota
	* <i>Thermococcus</i> sp. CL1	YP_006424652.1	Euryarchaeota
	* <i>Thermococcus zilligii</i>	WP_010479121.1	Euryarchaeota
	<i>Halorubrum</i> sp. SP3	KJ_865687.1	Halobacteria
	<i>Halorubrum</i> sp. SP9	KJ_865689.1	Halobacteria
<i>cdc21-b</i>	* <i>Cyanothecce</i> sp. PCC 7822	YP_003887897.1	Cyanobacteria
	<i>Halarchaeum acidiphilum</i>	WP_020220725.1	Halobacteria
	* <i>Candidatus Methanomassiliicoccus intestinalis</i> Issoire-Mx1	YP_008072558.1	Euryarchaeota
	* <i>Methanomassiliicoccus luminyensis</i>	WP_019178416.1	Euryarchaeota
	~ <i>Thermococcus barophilus</i>	YP_004070279.1	Euryarchaeota
	<i>Halorubrum</i> sp. SP3	KJ_865687.1	Halobacteria
	<i>Halorubrum</i> sp. SP7	KJ_865688.1	Halobacteria
<i>polB-d</i>	<i>Archaeoglobus profundus</i> DSM 5631	YP_003400528.1	Euryarchaeota
<i>polB-a</i>	~ <i>Salinibacter ruber</i> M8	YP_003572085.1	Bacteroidetes
	~ <i>Salinibacter ruber</i> DSM 13885	YP_446104.1	Bacteroidetes
	~ <i>Halarchaeum acidiphilum</i>	WP_020678478.1	Halobacteria
	~ <i>Methanoculleus bourgensis</i>	YP_006544623.1	Euryarchaeota
<i>polB-b</i>	<i>Halosimplex carlsbadense</i>	WP_006885382.1	Halobacteria
	* ~ <i>Salinibacter ruber</i> M8	YP_003572085.1	Bacteroidetes
	* ~ <i>Salinibacter ruber</i> DSM 13885	YP_446104.1	Bacteroidetes
	* ~ <i>Halanaerobium saccharolyticum</i>	WP_005489097.1	Firmicutes
	<i>Halarchaeum acidiphilum</i>	WP_020678478.1	Halobacteria
<i>polB-c</i>	* ~ <i>Thermus scotoductus</i>	YP_004202875.1	Deinococcus-Thermus
	* ~ <i>Methanotorris igneus</i> Kol 5	YP_004483799.1	Euryarchaeota
	<i>Halorubrum</i> sp. SP7	KJ_865686.1	Halobacteria
<i>polB-a</i>	<i>Archaeoglobus veneficus</i> SNP6	YP_004341738.1	Euryarchaeota
	<i>Halosimplex carlsbadense</i>	WP_006882195.1	Halobacteria
	* <i>Methanocaldococcus infernus</i> ME	YP_003616947.1	Euryarchaeota
	<i>Methanococcus aeolicus</i>	ABU41683.1	Euryarchaeota

(Continued)

**Table 3 | Continued**

Intein allele	Species name	Accession number	Phylum
	* <i>Methanoculleus bourgensis</i> MS2	YP_006544019.1	Euryarchaeota
	* <i>Methanoculleus marisnigri</i> JR-1	YP_001048029.1	Euryarchaeota
	<i>Methanofollis liminatans</i>	WP_004037227.1	Euryarchaeota
	* <i>Methanolinea tarda</i>	WP_007314808.1	Euryarchaeota
	* <i>Methanoplanus limicola</i>	WP_004076782.1	Euryarchaeota
	* <i>Methanoplanus petrolearius</i> DSM 11571	YP_003893638.1	Euryarchaeota
	<i>Methanoregula boonei</i> 6A8	YP_001403293.1	Euryarchaeota
	~ <i>Methanoregula fomicica</i> SMSP	YP_007242862.1	Euryarchaeota
	<i>Methanospaerula palustris</i> E1-9c	YP_002467270.1	Euryarchaeota
	* <i>Methanospirillum hungatei</i> JF-1	YP_503855.1	Euryarchaeota
	* <i>Pyrococcus horikoshii</i> OT3	NP_142130.1	Euryarchaeota
	* <i>Thermococcus gammatolerans</i> EJ3	YP_002958492.1	Euryarchaeota
	* <i>Thermococcus sibiricus</i> MM 739	YP_002994988.1	Euryarchaeota
	uncultured haloarchaeon	ABQ75865.1	Halobacteria
	<i>Halorubrum</i> sp. SP3	KJ_865692.1	Halobacteria
	<i>Halorubrum</i> sp. SP7	KJ_865690.1	Halobacteria
	<i>Halorubrum</i> sp. SP9	KJ_564691.1	Halobacteria
<i>pol-IIb</i>	<i>Halosimplex carlsbadense</i>	WP_006882195.1	Halobacteria
	* <i>Pyrococcus abyssi</i> GE5	YP_004624494.1	Euryarchaeota
	uncultured haloarchaeon	ABQ75865.1	Halobacteria
<i>gyrB</i>	<i>Allochromatium vinosum</i> DSM 180	YP_003443943.1	Gammaproteobacteria
	<i>Anabaena</i> sp. 90	YP_006997726	Cyanobacteria
	* <i>Anabaena</i> sp. PCC 7108	WP_016950132.1	Cyanobacteria
	<i>Bacillus subtilis</i> BEST7613	BAM51471.1	Firmicutes
	<i>Calothrix</i> sp. PCC 7103	WP_019489451.1	Cyanobacteria
	<i>Coleofasciculus chthonoplastes</i>	WP_006099284.1	Cyanobacteria
	* <i>Cylindrospermopsis reciborskii</i>	WP_006276716.1	Cyanobacteria
	* <i>Dactylococcopsis slamae</i> PCC 8305	YP_007173052.1	Cyanobacteria
	<i>Halarchaeum acidiphilum</i>	WP_021780646.1	Halobacteria
	<i>Methanomassiliicoccus luminyensis</i>	WP_019178436.1	Euryarchaeota
	<i>Microcystis aeruginosa</i>	WP_002774451.1	Cyanobacteria
	<i>Moorea producens</i>	WP_008190351.1	Cyanobacteria
	<i>Oscillatoria</i> sp. PCC 10802	WP_017715151.1	Cyanobacteria
	<i>Pleurocapsa</i> sp. PCC 7319	WP_019509077.1	Cyanobacteria
	<i>Prochlorothrix hollandica</i>	WP_017710941.1	Cyanobacteria
	<i>Raphidiopsis brookii</i>	WP_009342634.1	Cyanobacteria
	<i>Rivularia</i> sp. PCC 7116	YP_007054134.1	Cyanobacteria
	<i>Saccharothrix espanaensis</i> DSM 44229	YP_007037469.1	Actinobacteria
	<i>Synechocystis</i> sp. PCC 6803	NP_441040.1	Cyanobacteria
	<i>Trichodesmium erythraeum</i> IMS101	YP_723459.1	Cyanobacteria
	uncultured bacterium	EKD46222.1	
<i>helicase-b</i>	* <i>Bacillus amyloliquefaciens</i> TA208	YP_005540906.1	Firmicutes
	* <i>Bacillus subtilis</i>	WP_017696872.1	Firmicutes
	Nanoarchaeota archaeon SCGC AAA011-L22	WP_018204386.1	
<i>rfc-a</i>	<i>Methanocaldococcus jannaschii</i> DSM 2661	NP_248426.1	Euryarchaeota
	<i>Methanocaldococcus</i> sp. FS406	YP_003458055.1	Euryarchaeota
	<i>Methanothermococcus okinawensis</i> IH1	YP_004576337.1	Euryarchaeota
	* <i>Methanotorris formicicola</i>	WP_007044297.1	Euryarchaeota
	* <i>Pyrococcus abyssi</i> GE5	NP_125803.1	Euryarchaeota

(Continued)

**Table 3 | Continued**

Intein allele	Species name	Accession number	Phylum
	* <i>Pyrococcus furiosus</i> DSM 3638	NP_577822.1	Euryarchaeota
	* <i>Pyrococcus horikoshii</i> OT3	NP_142122.1	Euryarchaeota
	* <i>Pyrococcus</i> sp. ST04	YP_006353924.1	Euryarchaeota
	* <i>Thermococcus kodakorensis</i> KOD1	YP_184631.1	Euryarchaeota
	* <i>Thermococcus litoralis</i> DSM 5473	YP_008428897.1	Euryarchaeota
	<i>Thermococcus</i> sp. 4557	YP_004763272.1	Euryarchaeota
	<i>Thermococcus</i> sp. AM4	YP_002582171.1	Euryarchaeota
	* <i>Thermococcus</i> sp. CL1	YP_006425306.1	Euryarchaeota
<i>rpolA</i>	<i>Halorubrum</i> sp. SP3	KJ_865684.1	Halobacteria
	<i>Halorubrum</i> sp. SP9	KJ_865685.1	Halobacteria
<i>rir1-l</i>	<i>Chloroherpeton thalassium</i> ATCC 35110	YP_001995975.1	Chlorobi
	<i>Tepidanaerobacter acetatoxydans</i> Re1	YP_007273179.1	Firmicutes
	uncultured Chloroflexi bacterium	BAL53207.1	Chloroflexi
<i>rir1-k</i>	<i>Deinococcus peraridilitoris</i> DSM 19664	YP_007181218.1	Deinococcus-Thermus
<i>rir1-b</i>	<i>Acidovorax avenae</i> subsp. <i>avenae</i> ATCC 19860	YP_004233126.1	Betaproteobacteria
	<i>Acidovorax</i> sp. CF316	WP_007856012.1	Betaproteobacteria
	<i>Acidovorax</i> sp. NO-1	WP_008903130.1	Betaproteobacteria
	<i>Actinomadura atramentaria</i>	WP_019631066.1	Actinobacteria
	<i>Alicyclobacillus pohliae</i>	WP_018131875.1	Firmicutes
	<i>Aminomonas paucivorans</i>	WP_006300529.1	Synergistetes
	<i>Ammonifex degensii</i> KC4	WP_006300529.1	Firmicutes
	<i>Arhodomonas aquaeolei</i>	WP_018718131.1	Gammaproteobacteria
	<i>Bacillus licheniformis</i>	WP_016885361.1	Firmicutes
	<i>Bacillus subtilis</i>	WP_017697104.1	Firmicutes
	<i>Calothrix</i> sp. PCC 6303	YP_007136749.1	Cyanobacteria
	<i>Candidatus Chloracidobacterium thermophilum</i> B	YP_004863563.1	Acidobacteria
	<i>Candidatus Desulforudis audaxviator</i> MP104C	YP_001717412.1	Firmicutes
	<i>Clostridiaceae bacterium</i> L21-TH-D2	WP_006314960.1	Firmicutes
	<i>Deinococcus radiodurans</i> R1	NP_296095.1	Deinococcus-Thermus
	<i>Delftia acidovorans</i>	WP_016451949.1	Betaproteobacteria
	<i>Delftia</i> sp. Cs1-4	YP_004490724.1	Betaproteobacteria
	<i>Desulfitobacterium hafniense</i>	WP_005810476.1	Firmicutes
	<i>Desulfovibrio magneticus</i> RS-1	YP_002955841.1	Deltaproteobacteria
	<i>Desulfovibrio</i> sp. U5L	WP_009106508.1	Deltaproteobacteria
	<i>Ferroplasma acidarmanus</i> fer1	YP_008141532.1	Euryarchaeota
	<i>Ferroplasma</i> sp. Type II	WP_021787573.1	Euryarchaeota
	<i>Halomonas antarctica</i>	WP_016418429.1	Gammaproteobacteria
	<i>Halomonas jeotgali</i>	WP_017429019.1	Gammaproteobacteria
	<i>Halomonas smyrnensis</i>	WP_016854101.1	Gammaproteobacteria
	<i>Mahella australiensis</i> 50-1 BON	YP_004462974.1	Firmicutes
	<i>Marinobacter lipolyticus</i>	WP_018405479.1	Gammaproteobacteria
	<i>Methanofollis liminatans</i>	WP_004040239.1	Euryarchaeota
	<i>Methylobacter marinus</i>	WP_020160338.1	Gammaproteobacteria
	<i>Methylococcus capsulatus</i>	WP_017366201.1	Gammaproteobacteria
	<i>Methylomicrobium buryatense</i>	WP_017841702.1	Gammaproteobacteria
	nanoarchaeote Nst1	WP_004578017.1	
	<i>Nocardiopsis halotolerans</i>	WP_017572347.1	Actinobacteria
	<i>Polaromonas</i> sp. JS666	CAJ57177.1	Cyanobacteria
	<i>Pseudanabaena</i> sp. PCC 6802	WP_019499030.1	Cyanobacteria

(Continued)

**Table 3 | Continued**

Intein allele	Species name	Accession number	Phylum
	<i>Pseudanabaena</i> sp. PCC 7367	YP_007101092.1	Cyanobacteria
	<i>Rhodanobacter fulvus</i>	WP_007082010.1	Gammaproteobacteria
	<i>Rhodanobacter</i> sp. 2APBS1	YP_007588821.1	Gammaproteobacteria
	<i>Rhodanobacter thiooxydans</i>	WP_008437232.1	Gammaproteobacteria
	<i>Rhodothermus marinus</i> SG0.5JP17-172	YP_004824118.1	Bacteroidetes
	<i>Staphylococcus aureus</i>	WP_016187732.1	Firmicutes
	<i>Synechococcus elongatus</i> PCC 6301	CAJ57178.1	Cyanobacteria
	<i>Synechococcus elongatus</i> PCC 7942	YP_400626.1	Cyanobacteria
	<i>Synechococcus</i> sp. PCC 6312	YP_007060778.1	Cyanobacteria
	<i>Thermoanaerobacterium saccharolyticum</i> JW/SL-YS485	YP_006391581.1	Firmicutes
	<i>Thermoanaerobacterium thermosaccharolyticum</i> DSM 571	YP_003851043.1	Firmicutes
	<i>Thermobrachium celere</i>	WP_018663796.1	Firmicutes
	<i>Thermococcus kodakarensis</i> KOD1	YP_184312.1	Euryarchaeota
	<i>Thermodesulfatator indicus</i> DSM 15286	YP_004625205.1	Thermodesulfobacteria
	<i>Thermovirga lienii</i> DSM 17291	YP_004932130.1	Deinococcus-Thermus
	<i>Thermus igniterrae</i>	WP_018110436.1	Deinococcus-Thermus
	<i>Thermus thermophilus</i> HB8	CAJ57170.1	Deinococcus-Thermus
	<i>Thioalkalivibrio</i> sp. ALE11	WP_019570879.1	Gammaproteobacteria
	<i>Thioalkalivibrio</i> sp. ALE30	WP_018881426.1	Gammaproteobacteria
	<i>Thioalkalivibrio</i> sp. HL-Eb18	WP_017926201.1	Gammaproteobacteria
	<i>Thioalkalivibrio</i> sp. K90mix	YP_003459507.1	Gammaproteobacteria
	uncultured bacterium	EKE25755.1	
	<i>Xanthomonas</i> sp. SHU199	WP_017907463.1	Gammaproteobacteria
	<i>Xanthomonas</i> sp. SHU308	WP_017915139.1	Gammaproteobacteria
	<i>zeta proteobacterium</i> SCGC AB-604-B04	WP_018280466.1	Zetaproteobacteria
<i>rir1-g</i>	<i>Chloroherpeton thalassium</i> ATCC 35110	YP_001995975.1	Chlorobi
<i>rir1-g</i>	<i>Deinococcus aquatillis</i>	WP_019011777.1	Deinococcus-Thermus
<i>rir1-g</i>	<i>Halotheca</i> sp. PCC 7418	YP_007166732.1	Cyanobacteria
<i>rir1-g</i>	<i>Klebsiella pneumoniae</i>	WP_021313783.1	Gammaproteobacteria
<i>rir1-g</i>	<i>Nocardiopsis dassonvillei</i> subsp. <i>Dassonvillei</i> DSM 43111	YP_003681238.1	Actinobacteria
<i>rir1-g</i>	<i>Nocardiopsis</i> sp. CNS639	WP_019609645.1	Actinobacteria
<i>rir1-g</i>	<i>Rhodothermus marinus</i> SG0.5JP17-172	YP_004826277.1	Bacteroidetes
<i>rir1-g</i>	<i>Tepidanaerobacter acetatoxydans</i> Re1	YP_007273179.1	Firmicutes
<i>rir1-g</i>	<i>Thermomonospora curvata</i> DSM 43183	YP_003299200.1	Actinobacteria
<i>rir1-g</i>	<i>Thermus thermophilus</i> HB27	YP_005899.1	Deinococcus-Thermus
<i>rir1-g</i>	<i>Thermus thermophilus</i> HB8	CAJ57173.1	Deinococcus-Thermus
<i>rir1-g</i>	<i>Thermus thermophilus</i> JL-18	YP_006059430.1	Deinococcus-Thermus
<i>rir1-g</i>	<i>Thermus thermophilus</i> SG0.5JP17-16	YP_005639869.1	Deinococcus-Thermus
<i>rir1-g</i>	<i>Trichodesmium erythraeum</i> IMS101	YP_720358.1	Cyanobacteria
<i>rir1-g</i>	uncultured Chloroflexi bacterium	BAL53207.1	Chloroflexi
<i>rir1-m</i>	<i>Thermus aquaticus</i>	WP_003044118.1	Deinococcus-Thermus
<i>rir1-m</i>	<i>Thermus thermophilus</i> HB-8	CAJ57173.1	Deinococcus-Thermus
<i>rir1-m</i>	<i>Thermus thermophilus</i> SG0.5JP17-16	YP_005639869.1	Deinococcus-Thermus
<i>rir1-m</i>	uncultured Chloroflexi bacterium	BAL53207.1	Chloroflexi
<i>udp</i>	<i>Fervidibacteria bacterium</i> JGI 0000001-G10	WP_020250137.1	
<i>udp</i>	<i>Dictyglomus thermophilum</i> H-6-12	YP_002250310.1	Dictyglomi
<i>udp</i>	<i>Methanocaldococcus jannaschii</i> DSM 2661	NP_248048.1	Euryarchaeota
<i>udp</i>	<i>Methanocaldococcus vulcanis</i> M7	YP_003246412.1	Euryarchaeota
<i>udp</i>	<i>Methanococcus aeolicus</i> Nankai-3	YP_001324612.1	Euryarchaeota
<i>udp</i>	<i>Methanothermococcus okinawensis</i> IH1	YP_004575831.1	Euryarchaeota

(Continued)

**Table 3 | Continued**

Intein allele	Species name	Accession number	Phylum
	<i>Methanotorris igneus</i> Kol 5	WP_007044255.1	Euryarchaeota
	<i>Thermococcus gammatolerans</i> EJ3	YP_002960518.1	Euryarchaeota
<i>topA</i>	<i>Methanotorris igneus</i> Kol 5	WP_007044255.1	Euryarchaeota
<i>top6B</i>	<i>Halarchaeum acidiphilum</i>	WP_021780130.1	Halobacterium

\*Indicates the intein detected is a mini-intein.

~Indicates taxa that grouped within the halobacterial intein sequences.

the cluster while the other clusters all contain intein alleles that are unique to certain members of the cluster, suggesting ongoing transfer of these inteins within the population. Additionally, three out of the twelve total clusters demonstrate unique branching orders compared to the reference tree, though only five of the clusters reflected in the reference tree have identical intein profiles. The lack of fixation for the intein alleles in the majority of clusters (seven out of twelve) indicates that a signal due to vertical inheritance may aid the formation of the clusters, but that HGT and its bias is the driving force for intein distribution. This analysis demonstrates the utility of intein sequences in distinguishing a population structure amongst genomes isolated from the same location, as demonstrated with the genomes isolated from Aran-Bidgol. These relationships are made evident through analyzing all of the signals from each of the intein alleles represented in the strains, and thus represent a collapsed view of the major gene sharing networks that have shaped the intein profiles of these strains over time. The collapsed networks indicate a higher rate of recombination within compared to between species and groups, a finding similar to the sexual outcrossing in fungal populations where inteins also thrive, as the semi-sexual lifestyle promotes intein homing (Giraldo-Perez and Goddard, 2013).

It is tempting to speculate that strains that harbor an abundance of intein alleles partake in more gene transfer than their counterparts without as many inteins; however, these two phenomena should not be expected to have a strict correlation as HGT between strains that possess only one intein each cannot produce hybrids with more than two inteins each. The number of inteins present in a group of different strains and species may be more reflective of transfers with divergent organisms than within-group transfer frequency.

The presented research demonstrates the utility of intein sequences to follow gene flow within and between populations. Improved reliability to assess the presence and activity of the HEN domain intein will provide a better distinction between vertical and horizontal inheritance of inteins. The overall utility of inteins improves as new intein alleles and new host proteins are reported, increasing the distribution of samples and improving statistical robustness of studies like the one done here. Prior to this work, nine proteins had been reported to contain inteins in the Halobacteria. This work established seven new intein alleles in the Halobacteria, including two proteins not previously reported to contain inteins. The presence of inteins is especially useful in populations where high rates of recombination and widely

distributed populations may facilitate the maintenance of intein sequences over long periods of time (Gogarten and Hilario, 2006) and provide a means for distinguishing closely related partners involved in genetic transfers. The phylogenetic distribution of intein alleles, combined with the changing state within intein alleles, and the rapid substitution rate of inteins relative to the extein host sequences (Swithers et al., 2013) will provide a valuable tool to infer gene flow dynamics in and between sampled populations.

## AUTHOR CONTRIBUTIONS

Johann Peter Gogarten and Shannon M. Soucy participated in the design of this study and helped to draft the manuscript. Shannon M. Soucy performed the research and all authors contributed to data analysis. All authors read and approved the final manuscript.

## ACKNOWLEDGMENTS

The UConn Bioinformatics Facility provided computing resources for the analyses reported in this manuscript. The *Halorubrum* genomes provided by the Papke lab were sequenced in house by Andrea Makkay and Ryan Wheeler. We would like to thank them for their hard work, as well as acknowledge Dr. Elina Roine and Dennis Bamford (Helsinki University), and Dr. Antonio Ventosa (University of Sevilla) for supplying the sequenced strains. We would also like to recognize labs sequencing genomes and making them available in data repositories such as those hosted by the National Center for Biotechnology Information. This work was supported by the National Science Foundation Grant (DEB 0830024 and DEB0919290) and NASA Astrobiology: Exobiology and Evolutionary Biology Grants (NNX12AD70G and NNX13AI03G).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00299/abstract>

## REFERENCES

- Altschul, S. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Andam, C. P., Williams, D., and Gogarten, J. P. (2010). Biased gene transfer mimics patterns created through shared ancestry. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10679–10684. doi: 10.1073/pnas.1001418107

- Atanasova, N. S., Roine, E., Oren, A., Bamford, D. H., and Oksanen, H. M. (2012). Global network of specific virus-host interactions in hypersaline environments. *Environ. Microbiol.* 14, 426–440. doi: 10.1111/j.1462-2920.2011.02603.x
- Barzel, A., Obolski, U., Gogarten, J. P., Kupiec, M., and Hadany, L. (2011). Home and away—the evolutionary dynamics of homing endonucleases. *BMC Evol. Biol.* 11:324. doi: 10.1186/1471-2148-11-324
- Butler, M. I., Gray, J., Goodwin, T. J., and Poulter, R. T. (2006). The distribution and evolutionary history of the PRP8 intein. *BMC Evol. Biol.* 6:42. doi: 10.1186/1471-2148-6-42
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27, 1164–1165. doi: 10.1093/bioinformatics/btr088
- Demaere, M. Z., Williams, T. J., Allen, M. A., Brown, M. V., Gibson, J. A. E., Rich, J., et al. (2013). High level of intergenera gene exchange shapes the evolution of haloarchaea in an isolated Antarctic lake. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16939–16944. doi: 10.1073/pnas.1307090110
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Fullmer, M. S., Soucy, S. M., Swithers, K. S., Makkay, A. M., Wheeler, R., Ventosa, A., et al. (2014). Population and genomic analysis of the genus Halorubrum. *Front. Microbiol.* 5:140. doi: 10.3389/fmicb.2014.00140
- Gimble, F. S., and Thorner, J. (1992). Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*. *Nature* 357, 301–306. doi: 10.1038/357301a0
- Giraldo-Perez, P., and Goddard, M. R. (2013). A parasitic selfish gene that affects host promiscuity. *Proc. Biol. Sci.* 280:20131875. doi: 10.1098/rspb.2013.1875
- Goddard, M. R., and Burt, A. (1999). Recurrent invasion and extinction of a selfish gene. *Proc. Natl. Acad. Sci.* 96, 13880–13885. doi: 10.1073/pnas.96.24.13880
- Gogarten, J. P., and Hilario, E. (2006). Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evol. Biol.* 6:94. doi: 10.1186/1471-2148-6-94
- Gogarten, J. P., Senejani, A. G., Zhaxybayeva, O., Olendzenski, L., and Hilario, E. (2002). Inteins: structure, function, and evolution. *Annu. Rev. Microbiol.* 56, 263–287. doi: 10.1146/annurev.micro.56.012302.160741
- Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224. doi: 10.1093/molbev/msp259
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Hirata, R., Ohsumik, Y., Nakano, A., Kawasaki, H., Suzuki, K., and Anraku, Y. (1990). Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *J. Biol. Chem.* 265, 6726–6733.
- Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebel, M., and Stevens, T. H. (1990). Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* 250, 651–657. doi: 10.1126/science.2146742
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Khomyakova, M., Bükmmez, Ö., Thomas, L. K., Erb, T. J., and Berg, I. A. (2011). A methylaspartate cycle in haloarchaea. *Science* 331, 334–337. doi: 10.1126/science.1196544
- Lang, A. S., Zhaxybayeva, O., and Beatty, J. T. (2012). Gene transfer agents: phage-like elements of genetic exchange. *Nat. Rev. Microbiol.* 10, 472–482. doi: 10.1038/nrmicro2802
- Liu, K., Warnow, T. J., Holder, M. T., Nelesen, S. M., Yu, J., Stamatakis, A. P., et al. (2012). SATe-II: very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. *Syst. Biol.* 61, 90–106. doi: 10.1093/sysbio/syr095
- Ng, W. V., Kennedy, S. P., Mahairas, G. G., Berquist, B., Pan, M., Shukla, H. D., et al. (2000). Genome sequence of *Halobacterium* species NRC-1. *Proc. Natl. Acad. Sci. U.S.A.* 97, 12176–12181. doi: 10.1073/pnas.190337797
- Papke, R. T., and Gogarten, J. P. (2012). Ecology. How bacterial lineages emerge. *Science* 336, 45–46. doi: 10.1126/science.1219241
- Papke, R. T., Koenig, J. E., Rodriguez-Valera, F., and Doolittle, W. F. (2004). Frequent recombination in a saltern population of halorubrum. *Science* 306, 1928–1929. doi: 10.1126/science.1103289
- Papke, R. T., Zhaxybayeva, O., Feil, E. J., Sommerfeld, K., Muise, D., and Doolittle, W. F. (2007). Searching for species in haloarchaea. *Proc. Natl. Acad. Sci. U.S.A.* 104, 14092–14097. doi: 10.1073/pnas.0706358104
- Perler, F. B. (2002). InBase: the Intein Database. *Nucleic Acids Res.* 30, 383–384. doi: 10.1093/nar/30.1.383
- Perler, F. B., Olsen, G. J., and Adam, E. (1997). Compilation and analysis of intein sequences. *Nucleic Acids Res.* 25, 1087–1093. doi: 10.1093/nar/25.6.1087
- Pietrokovski, S. (2001). Intein spread and extinction in evolution. *Trends Genet.* 17, 465–472. doi: 10.1016/S0168-9525(01)02365-4
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539. doi: 10.1038/msb.2011.75
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Swithers, K. S., Senejani, A. G., Fournier, G. P., and Gogarten, J. P. (2009). Conservation of intron and intein insertion sites: implications for life histories of parasitic genetic elements. *BMC Evol. Biol.* 9:303. doi: 10.1186/1471-2148-9-303
- Swithers, K. S., Soucy, S. M., Lasek-Nesselquist, E., Lapierre, P., and Gogarten, J. P. (2013). Distribution and evolution of the mobile vma-1b intein. *Mol. Biol. Evol.* 30, 2676–2687. doi: 10.1093/molbev/mst164
- Williams, D., Gogarten, J. P., and Papke, R. T. (2012). Quantifying homologous replacement of loci between haloarchaeal species. *Genome Biol. Evol.* 4, 1223–1244. doi: 10.1093/gbe/evs098
- Yahara, K., Fukuyo, M., Sasaki, A., and Kobayashi, I. (2009). Evolutionary maintenance of selfish homing endonuclease genes in the absence of horizontal transfer. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18861–18866. doi: 10.1073/pnas.0909840106
- Zhaxybayeva, O., Gogarten, J. P., Charlebois, R. L., Doolittle, W. F., and Papke, R. T. (2006). Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res.* 16, 1099–1108. doi: 10.1101/gr.5322306

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 January 2014; accepted: 30 May 2014; published online: 26 June 2014.

Citation: Soucy SM, Fullmer MS, Papke RT and Gogarten JP (2014) Inteins as indicators of gene flow in the halobacteria. *Front. Microbiol.* 5:299. doi: 10.3389/fmicb.2014.00299

This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.

Copyright © 2014 Soucy, Fullmer, Papke and Gogarten. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Comparison of prokaryotic community structure from Mediterranean and Atlantic saltern concentrator ponds by a metagenomic approach

Ana B. Fernández<sup>1</sup>, Blanca Vera-Gargallo<sup>1</sup>, Cristina Sánchez-Porro<sup>1</sup>, Rohit Ghai<sup>2</sup>, R. Thane Papke<sup>3</sup>, Francisco Rodríguez-Valera<sup>2</sup> and Antonio Ventosa<sup>1\*</sup>

<sup>1</sup> Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Sevilla, Sevilla, Spain

<sup>2</sup> Evolutionary Genomics Group, Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández, San Juan de Alicante, Alicante, Spain

<sup>3</sup> Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT, USA

**Edited by:**

Jesse Dillon, California State University, Long Beach, USA

**Reviewed by:**

Eric E. Allen, Scripps Institution of Oceanography, USA

Jesse Dillon, California State University, Long Beach, USA

**\*Correspondence:**

Antonio Ventosa, Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Sevilla, Calle Profesor García González 2, 41012 Sevilla, Spain  
e-mail: ventosa@us.es

We analyzed the prokaryotic community structure of a saltern pond with 21% total salts located in Isla Cristina, Huelva, Southwest Spain, close to the Atlantic ocean coast. For this purpose, we constructed a metagenome (designated as IC21) obtained by pyrosequencing consisting of 486 Mb with an average read length of 397 bp and compared it with other metagenomic datasets obtained from ponds with 19, 33, and 37% total salts acquired from Santa Pola marine saltern, located in Alicante, East Spain, on the Mediterranean coast. Although the salinity in IC21 is closer to the pond with 19% total salts from Santa Pola saltern (designated as SS19), IC21 is more similar at higher taxonomic levels to the pond with 33% total salts from Santa Pola saltern (designated as SS33), since both are predominated by the phylum *Euryarchaeota*. However, there are significant differences at lower taxonomic levels where most sequences were related to the genus *Halorubrum* in IC21 and to *Haloquadratum* in SS33. Within the *Bacteroidetes*, the genus *Psychroflexus* is the most abundant in IC21 while *Salinibacter* dominates in SS33. Sequences related to bacteriorhodopsins and halorhodopsins correlate with the abundance of *Haloquadratum* in Santa Pola SS19 to SS33 and of *Halorubrum* in Isla Cristina IC21 dataset, respectively. Differences in composition might be attributed to local ecological conditions since IC21 showed a decrease in the number of sequences related to the synthesis of compatible solutes and in the utilization of phosphonate.

**Keywords:** metagenomics, haloarchaea, halophilic bacteria, saltern, prokaryotic diversity

## INTRODUCTION

Hypersaline habitats are characterized by high salt concentrations, in addition to other features, such as high or low temperatures, high pH, and/or low oxygen concentrations (Javor, 1989; Rodríguez-Valera, 1993). Hypersaline environments are often aquatic systems (thalassohaline, of marine origin, or athalassohaline, formed by dissolution of mineral salt deposits of continental origin) or saline soils (Walsh et al., 2005; Ventosa, 2006; Ventosa et al., 2008), but they are also represented by salt deposits, some desert plants, oilfield brines and a variety of salted foods, from seasoned fish or meat to fermented foods as well as animal hides (Grant et al., 1998; Ventosa, 2006). The best studied hypersaline habitats are aquatic hypersaline systems, such as salt lakes and salterns.

Salterns are excellent models for studying the ecology and diversity of microorganisms, given that they are composed by a series of ponds with widely different salinities that concentrate salt from seawater to the point of saturation and precipitation. Most saltern studies have been performed on the saturated brine crystallizer ponds (Antón et al., 1999, 2000; Benlloch et al., 2001; Pašić et al., 2005; Pasić et al., 2007; Legault et al., 2006; Oh et al., 2010). Some comparative reports performed on crystallizers ponds, as Pasić et al. (2007), compared haloarchaeal communities

from two Adriatic solar salterns, showing differences in the microbiota and that climate could play a role in the microbial community structure. Oh et al. (2010) examined the diversity of *Haloquadratum* and other haloarchaea in three coastal, but geographically distant saltern crystallizer ponds in Australia. The great majority of the 16S rRNA gene sequences recovered from these crystallizers were related to *H. walsbyi* and diverged by less than 2% from each other, and from the type strain of this genus (strain C23). However, our knowledge about intermediate salinity ponds is limited. Benlloch et al. (2002) analyzed 16S rRNA sequences by DGGE of three salt ponds (8, 22, and 32% total salts, respectively) from Santa Pola saltern in Eastern Spain. Most bacterial sequences in the 8% salt pond were related to organisms of marine origin belonging to representatives of the classes *Alpha*-, *Beta*-, *Gamma*-, and *Epsilon-Proteobacteria*, and the phyla *Bacteroidetes*, *Actinobacteria*, and *Cyanobacteria*. In the 22% salt pond were found *Alpha*- and *Gamma-Proteobacteria*, *Cyanobacteria*, and *Bacteroidetes*, and most of them were related to specialized halophiles. From the 32% salt pond, the only *Bacteria* found were sequences that clustered with *Salinibacter ruber*, an extremely halophilic *Bacteroidetes*. And in those three different salinity ponds, most of the clones were related to cultured strains of the archaeal

class *Halobacteria*. A metagenomic study of an intermediate salinity pond from Santa Pola saltern (SS19), revealed a low presence of bacterial species of *Halomonas*, *Chromohalobacter*, or *Salinivibrio* (Ghai et al., 2011), which are commonly isolated from those habitats (Arahal et al., 2001; Arenas et al., 2009) but abundant metagenomic reads affiliated to *Haloquadratum walsbyi* and *Salinibacter ruber* were also found. A novel representative of nanohaloarchaeota (“*Candidatus Haloredivivus*”) which was also found to be abundant, and its genome was partially assembled. The most abundant bacterium found in this 19% salt pond appeared to be a gammaproteobacterium closely related to *Alkalilimnicola* and *Nitrococcus*. This microbe has been recently cultured and its genome sequenced (Leon et al., 2013; López-Pérez et al., 2013). Besides, Ghai et al. (2011) found a large number of sequences related to presumably non-halophilic bacterial genera and, a group of low G+C *Actinobacteria* typical to freshwater habitats. A recent metagenomic study on a pond with 13% salts of Santa Pola saltern showed a large microbial diversity representing seven major higher taxa: *Euryarchaeota*, *Gammaproteobacteria*, *Alphaproteobacteria*, *Actinobacteria*, *Bacteroidetes*, *Verrucomicrobia*, and *Betaproteobacteria* (Fernández et al., 2014a). Community analysis of an intermediate salinity pond (18% total salts) from a saltern in Guerrero Negro, Mexico, demonstrated that the archaeal community was dominated by a single uncultured 16S rRNA phylotype with 99% similarity to sequences recovered from a Tunisian saltern and the most abundant bacterial sequences were 99% similar to an uncultured gammaproteobacterial clone from the Salton Sea (Dillon et al., 2013). Boujelben et al. (2012) explored the prokaryotic community in several ponds from Sfax saltern in Tunisia. They showed that some phylotypes, such as those related to *Haloquadratum* or representatives of *Bacteroidetes*, displayed a strong dependence of salinity and/or magnesium concentrations and that temperature was a strong factor structuring the prokaryotic community in the pond with 20% salinity, but not in the crystallizer pond, due to the seasonal changes. However, a survey about six salt lakes in Inner Mongolia, China, and a salt lake in Argentina showed that archaeal biogeography was influenced by  $\text{Na}^+$ ,  $\text{CO}_3^{2-}$ ,  $\text{HCO}_3^-$ , pH and temperature, and bacterial biogeography was influenced by  $\text{Na}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{HCO}_3^-$ , and pH as well as geographic distance (Pagaling et al., 2009; Grant et al., 2011).

Therefore, in order to learn more about the community structure of intermediate salinity ponds, we explored the phylogenetic and taxonomic differences as well as metabolic profiles of two geographically distant habitats from Spain: the Isla Cristina saltern that gets its water from the Atlantic Ocean and the Santa Pola saltern on the Mediterranean Sea.

## MATERIALS AND METHODS

### SAMPLE COLLECTION, DNA EXTRACTION, AND SEQUENCING

The metagenomic datasets analyzed in this study are derived from different saline systems: one dataset was obtained from a pond at the Isla Cristina saltern located on the Atlantic Ocean, in southwestern Spain with a salinity of 21% (Fernández et al., 2014b); four datasets were from the Santa Pola saltern located in eastern Spain on the Mediterranean Sea, of which three were from concentrator ponds (SS13, SS19, and SS33) and one from a crystallizer pond (SS37) with salt concentrations of 13, 19, 33, and 37%, respectively (Ghai et al., 2011; Fernández et al., 2013, 2014a); and two marine datasets, deep chlorophyll maximum from Mediterranean Sea (DCM3) and Mar Menor coastal lagoon (MM5) with salinities of 3.8 and 5%, respectively (Ghai et al., 2010, 2012). All databases were obtained using the same DNA extraction method and the samples were sequenced by pyrosequencing 454 (Martín-Cuadrado et al., 2007; Ghai et al., 2011). The accession numbers for the deposited databases are shown in Table 1.

### COMPARATIVE ANALYSIS OF METAGENOMIC READS

To estimate cumulative nucleotide differences between metagenomic datasets, we carried out BLASTN searches of the complete set of sequences from every dataset vs. all the others. Bitscores of the top high-scoring segment pairs (HSPs) from every sequence from one set vs. another were summed to yield a cumulative pairwise bitscore value (CPBV) that was normalized and used to construct a distance matrix. CPBVs were normalized by dividing each one by the cumulative bitscore value derived from the BLASTN of one dataset vs. itself. The distance matrix was analyzed using the Phylip package (Felsenstein, 1989) to obtain a neighbor-joining tree.

G+C contents were computed using the program geece in the EMBOSS package (Rice et al., 2000) and the amino acid frequency was calculated from a perl script. The metagenomic reads were

**Table 1 | Features of the different datasets from saline habitats used in this study.**

Datasets	Salinity (%)	Temperature (°C)	pH	Number of reads	Dataset size (Mb)	Average read length (bp)	Accession number/source	References
Deep Chlorophyll Maximum (DCM3)	3.8	15.9	8.1	1,204,321	312	259	SRP002017	Ghai et al., 2010
Mar Menor Coastal Lagoon (MM5)	5	19	8.4	730,997	243	335	<a href="http://camera.calit2.net">http://camera.calit2.net</a>	Ghai et al., 2012
Santa Pola Saltern (SS13)	13	29	8.0	1,443,593	441	305	SRP028290	Fernández et al., 2013, 2014a
Santa Pola Saltern (SS19)	19	30	8.0	1,315,302	475	361	SRP007685	Ghai et al., 2011
Santa Pola Saltern (SS33)	33	30	7.0	842,872	309	367	SRP028290	Fernández et al., 2013
Santa Pola Saltern (SS37)	37	41	7.1	760,740	309	417	SRP007685	Ghai et al., 2011
Isla Cristina Saltern (IC21)	21	25	7.5	1,223,923	486	397	SRP029970	Fernández et al., 2014b

annotated using UniProtKB database released in December 2013 (UniProt Consortium, 2014) through BLASTX search with a cutoff *e*-value of 1e-5.

16S ribosomal RNA genes were identified by comparing the datasets against the RDP database version 11.1 (Cole et al., 2014). All reads that matched a 16S rRNA sequence with an alignment length of more than 100 bp and an *e*-value lower than 1e-5 against the database were extracted. The best hit that was not described as unknown or unidentified was considered a reasonable closest attempt for classifying the 16S rRNA sequences. Sequences were assigned to a specific genus if they shared  $\geq 95\%$  16S rRNA sequence identity with a known representative.

### METAGENOMIC READS ASSEMBLY

Assembly of the metagenomic reads with greater than 100 bp was performed using stringent criteria requiring an overlap of at least 80 bp, 99% identity and at most a single gap in the alignment (using Geneious Pro 5.4). Next, assembled contigs that were less than 3 kb in length, and those with less than three predicted genes were discarded for further analysis. We retained only those contigs that provided consistent query hits to only single high level taxa (e.g., *Alphaproteobacteria*, *Euryarchaeota*, *Bacteroidetes*, *Actinobacteria*). To test if the assembly strategy produced authentic contigs from known organisms, we manually identified all contigs that belonged to *H. walsbyi*, one of the abundant organism in the datasets. The criterion was that all genes from a putative *H. walsbyi* contig must return best hits from that genome.

Tetranucleotide frequencies of the assembled contigs were computed using the wordfreq program in the EMBOSS package (Rice et al., 2000), and principal component analysis (PCA) was performed using the R package FactoMineR (Lê et al., 2008).

### CONSTRUCTION OF PHYLOGENETIC TREES

Maximum likelihood reference trees were constructed using RaxML as implemented in ARB software package (Ludwig et al., 2004) using reference 16S rRNA gene sequences with near full length ( $>1300$  nt) from cultured isolates. Later, partial 16S rRNA gene assembled metagenomic sequences and closely related environmental uncultured 16S rRNA gene sequences were inserted into reference trees without altering tree topology using maximum parsimony criterion and a 50% base frequency filter. Bootstrap values greater than 50% are indicated above nodes and the scale bar represents 10 base substitutions per 100 nt positions. The 16S rRNA gene sequences retrieved in this study were deposited in Genbank under accession numbers KJ546108–KJ546118, KJ588879–KJ588888, KJ588892–KJ588898 (archaeal) and KJ588890–KJ588891, KJ588899–KJ588905 (bacterial).

## RESULTS AND DISCUSSION

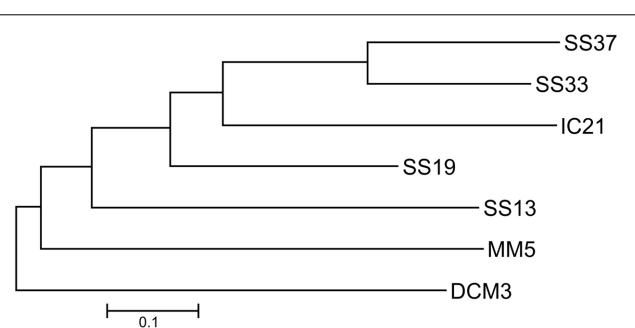
### FEATURES OF DATASETS

In this study several metagenomic datasets were examined (Table 1). The thalassosaline waters of the salterns in Santa Pola and Isla Cristina have a marine origin but they have a different source: the Mediterranean Sea and the Atlantic Ocean, respectively. Besides their high salt concentrations, these environments are subject to strong solar irradiation (Rodríguez-Valera et al.,

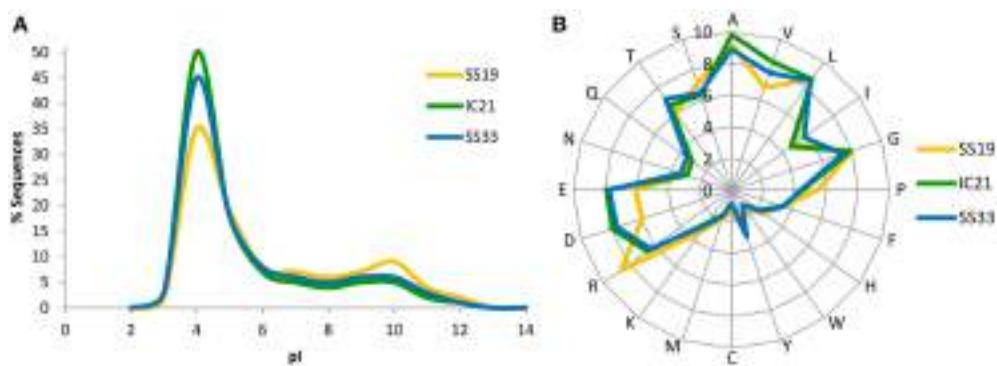
1985; Rodríguez-Valera, 1988). In order to determine how a gradient of salinity influences microbial communities, we compared metagenomes datasets of Santa Pola and Isla Cristina salterns and DCM3 and MM5 from marine sites (Figure 1). We expected marine derived datasets to form their own branch separate from the saltern datasets and that IC21 should be closer to SS19 than SS33, because salinity has been identified as the main factor determining the distribution of prokaryotic organisms in aquatic systems (Lozupone and Knight, 2007; Schapira et al., 2009). As expected, the phylogenetic tree showed that marine datasets, DCM3 and MM5 clearly differed from the saltern datasets, sharing a low number of sequences with them, and the saltern datasets were more similar to each other, making it possible to recognize the impact of a salinity gradient. Unexpectedly however, the community from IC21 was closer in structure to both the datasets, SS19 and SS33, although the salt concentration in IC21 was nearer to the dataset SS19, suggesting community composition is affected by local environmental characteristics. Thus, we focused this study on the intermediate salinity ponds and analyzed the datasets qualitatively and quantitatively to elucidate potential causes that might produce the observed differences.

A well-known adaptive feature for living at high salt concentration is the enrichment of acidic amino acids in proteins allowing them to properly function at high cytoplasmic salinities (Soppa, 2006). Therefore, we analyzed the isoelectric point and amino acids frequencies of proteins, which would reflect this adaptation. Surprisingly, IC21 proteins were more similar to those in SS33 than to SS19 (Figure 2). The dataset IC21 showed an increase in acidic amino acids compared to SS19 that therefore indicates a greater presence of microorganisms using the “salt-in” strategy to osmotically balance their cytoplasm with their environment (Oren, 2008, 2013).

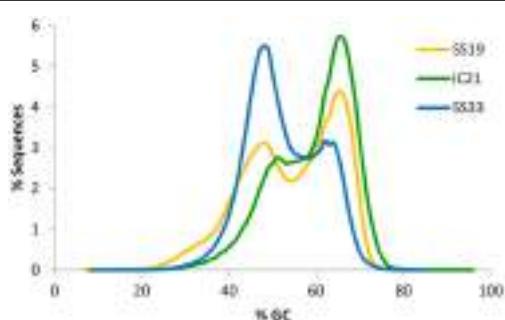
High G+C content is often associated with the presence of haloarchaea, except for the well-known exception *Haloquadratum* and is a useful predictor of community composition. Analysis showed that the G+C content of IC21 (Figure 3) had a bimodal distribution more similar to SS19 than to SS33, and the predominant peak in both is at  $\sim 65\%$ . This peak is consistent



**FIGURE 1 | Unrooted neighbor-joining phylogenetic tree based on a distance matrix calculated using “bit-scores” between the different metagenomes.** DCM3: Deep Chlorophyll Maximum, Mediterranean Sea (3.8% salinity), MM5: Mar Menor Coastal Lagoon (5% salinity), IC21: Isla Cristina saltern (21% salinity), SS13: Santa Pola saltern (13% salinity), SS19: Santa Pola saltern (19% salinity), SS33: Santa Pola saltern (33% salinity), and SS37: Santa Pola saltern (37% salinity).



**FIGURE 2 | Isoelectric point (pl) profiles (A) and amino acid frequency (B) of the predicted proteins in the metagenomic datasets SS19, IC21, and SS33.** The pl was computed for each translated read and is shown as a percentage of the dataset in intervals of bin width 2.



**FIGURE 3 | Comparison of G+C% of sequences from metagenomic datasets SS19, IC21, and SS33.** G+C% was computed for each read and the percentage of the dataset in intervals of bin width 10 is shown.

with the high G+C content associated with most halophilic archaea and bacteria described so far (Paul et al., 2008). Ponds from Santa Pola saltern (SS19 and SS33) exhibited a low G+C peak at 47.9%, which has previously been observed to increase at higher salinities (Ghai et al., 2011). This low G+C peak comes from *Haloquadratum walsbyi* (Bolhuis et al., 2006) and perhaps from the newly reported nanohaloarchaea (Ghai et al., 2011). In IC21 the low G+C peak is shifted to a value slightly higher than that observed in metagenomic datasets from Santa Pola ponds, SS19 and SS33, around 51–52%. This low G+C peak might correspond to genera containing representatives of halophilic bacterial genera such as *Halomonas*, *Salinicrobium*, or *Salinicoccus*. These genera have DNA G+C contents in the range of 52.0–74.3, 44.9–51.5, and 46–51.2%, respectively (de la Haba et al., 2011). Another possibility is that it corresponds to halophilic or halotolerant microorganism not yet described, or to AT-rich regions in haloarchaeal genomes (Ram Mohan et al., 2014). In members of the order *Halobacteriales* a “minor component” of the DNA (10–30% of the total DNA) with a G+C range of 51–59 mol% has been reported (Grant et al., 2001).

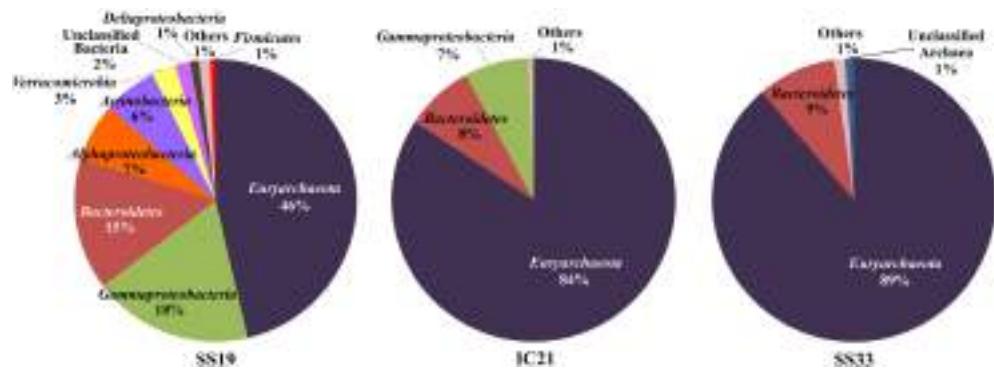
#### PROKARYOTIC COMMUNITY STRUCTURE

The taxonomic diversity was analyzed carrying out a search of the metagenomic sequences related to the 16S rRNA gene using

the RDP database and selecting those sequences with a minimum length of 100 bp and an identity over 80% for higher taxonomic levels (Figure 4) and 95% for the genus level (Table 2).

In Figure 4 it is observed that the bacterial community decreases sharply in IC21 compared to SS19, shifting to a largely archaeal community. Also identified are members of 14 higher taxa in SS19 but interestingly we found only six in IC21 and five in SS33. The phyla *Euryarchaeota* and *Bacteroidetes* and the class *Gammaproteobacteria* are shared by the three datasets, but there are numerically more sequences related to the phylum *Euryarchaeota* with increasing salinity and a concurrent decrease for the other two. Further, these data show that the phyla biodiversity in IC21 is more similar to SS33 than to SS19. Ghai et al. (2011) analyzed the changes of the biodiversity along a salinity gradient in two metagenomic datasets (with 19 and 37% salts) from Santa Pola saltern. The biodiversity detected in the crystallizer pond (SS37) is quite similar to that determined for SS33, due to their extreme salinities; however in the crystallizer pond only two phyla were found, corresponding to *Euryarchaeota* and *Bacteroidetes*.

The simplification of the prokaryotic community at higher salinities is also seen at the genus level: the number of sequences related to different genera decreases from 69 in SS19 to 28 and 16 in IC21 and SS33, respectively. The most abundant sequences in all three datasets are related to the archaeal genera *Halorubrum*, *Haloquadratum*, and *Natronomonas*, recruiting more *Halorubrum* sequences in SS19 and IC21 (12.5 and 65.8%, respectively) and *Haloquadratum* in SS33 (29.5%) (Table 2). The genus *Salinibacter* (belonging to the *Bacteroidetes*) is the next taxon highly represented in Santa Pola datasets with 6.4 and 4.7% of the sequences in SS19 and SS33, respectively. However, in IC21 the second predominant genus is *Psychroflexus*, also a member of the phylum *Bacteroidetes*, at 4.6% of the sequences. The six species comprising the genus *Psychroflexus* (*P. gondwanense*, *P. halocasei*, *P. salinarum*, *P. sediminis*, *P. torquis*, and *P. tropicus*) have been characterized as slightly or moderately halophilic bacteria and most of them have been isolated from saline environments (Bowman et al., 1998; Donachie et al., 2004; Chen et al., 2009; Yoon et al., 2009; Seiler et al., 2012). With respect to genera of the class *Gammaproteobacteria*, in SS19 and IC21, *Spiribacter* is identified as one of the most abundant. In spite of the strain “*Spiribacter*



**FIGURE 4 | Representation of high taxonomic levels affiliated to metagenomic rRNA reads.** Assigned sequences have an identity over 80% and a minimum length of 100 bp.

**Table 2 | Microbial diversity at genus level in the metagenomic datasets SS19, IC21, and SS33.**

Sample	SS19	IC21	SS33
<i>Halorubrum</i>	12.5%	<i>Halorubrum</i>	65.8%
<i>Haloquadratum</i>	7.9%	<i>Psychroflexus</i>	4.6%
<i>Salinibacter</i>	6.4%	<i>Natronomonas</i>	3.2%
<i>Natronomonas</i>	5.6%	<i>Haloquadratum</i>	1.9%
" <i>Spiribacter</i> "	4.4%	" <i>Spiribacter</i> "	1.1%
<i>Haloredivivus</i> *	3.2%		
<i>Puniceicoccus</i>	1.9%		
<i>Arhodomonas</i>	1.8%		
<i>Aquiluna</i> *	1.5%		
<i>Oceanicola</i>	1.1%		
Total number of sequences	1477	1538	888
Total number of genera	69	28	16

Assigned sequences have a 16S rRNA identity over 95% and a minimum length of 100 bp. Only those genera with more than 1% of assigned sequences are shown. SS19: Santa Pola saltern (19% salinity), IC21: Isla Cristina saltern (21% salinity), and SS33: Santa Pola saltern (33% salinity).

\*Candidatus.

*salinus*" M19–40 being isolated from Isla Cristina saltern (Leon et al., 2013; López-Pérez et al., 2013), it was more abundant in SS19 compared to IC21 (4.4% vs. 1.1%, respectively). In fact, Ghai et al. (2011) reported a great abundance of sequences belonging to *Gammaproteobacteria* related to the genus *Alkalilimnicola*, which later was assigned to "*S. salinus*" M19–40 (López-Pérez et al., 2013). Although at the phylum or class level IC21 is more similar to SS33 at the genus level the Isla Cristina dataset differs in biodiversity and abundance from both Santa Pola datasets, demonstrating an overrepresentation of the genus *Halorubrum* and decrease in the sequences (0.4%) related to *Salinibacter* in IC21. Ghai et al. (2011) observed that in Santa Pola datasets (SS19 and SS37) the species *Salinibacter ruber* appeared as an abundant microorganism, but in IC21 it is much lower.

Traditional culture methods carried out in Santa Pola saltern determined that in intermediate salinity ponds there were a variety of moderately halophilic microorganisms (Rodríguez-Valera et al., 1985). Subsequent studies performed by molecular techniques indicated that these prokaryotic

representatives belong in the groups *Gammaproteobacteria*, *Bacteroidetes*, and *Halobacteriaceae* and in a pond with 22% total salts 16S rDNA sequences were related to the genera *Psychroflexus*, *Halorubrum*, and *Natronomonas* (Benloch et al., 2002); similar results were obtained in the dataset IC21. In a recent study of a pond from the Exportadora de Sal (ESSA) evaporative saltern in Guerrero Negro (Mexico) with 18% total salts a *Halorubrum*-like sequence nearly identical (>99.5% similar) to environmental sequences from the Santa Pola saltern as well as sequences 97% similar to "*S. salinus*" were reported (Dillon et al., 2013). Recently, Podell et al. (2013) found that the microbial composition from Lake Tyrrell (Australia) was correlated with concentrations of potassium, magnesium, and sulfate, but not sodium, chloride, or calcium ions. Sequences related to *Haloquadratum* were positively correlated with potassium, magnesium, and sulfate ions while sequences related to *Halorubrum*, *Haloarcula*, *Halonotius*, *Halobaculum*, and *Salinibacter* were negatively correlated with them. In addition, *H. walsbyi* shows a higher tolerance to Mg<sup>2+</sup> than other halophilic archaea (Bolhuis

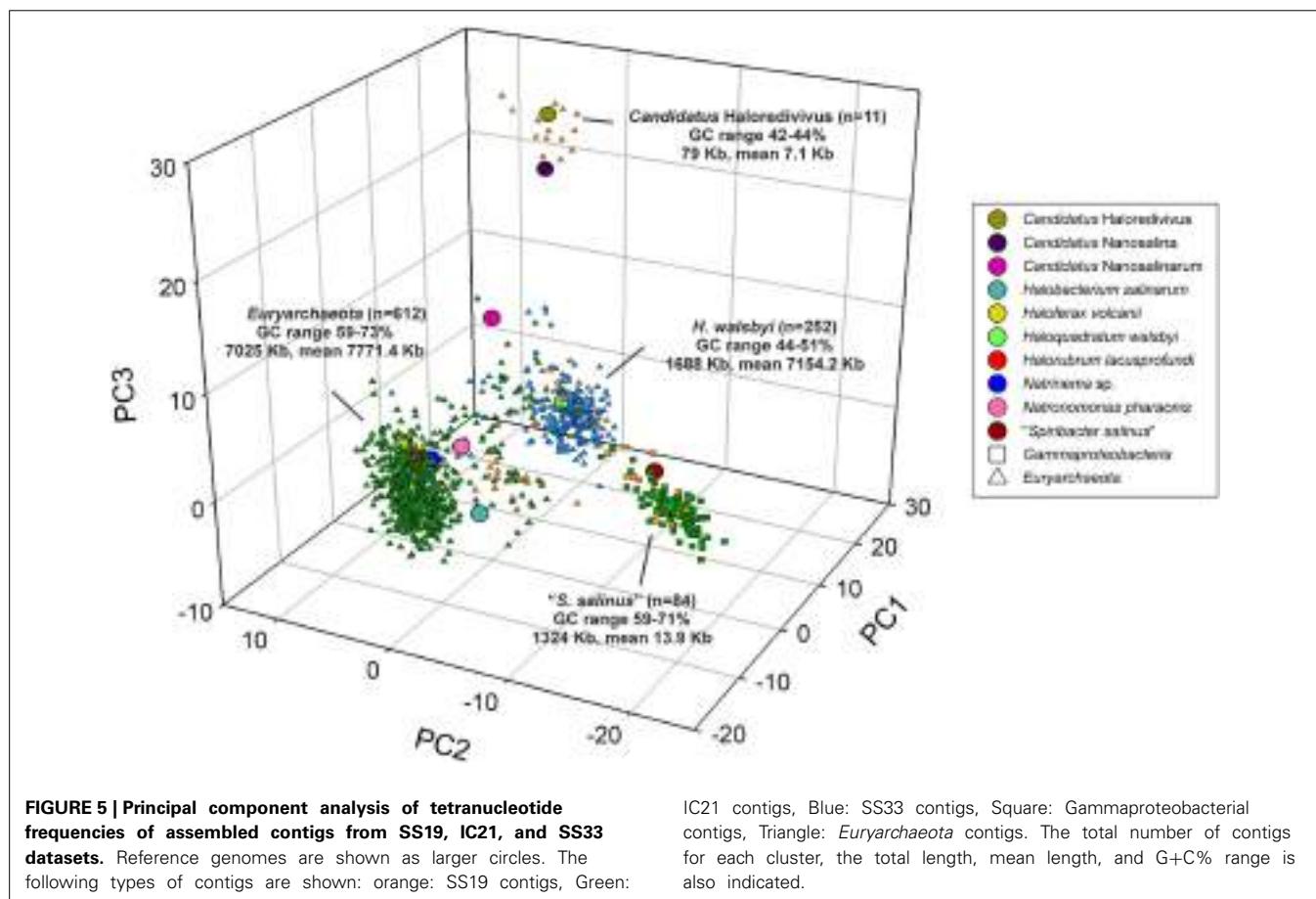
et al., 2004; Burns et al., 2004). The differences found among Isla Cristina and Santa Pola datasets, mainly due to the dominance of sequences related to *Halorubrum* in IC21 and to *Haloquadratum* in SS19 and SS33, might be explained by the difference in the ionic composition of both samples as previously reported in other hypersaline habitats (Pagaling et al., 2009; Grant et al., 2011; Boujelben et al., 2012; Podell et al., 2013).

### CONTIGS OF CONCENTRATOR PONDS

It is expected that contigs assembled from metagenomic reads will yield genomic fragments derived from the most abundant organisms in the sample (Ghai et al., 2011). Assembled contigs from the metagenomic datasets SS19, IC21, and SS33 were assembled and tested against the genome of an abundant organism in the datasets, *H. walsbyi*. In SS19, a total of 84 contigs larger than 5 kb were assembled, 69 could be assigned to the phylum *Euryarchaeota* and 15 to the class *Gammaproteobacteria* (Ghai et al., 2011). In IC21 710 contigs with at least 5 kb were obtained; 618 were assigned to the phylum *Euryarchaeota*, 85 contigs to the class *Gammaproteobacteria* and 7 contigs to viruses. With respect to SS33 a total of 248 contigs were assembled, 247 were assigned to *Euryarchaeota* and 1 to the phylum *Bacteroidetes*. A PCA on the normalized tetranucleotide frequencies of the contigs belonging to the most abundant groups was carried out (Figure 5). The contigs from SS19 and IC21 had consistent hits

to taxa within the *Euryarchaeota* and *Gammaproteobacteria* and from SS33 the majority were *Euryarchaeota*. Contigs related to the phylum *Euryarchaeota* are grouped in two different clusters with low G+C content, one of them closely related to *H. walsbyi*, comprising 30 and 222 contigs of SS19 and SS33, respectively, and the second cluster related to the nanohaloarchaeon “*Candidatus Haloredivivus*” forming 11 contigs of the SS19. Additionally, we observed a third cluster including *Euryarchaeota* of high G+C content, comprising contigs from the three datasets that were related to genomes of extremely halophilic archaea. On the other hand, many contigs related to *Euryarchaeota* from the three datasets did not cluster with contigs related to *H. walsbyi* nor *Euryarchaeota* of high G+C content. Their G+C content values are between the extremely halophilic archaeal reference genomes and *H. walsbyi*; possibly these contigs belong to unknown hyperhalophilic archaea. A fourth cluster of *Gammaproteobacteria* contigs from SS19 (13 contigs) and IC21 (71 contigs) was closely related to the genome of “*Spiribacter salinus*” M19–40 (López-Pérez et al., 2013).

Because of the unique G+C content, the contigs with a 51–52% G+C in IC21 were further examined (Figure 3). Contigs associated with *Gammaproteobacteria* and viruses did not have this G+C content, ruling them out as contributors. However, among the nine contigs related to *Euryarchaeota*, three could clearly be assigned to *Halorubrum* and also had similar G+C



**FIGURE 5 | Principal component analysis of tetranucleotide frequencies of assembled contigs from SS19, IC21, and SS33 datasets.** Reference genomes are shown as larger circles. The following types of contigs are shown: orange: SS19 contigs, Green:

IC21 contigs, Blue: SS33 contigs, Square: Gammaproteobacterial contigs, Triangle: *Euryarchaeota* contigs. The total number of contigs for each cluster, the total length, mean length, and G+C% range is also indicated.

content. These contigs mainly contain genes for ABC transporters, metallophosphoesterase, multi-sensor signal transduction histidine kinase, and hypothetical proteins.

Further taxonomic analysis was performed on 16S rRNA genes from assembled metagenomic reads longer than 500 bp. We found eight 16S rRNA sequences from each of the datasets SS33 and IC21, and 11 sequences from the dataset SS19 that were analyzed using BLAST (Figure 6). One assembled sequence from each of the datasets grouped within the *Haloquadratum* cluster together with uncultured archaeal sequences from Lake Tyrrell, VIC, Australia (with 29% salinity). The 16S rRNA sequences from SS19, IC21, and SS33 had high similarity with *H. walsbyi* HBSQ001 of 100, 99.9, and 98.9%, respectively. Only one sequence from SS33 was found within the *Haloplanus* cluster with a similarity of 98.0% to *Haloplanus natans* and 97.8% to an uncultured archaeon sequence of a saline soil from Jiangsu (China). Three sequences from IC21 and one sequence from SS19 were included into the *Halorubrum* cluster. One sequence from IC21 shared with *Hrr. chaoviator* a similarity of 99.2%, another sequence from IC21 had a similarity of 99.4 and 99.3% with two different sequences of uncultured archaea from Aran-Bidgol salt lake (salinity over 30%) and other sequence from IC21 shared a similarity of 99.5% with a sequence from SS19 and with *Hrr. orientale*. Within the *Halohasta* cluster one sequence from IC21 had a similarity of 98.8% to *Halohasta litorea*. The “*Candidatus Haloredivivus*” cluster was represented by one sequence from SS19. The rest of the contigs could not be classified into known-archaeal genera. The assembled 16S rRNA sequences from the three datasets showing the presence of potential microorganisms related to *Haloquadratum* (although in IC21 representatives of this genus do not appear to be very abundant) and uncultured archaea in clusters 3 and 5. In IC21 the assembled sequences show a high abundance of members of the genus *Halorubrum*. A pattern regarding the community composition among the datasets studied appears to be absent, but we do find evidence that some sequences from different hypersaline environments are related to them. This suggests unknown environmental parameters affect community composition, and/or perhaps random dispersal and horizontal gene transfer of key adaptive genes play a role. For instance, Parnell et al. (2010) demonstrated that genes providing adaptation to their niche, rather than the taxa living there, structured halophilic communities in the Great Salt Lake.

Additionally, we found 16S rRNA assembled metagenomic reads longer than 500 bp related to the phylum *Bacteroidetes*: four from SS19, two from IC21 and one from SS33 (Figure 7). We used BLAST to search in the nr/nt database the sequences showing a higher similarity to the 16S rRNA assembled sequences of *Bacteroidetes*. A first cluster related to *Psychroflexus* was detected including one sequence from IC21. This sequence was similar to an uncultured bacterium from a Tibetan hypersaline lake (96.4%) and *Psychroflexus sediminis* (95.7%). One sequence from SS19 was within the *Salinibacter* cluster, showing a 92.1% similarity to *Salinibacter ruber*. The rest of the sequences from SS19, IC21, and SS33 could not be classified into any cluster containing cultured microorganisms. The 16S rRNA assembled metagenomic reads of *Bacteroidetes* show that in SS19 this taxon is more abundant than

IC21 and SS33, with predominance of sequences related to uncultured bacteria, while in the dataset IC21 *Psychroflexus* was the genus that recruited more 16S rRNA sequences (Table 2). Neither of the 16S rRNA assembled sequences related to *Bacteroidetes* in IC21 are related to any of the assembled sequences of SS19 and SS33.

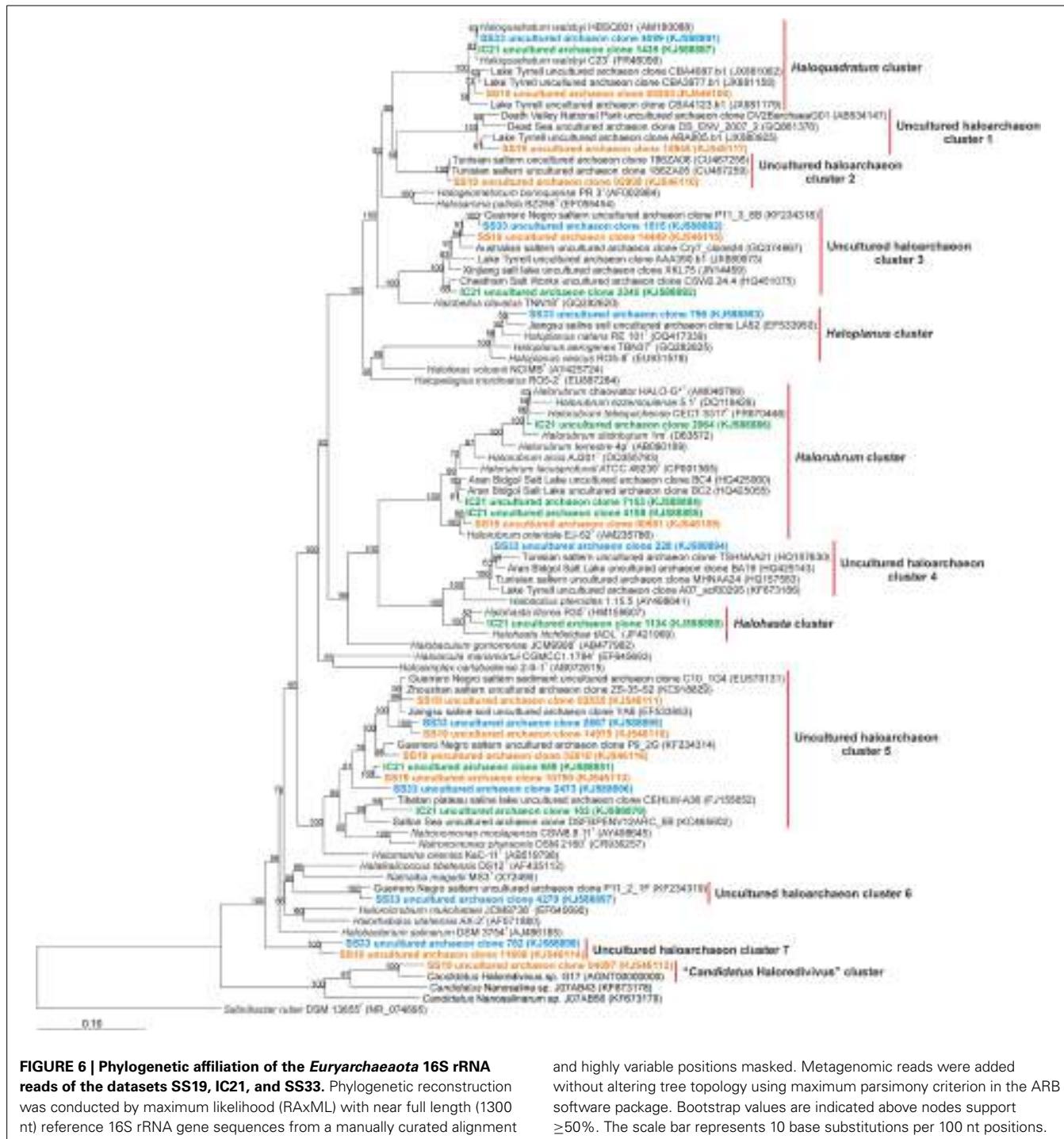
Although at high taxonomic levels IC21 is more similar to SS33 than SS19, the fact is that IC21 and SS33 are different with respect to the genera and their abundance observed as well as to the presence of different uncultured taxa. In saline and alkaline lakes in Iran a similar composition of microbial communities but differing in community structure has been reported (Makhdoumi-Kakhki et al., 2012).

## METABOLIC PROFILE

To analyze the metabolic diversity in the dataset IC21, we determined the relative abundance of individual genes involved in metabolic pathways of the datasets SS19, SS33, and IC21 (Supplementary Table 1) by searching predicted metagenomic proteins with the UniprotKB database (UniProt Consortium, 2014) using BLASTX (Camacho et al., 2008).

Metabolism based on energy from light was queried in our data sets. Because our screening protocol removed eukaryotic microorganisms, only prokaryotic microorganisms were considered in this analysis. The photosynthetic reaction centers, *psbA* and *psbD* genes, were not detected in the three datasets analyzed and therefore we can conclude that photosynthesis by prokaryotes was absent in these datasets. Instead, genes coding different types of rhodopsins were found suggesting that light is widely used as an energy source in these conditions, just not for carbon fixation. It was observed that bacteriorhodopsins and halorhodopsins increased in frequency at higher salinities. Bacteriorhodopsins associated with *Haloquadratum* were in a greater proportion in Santa Pola saltern datasets while in IC21 they were associated with *Halobacterium* and *Halorubrum*. Halorhodopsins related to *Halobacterium* were found in high proportion in all datasets but the number of sequences belonging to *Haloquadratum* is greater in Santa Pola saltern and those from *Halorubrum* are higher in IC21. Sequences related to bacteriorhodopsins and halorhodopsins show clearly the abundance of *Haloquadratum* in Santa Pola SS19 and SS37 datasets and *Halorubrum* in Isla Cristina IC21 dataset.

Microorganisms under osmotic stress conditions use different survival strategies. Most bacteria maintain cell integrity through accumulation of compatible solutes (“salt-out” strategy). Sequences related to genes of compatible solutes were mainly glutamate synthase, betaine transporters, glycerol kinase, and glycerol-3-phosphate dehydrogenase, and at lower frequencies were glycerol and glutamate transporters. In IC21 lower number of sequences related to choline dehydrogenase, glutamate synthase, and trehalose synthase were observed compared to Santa Pola datasets, suggesting a higher synthesis of these compatible solutes in Santa Pola saltern. The proportion of sequences related to compatible solutes decreased with the salinity except for glycerol degradation (glycerol kinase, glycerol-3-phosphate dehydrogenase, and dihydroxyacetone kinase), increasing the number of sequences in the dataset SS33 compared to SS19

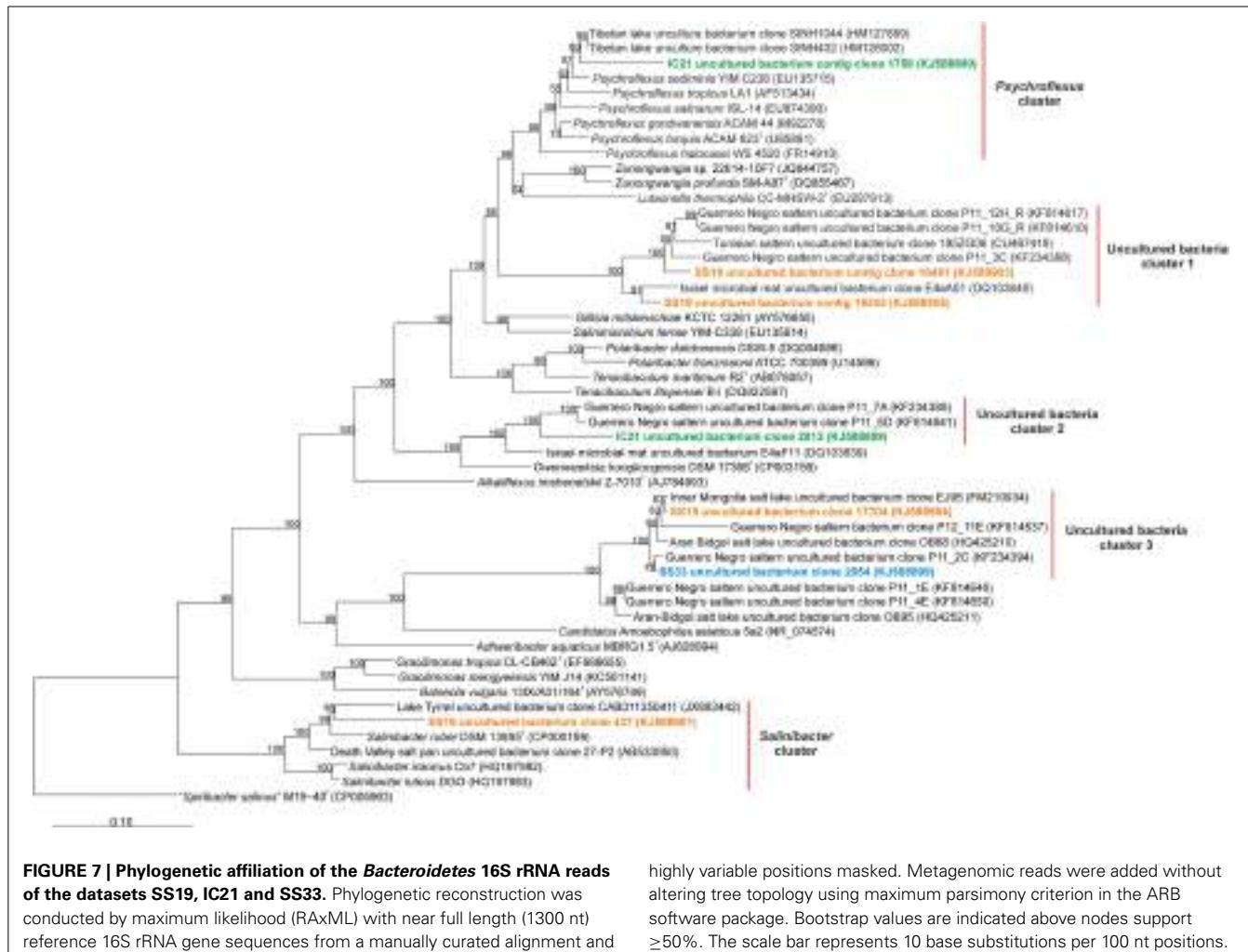


(number of sequences relative to the total number of metagenomic sequences). This result could indicate a higher presence of the algae *Dunaliella* or a higher primary production of glycerol, which is the predominant compatible solute in *Dunaliella* or alternatively, the oxidation of glycerol to dihydroxyacetone, that has been studied in the species *Salinibacter ruber* and is used as a growth substrate by *H. walsbyi* and *Haloferax volcanii* (Elevi Bardavid and Oren, 2008; Ouellette et al., 2013). Therefore,

and highly variable positions masked. Metagenomic reads were added without altering tree topology using maximum parsimony criterion in the ARB software package. Bootstrap values are indicated above nodes support  $\geq 50\%$ . The scale bar represents 10 base substitutions per 100 nt positions.

glycerol and probably dihydroxyacetone are considered as the main carbon source and energy for the heterotrophic community in salterns (Borowitzka and Brown, 1974; Borowitzka et al., 1977; Ouellette et al., 2013).

With respect to the nitrogen cycle, it seemed to be simplified in the saltern datasets analyzed, with a decrease of the number of sequences involved in the reduction of nitrate to nitrite by nitrate reductase and nitrite to nitric oxide by nitrite reductase in SS33.



**FIGURE 7 | Phylogenetic affiliation of the *Bacteroidetes* 16S rRNA reads of the datasets SS19, IC21 and SS33.** Phylogenetic reconstruction was conducted by maximum likelihood (RAxML) with near full length (1300 nt) reference 16S rRNA gene sequences from a manually curated alignment and

highly variable positions masked. Metagenomic reads were added without altering tree topology using maximum parsimony criterion in the ARB software package. Bootstrap values are indicated above nodes support  $\geq 50\%$ . The scale bar represents 10 base substitutions per 100 nt positions.

In salterns, sulfate is concentrated along the ponds until its saturation and precipitates forming calcium sulfate (gypsum) (Landry and Jaccard, 1984). Dissimilatory sulfate reduction has been reported until salt concentrations of 24% NaCl (Oren, 1988). We detected sequences of genes involved in a complete dissimilatory sulfate reduction (sulfate adenylyltransferase, adenylylsulfate kinase, phosphoadenylylsulfate reductase, and sulfite reductase) in all datasets studied, except for adenylylsulfate kinase in SS33 dataset. The microbial communities of salterns are traditionally considered heterotrophic, but chemolithotrophic bacteria can also be abundant and active in extreme conditions. In particular, chemolithotrophic sulfur-oxidizing bacteria are able to adapt well to hypersaline conditions because the complete oxidation of sulfide or thiosulfate to sulfate has high energy efficiency (Oren, 1999). In spite of this, we only found some sequences of sulfide dehydrogenase that oxidizes sulfide to sulfate in SS19. Therefore, an incomplete cycle of sulfate is observed in the datasets. In fact, salterns are generally considered eutrophic media and so the pressure of selection that favors organisms with biosynthetic capabilities with full paths is probably weak (Rodríguez-Valera et al., 1981). Additionally, this pathway may be

carried out by phototrophic sulfur-oxidizing bacteria in anoxic sediments, as hydrogen sulfide is considered an important transporter of electrons between the aerobic and anaerobic habitats (Jørgensen, 1982).

Phosphate regulon (Pho) plays a key role in phosphate homeostasis, products are involved in the transport and use of several forms of phosphates (Torriani and Ludtke, 1985; Shinagawa et al., 1987; Wanner, 1987, 1993). In the datasets studied a scarce number of sequences related to genes included in Pho regulon, as *phoR* (environmental phosphate sensor) and *phoB* (regulon activator) were found. However, in the dataset IC21 more sequences corresponding to the negative regulator protein of Pho regulon, PhoU and less for genes involved in the utilization of phosphate were detected compared to the datasets SS19 and SS33. In Santa Pola saltern the total phosphorus concentration increases with the salinity (Rodríguez-Valera et al., 1985). However, the brines support high Mg<sup>2+</sup> concentrations limiting the availability of inorganic phosphate (Bolhuis et al., 2006) and induce the use of phosphonate (Fox and Mendz, 2006). A recent study suggests the utilization of DNA as a phosphate source (Chimileski et al., 2014).

Overall, the main differences found among Isla Cristina and Santa Pola datasets at the metabolic level were a higher number of sequences related to genes involved in the synthesis of compatible solutes (such as choline dehydrogenase, glutamate synthase, and trehalose synthase) and in the utilization of phosphonate in Santa Pola datasets with respect to IC21. This is related to the microbial strategies of haloadaptation to these extreme environments by the different microbial communities present of these habitats as well as the ionic composition of the samples.

## CONCLUSIONS

Santa Pola saltern was built in 1890 over an ancient freshwater lake and close to the Mediterranean Sea (Dulau, 1983). By contrast, Isla Cristina saltern was built in 1955 over wetlands at the marsh of the river Carreras in the village of Isla Cristina, closely located to food-processing industries (Moreno et al., 2009). Santa Pola saltern is subjected to climatic conditions characterized by low annual rainfall and moderate temperatures, with little fluctuation between summer and winter (Rodríguez-Valera et al., 1985). However, Isla Cristina saltern is subjected to rainfall seasons, high solar radiation, and larger temperature fluctuations between day and night (Moreno et al., 2009). Although salinity has been considered the main factor involved in the structure of microbial biodiversity in saline aquatic systems, the data from our study strongly suggests that other factors may influence the composition of hypersaline aquatic microbial communities such as geographic locations (Naor et al., 2012; Zhaxybayeva et al., 2013) and environmental characteristics. The environmental conditions in Santa Pola saltern are more stable than in Isla Cristina saltern, therefore, the prokaryotic community in Santa Pola saltern is more stable over time, and in Isla Cristina saltern the habitat might be continuously recolonizing. Therefore, the differences between these two environments might be due to stable or unstable environmental conditions. The IC21 dataset looks more similar to SS33 dataset because they are mainly composed of representatives of the phylum *Euryarchaeota*, but the reality is that the community structure in IC21 is different because the most abundant genus in IC21 is *Halorubrum*, in contrast to *Haloquadratrum*, which predominates in SS33 dataset. The phylum *Bacteroidetes* is present in all datasets, but in the Santa Pola saltern datasets the most abundant genus of this phylum is *Salinibacter*, while in Isla Cristina saltern dataset is the genus *Psychroflexus*. Additionally, in Santa Pola datasets there are a higher number of sequences related to genes involved in the synthesis of compatible solutes and in the utilization of phosphonate, indicating some differences in the functional activity. Despite the results obtained, it is not clear what is causing the variations between these salterns; a detailed physico-chemical comparative study would be required to elucidate if the microbial structure is being influence by abiotic factors or by biogeographic situation.

## AUTHOR CONTRIBUTIONS

Antonio Ventosa, R. Thane Papke, Francisco Rodriguez-Valera, and Ana B. Fernández conceived the study. Ana B. Fernández and Cristina Sánchez-Porro obtained the metagenome. Ana B. Fernández, Blanca Vera-Gargallo, and Rohit Ghai performed the

analysis. Ana B. Fernández, R. Thane Papke, Francisco Rodriguez-Valera, and Antonio Ventosa wrote the manuscript. All authors read and approved the final manuscript.

## ACKNOWLEDGMENTS

This work was supported by grants from the Spanish Ministry of Science and Innovation (CGL2010-19303), the National Science Foundation (award numbers DEB0919290 and DEB0830024), MICROGEN (Programa CONSOLIDER-INGENIO 2010 CDS2009-00006), NASA Astrobiology: Exobiology and Evolutionary Biology Program Element (Grant Number NNX12AD70G), the Generalitat Valenciana (DIMEGEN PROMETEO/2010/089 and ACOMP/2009/155), MaCuMBA Project 311975 of the European Commission FP7 and the Junta de Andalucía (P07-CVI-03150 and P10-CVI-6226). FEDER funds also supported this project. Ana B. Fernández was supported by a postdoctoral contract from the Junta de Andalucía and Rohit Ghai was supported by a Juan de la Cierva scholarship from the Spanish Ministry of Science and Innovation.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00196/abstract>

## REFERENCES

- Antón, J., Llobet-Brossa, E., Rodríguez-Valera, F., and Amann, R. (1999). Fluorescence *in situ* hybridization analysis of the prokaryotic community inhabiting crystallizer ponds. *Environ. Microbiol.* 1, 517–523. doi: 10.1046/j.1462-2920.1999.00065.x
- Antón, J., Rosselló-Mora, R., Rodríguez-Valera, F., and Amann, R. (2000). Extremely halophilic bacteria in crystallizer ponds from solar salterns. *Appl. Environ. Microbiol.* 66, 3052–3057. doi: 10.1128/AEM.66.7.3052-3057.2000
- Arahal, D. R., García, M. T., Vargas, C., Cánovas, D., Nieto, J. J., and Ventosa, A. (2001). *Chromohalobacter salexigens* sp. nov., a moderately halophilic species that includes *Halomonas elongata* DSM 3043 and ATCC 33174. *Int. J. Syst. Evol. Microbiol.* 51, 1457–1462. doi: 10.1099/ijs.00207713-51-4-1457
- Arenas, M., Bañón, P. I., Copa-Patiño, J. L., Sánchez-Porro, C., Ventosa, A., and Soliveri, J. (2009). *Halomonas ilicicola* sp. nov., a moderately halophilic bacterium isolated from a saltern. *Int. J. Syst. Evol. Microbiol.* 59, 578–582. doi: 10.1099/ijss.0.003509-0
- Benlloch, S., Acinas, S. G., Antón, J., López-López, A., Luz, S. P., and Rodríguez-Valera, F. (2001). Archaeal biodiversity in crystallizer ponds from a solar saltern: culture versus PCR. *Microb. Ecol.* 41, 12–19. doi: 10.1007/s002480000069
- Benlloch, S., López-López, A., Casamayor, E. O., Ovreas, L., Goddard, V., Daee, F. L., et al. (2002). Prokaryotic genetic diversity throughout the salinity gradient of a coastal solar saltern. *Environ. Microbiol.* 4, 349–360. doi: 10.1046/j.1462-2920.2002.00306.x
- Bolhuis, H., Palm, P., Wende, A., Falb, M., Rampp, M., Rodríguez-Valera, F., et al. (2006). The genome of the square archaeon *Haloquadratum walsbyi*: life at the limits of water activity. *BMC Genomics* 7:169. doi: 10.1186/1471-2164-7-169
- Bolhuis, H., Poole, E. M., and Rodriguez-Valera, F. (2004). Isolation and cultivation of Walsby's square archaeon. *Environ. Microbiol.* 6, 1287–1291. doi: 10.1111/j.1462-2920.2004.00692.x
- Borowitzka, L. J., and Brown, A. D. (1974). The salt relation of marine and halophilic species of the unicellular green alga, *Dunaliella*. The role of glycerol as compatible solute. *Arch. Mikrobiol.* 96, 37–52. doi: 10.1007/BF00590161
- Borowitzka, L. J., Kessly, D. S., and Brown, A. D. (1977). The salt relations of *Dunaliella*. Further observations on glycerol production and its regulation. *Arch. Mikrobiol.* 113, 131–138. doi: 10.1007/BF00428592
- Boujelben, I., Gomariz, M., Martínez-García, M., Santos, F., Peña, A., López, C., et al. (2012). Spatial and seasonal prokaryotic community dynamics in ponds of increasing salinity of Sfax solar saltern in Tunisia. *Antonie van Leeuwenhoek* 101, 845–857. doi: 10.1007/s10482-012-9701-7

- Bowman, J. P., McCammon, S. A., Lewis, T., Skerratt, J. H., Brown, J. L., Nichols, D. S., et al. (1998). *Psychroflexus torquis* gen. nov., sp. nov., a psychrophilic species from Antarctic sea ice, and reclassification of *Flavobacterium gondwanense* (Dobson et al. 1993) as *Psychroflexus gondwanense* gen. nov., comb. nov. *Microbiology* 144, 1601–1609. doi: 10.1099/00221287-144-6-1601
- Burns, D. G., Camakaris, H. M., Janssen, P. H., and Dyall-Smith, M. L. (2004). Cultivation of Walsby's square haloarchaeon. *FEMS Microbiol. Lett.* 238, 469–473. doi: 10.1016/j.femsle.2004.08.016
- Camacho, C., Madden, T., Coulouris, G., Ma, N., Tao, T., and Agarwala, R. (2008). *BLAST Command Line Applications User Manual. Bookshelf ID: NBK 1763*.
- Chen, Y. G., Cui, X. L., Wang, Y. X., Tang, S. K., Zhang, Y. Q., Li, W. J., et al. (2009). *Psychroflexus sediminis* sp. nov., a mesophilic bacterium isolated from salt lake sediment in China. *Int. J. Syst. Evol. Microbiol.* 59, 569–573. doi: 10.1099/ijss.0.003269-0
- Chimileski, S., Dolas, K., Naor, A., Gophna, U., and Papke, R. T. (2014). Extracellular DNA metabolism in *Haloflexax volcanii*. *Front. Microbiol.* 5:57. doi: 10.3389/fmicb.2014.00057
- Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., et al. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 41, D633–D642. doi: 10.1093/nar/gkt1244
- de la Haba, R. R., Sánchez-Porro, C., Márquez, M. C., and Ventosa, V. (2011). “Taxonomy of halophiles,” in *Extremophiles Handbook*, eds K. Horikoshi, G. Antranikian, A. Bull, F. Robb, and K. Stetter (Heidelberg: Springer-Verlag), 255–308.
- Dillon, J. G., Carlin, M., Gutierrez, A., Nguyen, V., and McLain, N. (2013). Patterns of microbial diversity along a salinity gradient in the Guerrero Negro solar saltern, Baja CA Sur, Mexico. *Front. Microbiol.* 4:399. doi: 10.3389/fmicb.2013.00399
- Donachie, S. P., Bowman, J. P., and Alam, M. (2004). *Psychroflexus tropicus* sp. nov., an obligately halophilic *Cytophaga-Flavobacterium-Bacteroides* group bacterium from an Hawaiian hypersaline lake. *Int. J. Syst. Evol. Microbiol.* 54, 935–940. doi: 10.1099/ijss.0.02733-0
- Dulau, N. (1983). *Les Domaines Sédimentaires Préhalitiques des Marais Salants de la Région de Salin-de-Giraud (France) et de Santa Pola (Espagne)*. PhD thesis, 3ème Cycle-Université Louis Pasteur, Strasbourg.
- Elevi Bardavid, R., and Oren, A. (2008). Dihydroxyacetone metabolism in *Salinibacter ruber* and in *Halococcus walsbyi*. *Extremophiles* 12, 125–131. doi: 10.1007/s00792-007-0114-x
- Felsenstein, J. (1989). PHYLIP – phylogeny inference package (Version 3.2). *Cladistics* 5, 164–166.
- Fernández, A. B., Ghai, R., Martín-Cuadrado, A. B., Sánchez-Porro, C., Rodríguez-Valera, F., and Ventosa, A. (2013). Metagenome sequencing of prokaryotic microbiota from two hypersaline ponds of a marine saltern in Santa Pola, Spain. *Genome Announc.* 1:e00933-13. doi: 10.1128/genomeA.00933-13
- Fernández, A. B., Ghai, R., Martín-Cuadrado, A. B., Sánchez-Porro, C., Rodríguez-Valera, F., and Ventosa, A. (2014a). Prokaryotic taxonomic and metabolic diversity of an intermediate salinity hypersaline habitat assessed by metagenomics. *FEMS Microbiol. Ecol.* doi: 10.1111/1574-6941.12329. (In press).
- Fernández, A. B., León, M. J., Vera, B., Sánchez-Porro, C., and Ventosa, A. (2014b). Metagenomic sequence of prokaryotic microbiota from an intermediate salinity pond of a saltern in Isla Cristina, Spain. *Genome Announc.* 2:e00045-14. doi: 10.1128/genomeA.00045-14
- Fox, E. M., and Mendz, G. L. (2006). Phosphonate degradation in microorganisms. *Enzyme Microb. Technol.* 40, 145–150. doi: 10.1016/j.enzmictec.2005.10.047
- Ghai, R., Hernandez, C. M., Picazo, A., Mizuno, C. M., Ininbergs, K., Diez, B., et al. (2012). Metagenomes of Mediterranean coastal lagoons. *Sci. Rep.* 2:490. doi: 10.1038/srep00490
- Ghai, R., Martín-Cuadrado, A. B., Molto, A. G., Heredia, I. G., Cabrera, R., Martín, J., et al. (2010). Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing. *ISME J.* 4, 1154–1166. doi: 10.1038/ismej.2010.44
- Ghai, R., Pasic, L., Fernandez, A. B., Martín-Cuadrado, A. B., Mizuno, C. M., McMahon, K. D., et al. (2011). New abundant microbial groups in aquatic hypersaline environments. *Sci. Rep.* 1:135. doi: 10.1038/srep00135
- Grant, W. D., Gemmell, R. T., and McGenity, T. J. (1998). “Halophiles,” in *Extremophiles: Microbial Life in Extreme Environment*, eds K. Horikoshi and W. D. Grant (New York, NY: Wiley-Liss), 93–132.
- Grant, W. D., Kamekura, M., McGenity, T. J., and Ventosa, A. (2001). “Class III. Halobacteria class. nov.,” in *Bergey's Manual of Systematic Bacteriology, 2nd Edn.*, Vol. 1, eds D. R. Boone, R. W. Castenholz, and G. M. Garrity (New York, NY: Springer), 294–301.
- Grant, W. D., Pagaling, E., Marquez, M. C., Gutierrez, M. C., Cowan, D. A., Ma, Y., et al. (2011). “The hypersaline lakes of Inner Mongolia: the MGATech Project,” in *Halophiles and Hypersaline Environments*, eds A. Ventosa, A. Oren, and Y. Ma (Heidelberg: Springer), 65–107. doi: 10.1007/978-3-642-20198-1\_4
- Javor, B. J. (1989). “Solar salterns,” in *Hypersaline Environments: Microbiology and Biogeochemistry*, ed B. Javor (Berlin: Springer-Verlag), 189–204.
- Jørgensen, B. B. (1982). Ecology of the bacteria of the sulphur cycle with special reference to anoxic-oxic interface environments. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 298, 543–561. doi: 10.1080/rstb.1982.0096
- Landry, J. C., and Jaccard, J. (1984). Chimie des eaux libres dans le marais salant de Santa-Pola, salina de Bras del Port. *Rev. Geol.* 38/39, 37–53.
- Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: an R package for multivariate analysis. *J. Statist. Soft.* 25, 1–18.
- Legault, A. B., Lopez-Lopez, A., Alba-Casado, J. C., Doolittle, W. F., Bolhuis, H., Rodriguez-Valera, F., et al. (2006). Environmental genomics of “*Halococcus walsbyi*” in a saltern crystallizer indicates a large pool of accessory genes in an otherwise coherent species. *BMC Genomics* 7:171. doi: 10.1186/1471-2164-7-171
- Leon, M. J., Ghai, R., Fernandez, A. B., Sanchez-Porro, C., Rodriguez-Valera, F., and Ventosa, A. (2013). Draft genome of *Spiribacter salinus* M19-40, an abundant gammaproteobacterium in aquatic hypersaline environments. *Genome Announc.* 1:e00179-12. doi: 10.1128/genomeA.00179-12
- López-Pérez, M., Ghai, R., Leon, M. J., Rodríguez-Olmos, A., Copa-Patiño, J. L., Soliveri, J., et al. (2013). Genomes of “*Spiribacter*,” a streamlined, successful halophilic bacterium. *BMC Genomics* 14:787. doi: 10.1186/1471-2164-14-787
- Lozupone, C. A., and Knight, R. (2007). Global patterns in bacterial diversity. *Proc. Natl. Acad. Sci. U.S.A.* 104, 11436–11440. doi: 10.1073/pnas.0611525104
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, et al. (2004). ARB: a software environment for sequence data. *Nucleic Acids Res.* 32, 1363–1371. doi: 10.1093/nar/gkh293
- Makhdoomi-Kakhki, A., Amoozegar, M. A., Kazemi, B., Pašić, L., and Ventosa, A. (2012). Prokaryotic diversity in Aran-Bidgol salt lake, the largest hypersaline playa in Iran. *Microbes Environ.* 27, 87–93. doi: 10.1264/jsme2.ME11267
- Martín-Cuadrado, A. B., López-García, P., Alba, J. C., Moreira, D., Monticelli, L., Strittmatter, A., et al. (2007). Metagenomics of the deep Mediterranean, a warm bathypelagic habitat. *PLoS ONE* 2:e914. doi: 10.1371/journal.pone.0000914
- Moreno Mde, L., García, M. T., Ventosa, A., and Mellado, E. (2009). Characterization of *Salicola* sp. IC10, a lipase- and protease-producing extreme halophile. *FEMS Microbiol. Ecol.* 68, 59–71. doi: 10.1111/j.1574-6941.2009.00651.x
- Naor, A., Lapierre, P., Mevarech, M., Papke, R. T., and Gophna, U. (2012). Low species barriers in halophilic archaea and the formation of recombinant hybrids. *Curr. Biol.* 22, 1444–1448. doi: 10.1016/j.cub.2012.05.056
- Oh, D., Porter, K., Russ, B., Burns, D., and Dyall-Smith, M. (2010). Diversity of *Halococcus* and other haloarchaea in three, geographically distant, Australian saltern crystallizer ponds. *Extremophiles* 14, 161–169. doi: 10.1007/s00792-009-0295-6
- Oren, A. (1988). Anaerobic degradation of organic compounds at high salt concentrations. *Antonie van Leeuwenhoek* 54, 267–277. doi: 10.1007/BF00443585
- Oren, A. (1999). Bioenergetic aspects of halophilism. *Microbiol. Mol. Biol. Rev.* 63, 334–348.
- Oren, A. (2008). Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Syst.* 4:2. doi: 10.1186/1746-1448-4-2
- Oren, A. (2013). Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. *Front. Microbiol.* 4:315. doi: 10.3389/fmicb.2013.00315
- Ouellette, M., Makkay, A. M., and Papke, R. T. (2013). Dihydroxyacetone metabolism in *Halococcus volcanii*. *Front. Microbiol.* 4:376. doi: 10.3389/fmicb.2013.00376
- Pagaling, E., Wang, H., Venables, M., Wallace, A., Grant, W. D., Cowan, D. A., et al. (2009). Microbial biogeography of six salt lakes in Inner Mongolia, China, and a salt lake in Argentina. *Appl. Environ. Microbiol.* 75, 5750–5760. doi: 10.1128/AEM.00040-09
- Parnell, J. J., Rompato, G., Latta, L. C. 4th., Pfrender, M. E., Van Nostrand, J. D., He, Z., et al. (2010). Functional biogeography as evidence of gene transfer in hypersaline microbial communities. *PLoS ONE* 5:e12919. doi: 10.1371/journal.pone.0012919

- Pasić, L., Bartual, S. G., Ulrih, N. P., Grabnar, M., and Velikonja, B. H. (2005). Diversity of halophilic archaea in the crystallizers of an Adriatic solar saltern. *FEMS Microbiol. Ecol.* 54, 491–498. doi: 10.1016/j.femsec.2005.06.004
- Pasić, L., Ulrih, N. P., Crnigoj, M., Grabnar, M., and Velikonja, B. H. (2007). Haloarchaeal communities in the crystallizers of two Adriatic solar salterns. *Can. J. Microbiol.* 53, 8–18. doi: 10.1139/w06-091
- Paul, S., Bag, S. K., Das, S., Harvill, E. T., and Dutta, C. (2008). Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. *Genome Biol.* 9:R70. doi: 10.1186/gb-2008-9-4-r70
- Podell, S., Emerson, J. B., Jones, C. M., Ugalde, J. A., Welch, S., Heidelberg, K. B., et al. (2013). Seasonal fluctuations in ionic concentrations drive microbial succession in a hypersaline lake community. *ISME J.* 8, e61692. doi: 10.1038/ismej.2013.221
- Ram Mohan, N., Fullmer, M. S., Makkay, A. M., Wheeler, R., Ventosa, A., Naor, A., et al. (2014). Evidence from phylogenetic and genome fingerprinting analyses suggests rapidly changing variation in *Halorubrum* and *Haloarcula* populations. *Front. Microbiol.* 5:143. doi: 10.3389/fmicb.2014.00143
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 16, 276–277. doi: 10.1016/S0168-9525(00)02024-2
- Rodríguez-Valera, F. (1988). “Characteristics and microbial ecology of hypersaline environments,” in *Halophilic Bacteria*, ed F. Rodríguez-Valera (Boca Raton, FL: CRC Press), 3–30.
- Rodríguez-Valera, F. (1993). “The biology of halophilic bacteria,” in *Introduction to Saline Environments*, eds R. H. Vreeland and L. Hochstein (Boca Raton, FL: CRC Press), 1–23.
- Rodríguez-Valera, F., Ruiz-Berraquero, F., and Ramos-Cormenzana, A. (1981). Characteristics of the heterotrophic bacterial populations in hypersaline environments of different salt concentrations. *Microb. Ecol.* 7, 235–243. doi: 10.1007/BF02010306
- Rodríguez-Valera, F., Ventosa, A., Juez, G., and Imhoff, J. F. (1985). Variation of environmental features and microbial populations with salt concentrations in a multipond saltern. *Microb. Ecol.* 11, 107–115. doi: 10.1007/BF02010483
- Schapira, M., Buscot, M. J., Leterme, S. C., Pollet, T., Chapperon, C., and Seuront, L. (2009). Distribution of heterotrophic bacteria and virus-like particles along a salinity gradient in a hypersaline coastal lagoon. *Aquat. Microb. Ecol.* 54, 171–183. doi: 10.3354/ame01262
- Seiler, H., Bleicher, A., Busse, H. J., Hüfner, J., and Scherer, S. (2012). *Psychroflexus halocasei* sp. nov., isolated from a microbial consortium on a cheese. *Int. J. Syst. Evol. Microbiol.* 62, 1850–1856. doi: 10.1099/ijss.0.034801-0
- Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). “Structure and function of the regulatory genes for the phosphate regulon, in *Escherichia coli*,” in *Phosphate Metabolism and Cellular Regulation in Microorganisms*, eds A. Torriani-Gorini, F. G. Rothman, S. Silver, A. Wright, and E. Yagil (Washington, DC: American Society for Microbiology), 20–25.
- Soppa, J. (2006). From genomes to function: haloarchaea as model organisms. *Microbiology* 152, 585–590. doi: 10.1099/mic.0.28504-0
- Torriani, A., and Ludtke, D. N. (1985). “The pho regulon of *Escherichia coli*,” in *The Molecular Biology of Bacterial Growth*, eds M. Schaeffer, F. C. Neidhardt, J. Ingraham, and N. O. Kjeldgaard (Boston, MA: Jones and Bartlett Publishers), 224–242.
- UniProt Consortium. (2014). Activities at the Universal Protein Resource (UniProt). *Nucleic Acids. Res.* 42, D191–D198. doi: 10.1093/nar/gkt1140
- Ventosa, A. (2006). “Unusual micro-organisms from unusual habitats: hypersaline environments,” in *Prokaryotic Diversity: Mechanism and Significance*, eds N. A. Logan, H. M. Lappin-Scott, and P. C. F. Oyston (Cambridge: Cambridge University Press), 223–253.
- Ventosa, A., Mellado, E., Sanchez-Porro, C., and Marquez, M. C. (2008). “Halophilic and halotolerant micro-organisms from soils,” in *Microbiology of Extreme Soils*, eds P. Dion and C. S. Nautiyal (Berlin: Springer-Verlag), 87–115.
- Walsh, D. A., Papke, R. T., and Doolittle, W. F. (2005). Archaeal diversity along a soil salinity gradient prone to disturbance. *Environ. Microbiol.* 7, 1655–1666. doi: 10.1111/j.1462-2920.2005.00864.x
- Wanner, B. L. (1987). “Phosphate regulation of gene expression in *Escherichia coli*,” in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, Vol. 2, eds F. C. Neidhardt, J. Ingraham, K. B. Low, B. Magasanik, M. Schaeffer, and H. E. Umbarger (Washington, DC: American Society for Microbiology), 1326–1333.
- Wanner, B. L. (1993). Gene regulation by phosphate in enteric bacteria. *J. Cell. Biochem.* 51, 47–54. doi: 10.1002/jcb.240510110
- Yoon, J. H., Kang, S. J., Jung, Y. T., and Oh, T. K. (2009). *Psychroflexus salinarum* sp. nov., isolated from a marine solar saltern. *Int. J. Syst. Evol. Microbiol.* 59, 2404–2407. doi: 10.1099/ijss.0.008359-0
- Zhaxybayeva, O., Stepanauskas, R., Mohan, N. R., and Papke, R. T. (2013). Cell sorting analysis of geographically separated hypersaline environments. *Extremophiles* 17, 265–275. doi: 10.1007/s00792-013-0514-z

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 February 2014; accepted: 12 April 2014; published online: 08 May 2014.

Citation: Fernández AB, Vera-Gargallo B, Sánchez-Porro C, Ghai R, Papke RT, Rodriguez-Valera F and Ventosa A (2014) Comparison of prokaryotic community structure from Mediterranean and Atlantic saltern concentrator ponds by a metagenomic approach. *Front. Microbiol.* 5:196. doi: 10.3389/fmicb.2014.00196

This article was submitted to Extreme Microbiology, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Fernández, Vera-Gargallo, Sánchez-Porro, Ghai, Papke, Rodriguez-Valera and Ventosa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Raman spectroscopy in halophile research

Jan Jehlička<sup>1\*</sup> and Aharon Oren<sup>2</sup>

<sup>1</sup> Institute of Geochemistry, Mineralogy and Mineral Resources, Charles University in Prague, Prague, Czech Republic

<sup>2</sup> Department of Plant and Environmental Sciences, The Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel

**Edited by:**

Antonio Ventosa, University of Sevilla, Spain

**Reviewed by:**

Mohammad A. Amoozegar,  
University of Tehran, Iran

Antonio Ventosa, University of Sevilla, Spain

**\*Correspondence:**

Jan Jehlička, Institute of Geochemistry, Mineralogy and Mineral Resources, Charles University in Prague, Albertov 6, 12843 Prague, Czech Republic  
e-mail: jehlicka@natur.cuni.cz

Raman spectroscopy plays a major role in robust detection of biomolecules and mineral signatures in halophile research. An overview of Raman spectroscopic investigations in halophile research of the last decade is given here to show advantages of the approach, progress made as well as limits of the technique. Raman spectroscopy is an excellent tool to monitor and identify microbial pigments and other biomolecules in extant and extinct halophile biomass. Studies of bottom gypsum crusts from salterns, native evaporitic sediments, halite inclusions, and endoliths as well as cultures of halophilic microorganisms permitted to understand the content, distribution, and behavior of important molecular species. The first papers describing Raman spectroscopic detection of microbiological and geochemical key markers using portable instruments are highlighted as well.

**Keywords:** Raman spectroscopy, halophilic, salterns, gypsum crusts, compatible solutes, carotenoids

## INTRODUCTION

Laser Raman spectroscopy determines the wavelength and strength of inelastically dispersed light from molecules and their arrangement. Much of the laser light is scattered without change in wavelength (Rayleigh scattering), but part of light is dispersed differently and scattered light wavelength is shifted (Raman scattering). Measurements of these wavelength shifts create a unique Raman spectrum for each substance, providing information on chemical bonds and molecular and crystalline structure. Raman spectroscopy is therefore a rapid method of identifying organic compounds, it is generally considered to be non-destructive, and sample preparation is minimal (Schrader et al., 1999; Rösch et al., 2003; Schmitt and Popp, 2006; Jehlička and Edwards, 2008).

Many halophiles are pigmented by carotenoids or carotenoid-like pigments, and these have particularly strong Raman resonance signals (Merlin, 1985). Raman spectroscopy studies of pigmented halophiles in culture, have included red archaea of the family *Halobacteriaceae*, the  $\beta$ -carotene-rich alga *Dunaliella salina*, and *Salinibacter ruber*, a member of the phylum *Bacteroidetes*. The information gained was used to characterize natural communities of halophilic microorganisms in saltern brines and of halophiles trapped in fluid inclusions within halite crystals, recent as well as ancient. The high spatial resolution ( $\sim 1 \mu\text{m}$ ) at which data can be collected makes the method particularly useful in such studies. As halite does not give Raman signals, its presence does not interfere. Raman spectroscopy was also used to characterize the microbial communities colonizing evaporitic crusts such as gypsum precipitating in saltern evaporation ponds and halite in desert environments.

Another application for Raman spectroscopy is the assessment of the presence of organic osmotic solutes within halophilic microorganisms. Such organic, “compatible” solutes are expected to be present in high concentrations within the cells to provide osmotic equilibrium with the outside salt concentration (Roberts, 2000; Oren, 2013b). Therefore, Raman signals attributed to such

compounds are expected to be relatively easily detectable, even if typical osmotic solutes such as glycerol, glycine betaine, ectoine, etc. do not allow observing resonance Raman effects at the same sensitivity as carotenoid pigments.

Such studies are also highly relevant for astrobiology. Miniaturized Raman spectrometers are planned to be included in a Mars lander in the near future. Hypersaline and/or hyperarid environments on Earth are excellent model systems to explore the possibilities and limitations of Raman spectroscopy for the detection of life in general, and more specifically, for the detection of halophilic microorganisms, modern or ancient that may possess carotenoids or other protective pigments, and/or high concentrations of organic osmotic solutes (Vitek et al., 2012; Winters et al., 2013).

## RAMAN ANALYSIS OF MICROBIAL PIGMENTS IN EVAPORITES

The saltern ponds near Eilat (Red Sea coast of Israel) were used as a model system for several Raman spectroscopy studies. The upper layer (down to a few cm) of the gypsum crust that accumulates on the bottom of evaporation ponds of intermediate salinity (150–200 g/l total salts) is yellow-orange, due to the presence of unicellular cyanobacteria (*Aphanothecace*-type), below this a green layer with filamentous *Phormidium*-type cyanobacteria is found, and a third layer colored red-purple by purple sulfur bacteria (*Halochromatium*-type) (Oren et al., 1995, 2009). Pigments expected to be present include carotenoids such as echinenone and myxoxanthophyll in the cyanobacterial layers, chlorophyll *a* and phycocyanin (mainly in the green layers), and bacteriochlorophyll *a* and spirilloxanthin carotenoids in the purple layer (Oren et al., 1995; Oren, 2011). We have used a portable Raman spectrometer with 532 nm excitation laser for fast screening of these pigments in the field. The light handheld instrument allowed obtaining acceptable quality spectra fast and outdoors. Raman bands at 1513, 1153, 1005  $\text{cm}^{-1}$  recorded in the orange layer are

assumed to be related to echinenone. Chlorophyll *a* and phycyanin expected in the green layer were not detected with the green excitation. Raman spectra of the red layer contain major bands at 1510, 1151, and 1004 cm<sup>-1</sup>, documenting the presence of spirilloxanthin-like carotenoids (Jehlička and Oren, 2013).

Portable Raman spectrometers were also used to obtain information on microbial pigments within dry gypsum and halite evaporites in the hyperarid regions of the Atacama Desert, Chile. The halite samples were colonized by halophilic and halotolerant cyanobacteria; gypsum crusts contained cyanobacteria as well as *Trebouxia*-like algal cells. Measurements were performed directly on the rock as well as on the homogenized, powdered samples. Excitation at 532 nm was found superior for the analysis of powdered specimens due to its high sensitivity toward. However, the 785 nm excitation wavelength enabled more sensitive detection of the UV-protecting pigment scytonemin found in the sheath of many desert cyanobacteria (Edwards et al., 2005; Wierzchos et al., 2006; de los Ríos et al., 2010; Jehlička et al., 2011; Víttek et al., 2012, 2013). Raman spectroscopy combined with microscopic imaging of the phototrophic community in Atacama Desert gypsum crusts with cyanobacterial and algal colonization enabled the identification of β-carotene, other carotenoids and phycobiliproteins, and degradation of phycobiliproteins could be monitored (Víttek et al., 2013).

## RAMAN ANALYSIS OF PIGMENTED MICROORGANISMS IN SALTERN CRYSTALLIZER BRINES

NaCl-saturated brines of saltern crystallizers are typically inhabited by three types of carotenoid-rich microorganisms: the green alga *Dunaliella*, archaea of the family *Halobacteriaceae* (e.g., *Haloquadratum walsbyi*), and the bacterium *Salinibacter ruber* (Oren and Dubinsky, 1994; Antón et al., 2002; Oren, 2009, 2013a,c). Similar red brines are prominent in the northern arm of Great Salt Lake (Utah, USA) and in other hypersaline lakes (Post, 1977; Oren, 1988).

Although Raman spectra of carotenoids have many features in common, there are also differences that can be used to discriminate between compounds. The C=C band is an indicator of conjugated chain length, which can therefore be used for diagnostic purposes: the wavenumber of the C=C mode is decreased in value as the conjugation length increases. Bacterioruberin of haloarchaea has a rather different molecular structure with a primary conjugated isoprenoid chain length of 13 C=C units with no subsidiary conjugation arising from terminal groups, which contain four OH group functionalities only. A first Raman study haloarchaeal carotenoids was published by Marshall et al. (2007), who suggested that another haloarchaeal pigment, bacteriorhodopsin, may be of interest as a potential biomarker for exobiology purposes. The Raman spectrum of pure bacteriorhodopsin shows a C=C bonds-related band at 1536 cm<sup>-1</sup>. We recently have provided a comparison of different members of the *Halobacteriaceae* (*Halobacterium salinarum*, *Halorubrum sodomense*, *Haloarcula vallismortis*) (Jehlička et al., 2013), and *Salinibacter ruber*, which harbors salinixanthin, a C<sub>40</sub>-carotenoid acyl glycoside (Lutnæs et al., 2002). We did not find any Raman signatures of bacteriorhodopsin in our preparations. A similar observation was reported by Fendrihan et al. (2009). Probably the bacteriorhodopsin

content in the samples examined was too low for detection. In the salinixanthin spectrum we found signals corresponding to a carotenoid with 11 C=C in the primary isoprenoid chain conjugation plus one more in a terminal alicyclic ring; the three strongest bands in the spectrum arise from the isoprenoid chain. Other assignments of salinixanthin resonance bands included evidence for the glycosidic stretching mode of the associated sugar ring at 1187 cm<sup>-1</sup> and a COH stretching mode at 1041 cm<sup>-1</sup> (Jehlička et al., 2013).

Based on the information obtained with pure cultures, it now became possible to analyze biomass collected from saltern crystallizer brines by Raman spectroscopy. Planktonic biomass was collected by centrifugation, a treatment that yields pellets of prokaryotic cells (*Halobacteriaceae*, *Salinibacter*), but removes the major part of β-carotene-containing *Dunaliella* cells as these do not sink during centrifugation (Oren and Dubinsky, 1994; Oren, 2009). Bacterioruberin was the major carotenoid in a cell pellet collected. No distinctive signals for salinixanthin were detected. This was expected: salinixanthin is always a minor component of the community, and in Eilat it was found in much smaller amounts than e.g., in similar saltern ponds in Spain (Oren and Rodríguez-Valera, 2001).

Raman spectroscopy thus has a great potential in assessing presence of different carotenoid compounds of microbial communities living at the highest salt concentrations (Jehlička et al., 2013). However, due to the large extent of chemical similarities and the potential overlap of Raman bands, the technique must be complemented with other techniques of separation and identification of carotenoids such as HPLC.

## USE OF RAMAN SPECTROSCOPY TO CHARACTERIZE BIOMARKERS AND MICROORGANISMS TRAPPED WITHIN RECENT AND ANCIENT SALT CRYSTALS

When hypersaline brines are evaporated to yield halite, small amounts of brine are often included within the crystals. Halophilic microorganisms contained in the brine may become entrapped within the inclusions and survive there for prolonged periods (Schubert et al., 2009). There have been claims that microorganisms may be able to survive there for many millions of years.

As organisms inhabiting saturated brines are generally rich in carotenoid pigments, and as Raman microscopes have a high resolution to the scale of a few micrometers or less, and as halite does not interfere with the analysis, Raman spectroscopy is an excellent technique to monitor presence of microorganisms entrapped within salt crystals, modern as well as ancient. This idea was pioneered by Fendrihan et al. (2009), who compared Raman spectra of pure β-carotene to carotenoids within halophilic archaea (*Halococcus dombrowskii*, *Hcc. morrhuae*, *Halobacterium salinarum*, *Hbt. noricense*, *Haloarcula japonica*, *Halorubrum saccharovorum*, *Halorubrum chaoviator*) sealed in laboratory-grown halite crystals. The cells were embedded in halite by drying cell suspensions overnight on a glass microscope slide or on a quartz dish to simulate evaporitic conditions, and the crystals were investigated by FT-Raman spectroscopy using excitation at 1064 nm and dispersive microRaman spectroscopy at 514.5 nm. The spectra showed prominent peaks at 1507, 1152, and 1002 cm<sup>-1</sup>,

attributed to bacterioruberins. Similar studies permitted to show how Raman microspectrometry can be used for detecting amino acids (glycine, L-alanine,  $\beta$ -alanine, L-serine) in water solutions contained in synthetic inclusions in halite (Osterrothová and Jehlička, 2011).

As a preparation for future deployment of Raman instruments for the search for traces of life on Mars (see section Implications for the Search for Halophilic Life on Mars), Raman microspectrometry was tested as a non-destructive method of determining the lowest detectable  $\beta$ -carotene content within evaporitic matrices: gypsum, halite, and epsomite. Signals were obtained at  $\beta$ -carotene concentrations down to the  $0.1\text{--}10\text{ mg kg}^{-1}$  level in halite, the number Raman bands observed depending on the mineral matrix and the excitation wavelength used (514.5 nm being more sensitive than 785 nm for detection of carotenoids). Within gypsum and epsomite crystals, one order of magnitude higher concentrations were required for detection (Vítek et al., 2009).

Portable and miniaturized Raman spectrometers were also applied to the detection of microbial biomarkers in natural halite from the hyperarid region of the Atacama Desert. Measurements were performed directly on the rock and on homogenized, powdered rock samples. For carotenoid analysis, excitation at 532 nm excitation was found to be superior for the analysis of powdered specimens, but excitation at 785 nm enabled better detection of scytonemin, another pigment present in the material examined (Vítek et al., 2012).

A highly exciting application of Raman spectroscopy to detect life forms within salt crystals is the finding of carotenoids in brine inclusions in ancient halite, 9 ka to 1.44 Ma in age, from bore-hole cores from Death Valley, Saline Valley, and Searles Lake, CA. These carotenoids occurred within fluid inclusions as colorless to red-brown amorphous and crystalline masses associated with spheroidal algal cells resembling *Dunaliella*. Raman spectra of authentic carotenoid standards ( $\beta$ -carotene, lycopene, lutein) showed the same characteristic bands ( $1000\text{--}1020\text{ cm}^{-1}$ ,  $1150\text{--}1170\text{ cm}^{-1}$ , and  $1500\text{--}1550\text{ cm}^{-1}$ ) as found in the ancient halite. Carotenoids appear well-preserved in ancient salt, a finding that supports reports of the recovery of preserved DNA and even of live cells from fluid inclusions in buried halite deposits (Winters et al., 2013).

### RAMAN SPECTRA OF ORGANIC OSMOTIC SOLUTES IN BACTERIAL CULTURES AND IN MICROBIAL COMMUNITIES IN EVAPORITIC CRUSTS

A few halophiles (the haloarchaea, *Salinibacter*) use KCl to provide osmotic balance of their cytoplasm, but most other halophilic and halotolerant prokaryotic and eukaryotic microbes accumulate high concentrations of organic osmotic solutes ("compatible solutes") to protect the cells against extreme osmotic pressures. This strategy of osmotic adaptation does not require extensive changes in the structure of intracellular enzymes. A great diversity of organic osmotic solutes has been identified in the microbial world. Examples are glycerol produced by *Dunaliella*, simple sugars (sucrose, trehalose), amino acids (proline, glutamic acid,  $\beta$ -glutamine), and amino acid derivatives [e.g., glycine betaine, ectoine ((S)-2-methyl-1,4,5,6-tetrahydropyrimidine-4-carboxylic acid)], 5-hydroxyectoine, and

N $\epsilon$ -acetyl- $\beta$ -lysine (Galinski et al., 1985; Wohlfarth et al., 1990; Severin et al., 1992; Galinski, 1995; Welsh, 2000; Roberts, 2006). There are solutes found thus far in one organism only, such as N- $\alpha$ -carbamoyl-L-glutamine 1-amide from *Ectothiorhodospira marismortui* (Galinski and Oren, 1991). Many microorganisms can accumulate more than one compatible solute. Glycine betaine is widely found in photosynthetic prokaryotes, oxygenic as well as anoxygenic, but few heterotrophic prokaryotes are able of its *de novo* biosynthesis. However, when available (e.g., in media containing yeast extract), many heterotrophs accumulate glycine betaine from the growth medium (Imhoff and Rodriguez-Valera, 1984; Oren, 2013b). Many moderately halophilic/halotolerant cyanobacteria produce the heteroside glucosylglycerol [2-O- $\alpha$ -D-glucopyranosyl-(1 $\rightarrow$ 2)-glycerol], while the most salt tolerant strains accumulate glycine betaine (Mackay et al., 1984; Hagemann, 2011).

As osmotic solutes must be present in high concentrations intracellularly to provide the necessary osmotic equilibrium, we hypothesized that they may form suitable biomarkers to be identified by Raman spectroscopy. We therefore prepared a database of Raman spectra of commonly encountered solutes, including ectoine, hydroxyectoine, glycine betaine, glucosylglycerol, mannosylglycerate (potassium salt), and di-myo-inositol phosphate, complementing existing information on the Raman spectra of solutes such as glycerol, sucrose, and trehalose (Jehlička et al., 2012a) (Table 1). The purpose was to establish a tool for the rapid assessment of the osmotic adaptation mechanism used by halophiles in culture and in natural settings, including the use of miniaturized portable Raman spectrometers to probe halophilic microbial communities in the field.

We then tested for the presence of organic compatible solutes within cells of two model organisms, both members of the *Gammaproteobacteria*: the heterotrophic *Halomonas elongata* and the anoxygenic phototroph *Ectothiorhodospira marismortui*. We did not succeed detecting organic osmotic solutes directly in cell pellets, and therefore extraction and purification methods were found essential. Perchlorate extraction, followed by desalting on an ion retardation column and lyophilization, a method also recommended for analysis of compatible solutes by HPLC and NMR techniques (Wohlfarth et al., 1990; Severin et al., 1992; Roberts, 2006) yielded good results (Jehlička et al., 2012b). Because of the time and effort needed for sample preparation some of the normal advantages of Raman spectroscopy over other methods do not apply here. However, the rapid analysis by Raman spectroscopy, typically a few minutes only, remains a clear advantage. Near-infrared 785 nm lasers were found especially suitable for the identification of compatible solutes.

*H. elongata* is known to accumulate glycine betaine when grown in complex media containing yeast extract, but to produce ectoine in defined media on e.g., glucose as carbon and energy source (Imhoff and Rodriguez-Valera, 1984; Wohlfarth et al., 1990; Severin et al., 1992). Raman signals attributable to glycine betaine were found in extracts of cells grown in rich media. Extracts of *H. elongata* cells grown on glucose, in contrast, showed characteristic medium- and strong intensity bands permitting the unambiguous identification of the compatible solute present as ectoine (Jehlička et al., 2012b).

**Table 1 | Main Raman bands and intensities for pigments and compatible solutes.**

Compound	Origin	Prominent Raman bands observed	References
<b>PIGMENTS</b>			
Bacterioruberin	<i>Halobacterium salinarum</i> NRC-1 culture	1505 s, 1152 s, 1000 s	Marshall et al., 2007
	<i>Halococcus dombrowskii</i> culture	1507 s, 1152 s, 1002 s	Fendrihan et al., 2009
	<i>Halorubrum sodomense</i> culture	1506 s, 1152 s, 1001 s	Jehlička et al., 2013
	<i>Haloarcula vallismortis</i> culture	1506 s, 1151 s, 1000 s	Jehlička et al., 2013
Salinixanthin	<i>Salinibacter ruber</i> culture	1512 s, 1155 s, 1003 s	Jehlička et al., 2013
Bacteriorhodopsin	Standard	1526 s, 1199 m, 1168 m, 1150 m, 1007 s	Marshall et al., 2007
Chlorophyll	Native crusts Atacama cyanobacteria	1445 m, 1308 m, 745 m, 508 m	Vítek et al., 2013
Scytonemin	<i>Lyngbya aestuarii</i> from native mat	1595 m, 1554 m, 1173 m	Edwards et al., 2000
	Native crusts Atacama	1596 s, 1558 s, 1321 s, 1172 m	Vítek et al., 2010
Phycobiliproteins	Native crusts Atacama	1631 w, 1583 w, 1370 w, 1283 w, 1235 w, 874 w, 815 w, 665 w	Vítek et al., 2013
<b>COMPATIBLE SOLUTES</b>			
Glycerol	Standard	1465 m, 1315 m, 1257 m, 1110 s, 1055 s, 923 m, 850 s, 820 s, 675 m, 485 m, 416 m, 392 s	De Gelder et al., 2007
Trehalose	Standard	1455 m, 1411 m, 1386 m, 1371 m, 1358 vs, 1330 m, 1149 vs, 1120 vs, 1102 s, 1080 m, 1061 m, 1018 m, 912 vs, 838 vs, 804 m, 697 m, 671 m, 604 m, 540 s, 523 vs, 450 s, 430 m, 407 vs, 369 vs	De Gelder et al., 2007
Glutamate	Standard	1434 m, 1421 s, 1401 vs, 1353 m, 1341 s, 1317 s, 1293 m, 1282 m, 1142 m, 1095 m, 942 vs, 925 m, 876 s, 857 vs, 775 m, 529 m, 499 m	De Gelder et al., 2007
L-proline	Standard	1454 s, 1389 m, 1378 m, 1286 m, 1267 m, 1240 m, 1083 m, 1045 m, 1035 m, 994 m, 947 m, 930 m, 916 vs, 899 vs, 877 m, 850 s, 842 vs, 834 vs, 642 m, 452 vs	De Gelder et al., 2007
Ectoine	Standard	3009 m, 2981 m, 2944 s, 2895 m, 1469 m, 1444 m, 1392 m, 1366 w, 1316 ms, 1280 ms, 1215 w, 1207 vw, 1137 m, 1072 w, 1032 w, 999 w, 850 s, 804 vw, 692 vw, br, 657 vw, 599 w, 570 w, 498 w, 432 w	Jehlička et al., 2012a
	<i>Halomonas elongata</i> culture (minimal medium)	2985 m, sh, 2937 s, br, 2888 m, br, 1633 w, 1465 vw, sh, 1439 w, sh, 1394 s, 1319 vw, sh, 1275 m, 1212 w, sh, 1136 w, 1030 vw, 1000 vw, 903 vw, 845 m, 803 vw, 686 vw, 655 vw, 596 w, 571 w, 435 vw	Jehlička et al., 2012b
Hydroxyectoine	Standard	3017 m, 2935 s, 861 w, 1640 w, 1613 m, 1438 m, 1349 m, 1318 ms, 1263 s, 1213 m, 1151 m, 1092 ms, 1014 s, 970 w, 901 m, 890 m, 828 ms, 751 s, 631 m, 558 m, 395 m, 381 s, 359 m, 311 m	Jehlička et al., 2012a
Glycine betaine	Standard	3375 m, 3290 m, 3053 m, 3034 m, 2982 m, 2969 s, 2867 w, 2937 m, 1650 w, 1462 s, 1427 m, 1412 w, 1343 m, 1132 w, 953 w, 939 w, 895 m, 783 s, 539 m, 363 m	Ilczyszyn and Ratajczak, 1996

(Continued)

**Table 1 | Continued**

Compound	Origin	Prominent Raman bands observed	References
	<i>Halomonas elongata</i> culture (complex medium)	3050 w, 3031 m, 2978 m, 2965 s, 2942 w, 2935 m, 2885 w, 2880 w, 2864 w, 2784 m, 1613 vw, 1458 m, 1423 w, 1410 w, 1340 w, 1129 w, 974 m, 952 m, 933 s, 896 w, 780 s, 725 w, 538 m, 457 w, 361 m	Jehlička et al., 2012b
	<i>Ectothiorhodospira marismortui</i> culture	3059 w, 3032 m, 2966 m, 2936 m, 1645 vw, 1625 vw, 1457 m, 1410 vw, 1394 vw, 1342 vw, 1211 vw, 1127 vw, br, 976 w, 954 w, 933 s, 897 w, 779 s, 540 m, 456 w, 422 w, 366 w, 338 w	Jehlička et al., 2012b
	Native gypsum crust (Eilat, Israel, red layer)	3058 w, br, 3033 m, 2969 s, 2939 ms, 2880 m, 1458 m, 1452 m, 1421 w, sh, 1337 mW, 1212 w, 1131 vw, 978 w, 956 m, 934 s, 911 mW, 899 mW, 780 s, 730 vw, br, 684 vw, 542 mW, 458 w, 369 w, 339 vw	Oren et al., 2013
Glucosylglycerol		2930 br, 2907 s, 1460 s, 1366 s, br, 1340 s, br, 1265 w, br, 1127 m, 1103 w, 955 w, 897 m, 842 s, 755 w, 707 w, 666 w, 544 m, 514 m, 435 w	Jehlička et al., 2012a
Mannosylglycerate (K salt)		3585 m, 3405 m, 2975 m, 2956 m, 2938 s, 2931 s, 2906 m, 1461 m, 1353 m, 1308 m, 1153 m, 1135 m, 1107 m, 1093 m, 1069 s, 1055 m, 1004 m, 958 m, 892 s, 825 s, 671 m, 589 m, 517 s, 429 w, 390 m, 350 m	Jehlička et al., 2012a

s, strong bands; m, medium bands; vw, very weak bands; mw, weak bands; w, weak bands; br, broad bands; sh, shoulder.

In desalinated perchlorate extracts of *E. marismortui* a complete series of sharp bands corresponding to glycine betaine was recorded, including bands in the second order area. Trehalose was not found, nor were any signals obtained that may have corresponded to the minor osmotic solute *N*- $\alpha$ -carbamoyl-L-glutamine 1-amide (a compound not included yet in the library of Raman resonance signals) (Jehlička et al., 2012b).

The methodology developed for extraction and Raman analysis of compatible solutes in bacterial cultures (*Halomonas*, *Ectothiorhodospira*) (Jehlička et al., 2012b) was also applied to the analysis of the stratified microbial communities in an evaporitic gypsum crust found in an evaporation pond (~194 g/l total dissolved salts) of the salterns of Salt of the Earth Eilat (Oren et al., 1995; see also section Raman Analysis of Microbial Pigments in Evaporites). Solutes were extracted from the upper yellow-orange layer dominated by unicellular cyanobacteria, the green layer with filamentous cyanobacteria, and the layer colored red-purple by purple sulfur bacteria, all accompanied by dense communities of heterotrophic bacteria. Extracts were analyzed by Raman spectroscopy, by  $^1\text{H}$  and  $^{13}\text{C}$  nuclear magnetic resonance and by HPLC. The only osmotic solute detected in all three layers was glycine betaine; ectoine and other known solutes were not found. Most heterotrophic bacteria cannot produce glycine betaine but preferentially accumulate it when it is supplied. Presence of glycine betaine produced by the photoautotrophic members of the community may therefore relieve the heterotrophs from the need to synthesize other compounds at a high energetic cost (Oren et al., 2013).

## IMPLICATIONS FOR THE SEARCH FOR HALOPHILIC LIFE ON MARS

A miniaturized Raman spectrometer is scheduled to fly as part of the analytical instrumentation on an ESA remote robotic lander

in the ESA/Roscosmos ExoMars mission, planned to be launched in 2018. Salts are present on Mars, including ancient salts formed in periods when the climate may have been more suitable for life than it is today. In view of the longevity of halophiles on Earth when entrapped within salt crystals, and the presence of carotenoid pigments detectable by Raman spectroscopy in such ancient salt samples (Fendrihan et al., 2009; Winters et al., 2013), Raman techniques may be excellent tools for life detection systems to be sent to Mars (Edwards et al., 2005, 2013; Vítek et al., 2009, 2012). Whether halophilic microorganisms resembling the current ones on Earth are or were present in the past on Mars is yet unknown, but if they are, Raman spectroscopy may well help us discovering them.

## ACKNOWLEDGMENTS

We thank Lily Mana and Rahel Elevi Bardavid for their contributions to work, Salt of the Earth Eilat Ltd., for allowing access to the Eilat salterns, and the Interuniversity Institute for Marine Sciences of Eilat for logistic support. This study was supported by grant no. P210/10/0467 from the Grant Agency of the Czech Republic, by institutional support MSM0021620855 from the Ministry of Education of the Czech Republic, and by grant no. 1103/10 from the Israel Science Foundation.

## REFERENCES

- Antón, J., Oren, A., Benloch, S., Rodríguez-Valera, F., Amann, R., and Rosselló-Móra, R. (2002). *Salinibacter ruber* gen. nov., sp. nov., a novel extreme halophilic member of the Bacteria from saltern crystallizer ponds. *Int. J. Syst. Evol. Microbiol.* 52, 485–491. doi: 10.1099/ijss.0.01913-0
- De Gelder, J., de Gussem, K., Vandenebeele, P., and Moens, L. (2007). Reference database of Raman spectra of biological molecules. *J. Raman Spectrosc.* 38, 1133–1147. doi: 10.1002/jrs.1734
- de los Ríos, A., Valea, S., Ascaso, C., Davila, A., Kaštovský, J., McKay, C. P., et al. (2010). Comparative analysis of the microbial communities inhabiting

- halite evaporites of the Atacama Desert. *Int. Microbiol.* 13, 79–89. doi: 10.2436/20.1501.01.113
- Edwards, H. G. M., Garcia-Pichel, E., Newton, E. M., and Wynn-Williams, D. D. (2000). Vibrational Raman spectroscopic study of scytonemin, the UV-protective cyanobacterial pigment. *Spectrochim. Acta A* 56, 193–200. doi: 10.1016/S1386-1425(99)00218-8
- Edwards, H. G. M., Hutchinson, I. B., Ingleby, R., Parnell, J., Vitek, P., and Jehlička, J. (2013). Raman spectroscopic analysis of geological and biogeological specimens of relevance to the ExoMars mission. *Astrobiology* 13, 543–549. doi: 10.1089/ast.2012.0872
- Edwards, H. G. M., Jorge Villar, S. E., Parnell, J., Cockell, C. S., and Lee, P. (2005). Raman spectroscopic analysis of cyanobacterial gypsum halotrophs and relevance for sulfate deposits on Mars. *Analyst* 130, 917–923. doi: 10.1039/b503533c
- Fendrihan, S., Musso, M., and Stan-Lotter, H. (2009). Raman spectroscopy as a potential method for the detection of extremely halophilic archaea embedded in halite in terrestrial and possibly extraterrestrial samples. *J. Raman Spectrosc.* 40, 1996–2003. doi: 10.1002/jrs.2357
- Galinski, E. A. (1995). Osmoadaptation in bacteria. *Adv. Microb. Physiol.* 37, 272–328. doi: 10.1016/S0065-2911(08)60148-4
- Galinski, E. A., and Oren, A. (1991). Isolation and structure determination of a novel compatible solute from the moderately halophilic purple sulfur bacterium *Ectothiorhodospira marismortui*. *Eur. J. Biochem.* 198, 593–598. doi: 10.1111/j.1432-1033.1991.tb16055.x
- Galinski, E. A., Pfeiffer, H.-P., and Trüper, H. G. (1985). 1,4,5,6-tetrahydro-2-methyl-4-pyrimidinecarboxylic acid. A novel cyclic amino acid from halophilic phototrophic bacteria of the genus *Ectothiorhodospira*. *Eur. J. Biochem.* 149, 135–139. doi: 10.1111/j.1432-1033.1985.tb08903.x
- Hagemann, M. (2011). Molecular biology of cyanobacterial salt acclimation. *FEMS Microbiol. Rev.* 35, 87–123. doi: 10.1111/j.1574-6976.2010.00234.x
- Ilczyszyn, M. M., and Ratajczak, H. (1996). Polarized vibrational spectra of betaine monohydrate single crystal. *Vibrat. Spectrosc.* 10, 177–184. doi: 10.1016/0924-2031(95)00040-2
- Imhoff, J. F., and Rodriguez-Valera, F. (1984). Betaine is the main compatible solute of halophilic eubacteria. *J. Bacteriol.* 160, 478–479.
- Jehlička, J., and Edwards, H. G. M. (2008). Raman spectroscopy as a tool for the non-destructive identification of organic minerals in the geological record. *Org. Geochem.* 39, 371–386. doi: 10.1016/j.orggeochem.2008.01.005
- Jehlička, J., and Oren, A. (2013). Use of a handheld Raman spectrometer for fast screening of microbial pigments in cultures of halophilic microorganisms and in microbial communities in hypersaline environments in nature. *J. Raman Spectrosc.* 43, 1285–1291. doi: 10.1002/jrs.4362
- Jehlička, J., Oren, A., and Edwards, H. G. M. (2012a). Raman spectra of compatible osmotic solutes of halophiles. *J. Raman Spectrosc.* 43, 1134–1140. doi: 10.1002/jrs.3136
- Jehlička, J., Oren, A., and Vitek, P. (2012b). Use of Raman spectroscopy for identification of compatible solutes in halophilic bacteria. *Extremophiles* 16, 507–514. doi: 10.1007/s00792-012-0450-3
- Jehlička, J., Oren, A., and Edwards, H. G. M. (2013). Bacterioruberin and salinixanthin carotenoids of extremely halophilic Archaea and Bacteria: a Raman spectroscopic study. *Spectrochim. Acta A* 106, 99–103. doi: 10.1016/j.saa.2012.12.081
- Jehlička, J., Vitek, P., Oren, A., and Edwards, H. G. M. (2011). Raman spectroscopic identification of evaporitic minerals and biomarkers using miniaturised portable devices. *Mineral. Mag.* 75, 1107. ISSN: 0026-461X.
- Lutnaes, B. E., Oren, A., and Lælaen-Jensen, S. (2002). New C<sub>40</sub>-carotenoid acyl glycoside as principal carotenoid of *Salinibacter ruber*, an extremely halophilic eubacterium. *J. Nat. Prod.* 65, 1340–1343. doi: 10.1021/np020125c
- Mackay, M. A., Norton, R. S., and Borowitzka, L. J. (1984). Organic osmoregulatory solutes in cyanobacteria. *J. Gen. Microbiol.* 130, 2177–2191. doi: 10.1099/00221287-130-9-2177
- Marshall, C. P., Leuko, S., Coyle, C. M., Walter, M. R., Burns, B. P., and Neilan, B. A. (2007). Carotenoid analysis of halophilic Archaea by resonance Raman Spectroscopy. *Astrobiology* 7, 631–643. doi: 10.1089/ast.2006.0097
- Merlin, J. C. (1985). Resonance Raman spectroscopy of carotenoids and carotenoid-containing systems. *Pure Appl. Chem.* 57, 785–792. doi: 10.1351/pac198557050785
- Oren, A. (1988). “The microbial ecology of the Dead Sea,” in *Advances in Microbial Ecology*, Vol. 10, ed K. C. Marshall (New York, NY: Plenum Publishing Company), 193–229. doi: 10.1007/978-1-4684-5409-3\_6
- Oren, A. (2009). “Microbial diversity and microbial abundance in salt-saturated brines: why are the waters of hypersaline lakes red?” in *Saline Lakes Around the World: Unique Systems with Unique Values*, eds A. Oren, D. L. Naftz, P. Palacios, and W. A. Wurtsbaugh (Salt Lake City: Utah State University, The S. J. and Jessie E. Quinney Natural Resources Research Library, College of Natural Resources), 247–255.
- Oren, A. (2011). “Characterization of pigments of prokaryotes and their use in taxonomy and classification,” in *Taxonomy of Prokaryotes - Methods in Microbiology*, Vol. 38, eds F. A. Rainey and A. Oren (Amsterdam: Elsevier/Academic Press), 262–283. doi: 10.1016/B978-0-12-387730-7.00012-7
- Oren, A. (2013a). “The family *Halobacteriaceae*,” in *The Prokaryotes. A Handbook on the Biology of Bacteria: Ecophysiology and Biochemistry*, 4th Edn., eds E. Rosenberg, E. F. DeLong, F. Thompson, S. Lory, and E. Stackebrandt (New York, NY: Springer).
- Oren, A. (2013b). “Life at high salt concentrations,” in *The Prokaryotes. A Handbook on the Biology of Bacteria: Ecophysiology and Biochemistry*, 4th Edn., eds E. Rosenberg, E. F. DeLong, F. Thompson, S. Lory, and E. Stackebrandt (New York, NY: Springer).
- Oren, A. (2013c). *Salinibacter*, an extremely halophilic bacterium with archaeal properties. *FEMS Microbiol. Lett.* 342, 1–9. doi: 10.1111/1574-6968.12094
- Oren, A., and Dubinsky, Z. (1994). On the red coloration of saltern crystallizer ponds. II. Additional evidence for the contribution of halobacterial pigments. *Int. J. Salt Lake Res.* 3, 9–13. doi: 10.1007/BF01990638
- Oren, A., Elevi Bardavid, R., Kandel, N., Aizenshtat, Z., and Jehlička, J. (2013). Glycine betaine is the main organic osmotic solute in a stratified microbial community in a hypersaline evaporitic gypsum crust. *Extremophiles* 17, 445–451. doi: 10.1007/s00792-013-0522-z
- Oren, A., Kühl, M., and Karsten, U. (1995). An endoevaporitic microbial mat within a gypsum crust: zonation of phototrophs, photopigments, and light penetration. *Mar. Ecol. Prog. Ser.* 128, 151–159. doi: 10.3354/meps128151
- Oren, A., and Rodriguez-Valera, F. (2001). The contribution of *Salinibacter* species to the red coloration of saltern crystallizer ponds. *FEMS Microbiol. Ecol.* 36, 123–130.
- Oren, A., Sørensen, K. B., Canfield, D. E., Teske, A. P., Ionescu, D., Lipski, A., et al. (2009). Microbial communities and processes within a hypersaline gypsum crust in a saltern evaporation pond (Eilat, Israel). *Hydrobiologia* 626, 15–26. doi: 10.1007/s10750-009-9734-8
- Osterrothová, K., and Jehlička, J. (2011). Investigation of biomolecules trapped in fluid inclusions inside halite crystals by Raman spectroscopy. *Spectrochim. Acta A* 83, 288–296. doi: 10.1016/j.saa.2011.08.032
- Post, F. J. (1977). The microbial ecology of the Great Salt Lake. *Microb. Ecol.* 3, 143–165. doi: 10.1007/BF02010403
- Roberts, M. F. (2000). Osmoadaptation and osmoregulation in Archaea. *Front. Biosci.* 5, d796–d812. doi: 10.2741/roberts
- Roberts, M. F. (2006). “Characterization of organic compatible solutes of halotolerant and halophilic microorganisms,” in *Extremophiles - Methods in Microbiology*, Vol. 35, eds F. A. Rainey and A. Oren (Amsterdam: Elsevier/Academic Press), 615–627.
- Rösch, P., Schmitt, M., Kiefer, W., and Popp, J. (2003). The identification of microorganisms by micro-Raman spectroscopy. *J. Mol. Struct.* 661–662, 363–369. doi: 10.1016/j.molstruc.2003.06.004
- Schmitt, M., and Popp, J. (2006). Raman spectroscopy at the beginning of the twenty-first century. *J. Raman Spectrosc.* 37, 20–28. doi: 10.1002/jrs.1486
- Schrader, B., Klump, H. H., Schenzel, K., and Schulz, H. (1999). Non-destructive NIR FT Raman analysis of plants. *J. Mol. Struct.* 509, 201–212. doi: 10.1016/S0022-2860(99)00221-5
- Schubert, B. A., Lowenstein, T. K., and Timofeeff, M. N. (2009). Microscopic identification of prokaryotes in modern and ancient halite, Saline Valley and Death Valley, California. *Astrobiology* 9, 467–482. doi: 10.1089/ast.2008.0282
- Severin, J., Wohlforth, A., and Galinski, E. A. (1992). The predominant role of recently discovered tetrahydropyrimidines for the osmoadaptation of halophilic eubacteria. *J. Gen. Microbiol.* 138, 1629–1638. doi: 10.1099/00221287-138-8-1629
- Vitek, P., Cámará-Gallego, B., Edwards, H. G. M., Jehlička, J., Ascaso, C., and Wierzchos, J. (2013). Phototrophic community in gypsum crust from the Atacama Desert studied by Raman spectroscopy and microscopic imaging. *Geomicrobiol. J.* 30, 399–410. doi: 10.1080/01490451.2012.697976
- Vitek, P., Edwards, H. G. M., Jehlička, J., Ascaso, C., De los Ríos, A., Valea, S., et al. (2010). Microbial colonization of halite from the hyper-arid Atacama

- desert studied by Raman spectroscopy. *Philos. Trans. A Math. Phys. Eng. Sci.* 368, 3205–3221. doi: 10.1098/rsta.2010.0059
- Vítěk, P., Jehlička, J., Edwards, H. G. M., Hutchinson, I., Ascaso, C., and Wierzchos, J. (2012). The miniaturized Raman system and detection of traces of life in halite from the Atacama Desert: some considerations for the search for life signatures on Mars. *Astrobiology* 12, 1095–1099. doi: 10.1089/ast.2012.0879
- Vítěk, P., Jehlička, J., Edwards, H. G. M., and Osterrothová, K. (2009). Identification of  $\beta$ -carotene in an evaporitic matrix—evaluation of Raman spectroscopic analysis for astrobiological research on Mars. *Anal. Bioanal. Chem.* 393, 1967–1975. doi: 10.1007/s00216-009-2677-0
- Welsh, D. T. (2000). Ecological significance of compatible solute accumulation by micro-organisms: from single cells to global climate. *FEMS Microbiol. Rev.* 24, 263–290. doi: 10.1111/j.1574-6976.2000.tb00542.x
- Wierzchos, J., Ascaso, C., and McKay, C. P. (2006). Endolithic cyanobacteria in halite rocks from the hyperarid core of the Atacama Desert. *Astrobiology* 6, 415–422. doi: 10.1089/ast.2006.6.415
- Winters, Y. D., Lowenstein, T. K., and Timofeff, M. N. (2013). Identification of Carotenoids in Ancient Salt from Death Valley, Saline Valley, and Searles Lake, California using Laser Raman Spectroscopy. *Astrobiology* 13, 1065–1080. doi: 10.1089/ast.2012.0952
- Wohlfarth, A., Severin, J., and Galinski, E. A. (1990). The spectrum of compatible solutes in heterotrophic halophilic eubacteria of the family *Halomonadaceae*. *J. Gen. Microbiol.* 136, 705–712. doi: 10.1099/00221287-136-4-705

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Received: 02 September 2013; paper pending published: 10 November 2013; accepted: 22 November 2013; published online: 10 December 2013.*

*Citation: Jehlička J and Oren A (2013) Raman spectroscopy in halophile research. *Front. Microbiol.* 4:380. doi: 10.3389/fmicb.2013.00380*

*This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.*

*Copyright © 2013 Jehlička and Oren. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# A Glimpse of the genomic diversity of haloarchaeal tailed viruses

Ana Senčilo and Elina Roine \*

Department of Biosciences and Institute of Biotechnology, University of Helsinki, Helsinki, Finland

**Edited by:**

R. Thane Papke, University of Connecticut, USA

**Reviewed by:**

Antonio Ventosa, University of Sevilla, Spain

Takuro Nunoura, Japan Agency for Marine-Earth Science & Technology, Japan

**\*Correspondence:**

Elina Roine, Department of Biosciences and Institute of Biotechnology, University of Helsinki, PO Box 56 (Viikinkaari 5), 00014 Helsinki, Finland  
e-mail: elina.roine@helsinki.fi

Tailed viruses are the most common isolates infecting prokaryotic hosts residing in hypersaline environments. Archaeal tailed viruses represent only a small portion of all characterized tailed viruses of prokaryotes. But even this small dataset revealed that archaeal tailed viruses have many similarities to their counterparts infecting bacteria, the bacteriophages. Shared functional homologs and similar genome organizations suggested that all microbial tailed viruses have common virion architectural and assembly principles. Recent structural studies have provided evidence justifying this thereby grouping archaeal and bacterial tailed viruses into a single lineage. Currently there are 17 haloarchaeal tailed viruses with entirely sequenced genomes. Nine viruses have at least one close relative among the 17 viruses and, according to the similarities, can be divided into three groups. Two other viruses share some homologs and therefore are distantly related, whereas the rest of the viruses are rather divergent (or singletons). Comparative genomics analysis of these viruses offers a glimpse into the genetic diversity and structure of haloarchaeal tailed virus communities.

**Keywords:** archaea, haloarchaea, tailed virus, genome, bacteriophage

Viruses infecting haloarchaea come in a variety of virion morphotypes: spindle-shaped, pleomorphic, icosahedral and head-and-tail (or tailed) (Roine and Oksanen, 2011; Atanasova et al., 2012; Pietilä et al., 2013a). Yet, tailed viruses comprise the majority of the studied viruses infecting haloarchaea (Table 1). Despite the many early studies on  $\phi$ H genome and its rearrangements (Reiter et al., 1988) as well as detailed studies on  $\phi$ Ch1 virus (Witte et al., 1997; Baranyi et al., 2000; Klein et al., 2002; Rössler et al., 2004) we have had relatively little in-depth information about the haloarchaeal tailed virus genomes until recently (Klein et al., 2012; Pietilä et al., 2013b,c; Senčilo et al., 2013). The situation changed partly due to the recent technological advancements that have made for instance the sequencing of viral genomes much cheaper and faster than before. This caused an exponential increase in the number of sequencing projects focusing on separate virus genomes or on metaviromes from hypersaline environments (Santos et al., 2010; Boujelben et al., 2012; Garcia-Heredia et al., 2012; Pietilä et al., 2013b,c; Senčilo et al., 2013). While metaviromes revealed the richness and diversity of the viral communities present in hypersaline environments, whole-genome sequencing of isolated viruses provided more complete genomic information embedded in a clear biological context. The aim of this review is to summarize the findings on the 13 new complete haloarchaeal tailed virus genomes that were published in three separate papers (Pietilä et al., 2013b,c; Senčilo et al., 2013) and to combine these data with the previous knowledge of the complete genomes of haloarchaeal tailed viruses.

## CLASSIFICATION OF PROKARYOTIC TAILED VIRUSES

Tailed euryarchaeal (including haloarchaeal) viruses have been shown to have many properties in common with their bacterial

counterparts, the bacteriophages, starting from the morphology and the genome structure to gene regulation and some protein homologs (Torsvik and Dundas, 1974; Stolt and Zillig, 1994; Porter et al., 2007). Tailed bacteriophages are classified into order *Caudovirales*, which is further divided into three families according to the tail morphology: *Myoviridae* characterized by long contractile tails, *Siphoviridae* (long, non-contractile, but flexible tails) and *Podoviridae* (short non-contractile tails) (King et al., 2012). Some of the haloarchaeal tailed viruses have also been classified according to the criteria of the International Committee on Taxonomy of Viruses (ICTV) (King et al., 2012). The genus “PhiH-like viruses” belongs to the family *Myoviridae* and contains the species *Halobacterium* phage  $\phi$ H and a candidate *Halobacterium* phage Hs1 (King et al., 2012). Also HF2 has been added as a putative member of the *Myoviridae* family (King et al., 2012).

Before the times of having the means to generate massive amounts of sequence data, viral classification mainly based on virion morphology, the genome type (circular or linear ss/dsDNA or RNA) and host range, seemed rather straightforward. The current ease of genome sequencing revealed the Pandora’s box of the prokaryotic virus genomes. First of all, at the nucleotide sequence level the genomes are often very different from each other with no sequence similarity at all. In addition, mosaicism, the inherent feature of the prokaryotic viral genomes (Hendrix et al., 1999; Juhala et al., 2000; Lawrence et al., 2002; Krupovič et al., 2011), raises serious questions about the criteria to be used in classification. It has been proposed that in the absence of nucleotide or amino acid sequence similarity, the higher order classification of viruses should be based on the virion morphology and the major capsid protein fold (MCP) (Bamford et al., 2002, 2005;

**Table 1 | Characteristics of haloarchaeal tailed viruses with completely sequenced genomes.**

Morphology	Subgroup (if applicable)	Virus	Origin	Isolation host	Genome type	Genome size (bp)	Nr of ORFs	Nr of tRNAs	References
myo	HF2-like	HF1	Cheetham saltworks, Australia	<i>Hfx. lucentense</i>	Linear dsDNA, DTR (306 bp)	75,898	112	5	Nuttall and Dyall-Smith, 1993; Tang et al., 2004
	HF2		Cheetham saltworks, Australia	<i>Hrr. coriense</i>	Linear dsDNA, DTR (306 bp)	77,670	116	5	Nuttall and Dyall-Smith, 1993; Tang et al., 2002
	HRTV-5		Margherita di Savoia, Italy	<i>Hrr. sp. s5a-3</i>	Linear dsDNA, DTR (271 bp)	76,134	118	4	Atanasova et al., 2012; Senčilo et al., 2013
	HRTV-8		Samut Sakhon, Thailand	<i>Hrr. sp. B2-2</i>	Linear dsDNA, DTR (346 bp)	74,519	124	4	Atanasova et al., 2012; Senčilo et al., 2013
HRTV-7 like	HRTV-7		Margherita di Savoia, Italy	<i>Hrr. sp. B2-2</i>	linear dsDNA, DTR (340 bp)	69,048	105	1	Atanasova et al., 2012; Senčilo et al., 2013
	HSTV-2		Eilat, Israel	<i>Hrr. sodomense</i>	Linear dsDNA, DTR (340 bp)	68,187	103	1	Atanasova et al., 2012; Pietilä et al., 2013b
φCH1		Spontaneous release from <i>Nab. magadii</i> (L13)		<i>Nab. magadii</i> (L13)	circ. perm. dsDNA (+80,700 nt RNA)	58,498 143,855	98 281	— 36	Witte et al., 1997; Klein et al., 2002
HGTV-1		<i>Nab. magadii</i>	Samut Sakhon, Thailand	<i>Hgn. sp. S55-1</i>	circ. perm. dsDNA				Atanasova et al., 2012; Senčilo et al., 2013
sipho	HCTV-1 like	HCTV-1	Margherita di Savoia, Italy	<i>Har. californiae</i>	Linear dsDNA, DTR (739 bp)	103,257	160	1	Kukkaro and Bamford, 2009; Senčilo et al., 2013
	HCTV-5		Samut Sakhon, Thailand	<i>Har. californiae</i>	Linear dsDNA, DTR (583 bp)	102,105	166	1	Atanasova et al., 2012; Senčilo et al., 2013
	HVTV-1		Samut Sakhon, Thailand	<i>Har. valismortis</i>	Linear dsDNA, DTR (585 bp)	101,734	173	1	Atanasova et al., 2012; Pietilä et al., 2013b
	HCTV-2		Samut Sakhon, Thailand	<i>Har. californiae</i>	circ. perm. dsDNA	54,291	86	—	Atanasova et al., 2012; Senčilo et al., 2013
	HHTV-2		Samut Sakhon, Thailand	<i>Har. hispanica</i>	circ. perm. dsDNA	52,643	88	—	Atanasova et al., 2012; Senčilo et al., 2013
	BJ1		Lake Bagæjñor, China	<i>Hrr. sp. BJ1 B11</i>	Linear dsDNA*	42,271	70	1	Pagaling et al., 2007
	HHTV-1		Margherita di Savoia, Italy	<i>Har. hispanica</i>	circ. perm. dsDNA	49,107	74	—	Kukkaro and Bamford, 2009; Senčilo et al., 2013
	HRTV-4		Margherita di Savoia, Italy	<i>Hrr. sp. s5a-3</i>	circ. perm. dsDNA	35,722	73	—	Atanasova et al., 2012; Senčilo et al., 2013
podo	HSTV-1		Margherita di Savoia, Italy	<i>Har. sinaiensis</i>	circ. perm. dsDNA	32,189	53	—	Atanasova et al., 2012; Pietilä et al., 2013c

\*The genome type of BJ1 has not been clearly established; Hrr, *Haloteras*; Nab, *Natrialba*; Hfx, *Haloflexus*; Hgn, *Halogramnum*; circ. perm., circularly permuted; DTR, direct terminal repeats; myo, myovirus; siphovirus; podo, podovirus.

Abrescia et al., 2012). Viruses having the same MCP fold could then be grouped into lineages, and tailed bacteriophages were suggested to belong to the so called Hong Kong 97 (HK97)-like lineage together with the herpesviruses (Bamford, 2003; Bamford et al., 2005; Abrescia et al., 2012). Recent structural studies on haloarchaeal podovirus HSTV-1 suggested that it also has the HK97 MCP fold thereby justifying the placement of archaeal tailed viruses into HK97-like lineage (Pietilä et al., 2013c).

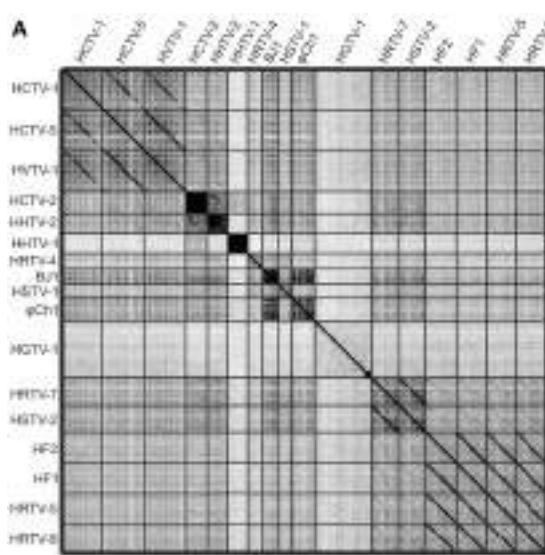
## CHARACTERISTICS OF HALOARCHAEAL TAILED VIRUS GENOMES

At the moment there are 43 haloarchaeal tailed viruses reported (Kukkaro and Bamford, 2009; Atanasova et al., 2012; Sabet, 2012) and 17 completely sequenced genomes comprise approximately 1.2 Mb of sequence information (Klein et al., 2002; Tang et al., 2002, 2004; Pagaling et al., 2007; Pietilä et al., 2013b,c; Senčilo et al., 2013). Also approximately 58 kb of the φH genome has been sequenced (Porter et al., 2007). In addition to that, several proviral regions found in haloarchaeal genomes extend our knowledge of the gene pool of haloarchaeal tailed viruses (Krupovič et al., 2010; Senčilo et al., 2013). Complete genomes of haloarchaeal tailed viruses range from approximately 32 to 144 kb in size (**Table 1**). Similarly to tailed bacteriophages, the genomes of haloarchaeal tailed viruses are either circularly permuted or non-permuted dsDNA molecules with direct terminal repeats (Klein et al., 2002; Tang et al., 2002, 2004; Pagaling et al., 2007; Pietilä et al., 2013b,c; Senčilo et al., 2013). The genomes have rather high GC percentage (above 50% on average), which is also characteristic of haloarchaea (Klein et al., 2002; Tang et al., 2002, 2004;

Oren, 2006; Pagaling et al., 2007; Pietilä et al., 2013b,c; Senčilo et al., 2013). Similar GC percentages suggest that the viruses are well-adapted to the codon usage of their hosts.

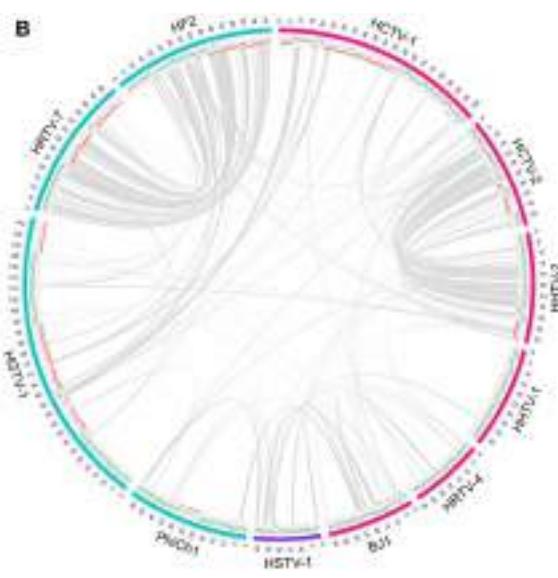
Annotation of the haloarchaeal tailed virus genomes is very often based on the similarity to bacteriophage genes (Klein et al., 2002; Tang et al., 2002, 2004; Pagaling et al., 2007; Krupovič et al., 2011; Pietilä et al., 2013b,c; Senčilo et al., 2013). Indeed, haloarchaeal tailed viruses share many similarities with bacteriophages both in terms of genome content and organization (Krupovič et al., 2011). In general, however, putative function can be assigned to no more than 20% of the new haloarchaeal tailed virus genes (Pagaling et al., 2007; Pietilä et al., 2013b,c; Senčilo et al., 2013). Large terminase subunit is among the most conserved proteins of prokaryotic tailed viruses and it was annotated in all haloarchaeal tailed virus genomes described to date (Klein et al., 2002; Tang et al., 2002, 2004; Pagaling et al., 2007; Pietilä et al., 2013b,c; Senčilo et al., 2013).

While the genomes of some haloarchaeal tailed viruses are collinear and highly similar at the nucleotide level, other viruses share up to several distant protein homologs at most (**Figure 1**). None of the completely sequenced genomes displayed close similarity to the putative proviral regions identified in the haloarchaeal genomes (Krupović et al., 2010; Senčilo et al., 2013). Among the 17 haloarchaeal tailed viruses, three groups of closely related viruses can be delineated based on the nucleotide sequence alignments (**Figure 1A**). Here we name these groups according to the first described representative: HF2-like, HRTV-7-like and HCTV-1-like groups (Nuttall and Dyall-Smith, 1993; Atanasova et al., 2012; Senčilo et al., 2013).



**FIGURE 1 | Genomic comparisons of the haloarchaeal tailed viruses with completely sequenced genomes. (A)** Dotplot alignment of the genomes.

Reverse complements of HF1 and HF2 genome sequences were used for the analyses in order to conform to the structure of the other haloarchaeal tailed virus genomes. The image was generated using the Gepard software (Krumbsiek et al., 2007). **(B)** Circular visualization of the homologous proteins shared between the selected virus representatives from each of the delineated groups and singletons. The outermost track represents the



genome maps with the coordinates (kbp). The myoviral genomes are marked in blue, siphoviral in pink and podoviral in violet. The following track displays the annotated ORFs (marked in green and red on the positive and the negative strands, respectively). Gray lines link pairs of genes coding for the putative homologs. Here proteins are defined as homologous if they share over 30% amino acid identity when aligned with EMBOSS Needle tool (Needleman and Wunsch, 1970). The image was generated using Circos software (Krzywinski et al., 2009).

## HF2-LIKE VIRUSES

The biggest group is HF2-like myovirus group, which, besides HF2, includes HF1, HRTV-5, and HRTV-8 viruses (**Figure 1A**) (Nuttall and Dyall-Smith, 1993; Atanasova et al., 2012; Senčilo et al., 2013). HF2-like viruses originate from spatially and temporally different environmental samplings (Nuttall and Dyall-Smith, 1993; Atanasova et al., 2012). Nevertheless, viruses share extensive similarity at the nucleotide level and subsequently most of their encoded proteins are homologous (Tang et al., 2002, 2004; Senčilo et al., 2013). Highly similar genomic regions are interrupted by non-homologous regions suggestive of the mosaic nature of HF2-like virus genomes (Tang et al., 2002, 2004; Senčilo et al., 2013). The clearest example is provided by HF1 and HF2 virus genomes, which are almost identical over 48 kb followed by a more diverged 28 kb region (Tang et al., 2004). The divergent region, among other putative proteins, codes for the tail fiber protein, which may be responsible for different host specificities of these two viruses (Tang et al., 2004). Majority of the non-conserved proteins in HF2-like viruses have no predicted function with an exception of putative restriction endonuclease and methylase (HF2p074 gene in HF2) found in all viruses except for HRTV-8, and HNH endonuclease found only in HRTV-8 (gene 43) (Tang et al., 2002, 2004; Senčilo et al., 2013).

## HRTV-7-LIKE VIRUSES

HF2-like viruses share some similarities with HRTV-7-like myoviruses, HRTV-7 and HSTV-2 (**Figures 1A,B**) (Pietilä et al., 2013b; Senčilo et al., 2013). Homologous genome regions are mostly located in the gene cluster coding for structural and assembly proteins (Pietilä et al., 2013b; Senčilo et al., 2013). Cryo-electron microscopy studies on HSTV-2 virus revealed that its capsid has a  $T = 7$  symmetry (Pietilä et al., 2013b). However, known viruses having capsids with this T-number, such as P22, package smaller genomes than that of HSTV-2 (Parent et al., 2010; Pietilä et al., 2013b). Therefore it was suggested that HSTV-2 capsids accommodate minor proteins, which increase the capsid volume (Pietilä et al., 2013b). Since all HRTV-7-like and HF2-like viruses have homologous MCPs as well as hypothetical proteins suggested to act as minor capsid proteins, it is likely that the capsid structures of all these viruses are similar (Pietilä et al., 2013b).

## HCTV-1-LIKE AND OTHER RELATED SIPHOVIRUSES

HCTV-1, HCTV-5, and HVTV-1 viruses encompass the HCTV-1-like virus group and are the only closely related haloarchaeal siphoviruses described to date (**Figure 1A**) (Pietilä et al., 2013b; Senčilo et al., 2013). HVTV-1 and HCTV-5 show similarity throughout their genomes, whereas HCTV-1 has a diverged genome region coding for tail structural and assembly proteins (Pietilä et al., 2013b; Senčilo et al., 2013). Another notable difference is rather high abundance of homing endonuclease genes in HVTV-1 and HCTV-5 genomes compared to HCTV-1 (Pietilä et al., 2013b; Senčilo et al., 2013). Structural studies available only for HVTV-1 virus showed that its capsomers are arranged in a  $T = 13$  lattice (Pietilä et al., 2013b).

Siphoviruses HCTV-2 and HVTV-2 also show some similarity to each other at the nucleotide sequence level and share a number of protein homologs (**Figures 1A,B**) (Senčilo et al., 2013).

As is the case for HF2-like and HRTV-7-like groups of viruses, similarities among HCTV-2 and HVTV-2 are mostly concentrated within the cluster of head and tail structural and assembly proteins (**Figure 1B**) (Senčilo et al., 2013).

## SINGLETONS

Siphovirus HHTV-1 is the most divergent among the completely sequenced haloarchaeal tailed viruses (Senčilo et al., 2013). The only homolog it shares with other haloarchaeal tailed viruses is a putative PCNA, which is similar to HSTV-1 podoviral PCNA (**Figure 1B**). Other two siphoviruses having no close relatives among the entirely sequenced haloarchaeal tailed viruses are HRTV-4 and BJ1 (Pagaling et al., 2007; Senčilo et al., 2013). However, even in these four diverged siphoviruses some of the structural and assembly proteins as well as putative proteins involved in nucleic acid metabolism were annotated based on the similarities to their counterparts in bacteriophages (Pagaling et al., 2007; Senčilo et al., 2013). The genome of the siphovirus HRTV-4 (Senčilo et al., 2013) shows close relatedness to an environmental clone eHP-10 (Garcia-Heredia et al., 2012). The two sequences align along approximately half of the length with close to 80% nucleotide sequence identity.

Although  $\phi$ Ch1 is rather distinct from other fully sequenced haloarchaeal tailed viruses, it is one of the best characterized haloarchaeal viruses to date (Witte et al., 1997; Klein et al., 2002, 2012).  $\phi$ Ch1 is a temperate virus infecting *Natrialba (Nab.) magadii* cells (Witte et al., 1997). The most unusual feature of the  $\phi$ Ch1 virus is that its particles along with the genomic dsDNA contain 80–700 nt RNA molecules of host origin (Witte et al., 1997). A 12 kb region of  $\phi$ Ch1 genome is highly similar to the  $\phi$ H virus L-fragment (Gropp et al., 1992; Klein et al., 2002). This fragment of  $\phi$ H virus was shown to be capable of autonomous replication in a plasmid state (p $\phi$ HL) (Gropp et al., 1992). It contains genes coding for proteins involved in replication, plasmid stabilization and gene expression regulation (Gropp et al., 1992).

The  $\phi$ Ch1 genome region and p $\phi$ HL align along almost the whole length with an exception of 1.7 kb fragment, which is in the inverse orientations in the two (Klein et al., 2002). Direct repeats flanking the fragment suggested that the rearrangement was a result of recombination between these repeats (Klein et al., 2002).  $\phi$ Ch1 genome contains a number of inverted repeats, one pair of which is involved in a phase variation system (Rössler et al., 2004; Klein et al., 2012). This system results in the production of two different variants of  $\phi$ Ch1 tail fiber protein (Klein et al., 2012).

HGTV-1 myovirus currently holds the record for having the largest genome among all described archaeal viruses (Senčilo et al., 2013). The genome of this virus has at least two distinctive features. First, it encodes unusually high number of tRNAs (36 in total) for all universal amino acids (Senčilo et al., 2013). Second, majority of ORFs located in HGTV-1 left-hand side of the genome are preceded by a conserved DNA motif, containing TATA box-like region and an inverted repeat (Senčilo et al., 2013). Similarity of these structures to promoter stem loops (PesLs) of T4-type bacteriophages led to the suggestion that as in T4-like bacteriophages, these DNA motifs in HGTV-1 are responsible for transcription regulation and genome shuffling (Arbiol et al.,

2010; Senčilo et al., 2013). Therefore, the mechanism of generating genetic diversity may also be shared among bacterial and archaeal tailed viruses in addition to the already pronounced similarity of structural and assembly proteins (Senčilo et al., 2013).

To date, HSTV-1 is the only reported archaeal podovirus (Pietilä et al., 2013c). It is also the only archaeal tailed virus for which the MCP fold was determined (Pietilä et al., 2013c). Despite its podoviral morphotype, HSTV-1 shares a handful of homologs with haloarchaeal myo- and siphoviruses (**Figure 1B**). These include the MCM DNA helicase, terminase large subunit, PCNA as well as several hypothetical proteins (**Figure 1B**).

## CONCLUSION

The growing number of complete genomes of haloarchaeal tailed viruses allowed us to determine groups of related viruses with more than two members. As new sequences are added, the groups are increasing in size and number. In addition to that, new singletons appear. A similar trend was also noticed for the growing database of complete mycobacteriophage genomes (Hatfull, 2012). The 17 completely sequenced haloarchaeal tailed viruses can be currently divided into 3 groups of closely related viruses, a pair of more distantly related siphoviruses and 6 singletons. Comparative genomics analysis of these genomes further corroborated several observations made earlier. First, different levels of relatedness can be observed among the haloarchaeal tailed virus genomes. In general this relatedness correlates neither with the place nor with the time of sampling for the virus isolation. For example very closely related viruses such as HF2-like viruses, were isolated from geographically distant sources in the span of almost 20 years (Nuttall and Dyall-Smith, 1993; Atanasova et al., 2012). Second, virion structure and assembly proteins are generally more conserved among the viruses, as is apparent from the examples of HF2-like and HRTV-7-like groups of viruses as well as HCTV-2 and HHTV-2 viruses (Pietilä et al., 2013b; Senčilo et al., 2013). Finally, the analysis of the extended data set did not yield more information on some single divergent viruses such as HHTV-1. This case exemplifies the gaps in our knowledge and highlights the fact that more sequences are needed for the deeper understanding of genetic diversity and structure of the viral communities as well as evolutionary processes shaping them.

## ACKNOWLEDGMENTS

This work was supported by the Helsinki University 3 year grant (2010–2013) to Elina Roine. Ana Senčilo is a graduate student of the Doctoral Program in Microbiology and Biotechnology, University of Helsinki.

## REFERENCES

- Abrescia, N. G., Bamford, D. H., Grimes, J. M., and Stuart, D. I. (2012). Structure unifies the viral universe. *Annu. Rev. Biochem.* 81, 795–822. doi: 10.1146/annurev-biochem-060910-095130
- Arbiol, C., Comeau, A. M., Kutateladze, M., Adamia, R., and Krisch, H. M. (2010). Mobile regulatory cassettes mediate modular shuffling in T4-type phage genomes. *Genome Biol. Evol.* 2, 140–152. doi: 10.1093/gbe/evq006
- Atanasova, N. S., Roine, E., Oren, A., Bamford, D. H., and Oksanen, H. M. (2012). Global network of specific virus-host interactions in hypersaline environments. *Environ. Microbiol.* 14, 426–440. doi: 10.1111/j.1462-2920.2011.02603.x
- Bamford, D. H. (2003). Do viruses form lineages across different domains of life? *Res. Microbiol.* 154, 231–236. doi: 10.1016/S0923-2508(03)00065-2
- Bamford, D. H., Burnett, R. M., and Stuart, D. I. (2002). Evolution of viral structure. *Theor. Popul. Biol.* 61, 461–470. doi: 10.1006/tpbi.2002.1591
- Bamford, D. H., Grimes, J. M., and Stuart, D. I. (2005). What does structure tell us about virus evolution? *Curr. Opin. Struct. Biol.* 15, 655–663. doi: 10.1016/j.sbi.2005.10.012
- Baranyi, U., Klein, R., Lubitz, W., Krüger, D. H., and Witte, A. (2000). The archaeal halophilic virus-encoded Dam-like methyltransferase M. phiCh1-I methylates adenine residues and complements dam mutants in the low salt environment of *Escherichia coli*. *Mol. Microbiol.* 35, 1168–1179. doi: 10.1046/j.1365-2958.2000.01786.x
- Boujelben, I., Yarza, P., Almansa, C., Villamor, J., Maalej, S., Anton, J., et al. (2012). Viriplankton community structure in Tunisian solar salterns. *Appl. Environ. Microbiol.* 78, 7429–7437. doi: 10.1128/AEM.01793-12
- Garcia-Heredia, I., Martin-Cuadrado, A. B., Mojica, F. J., Santos, F., Mira, A., Anton, J., et al. (2012). Reconstructing viral genomes from the environment using fosmid clones: the case of haloviruses. *PLoS ONE* 7:e33802. doi: 10.1371/journal.pone.0033802
- Gropp, F., Grampp, B., Stolt, P., Palm, P., and Zillig, W. (1992). The immunity-conferring plasmid p phi HL from the *Halobacterium salinarium* phage phi H: nucleotide sequence and transcription. *Virology* 190, 45–54. doi: 10.1016/0042-6822(92)91191-V
- Hatfull, G. F. (2012). Complete genome sequences of 138 mycobacteriophages. *J. Virol.* 86, 2382–2384. doi: 10.1128/JVI.06870-11
- Hendrix, R. W., Smith, M. C., Burns, R. N., Ford, M. E., and Hatfull, G. F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc. Natl. Acad. Sci. U.S.A.* 96, 2192–2197. doi: 10.1073/pnas.96.5.2192
- Juhala, R. J., Ford, M. E., Duda, R. L., Youlton, A., Hatfull, G. F., and Hendrix, R. W. (2000). Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J. Mol. Biol.* 299, 27–51. doi: 10.1006/jmbi.2000.3729
- King, A. M. Q., Adams, M. J., Carstens, E. B., and Lefkowitz, E. J. (2012). *Ninth Report of the International Committee on Taxonomy of Viruses*. San Diego, CA: Elsevier Academic Press.
- Klein, R., Baranyi, U., Rössler, N., Greineder, B., Scholz, H., and Witte, A. (2002). *Natrialba magadii* virus phiCh1: first complete nucleotide sequence and functional organization of a virus infecting a haloalkaliphilic archaeon. *Mol. Microbiol.* 45, 851–863. doi: 10.1046/j.1365-2958.2002.03064.x
- Klein, R., Rössler, N., Iro, M., Scholz, H., and Witte, A. (2012). Haloarchaeal myovirus phiCh1 harbours a phase variation system for the production of protein variants with distinct cell surface adhesion specificities. *Mol. Microbiol.* 83, 137–150. doi: 10.1111/j.1365-2958.2011.07921.x
- Krumsiek, J., Arnold, R., and Rattei, T. (2007). Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 23, 1026–1028. doi: 10.1093/bioinformatics/btm039
- Krupović, M., Forterre, P., and Bamford, D. H. (2010). Comparative analysis of the mosaic genomes of tailed archaeal viruses and proviruses suggests common themes for virion architecture and assembly with tailed viruses of bacteria. *J. Mol. Biol.* 397, 144–160. doi: 10.1016/j.jmb.2010.01.037
- Krupović, M., Prangishvili, D., Hendrix, R. W., and Bamford, D. H. (2011). Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. *Microbiol. Mol. Biol. Rev.* 75, 610–635. doi: 10.1128/MMBR.00011-11
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Kukkaro, P., and Bamford, D. H. (2009). Virus-host interactions in environments with a wide range of ionic strengths. *Environ. Microbiol. Rep.* 1, 71–77. doi: 10.1111/j.1758-2229.2008.00007.x
- Lawrence, J. G., Hatfull, G. F., and Hendrix, R. W. (2002). Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. *J. Bacteriol.* 184, 4891–4905. doi: 10.1128/JB.184.17.4891-4905.2002
- Needleman, S. B., and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48, 443–453. doi: 10.1016/0022-2836(70)90057-4

- Nuttall, S. D., and Dyall-Smith, M. L. (1993). HF1 and HF2: novel bacteriophages of halophilic archaea. *Virology* 197, 678–684. doi: 10.1006/viro.1993.1643
- Oren, A. (2006). “The order halobacteriales,” in *The Prokaryotes*, eds M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt (Singapore: Springer), 113–164.
- Pagaling, E., Haigh, R. D., Grant, W. D., Cowan, D. A., Jones, B. E., Ma, Y., et al. (2007). Sequence analysis of an archaeal virus isolated from a hypersaline lake in Inner Mongolia, China. *BMC Genomics* 8:410. doi: 10.1186/1471-2164-8-410
- Parent, K. N., Khayat, R., Tu, L. H., Suhanovsky, M. M., Cortines, J. R., Teschke, C. M., et al. (2010). P22 coat protein structures reveal a novel mechanism for capsid maturation: stability without auxiliary proteins or chemical crosslinks. *Structure* 18, 390–401. doi: 10.1016/j.str.2009.12.014
- Pietilä, M. K., Atanasova, N. S., Oksanen, H. M., and Bamford, D. H. (2013a). Modified coat protein forms the flexible spindle-shaped virion of haloarchaeal virus His1. *Environ. Microbiol.* 15, 1674–1686. doi: 10.1111/1462-2920.12030
- Pietilä, M. K., Laurinmäki, P., Russell, D. A., Ko, C. C., Jacobs-Sera, D., Butcher, S. J., et al. (2013b). Insights into head-tailed viruses infecting extremely halophilic archaea. *J. Virol.* 87, 3248–3260. doi: 10.1128/JVI.03397-12
- Pietilä, M. K., Laurinmäki, P., Russell, D. A., Ko, C. C., Jacobs-Sera, D., Hendrix, R. W., et al. (2013c). Structure of the archaeal head-tailed virus HSTV-1 completes the HK97 fold story. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10604–10609. doi: 10.1073/pnas.1303047110
- Porter, K., Russ, B. E., and Dyall-Smith, M. L. (2007). Virus-host interactions in salt lakes. *Curr. Opin. Microbiol.* 10, 418–424. doi: 10.1016/j.mib.2007.05.017
- Reiter, W. D., Zillig, W., and Palm, P. (1988). Archaeabacterial viruses. *Adv. Virus Res.* 34, 143–188. doi: 10.1016/S0065-3527(08)60517-5
- Roine, E., and Oksanen, H. M. (2011). “Viruses from the hypersaline environments: current research and future trends,” in *Halophiles and Hypersaline Environments*, eds A. Ventosa, A. Oren, and Y. Ma (Heidelberg: Springer), 153–172. doi: 10.1007/978-3-642-20198-1\_8
- Rössler, N., Klein, R., Scholz, H., and Witte, A. (2004). Inversion within the haloalkaliphilic virus φCh1 DNA results in differential expression of structural proteins. *Mol. Microbiol.* 52, 413–426. doi: 10.1111/j.1365-2958.2003.03983.x
- Sabet, S. (2012). “Halophilic viruses,” in *Advances in Understanding the Biology of Halophilic Microorganisms*, ed R. Vreeland (New York, NY: Springer), 81–116. doi: 10.1007/978-94-007-5539-0\_4
- Santos, F., Yarza, P., Parro, V., Briones, C., and Anton, J. (2010). The metavirome of a hypersaline environment. *Environ. Microbiol.* 12, 2965–2976. doi: 10.1111/j.1462-2920.2010.02273.x
- Senčilo, A., Jacobs-Sera, D., Russell, D. A., Ko, C. C., Bowman, C. A., Atanasova, N. S., et al. (2013). Snapshot of haloarchaeal tailed virus genomes. *RNA Biol.* 10, 803–816. doi: 10.4161/rna.24045
- Stolt, P., and Zillig, W. (1994). Gene regulation in Halophage phiH; more than Promoters. *Syst. Appl. Microbiol.* 16, 591–596.
- Tang, S. L., Nuttall, S., and Dyall-Smith, M. (2004). Haloviruses HF1 and HF2: evidence for a recent and large recombination event. *J. Bacteriol.* 186, 2810–2817. doi: 10.1128/JB.186.9.2810-2817.2004
- Tang, S. L., Nuttall, S., Nguia, K., Fisher, C., Lopez, P., and Dyall-Smith, M. (2002). HF2: a double-stranded DNA tailed haloarchaeal virus with a mosaic genome. *Mol. Microbiol.* 44, 283–296. doi: 10.1046/j.1365-2958.2002.02890.x
- Torsvik, T., and Dundas, I. D. (1974). Bacteriophage of *Halobacterium salinarium*. *Nature* 248, 680–681. doi: 10.1038/248680a0
- Witte, A., Baranyi, U., Klein, R., Sulzner, M., Luo, C., Wanner, G., et al. (1997). Characterization of *Natronobacterium magadii* phage phi Ch1, a unique archaeal phage containing DNA and RNA. *Mol. Microbiol.* 23, 603–616. doi: 10.1046/j.1365-2958.1997.d01-1879.x

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 14 January 2014; accepted: 17 February 2014; published online: 12 March 2014.

Citation: Senčilo A and Roine E (2014) A Glimpse of the genomic diversity of haloarchaeal tailed viruses. *Front. Microbiol.* 5:84. doi: 10.3389/fmicb.2014.00084

This article was submitted to Extreme Microbiology, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Senčilo and Roine. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Halophilic *Aspergillus penicilliooides* from athalassohaline, thalassohaline, and polyhaline environments

Sarita W. Nazareth\* and Valerie Gonsalves

Department of Microbiology, Goa University, Taleigao Plateau, India

**Edited by:**

R. Thane Papke, University of Connecticut, USA

**Reviewed by:**

Ronald Oremland, United States Geological Survey, USA

Mohammad Ali Amoozegar, University of Tehran, Iran

**\*Correspondence:**

Sarita W. Nazareth, Department of Microbiology, Goa University, Taleigao Plateau, Goa 403206, India  
e-mail: saritanazareth@yahoo.com

*Aspergillus penicilliooides* is a true halophile, present in diverse econiches – from the hypersaline athalassohaline, and thalassohaline environments, to polyhaline systems, and in different geographical locations. Twenty seven isolates from these environments, were seen to be moderate halophiles, euryhaline in nature. They had an obligate need of a low  $a_w$  and were unable to grow on a regular defined medium such as Czapek Dox Agar, as well as on varied nutrient rich agar media such as Malt Extract, Potato Dextrose and Sabouraud Agar; however, growth was obtained on all these media when amended with 10% solar salt. In absence of added salt, the conidia either did not germinate, or when germinated, distortions and lysis were seen in the short mycelial forms; on media with salt, the mycelia and vesicles appeared normal.

**Keywords:** *Aspergillus penicilliooides*, obligate, halophile, hypersaline, polyhaline

## INTRODUCTION

*Aspergillus penicilliooides* growth is favoured by low  $a_w$ , and can grow even at an  $a_w$  of 0.68, which is inhibitory to most fungi (Tamura et al., 1999; Pitt and Hocking, 2009). *A. penicilliooides* has been found in diverse habitats of low  $a_w$ , such as the Dead Sea, solar salterns, mangroves, estuary (Wasser et al., 2003; Butinar et al., 2011; Gonsalves et al., 2012; Nayak et al., 2012; Nazareth et al., 2012), on foods such as grains, dried fruit, baked goods, salted fish and spices, as well as on binocular lenses and human skin (Andrews and Pitt, 1987; Tamura et al., 1999; Pitt and Hocking, 2009).

Organisms able to grow under conditions of low  $a_w$  and requiring NaCl, are known as halophiles (Kushner, 1978; Grant, 2004), distinguishing them from those merely able to grow at a low  $a_w$  caused by low moisture content, such as xerophiles (Andrews and Pitt, 1987; Tamura et al., 1999; Grant, 2004; Pitt and Hocking, 2009) or high osmotic pressures of sugar solutions, such as osmophiles (Tucker and Featherstone, 2011).

This paper reports *A. penicilliooides* as a true halophile, present in diverse econiches of athalassohaline, and thalassohaline hypersaline environments, as well as polyhaline systems.

## MATERIALS AND METHODS

### ISOLATES

Twenty seven strains of *A. penicilliooides* were tested, which were previously isolated from the Dead Sea water (DSw) and sediment (DSs) samples (Nazareth et al., 2012), from the estuary of Mandovi, Goa, on the West Coast of the Indian peninsula, surface and bottom waters (EMws and EMwb) and from sediment (EMs) samples (Gonsalves et al., 2012), from water samples from mangroves of Ribander, Goa (MRw) and from solar salterns at Santa Cruz (SCw), Goa, India (Nayak et al., 2012). The isolate numbers along with the sites of isolation are shown in Table 1.

### HALOTOLERANCE CURVES

Salt tolerance curves were performed as given by Nazareth et al. (2012). Conidial suspensions of isolates were spot-inoculated on CzA containing solar salt (0–30%). Growth was recorded in terms of colony diameter after 7 d incubation, or after 15 d for those showing delayed growth.

### DETERMINATION OF OBLIGATE REQUIREMENT OF SALT FOR GROWTH ON DIFFERENT MEDIA

Isolates were selected on the basis of their halotolerance curves, and conidial suspensions,  $10^3$  spores in 5  $\mu$ l, were spot inoculated, in triplicate, on Czapek Dox Agar (CzA), Malt Extract Agar (MEA), Potato Dextrose Agar (PDA) and Sabouraud Agar (SA; HI Media), each without and with 10% solar salt (S), to confirm the obligate requirement of salt for growth. Growth was measured in terms of colony diameter after 7 days incubation at 30°C.

### EXAMINATION OF CONIDIAL GERMINATION

Conidial suspensions of selected isolates were spread on to plates of CzA and CzA + 10% solar salt, and incubated in the dark, at 30°C. Three agar plugs were aseptically sampled from each plate of CzA, at 3 h intervals between 12 and 48 h. The agar plug was placed on a slide, stained with lactophenol cotton blue dye and examined microscopically. A total of 50 conidia per agar plug were counted; conidia were considered germinated when the germ-tube length was equal to, or longer than, the diameter of the conidium (Ramirez et al., 2004) and expressed in terms of % germination.

### MORPHOLOGICAL CHANGES IN RESPONSE TO PRESENCE OR ABSENCE OF SALT

Conidial suspensions were spot-inoculated on CzA and on CzA + 10% solar salt, and incubated at 30°C for 15 d. Wet mounts of the isolates prepared in 1:1 lactophenol cotton blue

**Table 1 | Isolates obtained from various econiches of different salinity.**

Sampling site and isolate number	Salinity %
<b>Dead Sea (DS)</b>	
Water: DSw6, DSw22	370
Sediment: DSS30, DSS34, DSS36, DSS38, DSS40, DSS46, DSS54	450
<b>Estuary of Mandovi (EM)</b>	
Station 2 water: EM2w <sub>b</sub> 107	35
Station 4 water: EM4w <sub>s</sub> 118, EM4w <sub>s</sub> 120, EM4w <sub>b</sub> 125	33
Station 5 water: EM5w <sub>s</sub> 130	31
Station 6 sediment: EM6s137	05
Station 7 water: EM7w <sub>s</sub> 139, EM7w <sub>b</sub> 143, sediment: EM7s145	25
Station 8 water: EM8w <sub>s</sub> 146, EM8w <sub>s</sub> 147, EM8w <sub>s</sub> 148, sediment: EM8s153	16
Station 9 sediment: EM9s156	10
Mangrove water: MRw204, MRw207	32
Salterns water : SCw255	230

dye (HI Media) were then viewed microscopically for morphological changes in the germination of the conidia, the mycelia and conidiating structures; where growth was not visible on agar media without salt, an agar plug obtained as detailed above, was used for microscopic examination.

## RESULTS

### SALT TOLERANCE CURVES

The salt tolerance curves of the *A. penicillioides* isolates are shown in **Figure 1**. The results indicate that most of the isolates tested had a minimum salt requirement of 2 or 5% for growth, while a few required 10%, which clearly demonstrated their true halophilic nature. Optimal growth for almost all isolates was obtained at a salt concentration of 10%, with a few growing best at a salt concentration of 5 or 15%, irrespective of the econiche or its hypersaline or polyhaline characteristic from which the isolates were obtained. These isolates were therefore termed as moderate halophiles, in accordance to the definition of Kushner (1978). The isolates were euryhaline in nature, able to adapt to a wide range of salt concentrations, with only one isolate showing a stenohaline nature, having growth over a short range of salt concentrations.

### DETERMINATION OF OBLIGATE REQUIREMENT OF SALT FOR GROWTH ON DIFFERENT MEDIA

One isolate from each of the econiches was selected for the study, on the basis of its greater halophilic nature, having the highest minimal salt concentration requirement for growth, and the highest limit of salt tolerance, all requiring 10% salt concentration for optimal growth: DSS40 which grew readily with

10–20% salt concentrations, with a delayed growth in presence of 5 and 25% salt; EM6s137 which grew with a range of 5–25% salt and delayed growth with 30% salt; MRw207 which grew with salt concentrations of 2–25 or of 30% with a delayed growth.

The growth of the isolates on various nutrient media in presence or absence of added solar salt is given in **Figure 2A**. In absence of added salt, all the isolates tested were unable to grow on chemically defined medium of CzA, as well as the more nutrient-rich media of MEA, PDA and SA. However, with addition of 10% salt, growth was visible on all these media.

### CONIDIAL GERMINATION AND MORPHOLOGICAL CHANGES IN RESPONSE TO SALT

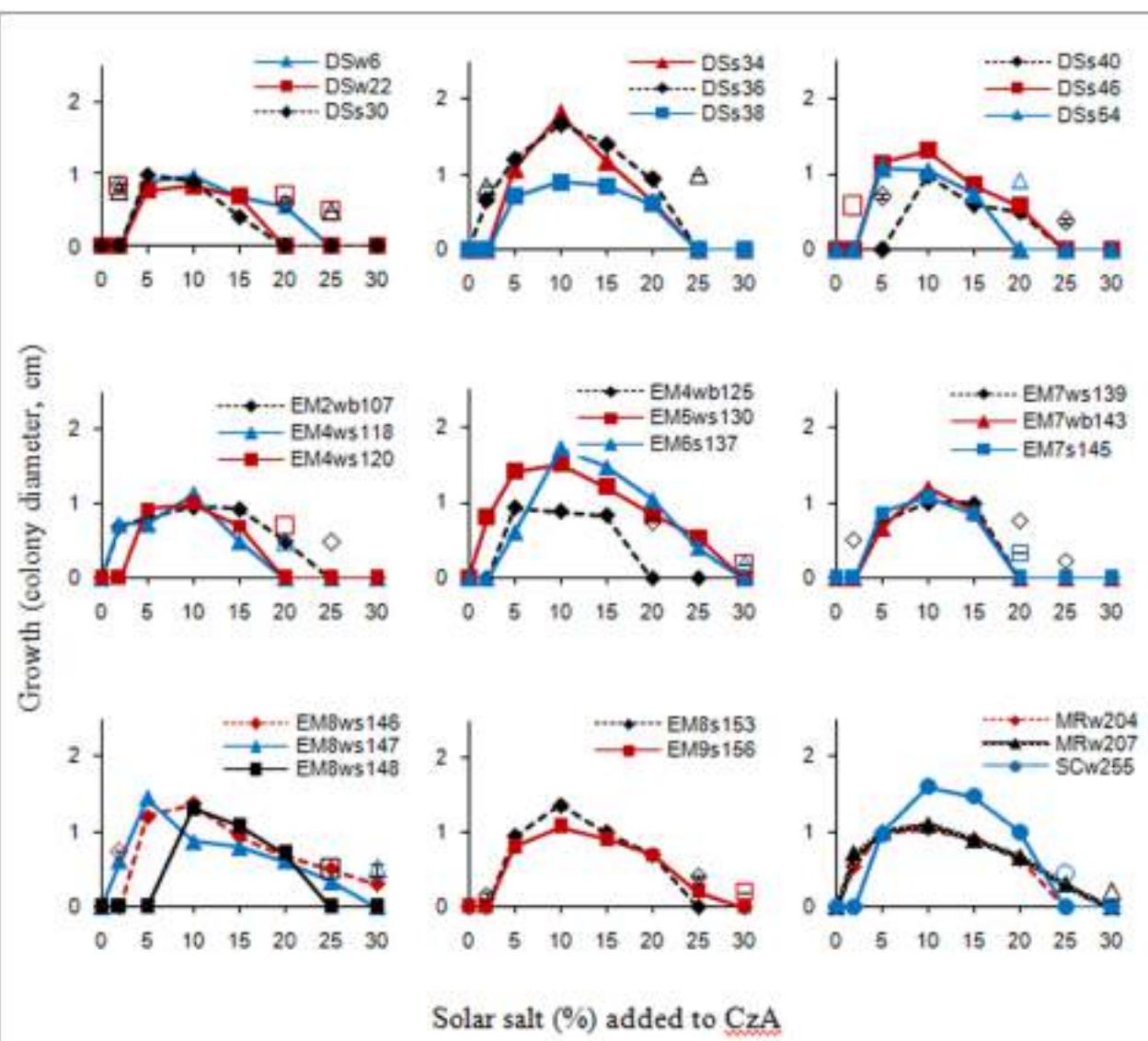
Germination curves and micromorphological examination of the selected isolates are shown in **Figure 2B**. Conidia of DSS40 did not germinate on media without solar salt within the 48 h tested, except for an occasional occurrence. EM6s137 conidia showed 2% germination which was initiated from 36 h; no germination was seen in conidia of MRw207. However, at 10% salt concentration which supported maximal growth, the conidia of all isolates showed germination. DSS40 and MRw207 conidial germination began within 15 h incubation, but while a number of germinated conidia even produced mycelial forms, some conidia had not yet germinated. As the mycelia covered the entire viewing field, making further observations impossible, the study was discontinued. The conidia of EM6s137 showed 100% germination within 33 h.

When incubated on medium without salt, DSS40 conidia remained ungerminated even after 15 days, but the conidia appeared swollen and distorted. Conidia of EM6s137 germinated, but became distorted without further mycelial growth. MRw207 conidia germinated, but formed distorted mycelia with very little cytoplasm, lysis at some parts and with oozing of the cytoplasm. However, at 10% salt concentration that supported optimal growth, micromorphological analysis revealed that not only had conidial germination occurred, but that there was formation of mycelia and vesicles with conidiation, which appeared normal.

## DISCUSSION

The strains of *A. penicillioides* had an obligate requirement of NaCl for growth, as was seen from the halotolerance curves, and were therefore classified as true halophiles. This was supported by the observation that the isolates did not show any growth on synthetic as well as nutrient rich agar media in absence of NaCl, but grew well on the same media when supplemented with 10% solar salt, irrespective of the synthetic or nutrient rich nature of the media, although the latter amended with salt, supported a better growth, possibly providing substrates for synthesis of osmolytes to combat the low  $a_w$  environment.

The absolute requirement for salt by the isolates was further confirmed by the lack of conidial germination and/or distortion of conidia or the germ tube when incubated in absence of salt, while germination and growth was observed to be normal when grown in presence of salt.



**FIGURE 1 |** Salt tolerance curves of the isolates as recorded after 7 d incubation; unconnected open symbols indicate delayed growth at respective salt concentrations, after 15 d incubation.

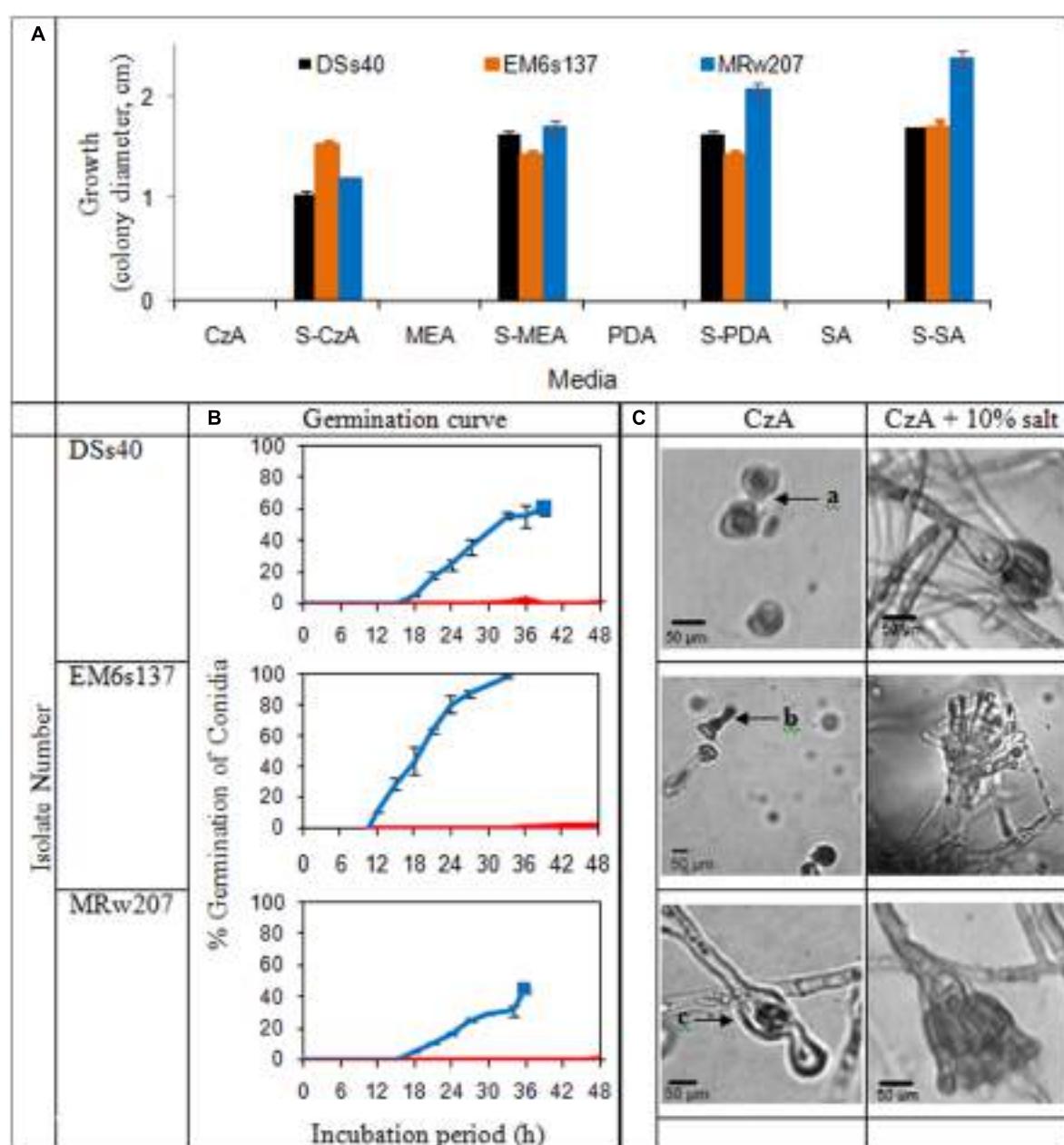
It has therefore been shown that salt and a low  $a_w$  are essential for the germination of the conidia, as well as germ tube elongation, prior to growth of the culture. The strains of *A. penicillioides* which were obtained from diverse evironments such as the hyper-saline athalassohaline Dead Sea and thalassohaline solar salterns, and the polyhaline estuary and mangroves, has thus been shown to be true halophiles.

The basidiomycete *Wallemia ichtyophaga* is another true halophile that requires a minimum of about 9% NaCl for growth and 15–20% NaCl for optimal growth (Plemenitas et al., 2014). However, the strains of *A. penicillioides* are the only asexual filamentous fungi reported thus far.

The osmoadaptation mechanism in true halophiles forms an intrinsic part of its metabolism. A mitogen-activated protein

kinase or MAPK pathway, involved in germ tube elongation, branching, and hyphal fusion events between conidial germlings (Pandey et al., 2004), has also been shown to be responsible for transcription of enzymes involved in glycerol synthesis and intracellular glycerol accumulation in response to osmotic stress (Jin et al., 2005). It appears therefore, that in true halophiles, the MAPK pathway is stimulated by conditions of low  $a_w$ , and hence in the absence of such stimulatory conditions, osmoadaptation for germination and/or germ tube elongation does not occur.

True or obligate halophiles can be termed as specialists, with their growth optimum shifted toward extreme values, and have a narrow ecological amplitude (Gostincar et al., 2010). *A. penicillioides* species does not have a sexual life cycle (Tamura et al.,



**FIGURE 2 | (A)** Colony characteristics of *Aspergillus penicilliooides* isolates on agar media: Czapek Dox (CzA), Malt Extract (MEA), potato dextrose (PDA), Sabouraud (SA), without and with 10% salt (S); **(B)** Curves of conidial germination on CzA amended with salt: 0% (red) and 10% (blue). Solid blocks indicate termination of experiment; **(C)** Micromorphology of the isolates after 15 d incubation on CzA and CzA + 10% salt, showing (a): ungerminated swollen and distorted conidia, (b): conidia germinated and distorted, (c): conidia germinated and mycelia distorted with little cytoplasm, and lysis

1999) which consequently inhibits gene flow. This will have caused a rapid fix of genetic information in these populations that have managed to adapt to saline habitats (Gostincar et al., 2010).

*A. penicilliooides* strains, by means of their absolute requirement for salt, are indigenous to the marine environment (Mackay et al., 1984) and have been shown to exist in diverse environments—from hypersaline to polyhaline systems, from athalassohaline to thalassohaline environments, and at a longitudinal distance of approximately 38.5° apart on the Asian continent. Hence

it can be expected to be found globally, in diverse saline environments.

*A. penicilliooides* has been described as xerophilic (Tamura et al., 1999) and osmophilic (Wasser et al., 2003). Its capacity to grow in environments in which the lowering of  $a_w$  is contributed by sodium chloride ions, as shown above, establishes these isolates as halophilic, as defined by Kushner (1978) and Grant (2004). It is therefore suggested that the species may be polyextremophilic in nature, the low  $a_w$  being the basic requirement for growth

of the species, whether contributed by low moisture, high sugar concentrations, or increased levels of sodium chloride.

## REFERENCES

- Andrews, S., and Pitt, J. I. (1987). Further studies on the water relations of xerophilic fungi, including some halophiles. *J. Gen. Appl. Microbiol.* 133, 233–238. doi: 10.1099/00221287-133-2-233
- Butinar, L., Frisvad, J. C., and Gunde-Cimerman, N. (2011). Hypersaline waters – a potential source of foodborne toxicigenic aspergilli and penicillia. *FEMS Microbiol. Ecol.* 77, 186–199. doi: 10.1111/j.1574-6941.2011.01108.x
- Gonsalves, V., Nayak, S., and Nazareth, S. (2012). Halophilic fungi in a polyhaline estuarine habitat. *J. Yeast Fungal Res.* 3, 30–36. doi: 10.5897/JYFR12.007
- Gostincar, C., Grube, M., De Hoog, S., Zalar, P., and Gunde-Cimerman, N. (2010). Extremotolerance in fungi: evolution on the edge. *FEMS Microbiol. Ecol.* 71, 2–11. doi: 10.1111/j.1574-6941.2009.00794.x
- Grant, W. D. (2004). Life at low water activity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 1249–1267. doi: 10.1098/rstb.2004.1502
- Jin, Y., Weining, S., and Nevo, E. (2005). A MAPK gene from Dead Sea fungus confers stress tolerance to lithium salt and freezing-thawing: prospects for saline agriculture. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18992–18997. doi: 10.1073/pnas.0509653102
- Kushner, D. J. (1978). “Life in high salt and solute concentrations,” in *Microbial Life in Extreme Environments*, ed. D. J. Kushner (London: Academic Press), 317–368.
- Mackay, M. A., Norton, R. S., and Borowitzka, L. J. (1984). Organic osmoregulatory solutes in cyanobacteria. *J. Gen. Microbiol.* 130, 2177–2191. doi: 10.1099/00221287-130-9-2177
- Nayak, S., Gonsalves, V., and Nazareth, S. (2012). Isolation and salt tolerance of halophilic fungi from mangroves and solar salterns in Goa – India. *Indian J. Mar. Sci.* 41, 164–172.
- Nazareth, S., Gonsalves, V., and Nayak, S. (2012). A first record of obligate halophilic aspergilli from the Dead Sea. *Indian J. Microbiol.* 52, 22–27. doi: 10.1007/s12088-011-0225-z
- Pandey, A., Roca, M. G., Read, N. D., and Glass, N. L. (2004). Role of a mitogen-activated protein kinase pathway during conidial germination and hyphal fusion in *Neurospora crassa*. *Eukaryot. Cell* 3, 348–358. doi: 10.1128/EC.3.2.348-358.2004
- Pitt, J. I., and Hocking, A. D. (2009). *Fungi and Food Spoilage*. doi: 10.1007/978-0-387-92207-2
- Plemenitas, A., Lenassi, M., Konte, T., Kejzar, A., Zajc, J., Gostinar, C., et al. (2014). Adaptation to high salt concentrations in halotolerant/halophilic fungi: a molecular perspective. *Front. Microbiol.* 5:199. doi: 10.3389/fmicb.2014.00199
- Ramirez, M. L., Chulze, S. N., and Magan, N. (2004). Impact of osmotic and matric water stress on germination, growth, mycelia water potentials and endogenous accumulation of sugars and sugar alcohols in *Fusarium graminearum*. *Mycologia* 96, 470–478. doi: 10.2307/3762167
- Tamura, M., Kawasaki, H., and Sugiyama, J. (1999). Identity of the xerophilic species *Aspergillus penicilliooides*: integrated analysis of the genotypic and phenotypic. *J. Gen. Appl. Microbiol.* 45, 29–37. doi: 10.2323/jgam.45.29
- Tucker, G., and Featherstone, S. (2011). *Essentials of Thermal Processing*. ISBN: 1444348035, 9781444348033.
- Wasser, S. P., Grishkan, I., Kis-Papo, T., Buchalo, A. S., Paul, A. V., Gunde-Cimerman, N., et al. (2003). “Species diversity of the Dead Sea fungi,” in *Fungal Life in the Dead Sea*, eds E. Nevo, A. Oren, and S. P. Wasser (Ruggel: Gantner Verlag), 203–270.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 03 April 2014; accepted: 21 July 2014; published online: 05 August 2014.

Citation: Nazareth SW and Gonsalves V (2014) Halophilic *Aspergillus penicilliooides* from athalassohaline, thalassohaline, and polyhaline environments. *Front. Microbiol.* 5:412. doi: 10.3389/fmicb.2014.00412

This article was submitted to Extreme Microbiology, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Nazareth and Gonsalves. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Extracellular DNA metabolism in *Haloferax volcanii*

Scott Chimileski<sup>1</sup>, Kunal Dolas<sup>1</sup>, Adit Naor<sup>2</sup>, Uri Gophna<sup>2</sup> and R. Thane Papke<sup>1\*</sup>

<sup>1</sup> Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT, USA

<sup>2</sup> Department of Molecular Microbiology and Biotechnology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel

**Edited by:**

Aharon Oren, The Hebrew University of Jerusalem, Israel

**Reviewed by:**

Aharon Oren, The Hebrew University of Jerusalem, Israel  
Alexander Beliaev, Pacific Northwest National Laboratory, USA

**\*Correspondence:**

R. Thane Papke, Department of Molecular and Cell Biology, University of Connecticut, 91 N. Eagleville Rd., Storrs, CT 06269, USA  
e-mail: thane@uconn.edu

Extracellular DNA is found in all environments and is a dynamic component of the microbial ecosystem. Microbial cells produce and interact with extracellular DNA through many endogenous mechanisms. Extracellular DNA is processed and internalized for use as genetic information and as a major source of macronutrients, and plays several key roles within prokaryotic biofilms. Hypersaline sites contain some of the highest extracellular DNA concentrations measured in nature—a potential rich source of carbon, nitrogen, and phosphorus for halophilic microorganisms. We conducted DNA growth studies for the halophilic archaeon *Haloferax volcanii* DS2 and show that this model Halobacteriales strain is capable of using exogenous double-stranded DNA as a nutrient. Further experiments with varying medium composition, DNA concentration, and DNA types revealed that DNA is utilized primarily as a phosphorus source, that growth on DNA is concentration-dependent, and that DNA isolated from different sources is metabolized selectively, with a bias against highly divergent methylated DNA. Additionally, fluorescence microscopy showed that labeled DNA co-localized with *H. volcanii* cells. The gene *Hvo\_1477* was also identified using a comparative genomic approach as a factor likely to be involved in DNA processing at the cell surface, and deletion of *Hvo\_1477* created a strain deficient in the ability to grow on extracellular DNA. Widespread distribution of *Hvo\_1477* homologs in archaea suggests metabolism of extracellular DNA may be of broad ecological and physiological relevance in this domain of life.

**Keywords:** extracellular DNA, *Haloferax volcanii*, DNA metabolism, Halobacteria, halophiles, archaea, natural competence, archaeal genetics

## INTRODUCTION

Extracellular DNA (eDNA) is present in every natural environment, amounting to a global molecular pool measured in gigatons (Dell'Anno and Danovaro, 2005). Beyond its sheer abundance, eDNA is a major component of the microbial ecosystem as a dynamic reservoir of carbon (C), nitrogen (N), phosphorus (P), nucleotides, and genetic information (Dell'Anno and Danovaro, 2005; Corinaldesi et al., 2007, 2008). eDNA is engaged by prokaryotes through often complex endogenous mechanisms, including degradation by unbound and cell-surface-bound secreted nucleases (Heins et al., 1967; Provvedi et al., 2001; Sakamoto et al., 2001; Schmidt et al., 2007; Godeke et al., 2011a), import and export systems mediating natural DNA uptake (Chen and Dubnau, 2004; Maier et al., 2004; Chen et al., 2005; Averhoff, 2009), and as a major component of biofilms (Steinberger and Holden, 2005; Godeke et al., 2011b; Kiedrowski et al., 2011; Gloag et al., 2013).

Hypersaline environments contain some of the highest measured levels of eDNA (Dell'Anno and Corinaldesi, 2004), possibly due to the preservative effect salt has on nucleic acids and other macromolecules (Tehei et al., 2002). Over three hundred micrograms of eDNA per gram of sediment was measured within the first centimeter beneath the water column of a hypersaline lake (Danovaro et al., 2005). However, the mechanisms through which the halophilic microorganisms living in this environment

interact, process, and exploit this potential cellular resource are unknown.

Brines above 15% salinity and ranging up to saturation are rich in microbial life and are predominated by euryarchaeal species of the order Halobacteriales, a group commonly known as haloarchaea (Benloch et al., 2001; Papke et al., 2003, 2004; Pasić et al., 2007; Dassarma and Dassarma, 2008; Andrei et al., 2012). A representative species *Haloferax volcanii* first isolated from Dead Sea sediments in 1975 (Mullakhanbhai and Larsen, 1975) was chosen for this study because it is genetically modifiable, and is a predominant model halobacterial species, and a top archaeal model in general (Bitan-Banin et al., 2003; Allers and Mevarech, 2005; Soppa, 2006; Allers et al., 2010; Blaby et al., 2010; Hartman et al., 2010; Leigh et al., 2011; Atomi et al., 2012). Here we describe an improved understanding of eDNA metabolism in hypersaline environments and the Halobacteria.

## MATERIALS AND METHODS

### STRAINS AND CULTURE MEDIA

*Haloferax volcanii* strains (**Table 1**) were provided by Thorsten Allers of the University of Nottingham, UK and grown in media as previously described (Allers et al., 2004). Hv-YPC contained 144 g NaCl, 21 g MgSO<sub>4</sub> × 7H<sub>2</sub>O, 18 g MgCl<sub>2</sub> × 6H<sub>2</sub>O, 4.2 g KCl, 12 mM Tris-HCl pH 7.5, 3.125 ml 1 M CaCl<sub>2</sub> solution, 1.0 ml trace element solution, 5.0 g yeast extract (Fisher, BP1422), 1.0 g

**Table 1 | Strains and plasmids.**

Strain or plasmid	Relevant properties	References or sources
<b>PLASMIDS</b>		
pTA131	Integrative vector based on pBluescript II; AmpR with BamHI-XbaI fragment from pGB70 containing <i>pfdx-pyrE2</i>	Allers et al., 2004
pTA409	Shuttle vector based on pBluescript II, with <i>pyrE2</i> and <i>hdrB</i> markers and <i>ori-pHV1/4</i> replication origin	Hölzle et al., 2008
pKD131_Δ1477	pTA131 with HindIII-NotI fragment containing <i>Hvo_1477</i> flanking regions for pop-in/pop-out gene deletion	This study
pKD409_1477c	pTA409 with BamHI-EcoRI fragment containing <i>Hvo_1477</i> and native upstream promoter	This study
<b>STRAINS</b>		
<i>E. coli</i>		
DH5α	Used for subcloning and as source of methylated <i>E. coli</i> DNA	Invitrogen, 18263012
K12	<i>dam</i> <sup>-</sup> / <i>dcm</i> <sup>-</sup> strain; used for source of unmethylated DNA	New England Biolabs, C2925I
<i>H. volcanii</i>		
DS2	Wild-type	Mullakhambhai and Larsen, 1975
H26	Δ <i>pyrE2</i>	Allers et al., 2004
Δ <i>Hvo_1477</i>	<i>Hvo_1477</i> deletion in H26 background	This study
Δ <i>Hvo_1477c</i>	Δ <i>Hvo_1477</i> with pKD409_1477c	This study

casamino acids (Fisher, BP1424), and 1.0 g peptone per liter (Fisher, BP1420). Casamino acids medium (Hv-Ca) contained all of the above components except for yeast extract and peptone. Starvation medium (used for starve conditions) contained all components in Hv-YPC other than yeast extract, casamino acids, and peptone. Minimal medium (Hv-min) also shared the same concentration of salts and basic constituents, with the addition of 4.25 ml 60% sodium lactate ( $\text{NaC}_3\text{H}_5\text{O}_3$ ), 5.0 ml 1 M ammonium chloride ( $\text{NH}_4\text{Cl}$ ) solution, and 2.0 ml 0.5 M monopotassium phosphate ( $\text{KH}_2\text{PO}_4$ ) buffer (pH 7.5). *Escherichia coli* cloning strains and additional bacterial strains used as DNA source and as extracellular nuclease positive control are listed (Table 1) and were grown using standard growth media and conditions.

All media used in growth experiments with variable C, N and P availability were derivatives of Hv-min. Conditions denoted as CNP contained C, N and P sources ( $\text{NaC}_3\text{H}_5\text{O}_3$ ,  $\text{NH}_4\text{Cl}$ , and  $\text{KH}_2\text{PO}_4$ , respectively) at the same final concentration as Hv-min medium above. Additional conditions/media derivatives lacked either sodium lactate, ammonium chloride, and/or potassium phosphate (e.g., NP medium contains  $\text{NH}_4\text{Cl}$ , and  $\text{KH}_2\text{PO}_4$ , but no carbon source).

#### DNA EXTRACTION AND PURIFICATION

The purity and integrity of supplemented high molecular weight (HMW) DNA was a primary consideration. Chromosomal DNA for supplementation was isolated from source species through standard lysis methods, followed by proteinase K digestion and ethanol precipitation. DNA was further purified through multiple phenol/chloroform/isoamyl alcohol (pH 8.0) extractions until no protein-rich interphase was present, followed by three chloroform/isoamyl alcohol extractions to remove trace phenol and an additional ethanol precipitation. DNA was then dissolved in 10 mM Tris-Cl solution (pH 8, in DNA-grade water) and passed through a mini polyacrylamide gel filtration spin column according to the manufacturer protocol (Bio-Rad Bio-Spin

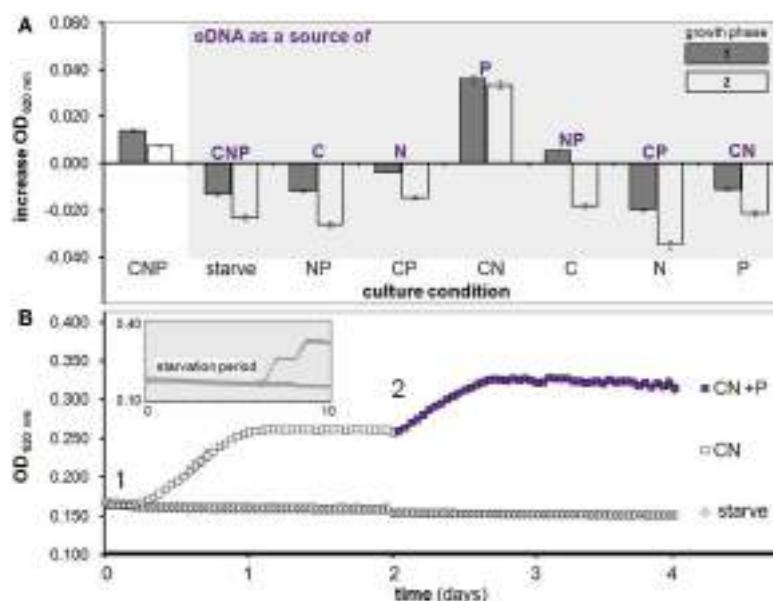
P-30, in Tris buffer, 732–6231) to remove small molecules including free nucleotides and oligonucleotides <20 bp in length. Purified DNA was used fresh for growth experiments to reduce subsequent hydrolysis, and was sterilized with a 0.22  $\mu\text{m}$  filter prior to supplementation. RNA was degraded within DNA samples using RNase I (Thermo, FEREN0601) according to manufacturer's protocol, followed by heat inactivation, ethanol precipitation, and resuspension. DNase digested DNA used in growth experiments was digested with DNaseI (Invitrogen, 18068–015) for 12 h at 37°C. Herring sperm DNA was from Promega (D1811).

DNA concentration and purity was determined using a Nanodrop ND-1000 ( $\text{OD}_{260\text{ nm}}/\text{OD}_{280\text{ nm}} = 1.8$ ) or a Qubit 2.0 fluorometer (Q32866) with dsDNA High Sensitivity kit (Invitrogen, Q32854). HMW DNA was visualized on agarose gels prior to supplementation.

#### eDNA METABOLISM STUDIES

For all growth experiments a minimum of three replicate cultures per condition began with an individual *H. volcanii* DS2 colony grown to mid-exponential phase ( $\text{OD}_{600\text{ nm}} \sim 0.4$ ) at 42°C in liquid Hv-YPC, washed three times with starvation medium, and diluted in medium specific to the experiment (e.g., starvation medium or Hv-min derivative). When cultures were starved, starvation occurred at 42°C for a ≥ 5 day period prior to supplementation to allow for depletion of internal nutrient stores (particularly phosphorus, see Figure 1B).

Growth on eDNA was tested in several ways: 200  $\mu\text{l}$  of culture within 96-well plates (sealed with transparent plastic film to avoid evaporation and salt precipitation), 5 ml of culture within 50 ml plastic culture tubes, 10 ml of culture in glass anaerobic culture tubes (with rubber stoppers and aluminum seals, headspace displaced with  $\text{N}_2$  and supplemented with 50 mM sodium nitrate), and 20 ml of culture in 125 ml cell culture flasks (baffled and unbaffled). All experiments were conducted at 42°C.



**FIGURE 1 |** *H. volcanii* growth on eDNA as a C, N and/or P source.

(A) Increase in OD during two growth phases after starvation (as in growth curve below) with eDNA supplementation in Hv-min medium containing all possible combinations of C, N and P. (B) After a 6 day starvation period (shown in smaller gray box) cultures were supplemented

with C and N (squares) or a negative control starvation solution (circles) at  $T = 0$  days (point 1) with OD measurements taken every hour. After one round of growth to stationary phase,  $\text{KH}_2\text{PO}_4$  (purple filled squares) was added to CN cultures as a P source at point 2 and induced a second growth phase. Error bars are SD of replicate cultures.

Cultures were shaken at 180 rpm, other than anaerobic tubes (not shaken).

For experiments with either eDNA and/or C, N, and P sources added after a starvation period (no C, N or P), supplementation occurred through a 10% v/v addition of purified DNA solution, and/or 10× sodium lactate, ammonium chloride or monopotassium phosphate solution (in DNA solvent, 10 mM Tris-Cl, pH 8.0), or negative control solution (10 mM Tris-Cl, pH 8.0). Optical density measurements of replicate cultures were taken on a shaking and incubating microplate reader (ThermoFisher Multiskan FC, measurement filter 620 nm) or a Biorad SmartSpec Plus (600 nm) over a 5–10 day period (depending on length of starvation, or induction of several growth phases as in Figure 1). All OD values from culture tube or flask experiments were measured using the Biorad SmartSpec Plus and are greater than identical samples read in 96-well microplates on Multiskan FC; this is caused by the difference in measured wavelength (600 and 620 nm) and a volume added to each microplate well (200  $\mu\text{l}$ ) less than the full path length (1 cm). Therefore, there are non-random differences in OD values between experiments measured by OD at 600 and 620 nm.

Viable cell count experiments were conducted within 96-well microplates. Triplicate *H. volcanii* DS2 cultures were starved of  $\text{KH}_2\text{PO}_4$  for 5 days in Hv-min CN medium after which samples were removed at indicated time points. The  $T = 0$  sample was taken after the addition of DNA solvent or unmethylated *E. coli* DNA to a final concentration of 500  $\mu\text{g}/\text{ml}$ , and optical density was monitored simultaneously. After 4 days the stationary phase sample was taken and cell titers were quantified through a serial dilution of each culture and plating for colony forming units (CFUs) on Hv-YPC medium.

Where significant difference is noted between conditions for growth experiment results, significance of difference between data series was determined by One-Way ANOVA test of the mean for each replicate set ( $n \geq 3$  biological replicates). A significant difference between means was defined as a  $P \leq 0.05$ , while no significant difference between two means was defined as a  $P \geq 0.05$ . Where increase in OD is shown, average increase in OD was calculated by subtracting the initial OD value at time zero, or OD of control culture (e.g., value for CN medium alone from value for CN +eDNA), from the final value achieved after incubation for each replicate.

#### ASSESSMENT OF EXTRACELLULAR DNase ACTIVITY

Conditioned media (CM) was harvested from *H. volcanii* DS2 and *Staphylococcus aureus* (as a positive control) by centrifuging exponential phase ( $\text{OD}_{600\text{ nm}} \sim 0.6$ ) cells grown in rich medium and passing the supernatant through a 0.22  $\mu\text{m}$  filter. Medium had been previously inoculated with a single colony. CM harvested from the two species was supplemented with 30  $\mu\text{g}$  of unmethylated pTA131 plasmid DNA (final concentration 150  $\text{ng}/\text{ml}$ ), and incubated for 12 h at 37°C. During this time, secreted nuclease in the CM is exposed to and may degrade high molecular weight DNA fragments, resulting in smearing of lower molecular weight DNA detectable on an agarose gel stained with ethidium bromide. The DNase I positive control digestion was performed according to manufacturer protocol (Invitrogen, 18068–015).

#### FLUORESCENCE MICROSCOPY

One microgram of unmethylated *E. coli* DNA as prepared for growth studies was first digested for 10 min with DNase I (Invitrogen, 18068–015) to increase reaction efficiency and probe

fragmentation and labeled with the Ulysis Alexa Fluor® 488 Nucleic Acids Labeling Kit (Molecular Probes, U-21650). The labeling reaction occurs during an 80°C incubation step, creating an irreversible complex between the Alexa Fluor® fluorophore and guanine and adenine bases. Labeled DNA was purified from un-reacted probe using mini-gel filtration spin columns (Bio-Rad Bio-Spin P-30, in Tris buffer, 732–6231) as recommended.

*H. volcanii* DS2 cells were grown to mid-log phase (OD 0.4) in Hv-YPC medium, pelleted, washed three-times and resuspended in a basal salts medium (Hv-starve). Cells were incubated with freshly labeled and purified DNA at a final concentration of 10 ng/μl for 1 h at 42°C, after which they were pelleted once again and suspended in Hv-starve medium to remove excess probe. Preparations of live cells were visualized immediately using a Nikon ECLIPSE TE-300 inverted fluorescence microscope. Photographs of labeled cells viewed at 600× total magnification were collected under white light and with excitation at 488 nm (pseudocolored green). Cells did not autofluoresce at the tested excitation wavelength as verified by no detectable signal in identically prepared unlabeled cells.

#### IDENTIFICATION OF PUTATIVE DNA METABOLISM GENE *Hvo\_1477* AND RELATED GENES

The *Hvo\_1477* protein (YP\_003535526) was identified as a putative membrane-bound nuclease involved in DNA metabolism through a BLASTP search (Altschul et al., 1990) of the *H. volcanii* genome (Hartman et al., 2010) using known bacterial proteins (**Table 2**) as queries. *Hvo\_1477* was targeted as a homolog of bacterial nuclease YokF (YP\_007534137). Clusters of Orthologous Groups or COGs (Tatusov, 1997), protein domains, and protein superfamilies (Gough and Chothia, 2002; Sigrist et al., 2013) were identified using the MicrobesOnLine portal (Dehal et al., 2010).

The phylogenetic tree of archaeal YokF/*Hvo\_1477* homologs was created using Seaview (Gouy et al., 2010). All homologs shown have an *E*-value of 1e-10 or lower in a pairwise BLASTP with YokF and were aligned with Clustal Omega (Sievers et al., 2011). The tree was constructed using PhyML (Guindon and Gascuel, 2003).

#### GENE DELETION AND COMPLEMENTATION

Deletion of *Hvo\_1477* was carried out using the pop-in pop-out method as previously described (Bitan-Banin et al., 2003; Allers et al., 2004). Briefly, the upstream flanking region of *Hvo\_1477* was amplified and restriction sites were incorporated into PCR products within primers (**Table 3**). The PCR products were purified using a Qiagen PCR purification kit (Qiagen, 28106) and digested with EcoRI (enzymes were purchased from New England Biolabs). The flanking region products were then purified using a Qiagen gel purification kit (Qiagen, 28706) and ligated with T4 ligase (Promega, M1804). The final product was verified using the forward primer of the upstream flanking regions and the reverse primer of the downstream flanking region (**Table 3**). Plasmid pTA131 (**Table 1**) and ligated flanking regions were digested with HindIII and NotI and the digested products were gel purified and ligated. The ligated plasmid (pTA131\_1477del) was transformed into competent *E. coli* cells from New England Biolabs (**Table 1**), grown on LB-ampicillin plates and extracted from

isolated ampicillin resistant colonies. Competent cells of *H. volcanii* strain H26 were made and transformed with the extracted plasmid via the standard polyethylene glycol method (Charlebois et al., 1987; Dyall-Smith, 2009) and plated on Hv-Ca medium for pop-in. Pop-in colonies were selected through colony PCR and plated on Hv-Ca with 5-FOA (Zymo Research, F9003) to counter-select for pop-outs. Final deletion mutants were identified through a colony PCR screen of 5-FOA resistant colonies.

For complementation plasmid construction, the native promoter for *Hvo\_1477* was predicted with the Neural Network Promoter Prediction site (<http://www.fruitfly.org/seq-tools/promoter.html>) and primers were constructed (**Table 3**) to include this region (beginning 125 bp from the start codon of *Hvo\_1477*). The product containing *Hvo\_1477* and promoter was ligated into pTA409 after digestion and gel purification of insert and plasmid with BamHI and EcoRI, creating the plasmid pKD409\_1477c (**Table 1**). The product was transformed into competent *E. coli* cells and colonies were selected and confirmed using PCR. Amplified product was purified, transformed into *H. volcanii* Δ*Hvo\_1477* cells (**Table 1**) and then plated for selection on Hv-Ca plates. Colonies that had regained uracil prototrophy and grew on Hv-Ca were grown in liquid medium and the final complemented strain (Δ*Hvo\_1477c*, **Table 1**) was confirmed using colony PCR.

## RESULTS

#### eDNA METABOLISM AND PRIMARY ROLE AS A PHOSPHORUS SOURCE

An ability to utilize eDNA to drive metabolic growth in *H. volcanii* was observed using several experimental platforms and methods. We first used 96-well microplates as a routine method for monitoring growth through optical density of replicate cultures, and discovered an increase in OD of *H. volcanii* cultures after supplementation with RNase treated, freshly precipitated and highly purified HMW double-stranded DNA (as in **Figure 1**). Initial experiments were followed by additional microplate-based studies whereby all possible combinations of typical C, N and P sources were supplemented after a starvation period (aimed at depletion of intracellular stores).

This next phase of starvation experiments lead to the principle findings that (i) *H. volcanii* stores phosphorus intracellularly, and (ii) eDNA is utilized primarily as a source of phosphorus. Internal P storage is demonstrated by the observation that cells starved of C, N and P for 6 days and then supplemented with C and N but not P were able to reproduce through approximately one growth phase (**Figure 1B**). A second growth phase was then induced in these same replicate cultures by the addition of KH<sub>2</sub>PO<sub>4</sub> alone, further indicating that cessation of growth was indeed due to P limitation (**Figure 1B**, purple squares). Likewise, eDNA's role as a P source is demonstrated by the observation that eDNA supplementation in starved cultures led to a significant growth advantage only when cultures were provided with C and N (CN medium, **Figure 1A**)—further verified with a second addition of eDNA after CN culture cells had reached the stationary phase—which caused a second phase of exponential growth, again only in CN medium (**Figure 1A**, white bars). DNaseI digested DNA was also tested and led to growth in CN medium equal to that of undigested DNA (data not shown), indicating *H. volcanii* can

**Table 2 | Hvo\_1477 protein homologs with known functions.**

Species	Name	Accession numbers	Functions and features	References	BLASTe-value
<i>Bacillus subtilis</i>	YokF	YP_007534137	Surface-bound nuclease, lipoprotein/involved in eDNA metabolism	Sakamoto et al., 2001	2e-22
	YhcR	EME06910	Sugar-non-specific surface-bound endonuclease, sortase substrate	Oussenko et al., 2004	4e-12
	YncB	YP_007533717	Nuclease, YokF paralog	Sakamoto et al., 2001	1e-27
<i>Staphylococcus</i> spp.	Nuc	BAH56528	Secreted nuclease, conditionally surface-bound/ modulator of biofilm formation	Chesneau and El Solh, 1994; Kiedrowski et al., 2011	5e-25

**Table 3 | Oligonucleotide primers used.**

Primer	Sequences (5' -3') <sup>a</sup>	Properties
Hvo_1477FR1_F	TTTAAGCTTCGTGCGCCCGATTCTTCT	<i>Hvo_1477</i> deletion, upstream region external, HindIII site
Hvo_1477FR1_R	TTT <u>GAA</u> TTCGATTCAACCGTTAGTCAGGG	<i>Hvo_1477</i> deletion, upstream region internal, EcoRI site
Hvo_1477FR2_F	AAAGAATT <u>CCG</u> ACGAAGTGCTCGCGTACA	<i>Hvo_1477</i> deletion, upstream downstream region internal, EcoRI site
Hvo_1477FR2_R	TTT <u>GCGCCG</u> CCGAGATGCGCGGGGAGGT	<i>Hvo_1477</i> deletion, downstream region external, NotI site
Hvo_1477P_F	AAAGGAT <u>CCG</u> TCCATTAAAAGTTCTGGT	Forward complementation, with native promoter, BamHI site
Hvo_1477_R	AAAGAATT <u>CC</u> CAGTGTCTCCCCAACAGCGG	Complementation, reverse, EcoRI site

<sup>a</sup>Restriction endonuclease sites are underlined.

utilize nucleotides and small oligonucleotide products in addition to HMW DNA. While microplate experiments are useful for high-throughput assays encompassing many conditions, concerns regarding growth limitation due to small culture volume and oxygen availability led us to validate observed trends using several independent culturing methods and conditions.

Further OD-based studies included a dose-dependence experiment, in addition to culturing in larger volumes within baffled and unbaffled flasks (**Figure 2**), and in culture tubes during both aerobic (**Figures 3A,B**) and anaerobic conditions (**Figure 3C**). As expected, a linear relationship between increasing eDNA concentration and OD<sub>600 nm</sub> was measured during growth in CN medium: with absorbance readings reaching 129% above control values at 250 µg/ml (**Figure 3B**). A scaled-up experiment with 20 ml of culture grown in culture flasks (100× greater volume than microplate wells) was conducted with OD readings taken after eDNA or DNA solvent was supplemented in NP, CP or CN medium inoculated with starved *H. volcanii* cells. As in microplate-based experiments, growth in each medium type without eDNA is indicative of a capacity for internal storage of the missing macronutrient. Relatively weak but significant growth (*P*-value of 0.002 when compared to starvation cultures) without eDNA was only measured in CN medium, affirming internal P storage, and suggesting insufficient internal C or N stores capable of driving cellular division (**Figure 2**). Also consistent with microplate experiments, eDNA supplementation led to a large increase in OD only in CN medium, and a small but significant increase in CP medium, confirming the use of eDNA as a P source and suggesting a role as a weak nitrogen source (**Figure 2**). Cell cultures grown in anaerobic tubes during

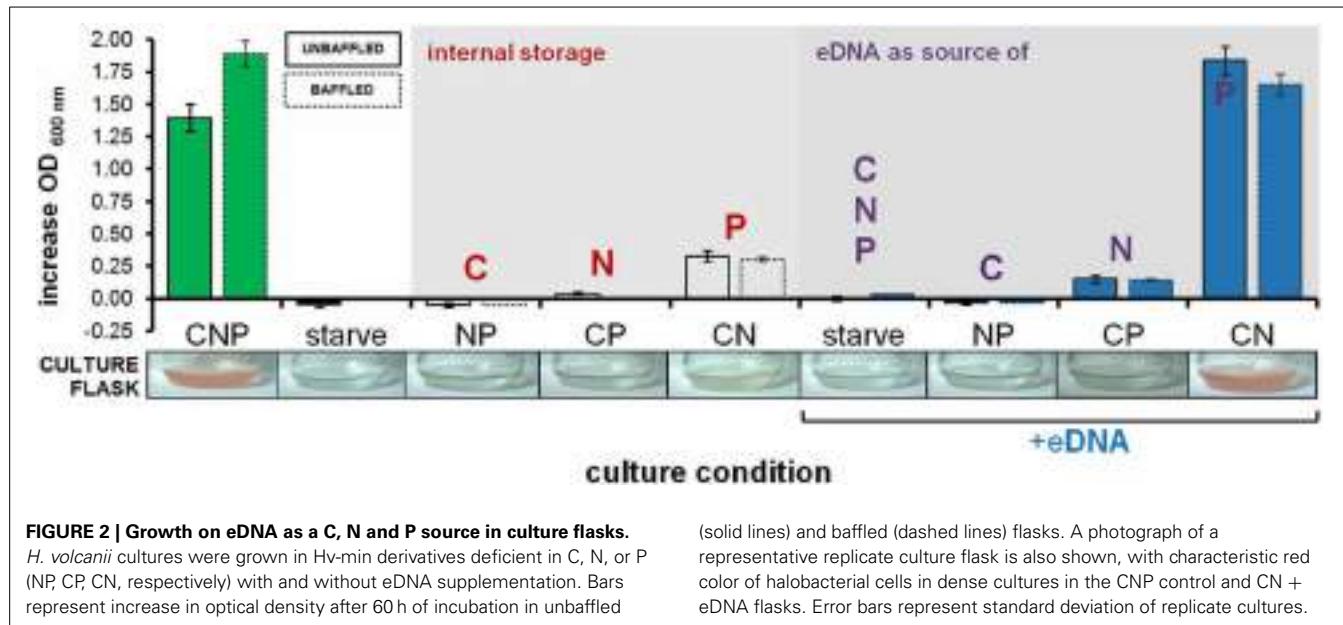
nitrate respiration were also able to utilize eDNA as a P source (**Figure 3C**).

Viable cell counts also verified growth on eDNA as a P source. Averaged CFUs at stationary phase for a culture starved of P (through growth in CN media, as in **Figure 1B**) within a microplate were over seventy times greater with eDNA supplementation as compared to control cultures (DNA solvent alone) (**Table 4**). The optical density-based growth curve for this culture indicated approximately one doubling during this same period (**Table 4**), typical of most eDNA supplementation experiments described here. This indicates that while OD measurements are useful for revealing overall trends, viable cell numbers are underestimated, likely due to difference in light scattering properties such as cell shape, size, and intracellular composition.

### SELECTIVE METABOLISM OF AVAILABLE eDNAs

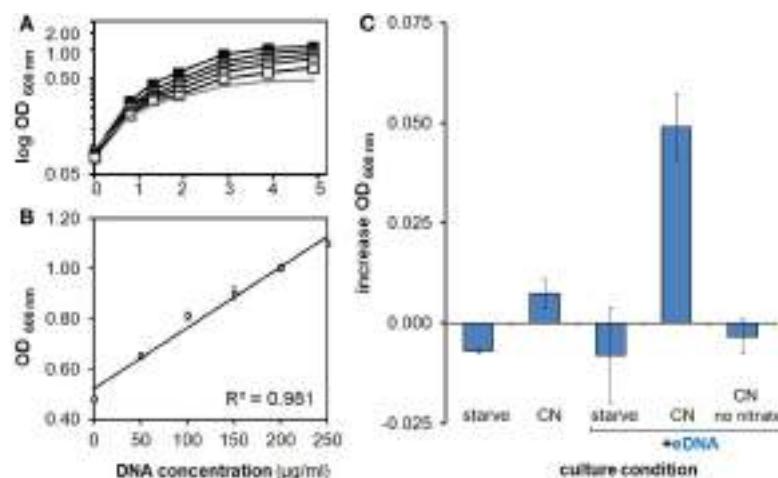
Our first observation of growth on eDNA occurred when supplementing *H. volcanii* with its own genomic DNA (i.e., conspecific DNA). However, we soon noticed an inability to metabolize certain DNA types when we attempted to grow *H. volcanii* on eDNA extracted from other DNA sources. This began with an inability to utilize herring sperm DNA and *E. coli* DNA (no growth advantage in CN media, **Figure 4**).

We then tested additional DNA types in order to identify any features or properties of available eDNA that could be discriminated by *H. volcanii* cells and any associated molecular components involved in DNA metabolism. After a bioinformatic search for DNA uptake signals like those found throughout the genomes of many competent Gram negative bacteria (Redfield et al., 2006) produced no putative short hyper-represented motifs,

**FIGURE 2 | Growth on eDNA as a C, N and P source in culture flasks.**

*H. volcanii* cultures were grown in Hv-min derivatives deficient in C, N, or P (NP, CP, CN, respectively) with and without eDNA supplementation. Bars represent increase in optical density after 60 h of incubation in unbaffled

(solid lines) and baffled (dashed lines) flasks. A photograph of a representative replicate culture flask is also shown, with characteristic red color of halobacterial cells in dense cultures in the CNP control and CN + eDNA flasks. Error bars represent standard deviation of replicate cultures.

**FIGURE 3 | Characterization of growth on eDNA as a phosphorus source.**

(A) Growth on eDNA as a P source is concentration dependent. *H. volcanii* DS2 cultures in minimal medium lacking a phosphorus source were provided with unmethylated *E. coli* DNA at final concentrations of 50, 100, 150, 200, and 250  $\mu\text{g}/\text{ml}$  (gray squares, increasing darkness). (B) Optical density achieved after 5 days of incubation at increasing DNA concentrations. (C)

Growth on eDNA after 48 h during anaerobic nitrate respiration. *H. volcanii* DS2 cultures in minimal medium lacking a phosphorus source were prepared and grown under anaerobic conditions with unmethylated *E. coli* DNA at a final concentration of 200  $\mu\text{g}/\text{ml}$ . All cultures other than the no nitrate eDNA control contained 50 mM sodium nitrate. Errors bars represent standard deviation of replicates.

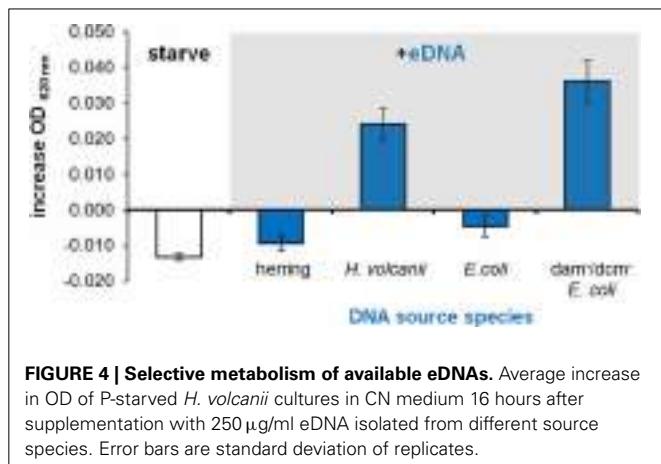
**Table 4 | Viable cell count of phosphorous-starved DNA supplemented *H. volcanii* DS2 cultures.**

	T = 0 (lag phase)		4 days (stationary phase)		Change in average CFU/ml
	CFU/ml	OD <sub>620 nm</sub> <sup>a</sup>	CFU/ml	OD <sub>620 nm</sub>	
DNA solvent	$4.47 \times 10^5 \pm 4.04 \times 10^3$	$0.1708 \pm 2.30 \times 10^{-3}$	$9.26 \times 10^4 \pm 1.33 \times 10^3$	$0.1648 \pm 1.10 \times 10^{-3}$	2.07-fold decrease
500 $\mu\text{g}/\text{ml}$ DNA	$3.93 \times 10^5 \pm 2.73 \times 10^3$	$0.1657 \pm 2.22 \times 10^{-3}$	$2.82 \times 10^7 \pm 1.20 \times 10^5$	$0.3015 \pm 4.45 \times 10^{-3}$	71.8-fold increase

<sup>a</sup>Corresponding optical density at time of sampling. Error shown is SD of counts from triplicate cultures.

we supplemented *H. volcanii* cultures with eDNA extracted from *Micrococcus luteus*, a divergent bacterial species with a high G-C content similar to that of the *Haloferax volcanii* genome (Hartman et al., 2010), and again observed no growth (data not shown). This decreased the likelihood that selectivity was primarily due to an uptake signal sequence feature or G-C content and we moved on to chemical modification of eDNA by methylation.

DNA extracted from an *E. coli* K12 strain with *dam* and *dcm* DNA methyltransferases deleted (Table 1) was tested, and we observed that along with conspecific DNA, unmethylated *E. coli* DNA led to a significant increase in OD at stationary phase (Figure 4). An *E. coli* strain with a single DNA methylation gene deleted (*dam*) was also tested and showed significant growth between that of fully methylated (DH5 $\alpha$ , no growth) and unmethylated DNA (data not shown). Unmethylated *E. coli* DNA fragments were also labeled with a fluorescent probe and incubated with *H. volcanii* in liquid culture to test for association of cells with eDNA. A majority of cells as visualized under white light (first panel) co-localized with strong signal from labeled eDNA

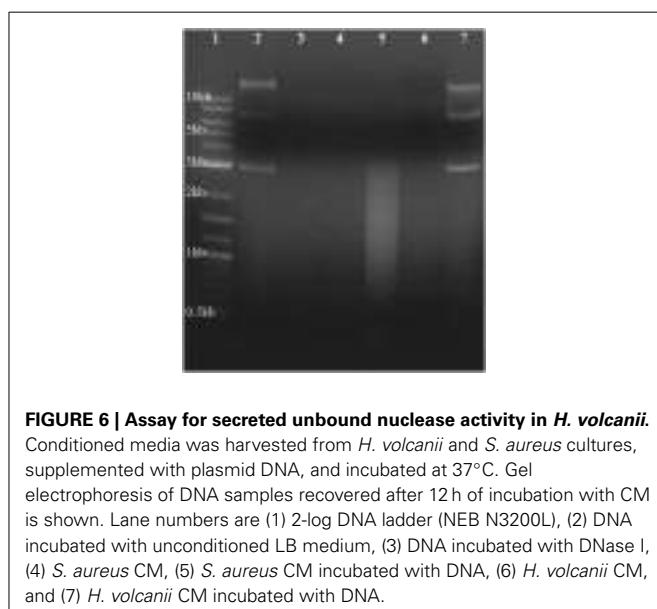


**FIGURE 4 | Selective metabolism of available eDNAs.** Average increase in OD of P-starved *H. volcanii* cultures in CN medium 16 hours after supplementation with 250  $\mu$ g/ml eDNA isolated from different source species. Error bars are standard deviation of replicates.

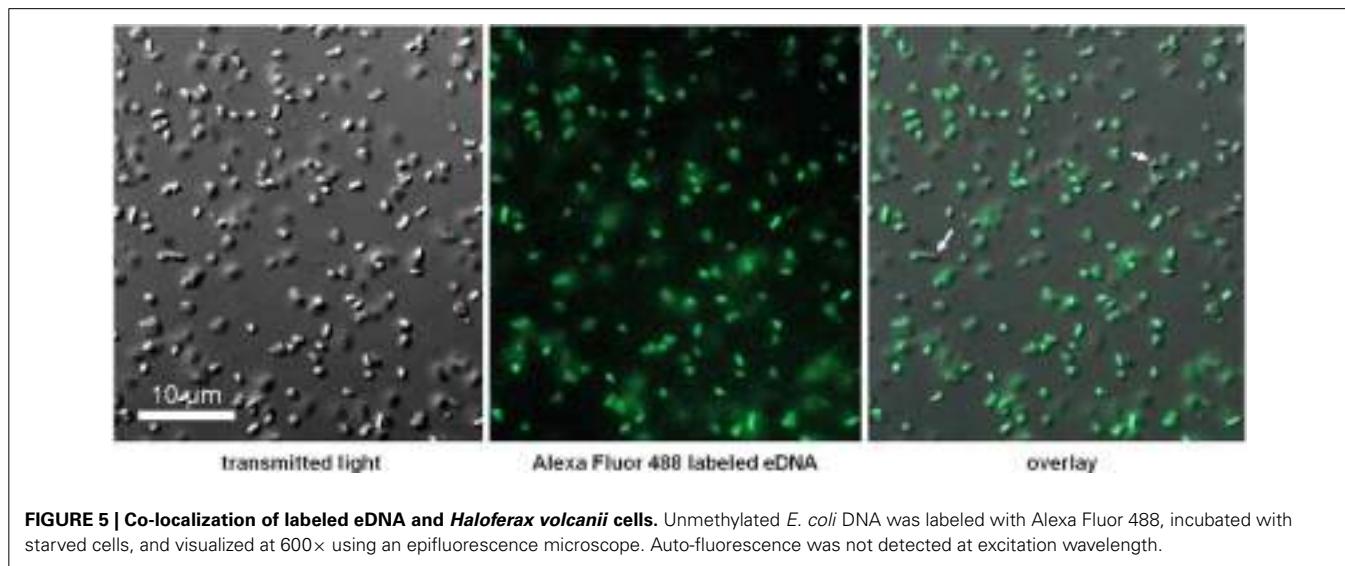
(Figure 5, third panel), while some visible cells in the focal plane appear not to co-localize with eDNA (Figure 5, third panel, white arrows).

#### SCREENS FOR HYDROLASE ACTIVITY WITHIN CONDITIONED MEDIUM INDICATE ABSENCE OF UNBOUND SECRETED NUCLEASE

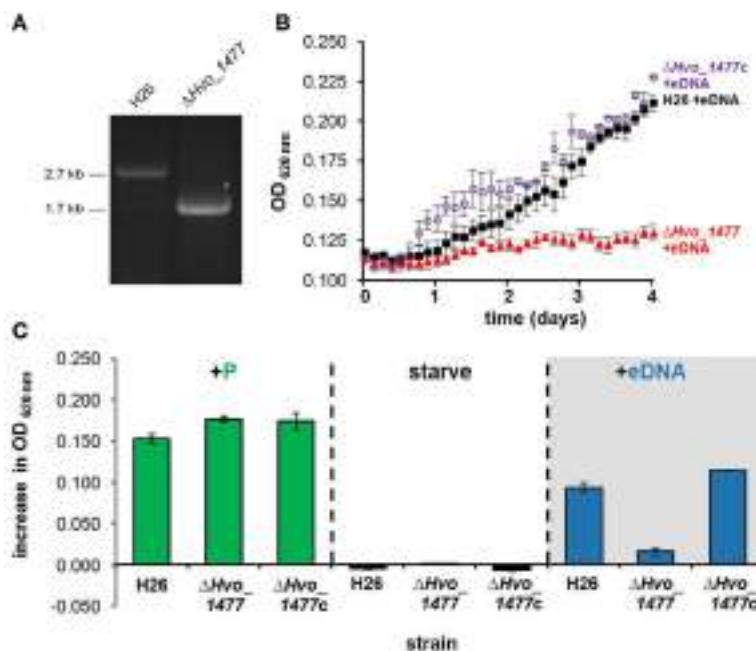
DNase activity from CM was assayed to evaluate the presence of hydrolytic enzyme secreted into the environment during growth. CM from the nuclease secreting bacterial species *S. aureus* (Figure 6, lane 5) produced the expected smear of DNA fragments ranging from  $\sim$ 2000 to 200 bp, with none of the original bands remaining. DNA within *H. volcanii* CM remained intact (Figure 6, lane 7), as in lane 2 in which DNA was added to non-conditioned medium, indicating an absence of evidence for eDNA degradation in *H. volcanii* CM. Previous studies have



**FIGURE 6 | Assay for secreted unbound nuclease activity in *H. volcanii*.** Conditioned media was harvested from *H. volcanii* and *S. aureus* cultures, supplemented with plasmid DNA, and incubated at 37°C. Gel electrophoresis of DNA samples recovered after 12 h of incubation with CM is shown. Lane numbers are (1) 2-log DNA ladder (NEB N3200L), (2) DNA incubated with unconditioned LB medium, (3) DNA incubated with DNase I, (4) *S. aureus* CM, (5) *S. aureus* CM incubated with DNA, (6) *H. volcanii* CM, and (7) *H. volcanii* CM incubated with DNA.



**FIGURE 5 | Co-localization of labeled eDNA and *Haloferax volcanii* cells.** Unmethylated *E. coli* DNA was labeled with Alexa Fluor 488, incubated with starved cells, and visualized at 600 $\times$  using an epifluorescence microscope. Auto-fluorescence was not detected at excitation wavelength.



**FIGURE 7 |** *Hvo\_1477* is required for growth on DNA. **(A)** Deletion of chromosomal gene *Hvo\_1477* in *H. volcanii* strain H26. PCR amplification of H26 and  $\Delta$ *Hvo\_1477* template DNA using the forward primer for the upstream *Hvo\_1477* flanking region, and the reverse primer of the downstream flanking region (*Hvo\_1477FR1\_F* and *Hvo\_1477FR2\_R*, Table 3). **(B)** Growth with eDNA in CN media for H26 (parental strain,

black filled squares),  $\Delta$ *Hvo\_1477* (red filled triangles) and  $\Delta$ *Hvo\_1477* with pKD409\_1477c complementation plasmid ( $\Delta$ *Hvo\_1477c*, purple circles). OD was measured every 3 h within a shaking and incubated 96-well plate reader. **(C)** Increase in optical density 96 h after supplementation with KH<sub>2</sub>PO<sub>4</sub>, DNA solvent (continued starvation), or eDNA. Error bars represent standard deviation of triplicate cultures.

**Table 5 |** Annotated features/domains of *Hvo\_1477*.

ID	Database*/domain ID	Start/end	Description
PROKAR_LIPOPROTEIN	PS51257	1/20	Lipobox motif
TNASE_3	PS50830	67/92	Thermostable nuclease domain
Micrococcal nuclease	COG1525	23/215	Micrococcal nuclease domain
Staphylococcal nuclease	SSF50199	55/203	Staphylococcal nuclease superfamily
Helix-turn-helix	HTH	171/192	DNA binding motif
Thermonuclease active site	PS01123	67/92	Active site of thermonuclease
Lamin A/C globular tail	SSF74853	236/326	No putative function

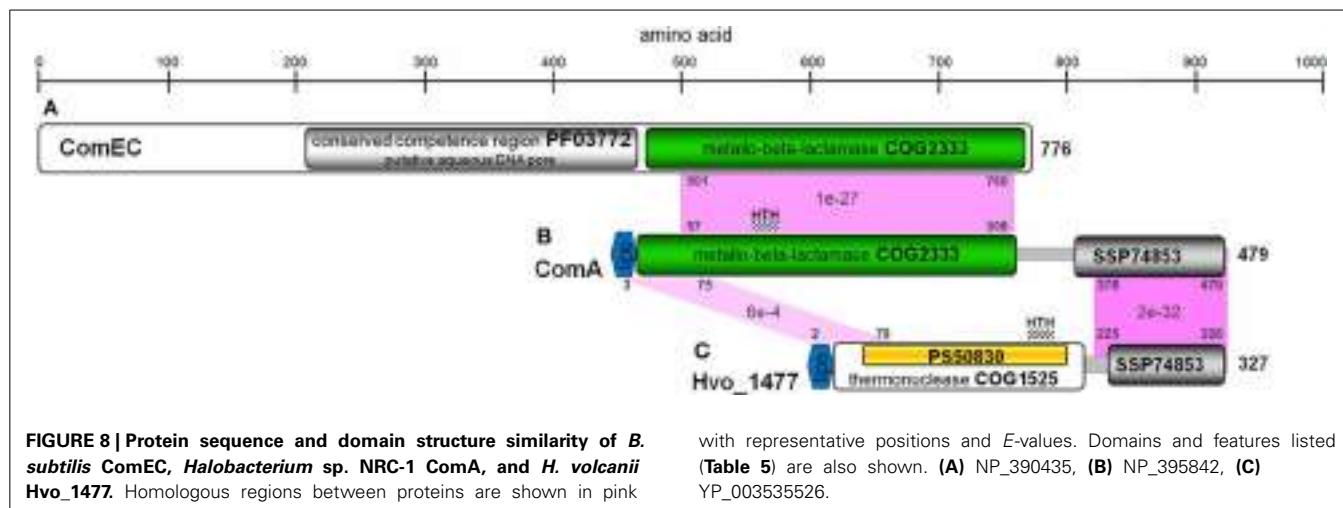
\*Database abbreviations: PS, PROSITE Database (Sigrist et al., 2013); COG, Cluster of Orthologous Groups (Tatusov, 1997); SS, SCOP Superfamily Database (Gough and Chothia, 2002).

also reported a lack of secreted nucleases in haloarchaeal species (Ventosa et al., 2005).

#### ***Hvo\_1477* IS IMPORTANT FOR DNA METABOLISM IN *H. volcanii***

Deletion of *Hvo\_1477* diminished growth on DNA (Figure 7). This phenotype was confirmed by complementation with plasmid pKD409\_1477 (Table 1) containing *Hvo\_1477* and its native promoter, and resulted in the restoration of growth on DNA to levels slightly greater ( $P = 0.049$ ) than that of the parental strain (H26, Figure 7), possibly due to multiple copies of the plasmid. No additional phenotype for the  $\Delta$ *Hvo\_1477* strain has been observed at the time of publication; growth rates in minimal medium (Hv-min) with sodium lactate as a carbon source are equal to that of H26 (Figure 7, +P).

*Hvo\_1477* is a 327 aa protein that has a predicted size of 34.1 kDa, several annotated sequence features (Table 5), and is homologous to known bacterial extracellular nuclease proteins YokF, YhcR, and YncB of *Bacillus subtilis*, and Nuc of staphylococcal species (Chesneau and El Solh, 1994) (Table 2). While the *Hvo\_1477* gene and associated haloarchaeal homologs are annotated as competence-like protein-encoding genes or *comA*, there is no homology between the *Hvo\_1477* protein sequence and bacterial competence protein orthologs ComA/ComEC, which contain multiple membrane-spanning regions and are known to form an aqueous DNA pore during natural DNA uptake (Facius and Meyer, 1993; Draskovic and Dubnau, 2005) (Figure 8). The basis for this annotation may be that some halobacterial homologs of *Hvo\_1477* (including *Halobacterium* sp. NRC-1, as



shown in **Figure 8**) are larger proteins containing an additional metallo-beta-lactamase domain (COG 2333) that does share a region of similarity with bacterial ComA/ComEC. These larger haloarchaeal bacterial nuclease homologs clustered together in the phylogenetic tree of Hvo\_1477 homologs (**Figure 9**, group 2), and some species such as *Haloarcula marismortui* have both the smaller thermonuclease containing (see **Figure 9**, group 1), and larger metallo-beta-lactamase containing version. However, homology of group 2 proteins (**Figure 9**) with ComEC/ComA appears to be based on only a single shared domain (COG 2333): all haloarchaeal homologs are missing the important putative DNA pore domain or “conserved competence region” (**Figure 8**).

## DISCUSSION

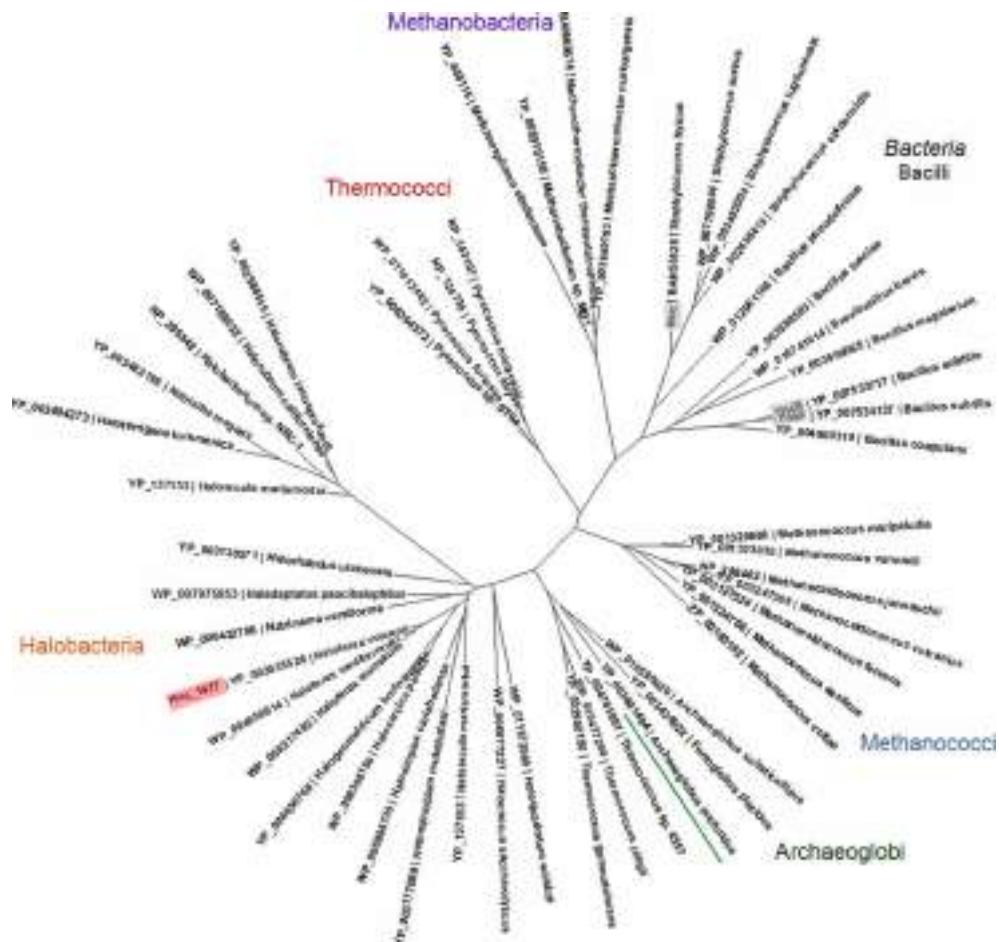
The discovery of an ability to metabolize HMW eDNA in the Halobacteria is a central finding of this work. While bacterial species are known to use eDNA as a nutrient source (Finkel and Kolter, 2001; Sakamoto et al., 2001; Palchevskiy and Finkel, 2006; Lennon, 2007; Pinchuk et al., 2008; Mulcahy et al., 2010), here we report this capacity in an archaeon. Most bacterial species known to use DNA as a nutrient metabolize DNA as a source of C, N and/or P; here we show that *H. volcanii* uses eDNA almost exclusively as a source of phosphorus (**Figures 1, 2**).

Our experimental demonstration of DNA metabolism as a P source in *H. volcanii* adds to previous reports showing that (i) eDNA concentrations are exceptionally high in hypersaline sample sites and (ii) organisms living in hypersaline environments are often limited by phosphorus (Oren and Shilo, 1982; Oren, 1983; Ludwig et al., 2006). We therefore propose that nutritional DNA uptake may be a primary mechanism through which haloarchaeal species obtain phosphorus and that DNA is likely a major currency of P exchange and storage in hypersaline environments. DNA is indeed a P-rich molecule (10% by weight), and has been shown to account for over 40% of P cycling in some environments (Dell'Anno and Danovaro, 2005; Corinaldesi et al., 2007, 2008). Interestingly, because *H. volcanii* is highly polyploid, intracellular DNA stores may also be important for biogeochemical systems in the environment (Soppa, 2013). Furthermore, the distribution

of Hvo\_1477 homologs throughout the Euryarchaeota (**Figure 9**) suggests DNA metabolism could be an important physiological ability relevant in many species and ecosystems.

We also discovered a bias in metabolism toward conspecific and unmethylated eDNAs, whereby highly divergent eDNA is only utilized when unmethylated (**Figure 4**). Because *H. volcanii* methylates its own DNA (Hartman et al., 2010), we suggest available eDNAs are processed through recognition of methylation patterns. This is the first report demonstrating the importance of methylation for eDNA discrimination and the extent of this characteristic among other prokaryotes is unknown. However, the presence of such a system for discrimination of eDNA offers a possible explanation to the finding that eDNA accumulates and remains preserved in environments despite high overall levels of DNase activity (Corinaldesi et al., 2008). High concentrations of eDNA found in a particular environment may reflect the inability of all organisms living there to utilize the available DNA because they cannot process and/or import it. For instance, there are many bacterial and eukaryal cells that live in hypersaline environments, and their DNA would not be methylated in a manner that *H. volcanii* can recognize and utilize.

Hvo\_1477 is the first eDNA processing/uptake related gene identified in an archaeon—a starting point toward understanding an archaeal eDNA degradation mechanism and associated phenotypes. Hvo\_1477 is not a ComA/ComEC homolog, but is instead a putative lipoprotein with a thermonuclease domain (**Tables 2, 5, Figures 8, 9**). Lipoproteins are secreted and attached to the cell surface in both bacterial and archaeal species (Szabo and Pohlschroder, 2012) and surface-bound nucleases in bacteria such as the YokF-related *B. subtilis* Hvo\_1477 homologs listed in **Table 2** (Sakamoto et al., 2001; Oussenko et al., 2004) and ExEM (not a Hvo\_1477 homolog) in *Shewanella* species (Godeke et al., 2011a) are known to be involved in DNA metabolism. Deleting *yokF* and its paralogs (e.g., *yncB*) in *B. subtilis* also greatly reduced but did not abolish growth levels on eDNA (Sakamoto et al., 2001). The use of DNA as a nutrient is considered a form of natural DNA uptake or natural competence (Finkel and Kolter, 2001) (NC); however, in strict terms, NC is defined by internalization of intact DNA fragments and by the presence of a complex



**FIGURE 9 | Phylogenetic tree of archaeal YokF/Nuc protein homologs.** A protein alignment was created with Clustal Omega (Sievers et al., 2011) and used to create a maximum likelihood tree using PhyML (Guindon and Gascuel, 2003). Proteins included in alignment were identified in a BLASTP

search of the non-redundant protein database and have an *E*-value of 1e-10 or lower in a pairwise BLASTP with YokF. Accession numbers and species names are shown; bacterial proteins with known functions (see also Table 2) are labeled gray.

molecular machine responsible for DNA binding, processing, and internalization (Chen and Dubnau, 2004). At this point, we have not conclusively demonstrated that HMW eDNA is imported across halobacterial membranes.

It remains unclear in *H. volcanii* whether Hvo\_1477 is associated with additional surface or transmembrane proteins and if eDNA is imported as HMW DNA into the cell. Other bacterial surface-bound nucleases, including EndA of *Streptococcus* species and NucA in *Bacillus* species are responsible for processing eDNA prior to internalization by additional proteins (Provvedi et al., 2001; Chen and Dubnau, 2004). It is possible then that Hvo\_1477 also acts as part of an unknown archaeal DNA uptake complex. The observation that *H. volcanii* is biased against most sources of DNA it can utilize for metabolism suggests that HMW DNA is indeed moved across the membrane and into the cell; if DNA hydrolysis occurred extracellularly, and only nucleotides were imported for growth, it is difficult to explain why cells would reject most DNA sources and undergo starvation. Fluorescence microscopy experiments revealed that labeled eDNA co-localizes with *H. volcanii* cells (Figure 5), consistent with the assumption

that HMW DNA associates with the cell during a multi-part process of DNA processing and metabolism. Because not all cells co-localized with DNA, it is possible that a fraction of cells within a given population do not express DNA binding factors (i.e., regulated expression) and are unable to associate with eDNA. The identification of protein-protein interactions, regulation, and dynamics of eDNA processing at the cell surface (including cellular binding assays with multiple DNA types), and further biochemical characterization of Hvo\_1477 are necessary for further insight.

Drawing from studies of surface-associated nucleases in bacterial species, it seems likely that Hvo\_1477 has additional important phenotypes in *H. volcanii*. DNA degradation for “food” is only one useful physiological function of a surface-bound nuclease in the DNA-rich milieu in which prokaryotes live. eDNA has been proposed as a structural element of bacterial biofilm structure (Dominiak et al., 2011; Godeke et al., 2011b), and plays additional roles within a biofilm such as aiding in attachment (Harmsen et al., 2010), self-organization (Gloag et al., 2013), and counteraction of antibiotic

action (Chiang et al., 2013). It is not surprising then that extracellular nucleases in bacteria (including Hvo\_1477 homolog Nuc, **Table 2**) modulate biofilm development (Kiedrowski et al., 2011). For example, the nucleases Dns in *Vibrio cholerae* and ExeM in *S. oneidensis* are involved in both eDNA processing for nutrition (ExeM) and/or natural transformation (Dns) and biofilm regulation (Blokesch and Schoolnik, 2008; Godeke et al., 2011a; Seper et al., 2011). Halobacteria form biofilms and like bacterial biofilms, high levels of eDNA are found in archaeal biofilms (Frols et al., 2012). It is possible then that Hvo\_1477 is also involved in biofilm lifecycle through its putative activity as a surface-bound nuclease (**Table 5**).

## AUTHOR CONTRIBUTIONS

Scott Chimileski, Uri Gophna, Kunal Dolas, Adit Naor, and R. Thane Papke conceived of the research and designed the experiments. Scott Chimileski, Kunal Dolas, and Adit Naor carried out and analyzed the experiments, and Scott Chimileski, Uri Gophna, Kunal Dolas, Adit Naor, and R. Thane Papke wrote the manuscript.

## ACKNOWLEDGMENTS

We wish to acknowledge the following agencies for funding this research: National Science Foundation (award numbers, 0919290 and 0830024), the U.S.–Israel Binational Science Foundation (award number 2007043) and NASA Astrobiology: Exobiology and Evolutionary Biology Program Element (Grant Number NNX12AD70G). We thank Dr. Thorsten Allers from the University of Nottingham for the *Haloferax* strains and plasmids.

## REFERENCES

- Allers, T., Barak, S., Liddell, S., Wardell, K., and Mevarech, M. (2010). Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in *Haloferax volcanii*. *Appl. Environ. Microbiol.* 76, 1759–1769. doi: 10.1128/AEM.02670-09
- Allers, T., and Mevarech, M. (2005). Archaeal genetics—the third way. *Nat. Rev. Genet.* 6, 58–73. doi: 10.1038/nrg1504
- Allers, T., Ngo, H. P., Mevarech, M., and Lloyd, R. G. (2004). Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the leuB and trpA genes. *Appl. Environ. Microbiol.* 70, 943–953. doi: 10.1128/AEM.70.2.943-953.2004
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1006/jmbi.1990.9999
- Andrei, A. S., Banciu, H. L., and Oren, A. (2012). Living with salt: metabolic and phylogenetic diversity of archaea inhabiting saline ecosystems. *FEMS Microbiol. Lett.* 330, 1–9. doi: 10.1111/j.1574-6968.2012.02526.x
- Atomi, H., Imanaka, T., and Fukui, T. (2012). Overview of the genetic tools in the Archaea. *Front. Microbiol.* 3:337. doi: 10.3389/fmicb.2012.00337
- Averhoff, B. (2009). Shuffling genes around in hot environments: the unique DNA transporter of *Thermus thermophilus*. *FEMS Microbiol. Rev.* 33, 611–626. doi: 10.1111/j.1574-6976.2008.00160.x
- Benloch, S., Acinas, S. G., Anton, J., Lopez-Lopez, A., Luz, S. P., and Rodriguez-Valera, F. (2001). Archaeal biodiversity in crystallizer ponds from a solar saltern: culture versus PCR. *Microb. Ecol.* 41, 12–19. doi: 10.1007/s002480000069
- Bitan-Banin, G., Ortenberg, R., and Mevarech, M. (2003). Development of a gene knockout system for the halophilic archaeon *Haloferax volcanii* by use of the pyrE gene. *J. Bacteriol.* 185, 772–778. doi: 10.1128/JB.185.3.772-778.2003
- Blaby, I. K., Phillips, G., Blaby-Haas, C. E., Gulig, K. S., El Yacoubi, B., and De Crecy-Lagard, V. (2010). Towards a systems approach in the genetic analysis of archaea: accelerating mutant construction and phenotypic analysis in *Haloferax volcanii*. *Archaea* 2010:426239. doi: 10.1155/2010/426239
- Blokesch, M., and Schoolnik, G. K. (2008). The extracellular nuclease Dns and its role in natural transformation of *Vibrio cholerae*. *J. Bacteriol.* 190, 7232–7240. doi: 10.1128/JB.00959-08
- Charlebois, R. L., Lam, W. L., Cline, S. W., and Doolittle, W. F. (1987). Characterization of pHV2 from *Halobacterium volcanii* and its use in demonstrating transformation of an archaeabacterium. *Proc. Natl. Acad. Sci. U.S.A.* 84, 8530–8534. doi: 10.1073/pnas.84.23.8530
- Chen, I., Christie, P. J., and Dubnau, D. (2005). The ins and outs of DNA transfer in bacteria. *Science* 310, 1456–1460. doi: 10.1126/science.1114021
- Chen, I., and Dubnau, D. (2004). DNA uptake during bacterial transformation. *Nat. Rev. Microbiol.* 2, 241–249. doi: 10.1038/nrmicro844
- Chesneau, O., and El Solh, N. (1994). Primary structure and biological features of a thermostable nuclease isolated from *Staphylococcus hyicus*. *Gene* 145, 41–47. doi: 10.1016/0378-1119(94)90320-4
- Chiang, W. C., Nilsson, M., Jensen, P. O., Hoiby, N., Nielsen, T. E., Givskov, M., et al. (2013). Extracellular DNA shields against aminoglycosides in *Pseudomonas aeruginosa* biofilms. *Antimicrob. Agents Chemother.* 57, 2352–2361. doi: 10.1128/AAC.00001-13
- Corinaldesi, C., Beolchini, F., and Dell’Anno, A. (2008). Damage and degradation rates of extracellular DNA in marine sediments: implications for the preservation of gene sequences. *Mol. Ecol.* 17, 3939–3951. doi: 10.1111/j.1365-294X.2008.03880.x
- Corinaldesi, C., Dell’Anno, A., and Danovaro, R. (2007). Early diagenesis and trophic role of extracellular DNA in different benthic ecosystems. *Limnol. Oceanogr.* 52, 1710–1717. doi: 10.4319/lo.2007.52.4.1710
- Danovaro, R., Corinaldesi, C., Dell’Anno, A., Fabiano, M., and Corselli, C. (2005). Viruses, prokaryotes and DNA in the sediments of a deep-hypersaline anoxic basin (DHAB) of the Mediterranean Sea. *Environ. Microbiol.* 7, 586–592. doi: 10.1111/j.1462-2920.2005.00727.x
- Dassarma, P., and Dassarma, S. (2008). On the origin of prokaryotic “species”: the taxonomy of halophilic Archaea. *Saline Syst.* 4:5. doi: 10.1186/1746-1448-4-5
- Dehal, P. S., Joachimiak, M. P., Price, M. N., Bates, J. T., Baumohl, J. K., Chivian, D., et al. (2010). MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res.* 38, D396–D400. doi: 10.1093/nar/gkp919
- Dell’Anno, A., and Corinaldesi, C. (2004). Degradation and turnover of extracellular DNA in marine sediments: ecological and methodological considerations. *Appl. Environ. Microbiol.* 70, 4384–4386. doi: 10.1128/AEM.70.7.4384-4386.2004
- Dell’Anno, A., and Danovaro, R. (2005). Extracellular DNA plays a key role in deep-sea ecosystem functioning. *Science* 309, 2179. doi: 10.1126/science.1117475
- Dominiak, D. M., Nielsen, J. L., and Nielsen, P. H. (2011). Extracellular DNA is abundant and important for microcolony strength in mixed microbial biofilms. *Environ. Microbiol.* 13, 710–721. doi: 10.1111/j.1462-2920.2010.02375.x
- Draskovic, I., and Dubnau, D. (2005). Biogenesis of a putative channel protein, ComEC, required for DNA uptake: membrane topology, oligomerization and formation of disulphide bonds. *Mol. Microbiol.* 55, 881–896. doi: 10.1111/j.1365-2958.2004.04430.x
- Dyall-Smith, M. (2009). *Protocols for Haloarchaeal Genetics: The Halohandbook*. Ver 7.1. Available online at: <http://www.haloarchaea.com/resources/halohandbook>
- Facius, D., and Meyer, T. F. (1993). A novel determinant (comA) essential for natural transformation competence in *Neisseria gonorrhoeae* and the effect of a comA defect on pilin variation. *Mol. Microbiol.* 10, 699–712. doi: 10.1111/j.1365-2958.1993.tb00942.x
- Finkel, S. E., and Kolter, R. (2001). DNA as a nutrient: novel role for bacterial competence gene homologs. *J. Bacteriol.* 183, 6288–6293. doi: 10.1128/JB.183.21.6288-6293.2001
- Frols, S., Dyall-Smith, M., and Pfeifer, F. (2012). Biofilm formation by haloarchaea. *Environ. Microbiol.* 14, 3159–3174. doi: 10.1111/j.1462-2920.2012.02895.x
- Gloag, E. S., Turnbull, L., Huang, A., Vallotton, P., Wang, H., Nolan, L. M., et al. (2013). Self-organization of bacterial biofilms is facilitated by extracellular DNA. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11541–11546. doi: 10.1073/pnas.1218898110
- Godeke, J., Heun, M., Bubendorfer, S., Paul, K., and Thormann, K. M. (2011a). Roles of two *Shewanella oneidensis* MR-1 extracellular endonucleases. *Appl. Environ. Microbiol.* 77, 5342–5351. doi: 10.1128/AEM.00643-11
- Godeke, J., Paul, K., Lassak, J., and Thormann, K. M. (2011b). Phage-induced lysis enhances biofilm formation in *Shewanella oneidensis* MR-1. *ISME J.* 5, 613–626. doi: 10.1038/ismej.2010.153

- Gough, J., and Chothia, C. (2002). SUPERFAMILY: HMMs representing all proteins of known structure. SCOP sequence searches, alignments and genome assignments. *Nucleic Acids Res.* 30, 268–272. doi: 10.1093/nar/30.1.268
- Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224. doi: 10.1093/molbev/msp259
- Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520
- Harmsen, M., Lappann, M., Knochel, S., and Molin, S. (2010). Role of extracellular DNA during biofilm formation by *Listeria monocytogenes*. *Appl. Environ. Microbiol.* 76, 2271–2279. doi: 10.1128/AEM.02361-09
- Hartman, A. L., Norais, C., Badger, J. H., Delmas, S., Haldenby, S., Madupu, R., et al. (2010). The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS ONE* 5:e9605. doi: 10.1371/journal.pone.0009605
- Hölzle, A., Fischer, S., Heyer, R., Schutz, S., Zacharias, M., Walther, P., et al. (2008). Maturation of the 5S rRNA 5' end is catalyzed *in vitro* by the endonuclease tRNase Z in the archaeon *H. volcanii*. *RNA* 14, 928–937. doi: 10.1261/rna.933208
- Heins, J. N., Suriano, R., Taniuchi, H., and Anfinsen, C. B. (1967). Characterization of a nuclease produced by *Staphylococcus aureus*. *J. Biol. Chem.* 242, 1016–1120.
- Kiedrowski, M. R., Kavanaugh, J. S., Malone, C. L., Mootz, J. M., Voyich, J. M., Smeltzer, M. S., et al. (2011). Nuclease modulates biofilm formation in community-associated methicillin-resistant *Staphylococcus aureus*. *PLoS ONE* 6:e26714. doi: 10.1371/journal.pone.0026714
- Leigh, J. A., Albers, S. V., Atomi, H., and Allers, T. (2011). Model organisms for genetics in the domain Archaea: methanogens, halophiles, *Thermococcales* and *Sulfolobales*. *FEMS Microbiol. Rev.* 35, 577–608. doi: 10.1111/j.1574-6976.2011.00265.x
- Lennon, J. T. (2007). Diversity and metabolism of marine bacteria cultivated on dissolved DNA. *Appl. Environ. Microbiol.* 73, 2799–2805. doi: 10.1128/AEM.02674-06
- Ludwig, R., Pringault, O., de Wit, R., de Beer, D., and Jonkers, H. M. (2006). Limitation of oxygenic photosynthesis and oxygen consumption by phosphate and organic nitrogen in a hypersaline microbial mat: a microsensor study. *FEMS Microbiol. Ecol.* 57, 9–17. doi: 10.1111/j.1574-6941.2006.00109.x
- Maier, B., Chen, I., Dubnau, D., and Sheetz, M. P. (2004). DNA transport into *Bacillus subtilis* requires proton motive force to generate large molecular forces. *Nat. Struct. Mol. Biol.* 11, 643–649. doi: 10.1038/nsmb783
- Mulcahy, H., Charron-Mazenod, L., and Lewenza, S. (2010). *Pseudomonas aeruginosa* produces an extracellular deoxyribonuclease that is required for utilization of DNA as a nutrient source. *Environ. Microbiol.* 12, 1621–1629. doi: 10.1111/j.1462-2920.2010.02208.x
- Mullakhanbhui, M. F., and Larsen, H. (1975). *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Arch. Microbiol.* 104, 207–214. doi: 10.1007/BF00447326
- Oren, A. (1983). Population dynamics of halobacteria in the Dead Sea water column. *Limnol. Oceanogr.* 28, 1094–1103. doi: 10.4319/lo.1983.28.6.1094
- Oren, A., and Shilo, M. (1982). Population dynamics of *Dunaliella parva* in Dead Sea. *Limnol. Oceanogr.* 27, 201–211. doi: 10.4319/lo.1982.27.2.0201
- Oussenko, I. A., Sanchez, R., and Bechhofer, D. H. (2004). *Bacillus subtilis* YhcR, a high-molecular-weight, nonspecific endonuclease with a unique domain structure. *J. Bacteriol.* 186, 5376–5383. doi: 10.1128/JB.186.16.5376-5383.2004
- Palchevskiy, V., and Finkel, S. E. (2006). *Escherichia coli* competence gene homologs are essential for competitive fitness and the use of DNA as a nutrient. *J. Bacteriol.* 188, 3902–3910. doi: 10.1128/JB.01974-05
- Papke, R. T., Douady, C. J., Doolittle, W. F., and Rodriguez-Valera, F. (2003). Diversity of bacteriorhodopsins in different hypersaline waters from a single Spanish saltern. *Environ. Microbiol.* 5, 1039–1045. doi: 10.1046/j.1462-2920.2003.00501.x
- Papke, R. T., Koenig, J. E., Rodriguez-Valera, F., and Doolittle, W. F. (2004). Frequent recombination in a saltern population of *Halorubrum*. *Science* 306, 1928–1929. doi: 10.1126/science.1103289
- Pasić, L., Ulrich, N. P., Crnigoj, M., Grabnar, M., and Velikonja, B. H. (2007). Haloarchaeal communities in the crystallizers of two adriatic solar salterns. *Can. J. Microbiol.* 53, 8–18. doi: 10.1139/w06-091
- Pinchuk, G. E., Ammons, C., Culley, D. E., Li, S. M., McLean, J. S., Romine, M. F., et al. (2008). Utilization of DNA as a sole source of phosphorus, carbon, and energy by *Shewanella* spp.: ecological and physiological implications for dissimilatory metal reduction. *Appl. Environ. Microbiol.* 74, 1198–1208. doi: 10.1128/AEM.02026-07
- Provvedi, R., Chen, I., and Dubnau, D. (2001). NucA is required for DNA cleavage during transformation of *Bacillus subtilis*. *Mol. Microbiol.* 40, 634–644. doi: 10.1046/j.1365-2958.2001.02406.x
- Redfield, R. J., Findlay, W. A., Bosse, J., Kroll, J. S., Cameron, A. D., and Nash, J. H. (2006). Evolution of competence and DNA uptake specificity in the *Pasteurellaceae*. *BMC Evol. Biol.* 6:82. doi: 10.1186/1471-2148-6-82
- Sakamoto, J. J., Sasaki, M., and Tsuchoo, T. (2001). Purification and characterization of a *Bacillus subtilis* 168 nuclease, YofK, involved in chromosomal DNA degradation and cell death caused by thermal shock treatments. *J. Biol. Chem.* 276, 47046–47051. doi: 10.1074/jbc.M106205200
- Schmidt, J. A., Browning, G. F., and Markham, P. F. (2007). *Mycoplasma hyopneumoniae* mhp379 is a Ca<sup>2+</sup>-dependent, sugar-nonspecific exonuclease exposed on the cell surface. *J. Bacteriol.* 189, 3414–3424. doi: 10.1128/JB.01835-06
- Seper, A., Fengler, V. H., Roier, S., Wolinski, H., Kohlwein, S. D., Bishop, A. L., et al. (2011). Extracellular nucleases and extracellular DNA play important roles in *Vibrio cholerae* biofilm formation. *Mol. Microbiol.* 82, 1015–1037. doi: 10.1111/j.1365-2958.2011.07867.x
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539. doi: 10.1038/msb.2011.75
- Sigrist, C. J., de Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., et al. (2013). New and continuing developments at PROSITE. *Nucleic Acids Res.* 41, D344–D347. doi: 10.1093/nar/gks1067
- Soppa, J. (2006). From genomes to function: haloarchaea as model organisms. *Microbiology* 152, 585–590. doi: 10.1099/mic.0.28504-0
- Soppa, J. (2013). Evolutionary advantages of polyploidy in halophilic archaea. *Biochem. Soc. Trans.* 41, 339–343. doi: 10.1042/BST20120315
- Steinberger, R. E., and Holden, P. A. (2005). Extracellular DNA in single- and multiple-species unsaturated biofilms. *Appl. Environ. Microbiol.* 71, 5404–5410. doi: 10.1128/AEM.71.9.5404-5410.2005
- Szabo, Z., and Pohlschroder, M. (2012). Diversity and subcellular distribution of archaeal secreted proteins. *Front. Microbiol.* 3:207. doi: 10.3389/fmicb.2012.00207
- Tatusov, R. L. (1997). A genomic perspective on protein families. *Science* 278, 631–637. doi: 10.1126/science.278.5338.631
- Tehei, M., Franzetti, B., Maurel, M. C., Vergne, J., Hountondji, C., and Zaccai, G. (2002). The search for traces of life: the protective effect of salt on biological macromolecules. *Extremophiles* 6, 427–430. doi: 10.1007/s00792-002-0275-6
- Ventosa, A., Sanchez-Porro, C., Martin, S., Mellado, E., and Gundl-Cimerman, N. (2005). “Halophilic archaea and bacteria as a source of extracellular hydrolytic enzymes,” in *Adaptation to Life at High Salt Concentrations in Archaea, Bacteria, and Eukarya*, eds A. Oren and A. Plemenitas (Dordrecht: Springer).
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received:** 31 December 2013; **paper pending published:** 20 January 2014; **accepted:** 29 January 2014; **published online:** 20 February 2014.
- Citation:** Chimileski S, Dolas K, Naor A, Gophna U and Papke RT (2014) Extracellular DNA metabolism in *Haloferax volcanii*. *Front. Microbiol.* 5:57. doi: 10.3389/fmicb.2014.00057
- This article was submitted to Extreme Microbiology, a section of the journal *Frontiers in Microbiology*.
- Copyright © 2014 Chimileski, Dolas, Naor, Gophna and Papke. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes

Aharon Oren\*

Department of Plant and Environmental Sciences, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel

**Edited by:**

Antonio Ventosa, University of Sevilla, Spain

**Reviewed by:**

Mohammad Ali Amoozegar, University of Tehran, Iran

Melanie R. Mormile, Missouri University of Science and Technology, USA

**\*Correspondence:**

Aharon Oren, Department of Plant and Environmental Sciences, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Edmond Safra Campus, Jerusalem 91904, Israel  
e-mail: aharon.oren@mail.huji.ac.il

Extremely halophilic microorganisms that accumulate KCl for osmotic balance (the *Halobacteriaceae*, *Salinibacter*) have a large excess of acidic amino acids in their proteins. This minireview explores the occurrence of acidic proteomes in halophiles of different physiology and phylogenetic affiliation. For fermentative bacteria of the order *Halanaerobiales*, known to accumulate KCl, an acidic proteome was predicted. However, this is not confirmed by genome analysis. The reported excess of acidic amino acids is due to a high content of Gln and Asn, which yield Glu and Asp upon acid hydrolysis. The closely related *Halorhodospira halophila* and *Halorhodospira halochloris* use different strategies to cope with high salt. The first has an acidic proteome and accumulates high KCl concentrations at high salt concentrations; the second does not accumulate KCl and lacks an acidic proteome. Acidic proteomes can be predicted from the genomes of some moderately halophilic aerobes that accumulate organic osmotic solutes (*Halomonas elongata*, *Chromohalobacter salexigens*) and some marine bacteria. Based on the information on cultured species it is possible to understand the pI profiles predicted from metagenomic data from hypersaline environments.

**Keywords:** acidic proteins, osmotic adaptation, halophilic, marine bacteria, anaerobic, *Halanaerobiales*

## INTRODUCTION

In a study of the proteins of *Halobacterium* and *Halococcus*, Reistad (1970) noted an unusual amino acids composition of the cells' bulk protein: a great excess of the acidic amino acids glutamate and aspartate compared to the basic amino acids lysine and arginine. Analysis of the genome of *Halobacterium* NRC-1 (Ng et al., 2000) and related organisms (Oren, 2013a) has confirmed the special properties of the proteins of this group of Archaea. The acidic proteins of the *Halobacteriaceae* typically require high salt concentrations for structural stability and activity, and the presence of such an acidic proteome was considered to be correlated with the accumulation of molar concentrations of KCl to provide osmotic balance to the cells (Lanyi, 1974; Mevarech et al., 2000; Oren, 2013b).

A different strategy of osmotic adaptation in halophilic and halotolerant microorganisms is the accumulation of organic osmotic solutes (glycine betaine, ectoine, glycerol, simple sugars, etc., often termed "compatible solutes"). Such molecules are generally uncharged or zwitterionic, and their presence in high intracellular concentrations does not require far-going adaptation of the proteins. The intracellular solute concentrations can be rapidly adjusted according to the outside salinity, so that microorganisms using this "low-salt-in" strategy can often adapt to life at a wide range of salinities (Galinski, 1995; Oren, 2002a,b, 2011, 2013b).

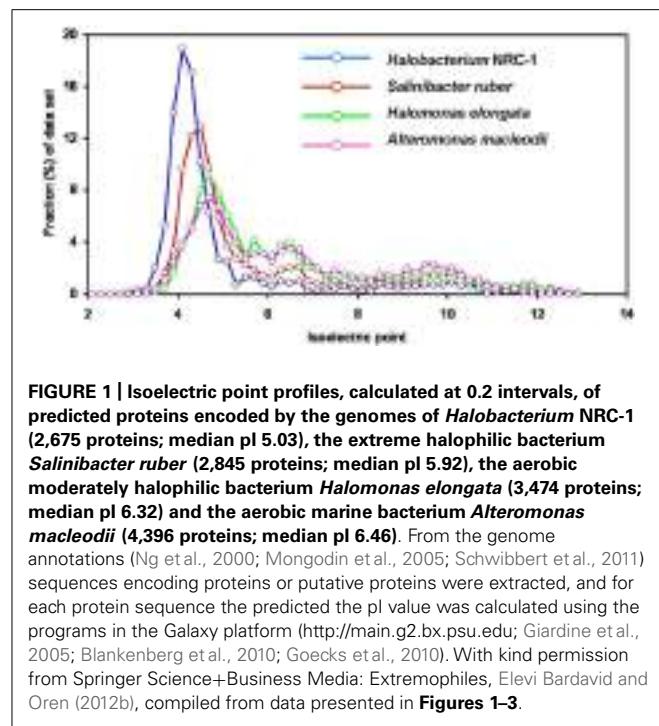
Early attempts to correlate the mode of osmotic adaptation used by halophilic microorganisms with their phylogenetic position yielded a relatively simple picture: the "high-salt-in" strategy was thought to be limited to the *Halobacteriaceae* within the Archaea domain. Among the Bacteria a single group was known that had high intracellular KCl concentrations and allegedly had a highly acidic proteome: the anaerobic fermentative

*Halanaerobiales* (*Firmicutes*). Other groups of Bacteria use organic osmotic solutes, generally in a pattern correlated with their phylogenetic affiliation; organic solutes are also found in halophilic methanogens (Archaea) and in salt-adapted eukaryotic microorganisms (Oren, 2008; Trüper et al., 1991). This represented the state of our knowledge up to turn of the century. Since then the relatively simple picture got complicated by new information and insights. Some of these new data are discussed below.

## *Salinibacter ruber*, ITS MODE OF OSMOTIC ADAPTATION AND THE PROPERTIES OF ITS PROTEINS

*Salinibacter ruber* (*Bacteroidetes*) is a red-pigmented aerobic heterotrophic extremely halophile, first isolated from Spanish saltern crystallizer ponds (Antón et al., 2002), but now known to be distributed worldwide in neutral-pH water bodies at or near salt saturation. This interesting organism shares many key properties with the *Halobacteriaceae* with which it shares its habitat (Oren, 2013c). These include the accumulation of molar concentrations of KCl intracellularly, insignificant concentrations of organic osmotic solutes (Oren et al., 2002), a highly acidic nature of the bulk protein, and a strict salt requirement of key enzymes (Oren and Mana, 2002). High intracellular KCl concentrations were also measured in the phylogenetically related *Salisaeta longa* (Vaisman and Oren, 2009; Vaisman and Oren, unpublished results).

Analysis of the *Salinibacter* genome (Mongodin et al., 2005) confirmed the highly acidic nature of most of its proteins. The median pI value of 5.92 for the proteins encoded by the *S. ruber* genome is slightly higher than that for *Halobacterium* NRC-1 (5.03; Figure 1). *Salinibacter* can be considered as an example of convergent evolution mediated by extensive gene exchange with archaeal halophiles found in the same habitat. The combination



of the “salt-in” strategy and the possession of salt-dependent, highly acidic proteins is thus not necessarily limited to the *Halobacteriaceae* lineage of aerobic halophilic Archaea.

### THE NATURE OF THE PROTEOME OF THE *Halanaerobiales* AND OTHER HALOPHILIC ANAEROBES WITHIN THE BACTERIAL DOMAIN

The order *Halanaerobiales*, families *Halanaerobiaceae* and *Halobacteroidaceae* (Rainey et al., 1995) forms a phylogenetically coherent group of anaerobic bacteria affiliated with the low G+C *Firmicutes* (Kivistö and Karp, 2011; Oren, 2013d). Members of the group have been found in sediments of Great Salt Lake, Utah, the Dead Sea, salterns, oil wells, and in alkaline hypersaline lakes (Lake Magadi, Kenya, Big Soda Lake, Nevada). Most species grow optimally at 10–15% NaCl, and some tolerate salt up to saturation. Most ferment sugars to acetate, ethanol, H<sub>2</sub>, CO<sub>2</sub>, and other products. One of these (*Halothermothrix orenii*, isolated from a warm saline lake in Tunisia), is a true thermophile (up to 68°C; optimum at 60°C) halophile (growth up to 20% NaCl; Cayol et al., 1994). Some genera (*Acetohalobium*, *Natroniella*) have a homoacetogenic metabolism. *Selenihalanaerobacter shriftii* grows by anaerobic respiration with selenite or nitrate as electron acceptor.

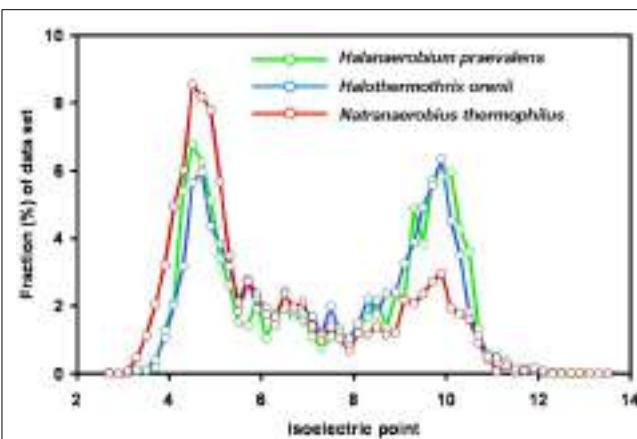
Examination of the cytoplasm of members of the *Halanaerobiales* did not show significant concentrations of organic osmotic solutes (Oren, 1986; Rengpipat et al., 1988), with the possible exception of the finding of glycine betaine in *Orenia salinaria* grown in media containing yeast extract (Mouné et al., 2000). However, high ionic concentrations (K<sup>+</sup>, Na<sup>+</sup>, Cl<sup>-</sup>), sufficient to provide osmotic balance, were measured in *Halanaerobium praevalens* (Oren, 1986; Oren et al., 1997), *Halanaerobium acetethylicum* (Rengpipat et al., 1988), *Halobacteroides halobius*

(Oren, 1986), and *Natroniella acetigena* (Detkova and Pusheva, 2006). Studies performed on glyceraldehyde-3-phosphate dehydrogenase, NAD-linked alcohol dehydrogenase, pyruvate dehydrogenase, and methyl viologen-linked hydrogenase from *H. acetethylicum* (Rengpipat et al., 1988), carbon monoxide dehydrogenase of *N. acetigena* (Detkova and Boltysanskaya, 2006; Detkova and Pusheva, 2006), the fatty acid synthetase complex of *H. praevalens* (Oren and Gurevich, 1993), and hydrogenase and carbon monoxide dehydrogenase of *Acetohalobium arabaticum* (Zavarzin et al., 1994) showed that all these enzymes function well in the presence of molar salt concentrations, and many need high salt for optimal activity. Therefore the “high-salt-in” strategy was assumed to be the mode of osmotic adaptation in this group (Kivistö and Karp, 2011; Oren, 2013d).

Based on these observations the proteome of the members of the *Halanaerobiales* was predicted to have a strongly acidic nature. Indeed, analysis of acid hydrolysates of *Halanaerobium praevalens*, *Halanaerobium saccharolyticum*, *Natroniella acetigena*, *Halobacteroides halobius*, and *Sporohalobacter loretii* suggested that the bulk protein of all these species may have a strongly acidic nature (Oren, 1986; Detkova and Boltysanskaya, 2006). However, it must be remembered that during acid hydrolysis, the neutral asparagine and glutamine are deaminated to form aspartate and glutamate.

The first evidence against a highly acidic proteome in members of the *Halanaerobiales* was published in 1987 when it was shown that the *H. praevalens* ribosomal A-protein is not particularly rich in acidic amino acids (Matheson et al., 1987). Today genome sequences of three members of the group are available: *H. praevalens* GSL<sup>T</sup> (Ivanova et al., 2011), a haloalkaliphilic strain from Soap Lake, WA, USA, described as “*Halanaerobium hydrogeniformans*” (Brown et al., 2011), and *Halothermothrix orenii* H168<sup>T</sup> (Mavromatis et al., 2009). Analysis of these three genomes did not show preferential use of acidic amino acids and no low content of basic amino acids (Elevi Bardavid and Oren, 2012a; Figure 2). It was earlier suggested that the proteins of *H. orenii* may lack a pronounced acidic nature as a special adaptation toward growth at high temperatures (Mijts and Patel, 2001; Mavromatis et al., 2009). The properties of the other two genomes show that also the mesophilic species lack an acidic proteome. The bimodal distribution of the pI values with peaks around 4.6–4.8 and 9.8–10.2 is similar to that of the non-halophiles *Bacteroides fragilis* and *Chlorobaculum tepidum* (Mongodin et al., 2005). The main reason for the apparent discrepancy between the bulk protein analyses, showing a pronounced acid nature, and the analysis of the proteins encoded by the genomes, is the high content of glutamine and asparagine, which lose their amide group during the acid hydrolysis procedure involved in sample preparation for amino acid analysis.

Still there is no reason to doubt the presence of high ionic concentrations within the cytoplasm to balance the osmotic pressure of the medium. Analysis of the three *Halanaerobiales* genomes did not show clear evidence for pathways leading to the synthesis of organic osmotic solutes. A gene for sucrose phosphate synthase was identified in *H. orenii*, which may point to the possibility of sucrose biosynthesis (Mavromatis et al., 2009). Whether indeed



**FIGURE 2 |** Isoelectric point profiles, calculated at 0.2 intervals, of predicted proteins encoded by the genomes of the three fermentative halophilic anaerobes: *Halanaerobium praevalens* GS<sub>L</sub><sup>T</sup> (2,110 proteins; median pI 7.42), the thermophilic *Halothermothrix orenii* H168<sup>T</sup> (2,365 proteins, median pI 7.41) and the alkaliphilic and thermophilic *Natranaerobius thermophilus* JW.NM-WN-LF<sup>T</sup> (2,906 proteins; median pI 6.27). From the genome annotations (Mavromatis et al., 2009; Ivanova et al., 2011; Zhao et al., 2011) sequences encoding proteins or putative proteins were extracted, and pI profiles were calculated as for **Figure 1**. With kind permission from Springer Science+Business Media: Extremophiles, compiled from data presented in Elevi Bardavid and Oren (2012a) – **Figure 1**, and Elevi Bardavid and Oren (2012b) – **Figure 4**.

sucrose is present in the cells at high concentrations, remains to be ascertained. The possibility must be taken into account that the anaerobic halophilic of the *Halanaerobiales* group use a “high-salt-in” strategy of osmotic adaptation but have not adopted the pattern of acidic, low-pI proteins commonly associated with haloadaptation in the aerobic halophiles (*Halobacteriaceae*, *Salinibacter*). A renewed study of the special properties of the *Halanaerobiales* may therefore provide new insights into the strategies available to the prokaryote world to thrive at high salt concentrations (Elevi Bardavid and Oren, 2012a).

The genomes of two anaerobic fermentative halophiles belonging to other phylogenetic lineages were recently sequenced. One is *Flexistipes sinusarabici* MAS 10<sup>T</sup>, isolated from a deep-sea brine pool on the bottom of the Red Sea (Fiala et al., 1990; Lapidus et al., 2011). It was classified as a member of the *Deferrribacteres*, a deep branch within the Bacteria; it grows between 3 and 10% salt and possibly higher. Its mode of osmotic adaptation is yet unknown. Its pI profile (bimodal, with a median pI of 7.47) resembles that of *Halanaerobium*. The second is *Natranaerobius thermophilus* JW.NM-WN-LF<sup>T</sup> an anaerobic halothermoalkaliphile isolated from the Wadi An Natrun lakes in Egypt. It was classified in the newly established order *Natranaerobiales* (*Clostridia*), requires 3.1–4.9 M Na<sup>+</sup>, and is markedly thermophilic (optimum growth at 53°C) and alkaliphilic (optimum pH 9.5; Mesbah et al., 2007). Analysis of its genome (Zhao et al., 2011) yielded a markedly acidic proteome (median pI 6.27; Elevi Bardavid and Oren, 2012b; **Figure 2**). Comparison of the proteins of five anaerobic halophiles of different phylogenetic lineages and with different temperature and pH optima thus shows great variations in the acidic nature of the proteome.

## DISPARATE OSMOTIC ADAPTATION STRATEGIES WITHIN THE GENUS *Halorhodospira*

The genus *Halorhodospira* currently contains four species: the type species *H. halophila*, *H. neutriphila*, *H. halochloris* and *H. abdelmalekii*. With respect to salt requirement and tolerance they are quite similar, and all tolerate NaCl at concentrations up to 25% or higher. They can be divided into two groups, phylogenetically separated on the basis of 16S rRNA gene sequences: *H. halophila* and *H. neutriphila* contain bacteriochlorophyll *a* and carotenoids of the spirilloxanthin group, while *H. halochloris* and *H. abdelmalekii* contain bacteriochlorophyll *b* and rhodopin carotenoids (Imhoff and Söling, 1996; Hirschler-Réa et al., 2003; Oren, 2013e). With respect to their mode of osmotic adaptation they were always considered to be a prime example of organisms that use organic compatible solutes. *H. halophila*, *H. halochloris* and *H. abdelmalekii* were all shown to produce glycine betaine as osmotic solute, with minor amounts of ectoine and trehalose (Trüper et al., 1991). Ectoine, now known to be the most widespread osmotic solute in the prokaryote world, was first discovered in *H. halochloris* (Galinski et al., 1985).

In view of their common phylogeny and documented content of organic osmotic solutes, the finding of an acidic proteome and of high intracellular KCl concentrations in *H. halophila* but not in *H. halochloris* (Deole et al., 2013) came as a big surprise. While the latter does not accumulate KCl, the first contains high KCl when grown at high salt (35%) but not at low salt (5%). The genus *Halorhodospira* thus presents a thus far unique case in which different combinations of KCl concentrations, production of organic osmotic solutes, and presence of acidic vs. non-acidic proteomes are used for osmotic adaptation in phylogenetically closely related species. The authors concluded that “proteome acidity is not driven by stabilizing interactions between K<sup>+</sup> ions and acidic side chains but by the need for maintaining sufficient solvation and hydration of the protein surface at high salinity . . .,” and they proposed that “obligate protein halophilicity is a non-adaptive property resulting from genetic drift in which constructive neutral evolution progressively incorporates weak K<sup>+</sup>-binding sites on an increasingly acidic protein surface” (Deole et al., 2013).

## ACIDIC PROTEOMES IN MODERATELY HALOPHILIC *Gammaproteobacteria*

There is no *a priori* reason to assume that moderately halophilic aerobic heterotrophic bacteria that synthesize and/or accumulate organic compatible solutes should have a high acidic proteome adapted to function in the presence of high intracellular salt concentrations. A first survey of the proteins of the gammaproteobacterium *Chromohalobacter salexigens* DSM 3043<sup>T</sup>, based on 238 out of the 3,319 proteins encoded by its genome, indeed showed that most selected proteins were no more acidic than comparable proteins from non-halophilic counterparts. A notable exception was found for periplasmic proteins exposed to the high medium salinity (Oren et al., 2005).

Analysis of the entire *C. salexigens* genome, together with that of the phylogenetically related moderate halophile *Halomonas elongata* 1H9<sup>T</sup> (Schwibbert et al., 2011) showed large peaks of acidic proteins (maximum at pI 4.4–5.0 and 4.5–5.1, respectively) in the pI profiles of the predicted proteins. The median pI values for

the proteins encoded by these genomes are 6.60 and 6.32, respectively (Elevi Bardavid and Oren, 2012b). These values are still in the low pI range, albeit somewhat higher than those reported for “high-salt-in” organisms such as *Halobacterium* and *Salinibacter* (**Figure 2**). Both organisms synthesize ectoine as compatible solute and accumulate glycine betaine when available in the medium.

Such acidic proteomes are found not only in halophilic and highly halotolerant members of the *Gammaproteobacteria*, but also in typically marine members of the group. Analysis of the pI distribution of the proteins predicted from the genomes of *Alteromonas macleodii* ATCC 27126<sup>T</sup> (Ivars-Martínez et al., 2008), a representative of a genus ubiquitous in the world’s oceans, and of the luminescent *Aliivibrio fischeri* strain MJ11 (Mandel et al., 2009) showed a pronounced peak in the acidic range (maximum at pI values of 4.6–4.8), with median pI values of 6.46 and 6.52, respectively (Elevi Bardavid and Oren, 2012b).

## ACIDIC METAPROTEOMES IN HYPERSALINE ENVIRONMENTS

Metagenomic data from saline and hypersaline environments can be subjected to analyses similar to those shown above for microbial isolates. As shown by Rhodes et al. (2010), there is a general trend of increased average protein acidity (as expressed by the ratio of acidic to basic amino acids) with increased salinity. The highest salinity environments (the Dead Sea, saltern crystallizer ponds) have the greatest excess of acidic amino acids in the proteins encoded by the recovered DNA ([Glu + Asp]/[Lys + Arg + His] = 1.42–1.26). This could be expected as high-salt-in strategists (species of *Halobacteriaceae*, *Salinibacter*) with highly acidic proteomes dominate their biota. Metagenomes from different samples from the marine environment gave values in the range 0.86–0.95, and the benthic microbial mats in the 9% salt lagoons of Guerrero Negro, Mexico (Kunin et al., 2008), yielded an intermediate value of 1.01 on the average.

The finding that the pI distribution of the proteins encoded by the metagenome of the Guerrero Negro microbial mats showed an acid-shifted proteome (major peak at pI 4.5–4.9, median pI 6.8) as compared to non-halophilic or marine environments was at first puzzling, as at that salinity microorganisms are expected to use organic osmotic solutes, without the need to adapt their proteins to high salt. Kunin et al. (2008) concluded that the enhanced acidic nature of the proteins is linked to the increased salinity, explained as an example of species-independent molecular convergence in a microbial community. However, as documented above, many moderate halophiles, also those that do not accumulate organic osmotic solutes, show a broad peak at pI values 4.5–4.9 (Elevi Bardavid and Oren, 2012b). In comparison to the proteins encoded by the genomes of moderately halophilic aerobes and even certain marine bacteria (**Figure 2**), the metaproteome encoded by the metagenome of the different layers within the 9% salt Guerrero Negro microbial mats is not conspicuously acid-shifted.

## REFERENCES

- Antón, J., Oren, A., Benlloch, S., Rodríguez-Valera, F., Amann, R., and Rosselló-Móra, R. (2002). *Salinibacter ruber* gen. nov., sp. nov., a novel extreme halophilic member of the Bacteria from saltern crystallizer ponds. *Int. J. Syst.*

The finding that marine metagenomes do not encode for metaproteomes enriched in acidic proteins (Rhodes et al., 2010) needs to be evaluated in view of the above-mentioned observation that some typically marine *Gammaproteobacteria* (*Alteromonas* and *Vibrio* spp.) are rich in low-pI proteins. A possible explanation is that other, possibly very abundant marine bacteria show the opposite trend. The small genome of “*Pelagibacter ubique*” (*Alphaproteobacteria*; the “SAR-11” phylotype; Giovannoni et al., 2005) encodes for 1,393 proteins with an overall excess of 2% (Lys + Arg) – (Glu + Asp). For comparison, *Halobacterium*, *Salinibacter* and *H. elongata* and *Halomonas* all show an excess of acidic amino acids (7.5, 4.1, and 2.8 mol %, respectively; Elevi Bardavid and Oren, 2012b). Most “*Pelagibacter*” proteins have pI values between 9.4 and 10.8, with a median value of 8.42.

## FINAL COMMENTS

The genomic and metagenomic data discussed above show that dominance of acidic proteins in halophilic microorganisms is by no means restricted to the *Halobacteriaceae* and to *Salinibacter* which resembles the *Halobacteriaceae* in many properties. Somewhat less acidic proteomes are found in many moderately halophilic and even in some marine bacteria, organisms that exclude salt from their cytoplasm to a large extent (Elevi Bardavid and Oren, 2012b). On the other hand, the analysis of the genomes of different anaerobic halophiles (*Halanaerobiales* and others) unexpectedly failed to show a highly acidic proteome (Elevi Bardavid and Oren, 2012a). The case of the two *Halorhodospira* species demonstrates that phylogenetically very closely related organisms may use completely different strategies for osmotic adaptation, and accordingly have highly different amino acid signatures of their proteins. The more or less coherent picture of a clear correlation between phylogenetic affiliation and modes of salt adaptation that was apparent in the past (Trüper et al., 1991; Oren, 2008) needs therefore drastic revision. We must rethink our concepts about the correlation between acidic proteomes, salt requirement and tolerance, accumulation of KCl, use of organic osmotic solutes, and microbial phylogeny and taxonomy.

A recently published analysis of the structure of primitive proteins that may have been formed from “prebiotic” amino acids expected to have been available at the time life originated on Earth showed that the predicted foldable proteins have a substantial acidification of pI and possessed halophilic properties (Longo et al., 2013). The question whether the environment for primordial life may have been hypersaline has been addressed earlier (Dundas, 1998). Therefore the issues discussed above may even have direct implications for our ideas on the origin of life and the properties of the earliest organisms that inhabited our planet.

## ACKNOWLEDGMENTS

I thank Rahel Elevi Bardavid and Omri Finkel for their contributions to the data evaluation. This study was supported by grants no. 1103/10 and 343/13 from the Israel Science Foundation.

- Evol. Microbiol.* 52, 485–491. doi: 10.1128/AEM.66.7.3052–3057.2000  
Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., et al. (2010). Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* Chapter 19, Unit 19.10.1–

21. doi: 10.1002/0471142727.mb19  
10s89
- Brown, S. D., Begemann, M. B., Mormile, M. R., Wall, J. D., Han, C. S., Goodwin, L. A., et al. (2011). Complete genome sequence of the haloalkaliphilic, hydrogen-producing bacterium *Halanaerobium hydrogeniformans*. *J. Bacteriol.* 193, 3682–3683. doi: 10.1128/JB.05209-11
- Cayol, J.-L., Ollivier, B., Patel, B. K. C., Prensier, G., Guezenec, J., and Garcia, J.-L. (1994). Isolation and characterization of *Halothermothrix orenii* gen. nov., sp. nov., a halophilic, thermophilic, fermentative, strictly anaerobic bacterium. *Int. J. Syst. Bacteriol.* 44, 534–540. doi: 10.1099/00207713-44-3-534
- Detkova, E. N., and Boltyanskaya, Y. V. (2006). Relationships between the osmoadaptation strategy, amino acid composition of bulk protein, and properties of certain enzymes of haloalkaliphilic bacteria. *Mikrobiologiya* 75, 259–265.
- Detkova, E. N., and Pusheva, M. A. (2006). Energy metabolism in halophilic and alkaliphilic acetogenic bacteria. *Mikrobiologiya* 75, 5–17. doi.org/10.1134/S0026261706010012
- Deole, R., Challacombe, J., Raiford, D. W., and Hoff, W. D. (2013). An extremely halophilic proteobacterium combines a highly acidic proteome with a low cytoplasmic potassium content. *J. Biol. Chem.* 288, 581–588. doi: 10.1074/jbc.M112.420505
- Dundas, I. (1998). Was the environment for primordial life hypersaline? *Extremophiles* 2, 375–377. doi: 10.1007/s007920050081
- Elevi Bardavid, R., and Oren, A. (2012a). The amino acid composition of proteins from anaerobic halophilic bacteria of the order *Halanaerobiales*. *Extremophiles* 16, 567–572. doi: 10.1007/s00792-012-0455-y.
- Elevi Bardavid, R., and Oren, A. (2012b). Acid-shifted isoelectric point profiles of the proteins in a hypersaline microbial mat – an adaptation to life at high salt concentrations? *Extremophiles* 16, 787–792. doi: 10.1007/s00792-012-0476-6
- Fiala, G., Woese, C. R., Langworthy, T. A., and Stetter, K. O. (1990). *Flexistipes sinusarabici*, a novel genus and species of eubacteria occurring in the Atlantis II Deep brines of the Red Sea. *Arch. Microbiol.* 154, 120–126. doi: 10.1007/BF00423320
- Galinski, E. A. (1995). Osmoadaptation in bacteria. *Adv. Microb. Physiol.* 37, 272–328.
- Galinski, E. A., Pfeiffer, H.-P., and Trüper, H. G. (1985). 1,4,5,6-tetrahydro-2-methyl-4-pyrimidine-carboxylic acid. A novel cyclic amino acid from halophilic phototrophic bacteria of the genus *Ectothiorhodospira*. *Eur. J. Biochem.* 149, 135–139. doi: 10.1111/j.1432-1033.1985.tb08903.x
- Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., et al. (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* 15, 1451–1455. doi: 10.1101/gr.408650
- Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., et al. (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science* 309, 1242–1245. doi: 10.1126/science.1114057
- Goecks, J., Nekrutenko, A., Taylor, J., and The Galaxy Team. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 11, R86. doi: 10.1186/gb-2010-11-8-r86
- Hirschler-Réa, A., Matheron, R., Riffaud, C., Mouné, S., Eatock, C., Herbert, R. A., et al. (2003). Isolation and characterization of spirilloid purple phototrophic bacteria forming red layers in microbial mats of Mediterranean salterns: description of *Halorhodospira neutriphila* sp. nov. and emendation of the genus *Halorhodospira*. *Int. J. Syst. Evol. Microbiol.* 53, 153–163. doi: 10.1099/ijss.0.02226-0
- Imhoff, J. F., and Söling, J. (1996). The phylogenetic relationship among *Ectothiorhodospiraceae*: a reevaluation of their taxonomy on the basis of 16S rDNA analyses. *Arch. Microbiol.* 165, 106–113. doi: 10.1007/s002030050304
- Ivanova, N., Sikorski, J., Chertkov, O., Nolan, M., Lucas, S., Hammon, N., et al. (2011). Complete genome sequence of the extremely halophilic *Halanaerobium praevalens* type strain (GSLT). *Stand. Genomic Sci.* 4, 312–321. doi: 10.4056/sigs.1824509
- Ivars-Martinez, E., D'Auria, G., Rodríguez-Valera, F., Sánchez-Porro, C., Ventosa, A., Ioint, I., et al. (2008). Biogeography of the ubiquitous marine bacterium *Alteromonas macleodii* determined by multilocus sequence analysis. *Mol. Ecol.* 17, 4092–4106. doi: 10.1111/j.1365-294X.2008.03883.x
- Kivistö, A. T., and Karp, M. (2011). Halophilic anaerobic fermentative bacteria. *J. Biotechnol.* 152, 114–124. doi: 10.1016/j.jbiotec.2010.08.014
- Kunin, V., Raes, J., Harris, J. K., Spear, J. R., Walker, J. J., Ivanova, N., et al. (2008). Millimeter-scale genetic gradients and community-level molecular convergence in a hypersaline microbial mat. *Mol. Syst. Biol.* 4, 198. doi: 10.1038/msb.2008.35
- Lanyi, J. K. (1974). Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriol. Rev.* 38, 272–290.
- Lapidus, A., Chertkov, O., Nolan, M., Lucas, S., Hammon, N., Deshpande, S., et al. (2011). Genome sequence of the moderately thermophilic halophile *Flexistipes sinusarabici* strain (MAS10T). *Stand. Genomic Sci.* 5, 1. doi: 10.4056/sigs.2235024
- Longo, L. M., Lee, J., and Blaber, M. (2013). Simplified protein design biased for prebiotic acids yields a foldable, halophilic protein. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2135–2139. doi: 10.1073/pnas.1219530110
- Mandel, M. J., Wollenberg, M. S., Stabb, E. V., Visick, K. L., and Ruby, E. G. (2009). A single regulatory gene is sufficient to alter bacterial host range. *Nature* 458, 215–218. doi: 10.1038/nature07660
- Matheson, A. T., Louie, K. A., Tak, B. D., and Zuker, M. (1987). The primary structure of the ribosomal A-protein (L12) from the halophilic eubacterium *Halanaerobium praevalens*. *Biochimie* 69, 1013–1020. doi: 10.1016/0300-9084(87)90001-0
- Mavromatis, K., Ivanova, N., Anderson, I., Lykidis, A., Hooper, S. D., Sun, H., et al. (2009). Genome analysis of the anaerobic thermohalophilic bacterium *Halothermothrix orenii*. *PLoS ONE* 4:e4192. doi: 10.1371/journal.pone.0004192
- Mesbah, N. M., Hedrick, D. B., Peacock, A. D., Rohde, M., and Wiegel, J. (2007). *Natrananrobius thermophilus* gen. nov., sp. nov., a halophilic, alkali-lithophilic bacterium from soda lakes of the Wadi An Natrun, Egypt, and proposal of *Natrananobiaceae* fam. nov. and *Natrananrobiales* ord. nov. *Int. J. Syst. Evol. Microbiol.* 57, 2507–2512. doi: 10.1099/ijss.0.65068-0
- Mevarech, M., Frolow, F., and Gloss, L. M. (2000). Halophilic enzymes: proteins with a grain of salt. *Bioophys. Chem.* 86, 155–164. doi: 10.1016/S0301-4622(00)00126-5
- Mijts, B. M., and Patel, B. K. C. (2001). Random sequence analysis of genomic DNA of an anaerobic, thermophilic, halophilic bacterium, *Halothermothrix orenii*. *Extremophiles* 5, 61–69. doi: 10.1007/s007920000174
- Mongodin, E. F., Nelson, K. E., Daugherty, S., DeBoy, R. T., Wister, J., Khouri, H., et al. (2005). The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18147–18152. doi: 10.1073/pnas.0509073102
- Mouné, S., Eatock, C., Matheron, R., Willison, J. C., Hirschler, A., Herbert, R., et al. (2000). *Orenia salinaria* sp. nov., a fermentative bacterium isolated from anaerobic sediments of Mediterranean salterns. *Int. J. Syst. Evol. Microbiol.* 50, 721–729. doi: 10.1099/00207713-50-2-721
- Ng, W. V., Kennedy, S. P., Mahairas, G. G., Berquist, B., Pan, M., Shukla, H. D., et al. (2000). Genome sequence of *Halobacterium* species NRC-1. *Proc. Natl. Acad. Sci. U.S.A.* 97, 12176–12181. doi: 10.1073/pnas.190337797
- Oren, A. (1986). Intracellular salt concentrations of the anaerobic halophilic eubacteria *Halanaerobium praevalens* and *Halobacteroides halobius*. *Can. J. Microbiol.* 32, 4–9. doi: 10.1139/m86-002
- Oren, A. (2002a). *Halophilic Microorganisms and their Environments*. Dordrecht: Kluwer Scientific Publishers.
- Oren, A. (2002b). Diversity of halophilic microorganisms: environments, phylogeny, physiology, and applications. *J. Ind. Microbiol. Biotechnol.* 28, 56–63.
- Oren, A. (2008). Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Syst.* 4, 2. doi: 10.1186/1746-1448-4-2
- Oren, A. (2011). “Diversity of halophiles,” in *Extremophiles Handbook*, ed. K. Horikoshi (Tokyo: Springer), 309–325.
- Oren, A. (2013a). “The family *Halobacteriaceae*,” in *The Prokaryotes. A Handbook on the Biology of Bacteria: Ecophysiology and Biochemistry*, 4th Edn, eds E. Rosenberg, E. F. DeLong, F. Thompson, S. Lory, and E. Stackebrandt (New York, NY: Springer), in press.
- Oren, A. (2013b). “Life at high salt concentrations,” in *The Prokaryotes. A Handbook on the Biology of Bacteria: Ecophysiology and Biochemistry*, 4th Edn, eds E. Rosenberg, E. F. DeLong, F. Thompson, S. Lory, and E. Stackebrandt (New York, NY: Springer), in press.
- Oren, A. (2013c). *Salinibacter*, an extremely halophilic bacterium with archaeal properties. *FEMS Microbiol. Lett.* 342, 1–9.
- Oren, A. (2013d). “The order *Halanaerobiales*, families *Halanaerobiaceae* and *Halobacteroidaceae*,” in *The*

- Prokaryotes. A Handbook on the Biology of Bacteria: Ecophysiology and Biochemistry*, 4th Edn, eds E. Rosenberg, E. F. DeLong, F. Thompson, S. Lory, and E. Stackebrandt (New York, NY: Springer), in press.
- Oren, A. (2013e). "Family *Ectothiorhodospiraceae*," in *The Prokaryotes. A Handbook on the Biology of Bacteria: Ecophysiology and Biochemistry*, 4th Edn, eds E. Rosenberg, E. F. DeLong, F. Thompson, S. Lory, and E. Stackebrandt (New York, NY: Springer), in press.
- Oren, A., and Gurevich, P. (1993). The fatty acid synthetase complex of *Haloanaerobium praevalens* is not inhibited by salt. *FEMS Microbiol. Lett.* 108, 287–290. doi: 10.1111/j.1574-6968.1993.tb06117.x
- Oren, A., Heldal, M., and Norland, S. (1997). X-ray microanalysis of intracellular ions in the anaerobic halophilic eubacterium *Haloanaerobium praevalens*. *Can. J. Microbiol.* 43, 588–592. doi: 10.1139/m97-083
- Oren, A., Heldal, M., Norland, S., and Galinski, E. A. (2002). Intracellular ion and organic solute concentrations of the extremely halophilic bacterium *Salinibacter ruber*. *Extremophiles* 6, 491–498. doi: 10.1007/s00792-002-0286-3
- Oren, A., Larimer, F., Richardson, P., Lapidus, A., and Csonka, L. N. (2005). How to be moderately halophilic with a broad salt tolerance: clues from the genome of *Chromohalobacter salexigens*. *Extremophiles* 9, 275–279. doi: 10.1007/s00792-005-0442-7
- Oren, A., and Mana, L. (2002). Amino acid composition of bulk protein and salt relationships of selected enzymes of *Salinibacter ruber*, an extremely halophilic bacterium. *Extremophiles* 6, 217–223. doi: 10.1007/s007920100241
- Rainey, F. A., Zhilina, T. N., Boulygina, E. S., Stackebrandt, E., Tourova, T. P., and Zavarzin, G. A. (1995). The taxonomic status of the fermentative halophilic anaerobic bacteria: description of Halobacterales ord. nov., *Halobacteroidaceae* fam. nov., *Orenia* gen. nov. and further taxonomic rearrangements at the genus and species level. *Anaerobe* 1, 185–199. doi: 10.1006/anae.1995.1018
- Reistad, R. (1970). On the composition and nature of the bulk protein of extremely halophilic bacteria. *Arch. Mikrobiol.* 71, 353–360. doi: 10.1007/BF00417131
- Rengpipat, S., Lowe, S. E., and Zeikus, J. G. (1988). Effect of extreme salt concentrations on the physiology and biochemistry of *Halobacteroides acetotrophicus*. *J. Bacteriol.* 170, 3065–3071.
- Rhodes, M. E., Fitz-Gibbon, S., Oren, A., and House, C. H. (2010). Amino acid signatures of salinity on an environmental scale with a focus on the Dead Sea. *Environ. Microbiol.* 12, 2613–2623. doi: 10.1111/j.1462-2920.2010.02232.x
- Schwiibert, K., Marin-Sanguino, A., Bagyan, I., Heidrich, G., Lentzen, G., Seitz, H., et al. (2011). A blueprint of ectoine metabolism from the genome of the industrial producer *Halomonas* DSM 2581. *Environ. Microbiol.* 13, 1973–1994. doi: 10.1111/j.1462-2920.2010.02336.x
- Trüper, H. G., Severin, J., Wolforth, A., Müller, E., and Galinski, E. A. (1991). "Halophily, taxonomy, phylogeny and nomenclature," in *General and Applied Aspects of Halophilism*, ed. F. Rodriguez-Valera (New York, NY: Plenum Press), 3–7.
- Vaisman, N., and Oren, A. (2009). *Salisaeta longa* gen. nov., sp. nov., a red halophilic member of the *Bacteroidetes*. *Int. J. Syst. Evol. Microbiol.* 59, 2571–2574. doi: 10.1099/ijs.0.010892-0
- Zavarzin, G. A., Zhilina, T. A., and Pusheva, M. A. (1994). "Halophilic acetogenic bacteria," in *Acetogenesis*, ed. H. L. Drake (New York: Chapman & Hall), 432–444.
- Zhao, B., Mesbah, N. M., Dalin, E., Goodwin, L., Nolan, M., Pitluck, S., et al. (2011). Complete genome sequence of the anaerobic, halophilic alkalithermophilic *Natranaerobius thermophilus*. *J. Bacteriol.* 193, 4023–4024. doi: 10.1128/JB.05157-11
- Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 28 June 2013; paper pending published: 05 October 2013; accepted: 06 October 2013; published online: 05 November 2013.*
- Citation: Oren A (2013) Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. *Front. Microbiol.* 4:315. doi: 10.3389/fmicb.2013.00315*
- This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.*
- Copyright © 2013 Oren. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# Adaptation to high salt concentrations in halotolerant/halophilic fungi: a molecular perspective

Ana Plemenitaš<sup>1\*</sup>, Metka Lenassi<sup>1</sup>, Tilen Konte<sup>1</sup>, Anja Kejžar<sup>1</sup>, Janja Zajc<sup>2</sup>, Cene Gostinčar<sup>3</sup> and Nina Gunde-Cimerman<sup>2,4</sup>

<sup>1</sup> Faculty of Medicine, Institute of Biochemistry, University of Ljubljana, Ljubljana, Slovenia

<sup>2</sup> Biology Department, Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia

<sup>3</sup> Department of Biotechnology and Systems Biology, National Institute of Biology, Ljubljana, Slovenia

<sup>4</sup> Centre of Excellence for Integrated Approaches in Chemistry and Biology of Proteins (CIPKeBiP), Ljubljana, Slovenia

**Edited by:**

Aharon Oren, The Hebrew University of Jerusalem, Israel

**Reviewed by:**

Aharon Oren, The Hebrew University of Jerusalem, Israel  
Laszlo N. Csoka, Purdue University, USA

**\*Correspondence:**

Ana Plemenitaš, Faculty of Medicine, Institute of Biochemistry, University of Ljubljana, Vrazov trg 2, Ljubljana 1000, Slovenia  
e-mail: ana.plemenitas@mf.uni-lj.si

Molecular studies of salt tolerance of eukaryotic microorganisms have until recently been limited to the baker's yeast *Saccharomyces cerevisiae* and a few other moderately halotolerant yeast. Discovery of the extremely halotolerant and adaptable fungus *Hortaea werneckii* and the obligate halophile *Wallemia ichthyophaga* introduced two new model organisms into studies on the mechanisms of salt tolerance in eukaryotes. *H. werneckii* is unique in its adaptability to fluctuations in salt concentrations, as it can grow without NaCl as well as in the presence of up to 5 M NaCl. On the other hand, *W. ichthyophaga* requires at least 1.5 M NaCl for growth, but also grows in up to 5 M NaCl. Our studies have revealed the novel and intricate molecular mechanisms used by these fungi to combat high salt concentrations, which differ in many aspects between the extremely halotolerant *H. werneckii* and the halophilic *W. ichthyophaga*. Specifically, the high osmolarity glycerol signaling pathway that is important for sensing and responding to increased salt concentrations is here compared between *H. werneckii* and *W. ichthyophaga*. In both of these fungi, the key signaling components are conserved, but there are structural and regulation differences between these pathways in *H. werneckii* and *W. ichthyophaga*. We also address differences that have been revealed from analysis of their newly sequenced genomes. The most striking characteristics associated with *H. werneckii* are the large genetic redundancy, the expansion of genes encoding metal cation transporters, and a relatively recent whole genome duplication. In contrast, the genome of *W. ichthyophaga* is very compact, as only 4884 protein-coding genes are predicted, which cover almost three quarters of the sequence. Importantly, there has been a significant increase in their hydrophobins, cell-wall proteins that have multiple cellular functions.

**Keywords:** halophilic/halotolerant fungi, *Hortaea werneckii* genome, *Wallemia ichthyophaga* genome, HOG signaling pathway, ion homeostasis

## INTRODUCING *HORTAEA WERNECKII* AND *WALLEMIA ICHTHYOPHAGA*

Studies of fungal populations in hypersaline environments have revealed the high diversity of fungal species (Gunde-Cimerman et al., 2000), most of which do not require salt for growth, and have their growth optimum in the absence of salt. The dominant fungal group in the hypersaline waters of salterns are the melanized polymorphic black yeast, the most abundant and adapted species of which is *Hortaea werneckii*. *H. werneckii* is naturally adapted to fluctuating salt concentrations in its environment, and it can grow without salt and in up to saturated NaCl. Its optimum for growth is between 0.8 M and 1.7 M NaCl. Another successful survivor in these extremely salty environments is the basidiomycetous fungus *Wallemia ichthyophaga* (Zalar et al., 2005b), which does not grow without salt, and is therefore obligately halophilic.

Due to their different ecology and halotolerances, these two fungi represent highly relevant organisms for the study of eukaryotic adaptation to life at high salt. Studies of haloadaptation

mechanisms of *H. werneckii* started some 15 years ago (for reviews, see Petrović et al., 2002; Gunde-Cimerman and Plemenitaš, 2006; Plemenitaš et al., 2008; Gostinčar et al., 2011), while with *W. ichthyophaga*, these studies began later, and are thus less advanced.

## THE EXTREMELY HALOTOLERANT *HORTAEA WERNECKII*

*H. werneckii* (Horta) Nishim and Miyaji (Capnodiales, Dothideomycetes) is a melanized yeast-like ascomycete that is known as the causative agent of *tinea nigra*, a superficial mycotic infection of the human palm (de Hoog and Gerrits van den Ende, 1992). *H. werneckii* has been isolated from diverse environments with low water activity ( $a_w$ ), including salty food (Mok et al., 1981), seawater (Iwatsu and Udagawa, 1988), beach soil (de Hoog and Guého, 1998), rocks (Stanley et al., 1982), wood immersed in hypersaline waters (Wollenzien et al., 1995; Zalar et al., 2005a) and microbial mats (Cantrell et al., 2006). However, it appears that its primary habitat is hypersaline water in the evaporate

ponds of solar eutrophic salterns (Gunde-Cimerman et al., 2000).

While *H. werneckii* has been extensively described in our earlier review papers (Petrovič et al., 2002; Gunde-Cimerman and Plemenitaš, 2006; Plemenitaš et al., 2008; Gostinčar et al., 2011), *W. ichthyophaga* has not been reviewed to date, and thus it is presented below in more detail.

### THE HALOPHILIC *WALLEMIA ICHTHYOPHAGA*

*Wallemia* Johan-Olsen (Wallemiales, Wallemiomycetes) is a genus of cosmopolitan xerophilic fungi that can be found in a wide variety of environments that are characterized by low  $a_w$  (Samson et al., 2004; Zalar et al., 2005b). Its phylogenetic position was unclear until recently, and previously it has been placed in various positions in the Basidiomycota phylogenetic tree, from the root of basidiomycetes (Zalar et al., 2005b), to *incertae sedis* (Hibbett et al., 2007), to being a sister group of the Agaricomycotina and Ustilaginomycotina (Matheny et al., 2006). Genome sequencing has shown that it is indeed a sister group of the Agaricomycotina (Padamsee et al., 2012; Zajc et al., 2013). The Wallemiomycetes split from the Agaricomycotina ancestors an estimated 250 million years ago (Zajc et al., 2013). Initially the genus contained only one species, but it was later segregated into three species based on differences in conidial size, xerotolerance, and sequence data: *W. ichthyophaga*, *Wallemia sebi* and *Wallemia muriae* (Zalar et al., 2005b). To date, only around 20 strains of *W. ichthyophaga* have been isolated from hypersaline waters of solar salterns, bitterns (i.e., magnesium-rich residual solutions in salt production from sea water) and salted meat (Zalar et al., 2005b). In addition to phylogenetic differences, *W. ichthyophaga* is also distinguished from the other two representatives of this genus by its characteristic morphology and halophilic physiology (Zalar et al., 2005b; Kralj Kunčič et al., 2010).

Although xerotolerance is rare in the Basidiomycota, all three *Wallemia* spp. are among the most xerophilic fungal taxa known to date (Zalar et al., 2005b). However, while *W. sebi* and *W. muriae* strongly prefer high concentrations of non-ionic solutes over those of NaCl (Kralj Kunčič et al., 2013), the opposite is true for *W. ichthyophaga* (Zalar et al., 2005b). For growth, *W. ichthyophaga* requires at least 1.5 M NaCl, or some other osmolyte at an equivalent  $a_w$ . Such a narrow ecological amplitude of salt concentrations is common for specialized archaeal halophiles, but it is exceptional in the fungal kingdom. Hence, *W. ichthyophaga* is a rare fungal example of an obligate extremophilic specialist (Gostinčar et al., 2010), and it is considered to be the most halophilic fungus known to date. Although it even thrives in saturated NaCl solution, its *in vitro* growth optimum is between 2.6 M and 3.5 M NaCl, which is the highest described among fungi (Zajc et al., 2014). It also tolerates high concentrations of salts other than NaCl; e.g., MgCl<sub>2</sub> (our unpublished data).

### SENSING HYPEROSMOLARITY IN *H. WERNECKII* AND *W. ICHTHYOPHAGA*

Exposure to high salinity includes two different environmental stimuli for the cell: osmotic stress, and ionic stress. In general, hyperosmotic stress in non-adapted organisms causes immediate water efflux from the cell, which reduces the turgor pressure

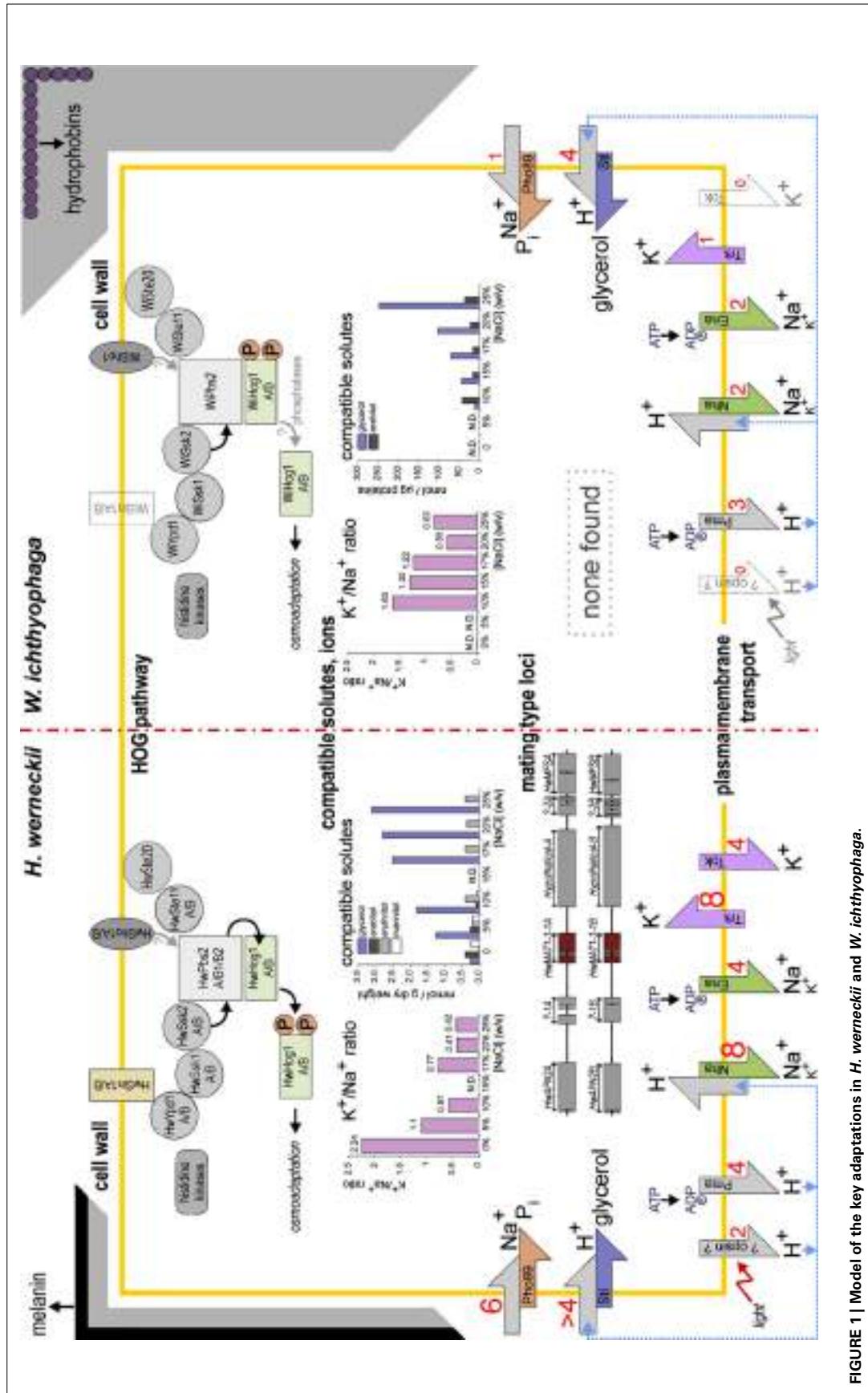
and triggers cytosol dehydration, thereby increasing the concentrations of the solutes in the cytoplasm (Petelenz-Kurdziel et al., 2011). In particular, under high ionic stress conditions, ions (e.g., Na<sup>+</sup>) enter the cell, which leads to increased intracellular ion concentrations, which subsequently damage the membranes as well as the cytosolic systems. The main survival strategies to counteract changes in turgor pressure for fungi that are adapted to life at low  $a_w$  are an accumulation of compatible solutes that do not interfere with vital cellular protein functions, and maintenance of intracellular concentrations of Na<sup>+</sup> below toxic levels (Blomberg and Adler, 1992). Both *H. werneckii* and *W. ichthyophaga* use the strategy of compatible organic solutes to maintain low intracellular Na<sup>+</sup> concentrations, with glycerol being the main solute used.

The main signaling pathway in fungi that is responsible for cellular stress responses is the high osmolarity glycerol (HOG) pathway, which has been extensively studied in the context of osmotic stress in *S. cerevisiae*. The production and homeostasis of the compatible solute glycerol is one of the main targets under the control of this signaling pathway (Hohmann et al., 2007). The core of the pathway is represented by the mitogen-activated protein kinase (MAPK) signaling module, which is known for its high evolutionary conservation and its activation by sequential phosphorylations (Widmann et al., 1999). This upstream part of the HOG pathway consists of two branches, which are functionally redundant but structurally distinct. They are known as the SHO1 and SLN1 branches and they converge at the MAPK kinase (MAPKK) Pbs2. Upon hyperosmotic shock, when the cell loses some of its water, the MAPK Hog1 is phosphorylated and activated by the upstream MAPKK Pbs2. The main effect of this HOG pathway activation is glycerol production, which restores the cellular osmotic balance. When turgor is re-established, Hog1 is dephosphorylated by phosphatases (Saito and Posas, 2012).

### THE HOG SIGNAL TRANSDUCTION IN *H. WERNECKII* AND *W. ICHTHYOPHAGA*

In *H. werneckii* and *W. ichthyophaga*, several components of the HOG pathway have been identified and characterized (Lenassi and Plemenitas, 2007; Fettich et al., 2011; Konte and Plemenitaš, 2013). Since their sequenced genomes became available, the presence of some novel HOG components has been confirmed through homology searches. Altogether, there are many similarities between the HOG pathways in *H. werneckii*, *W. ichthyophaga*, and *S. cerevisiae*, although there are also some important differences that might explain the different halotolerant/halophilic characters of *H. werneckii* and *W. ichthyophaga*.

The presence of homologs of Sho1, Ste20, and Ste11 has been confirmed for the genomes of *H. werneckii* and *W. ichthyophaga*, with two copies of each component in *H. werneckii* (Figure 1; Lenassi et al., 2013; Zajc et al., 2013; our unpublished data). Two isoforms of the *H. werneckii* putative osmosensor protein HwSho1A and HwSho1B fully complement the function of the homologous *S. cerevisiae* Sho1 protein and they can activate the HOG pathway under osmotic stress in *S. cerevisiae*. Structurally, when compared to other fungal Sho1 homologs, they contain a conserved SH3 domain and a divergent Ste11-binding motif (Fettich et al., 2011). On the other hand, the SH3 domain of



**FIGURE 1 |** Model of the key adaptations in *H. werneckii* and *W. ichthyocephala*.

*W. ichthyophaga* (Wi)Sho1 is functional when it is attached to the N-terminal part of *S. cerevisiae* Sho1, although the whole sequence of WiSho1 does not appear to function correctly in *S. cerevisiae* (our unpublished data). Regardless of the complementation of the HwSho1 protein and the WiSho1 SH3 domain, data from recent preliminary investigations addressing the role of the SHO1 branch pathway in osmo-adaptation do not support the involvement of this branch in the signal transfer downstream to heterologously expressed HwPbs2 and WiPbs2 in *S. cerevisiae*. WiSte11 also failed to complement ScSte11 in *S. cerevisiae* ste11Δssk2ssk22Δ cells, further supporting this hypothesis (our unpublished data).

The other branch of the HOG pathway, which was named as SLN1 after the transmembrane Sln1 hybrid histidine kinase, transmits its signals via a Sln1–Ypd1–Ssk1 phosphorelay. Sln1 kinase is inactive under hyperosmolar conditions, where Ssk1 is dephosphorylated and therefore binds to the autoinhibitory region of Ssk2 and Ssk22, which triggers their autophosphorylation (Saito and Posas, 2012). The *H. werneckii* histidine kinases HwHhk7A and HwHhk7B have been identified and characterized in more detail (Figure 1; Lenassi and Plemenitas, 2007). HwHhk7A and HwHhk7B lack the transmembrane domain, but otherwise they have a typical eukaryotic hybrid histidine-kinase-domain composition. Their transcription in *H. werneckii* depends on the extracellular salt concentration, and when they are expressed in *S. cerevisiae*, they increase its osmotolerance (Lenassi and Plemenitas, 2007). Both the *H. werneckii* and *W. ichthyophaga* genomes also contain homologs of the group III histidine kinases. This group of cytosolic histidine kinases can act as osmosensors through their HAMP domain repeats (Meena et al., 2010). On the other hand, the membrane-spanning Sln1-like histidine kinase is present in the genome of *H. werneckii*, again in two copies, but there is no evidence for it in the genome of *W. ichthyophaga* (Figure 1), which suggests that the group III histidine kinases are probably involved in osmosensing in this halophilic fungus. Of the other proteins involved in the SLN1 branch, two forms of each of the *S. cerevisiae* homologs have been found in the genome of *H. werneckii*, HwYpd1A/B, HwSsk1A/B, and two homologs of the MAPKK kinase Ssk2, HwSsk2A/B (Table 2) (Lenassi et al., 2013; our unpublished data). Also *W. ichthyophaga* has the proteins WiYpd1, WiSsk1, and WiSsk2 (Figure 1, Table 2; Zajc et al., 2013; our unpublished data).

The signals from both branches of the HOG pathway in *S. cerevisiae* converge at the MAPKK Pbs2 scaffold, which transmits the signals further to Hog1 (Saito and Posas, 2012). Two gene copies of the MAPKK HwPbs2 have been identified in *H. werneckii* and one in *W. ichthyophaga*, WiPbs2 (Figure 1; Lenassi et al., 2013; Zajc et al., 2013; our unpublished data). However, preliminary data show that the kinases HwPbs2 and WiPbs2 do not interact with the *S. cerevisiae* Sho1 protein (our unpublished data). This suggests that the SHO1 branch is not involved in HOG pathway activation in *H. werneckii* and *W. ichthyophaga*.

We had previously identified and characterized only one isoform of the final MAPK HwHog1 (Turk and Plemenitaš, 2002; Lenassi et al., 2007); however, the *H. werneckii* whole genome sequence revealed another copy of the *HwHOG1* gene, the characterization of which is currently in progress. We have also

identified two Hog1-like kinase paralogs in *W. ichthyophaga*, although all of the other HOG pathway components are represented by only single gene copies (Figure 1; Konte and Plemenitaš, 2013).

HwHog1A, HwHog1B, WiHog1A, and WiHog1B are all considerably shorter than ScHog1, although they contain the conserved domains and motifs that are characteristic of the MAPKs, such as the ATP-binding region, Asp in the active site, a TGY phosphorylation motif, a common docking domain, and a Pbs2-binding domain. While HwHog1A, HwHog1B, and WiHog1B are fully functional kinases in the *S. cerevisiae* *hog1*Δ background (Lenassi et al., 2007; Konte and Plemenitaš, 2013; our unpublished data), WiHog1A can only partly restore the osmotolerance of the *hog1*Δ strain. Lower phosphorylation levels, lower *GPD1* induction, and greater cross-talk with the mating pathway indicate that WiHog1A cannot interact optimally with the protein partners in *S. cerevisiae*. WiHog1B, on the other hand, is a fully functional kinase in the *S. cerevisiae* *hog1*Δ background. Moreover, WiHog1B even improves the salt tolerance of *S. cerevisiae* (Konte and Plemenitaš, 2013). We have also demonstrated that in contrast to *S. cerevisiae*, where the levels of *HOG1* mRNA remain unchanged when the cells are exposed to osmotic shock (Brewster et al., 1993), the transcript levels of *HOG1*-like kinases in *H. werneckii* and *W. ichthyophaga* are salt-dependent (Lenassi et al., 2007; Konte and Plemenitaš, 2013).

Upon hyperosmotic shock in *S. cerevisiae*, Hog1 is rapidly phosphorylated and it translocates into the nucleus. After the cell adapts to the higher osmolarity, Hog1 is dephosphorylated by phosphatases in a negative-feedback manner (Hohmann et al., 2007). Phosphorylation patterns in the extremely halotolerant *H. werneckii* and the obligate halophile *W. ichthyophaga* appear to be more complex. In *H. werneckii*, we have observed a phosphorylation mechanism that is similar to that of *S. cerevisiae*, although the HwHog1 kinase is noticeably phosphorylated only when the *H. werneckii* cells were exposed to  $\geq 3$  M NaCl (Turk and Plemenitaš, 2002). While in *S. cerevisiae* constitutive Hog1 phosphorylation is lethal (Maeda et al., 1994), in *W. ichthyophaga* this is not the case. Even more interestingly, *W. ichthyophaga* has a completely “opposite” phosphorylation pattern to that of *S. cerevisiae*: WiHog1 kinase is dephosphorylated after hypo-osmotic or hyperosmotic shock in *W. ichthyophaga*, and it is constitutively phosphorylated under optimal osmotic conditions (3.4 M NaCl). These data indicate an important role for the phosphatases in the regulation of the HOG pathway in *W. ichthyophaga* (Konte and Plemenitaš, 2013). This model has already been reported for *Cryptococcus neoformans*, where some serotypes show inverted *W. ichthyophaga*-like phosphorylation patterns (Bahn et al., 2007).

When activated, HwHog1 is translocated into the nucleus, where it associates with the chromatin of osmoreponsive genes and induces or represses their expression (Vaupotič and Plemenitaš, 2007). A transcriptional response to hyperosmolar stress of 95 differentially expressed genes has been reported for the comparison of moderately (3 M NaCl) and extremely (4.5 M) osmolar environments. Data from the ChIP method show that 36 of these genes physically interact with HwHog1 in long-term adaptation to extreme environments (Vaupotič and Plemenitaš, 2007). In 17 out of these 36 genes, simultaneous co-localization

of RNA polymerase II was seen. More than half of differentially expressed genes are related to general metabolism and energy production, and the other osmoresponsive genes are involved in the biogenesis of mitochondria, protein biosynthesis, protein quality control, transport facilitation, the cell cycle, and the cell wall (Vaupotič and Plemenitaš, 2007). Thirteen of these 95 genes could not be classified. Certain osmoresponsive genes controlled by MAPK HwHog1 have been studied in greater detail. Genes that code for the P-type ATPases HwEna1 and HwEna2 are the *S. cerevisiae* ENA1 homologs, and therefore they are believed to be involved in the maintenance of a low intracellular K<sup>+</sup>/Na<sup>+</sup> ratio. Their transcription is salt regulated (Gorjan and Plemenitaš, 2006). Two homologs of the *S. cerevisiae* key enzyme in glycerol biosynthesis, the glycerol-3-phosphate dehydrogenase Gpd1, have been characterized in *H. werneckii*. These both show similar transcription profiles in response to different salt concentrations (Lenassi et al., 2011). We have also demonstrated that the MAPK WiHog1 can up-regulate the transcription of *GPD1* in *S. cerevisiae*. The regulation of other osmoresponsive genes that are potential targets of WiHog1 remains to be defined.

## HORTAEA WERNECKII GENOME ANALYSIS

### WHOLE-GENOME DUPLICATION

The genome of *H. werneckii* was recently sequenced and it has been deposited at DDBJ/EMBL/GenBank under the accession number AIJO00000000 (Lenassi et al., 2013). The genome statistics are summarized in **Table 1**. The genome of *H. werneckii* has a size of 51.6 Mb, which is relatively large. In species belonging to the same order as *H. werneckii* (*Capnodiales*), the genome sizes are very variable, as they range from 21.88 to 74.12 Mb. The larger genome sizes are mostly due to a substantial amount of repetitive sequences. However, in *H. werneckii*, despite its large genome size, the proportion of repetitive sequences is only 1.02%. On the other hand, it contains 23,333 predicted genes, which is twice as many as the average number of predicted genes in other related fungi (approx. 11,955 genes) (Ohm et al., 2012). This large number of genes can be attributed to a relatively recent whole genome duplication, which resulted in two nearly identical copies of almost every protein of *H. werneckii* (Lenassi et al., 2013). This discovery is in line with our previous studies of several individual genes from *H. werneckii* that were present in two copies (Gorjan and Plemenitaš, 2006; Lenassi and Plemenitas, 2007; Fettich et al., 2011). In most cases, the expression of both of the gene copies is salt dependent, although their expression profiles differ (Lenassi and Plemenitas, 2007). It may well be that as a consequence of this whole genome duplication, *H. werneckii* can benefit from the potential advantages of large genetic redundancy, even though it is formally in a haploid stage (i.e., it is not a diploid that has resulted from the mating of two strains with opposite mating types, which would regain the haploid stage with meiosis before the next mating event; see below).

### MATING GENES

The *H. werneckii* genome sequence has offered the opportunity to gain insight into the genetic information on the mating type(s) and on the mating strategy. To date, no sexual cycle has been described for *H. werneckii*. Using *M. graminicola*

**Table 1 | Genome statistics for *W. ichthyophaga* and *H. werneckii* (after Lenassi et al., 2013; Zajc et al., 2013).**

Statistic	Wi	Hw
Coverage	>270×	~70×
Genome assembly size (Mbp)	9.63	51.6
Number of contigs	101	~20,000
Number of scaffolds	82	/
Gene models (n)	4884	23333
GC content (%)	45%	54%
GC content of exons (%)	48%	56%
Repeat content (%)	1.67%	1.02%

proteins that contain the alpha1 domain (Mat1-1-1) and the HMG domain (Mat1-1-2), we identified the putative *HwMAT1-1-1A* and *HwMAT1-1-1B* genes (**Figure 1**, **Table 2**), both of which are translated into 358 amino-acid proteins that contain the alpha1 domain and have an overall amino-acid sequence identity of 87.5% (Lenassi et al., 2013). Importantly, no homologs of the HMG-domain-containing Mat1-1-2 protein were found in *H. werneckii*, which indicates that this species is heterothallic, and that if it can still undergo sexual reproduction, this requires a strain that codes for the opposite mating type (in the case of the sequenced strain, this would be a strain with a Mat1-1-2 homolog; Lenassi et al., 2013).

### ALKALI-CATION TRANSPORT SYSTEMS

Eukaryotic microorganisms have developed numerous plasma-membrane transport systems to maintain their appropriate alkali cation levels, and in particular, to eliminate any surplus of toxic Na<sup>+</sup> ions. The alkali-cation transport systems in *S. cerevisiae* and in non-conventional yeast have recently been reviewed (Arino et al., 2010; Ramos et al., 2011).

In *S. cerevisiae*, the plasma-membrane transporters Trk1 and Trk2 for K<sup>+</sup> uptake, the Tok1 K<sup>+</sup> channel, the Pho98 inorganic phosphate (P<sub>i</sub>)-Na<sup>+</sup> symporter, the Ena Na<sup>+</sup>-ATPases, and the Nha1 Na<sup>+</sup>/H<sup>+</sup> antiporter have all been well characterized (Arino et al., 2010). Together with these, non-specific protein transporters (e.g., Pm3, Qdr2) have been described to be involved in K<sup>+</sup>/Na<sup>+</sup> fluxes across the plasma membrane (Arino et al., 2010). Trk transporters for K<sup>+</sup> uptake, Nha antiporters, Ena ATPases, and Tok1 channels have also been identified in non-conventional yeast, together with the Hak K<sup>+</sup>/H<sup>+</sup> symporters and the rare K<sup>+</sup>/Na<sup>+</sup>-uptake ATPase Acu (Ramos et al., 2011).

Physiological studies have shown that *H. werneckii* maintains very low intracellular K<sup>+</sup> and Na<sup>+</sup> levels (Kogej et al., 2005), even when it grows in the presence of 4.5 M NaCl, which suggested that it can effectively extrude Na<sup>+</sup> ions and also prevent their influx. Analysis of *H. werneckii* genome has revealed considerable expansion of families of genes that encode plasma-membrane metal cation transporters, as presented schematically in **Figure 1** and summarized in **Table 2** (Lenassi et al., 2013). We identified eight homologs of the Trk1 and Trk2 K<sup>+</sup> channels, with each containing the conserved TrkH domain that is typical for cation transport proteins. In general, they show low homology to the Trk1 protein, but the amino-acid sequence identity increases in the TrkH domain. We also identified four homologs of the Tok1 K<sup>+</sup>

**Table 2 | Major proteins, identified in *H. werneckii* and *W. ichthyophaga* and presumably involved in adaptation mechanisms to increased salinity.**

Description	<i>S. cerevisiae</i> *	<i>H. werneckii</i> *	<i>W. ichthyophaga</i> *
<b>HOG PATHWAY</b>			
transmembrane osmosensor	Sho1	HwSho1A/B	
MAP kinase kinase	Pbs2	HwPbs2A/B	WiPbs2
MAP kinase	Hog1 (ScHog1)	HwHog1A/B	WiHog1A/B
MAP kinase kinase kinase	Ste11	HwSte11A/B	WiSte11
phosphorelay response regulator	Ssk1	HwSsk1A/B	WiSsk1
MAP kinase kinase kinase	Ssk2	HwSsk2A/B	WiSsk2
hybrid histidine kinase		HwHhk7A/B	
tyrosine (Y) phosphatase dependent	Ypd1	HwYpd1A/B	WiYpd1
<b>MATING</b>			
mating-type protein	Mat1-1-1 or Mat1-2-1	HwMat1-1-1A/B	none
<b>COMPATIBLE SOLUTE MANAGEMENT</b>			
glycerol-3-phosphate dehydrogenase	Gpd1, Gpd2	HwGpd1A/B, HwGpd2A/B	WiGpd1, WiGpd2
glycerol-3-phosphatase	Gpp1, Gpp2		WiGpp1
glycerol proton symporter, plasma membrane	Stl1		4
aquaglyceroporin	Fps1		3
D-arabinitol-2-dehydrogenase			2
<b>ALKALI METAL ION CONCENTRATION MANAGEMENT</b>			
K <sup>+</sup> efflux antiporter, plasma membrane	Nha1	8	2
K <sup>+</sup> efflux channel, plasma membrane	Tok1	4	none
K <sup>+</sup> uptake uniporter, plasma membrane	Trk1,2	8	1
Na <sup>+</sup> (and Li <sup>+</sup> ) efflux P-type ATPase, plasma membrane	Ena1-5	4	2
Na <sup>+</sup> /Pi symporter, plasma membrane	Pho89	6	1
H <sup>+</sup> exporter P-type ATPase, plasma membrane	Pma1	4	3
K <sup>+</sup> /H <sup>+</sup> antiporter, Golgi apparatus	Kha1	2	2
Na <sup>+</sup> /H <sup>+</sup> antiporter, late endosomes	Nhx1	2	1
Na <sup>+</sup> , K <sup>+</sup> /H <sup>+</sup> antiporter, vacuole	Vnx1	8	1
H <sup>+</sup> V-type ATPase, subunit A	Vma1	2	1
K <sup>+</sup> /H <sup>+</sup> antiporter, mitochondria	Mrs7/Mdm38	2	1

\*The columns contain protein names or number of homologs of each protein.

channels, each of which contains two conserved transmembrane helices that are typical of this ion-channel family. Again, the homology to the Tok1 protein is low, but the identity is high in the transmembrane helices. The presence of eight homologs of the Nha1 Na<sup>+</sup>/K<sup>+</sup>, H<sup>+</sup> antiporters was demonstrated, each of which contains a transmembrane region at the N-terminal, which is conserved through the Na<sup>+</sup>/K<sup>+</sup>, H<sup>+</sup> exchanger family, and only two of them additionally contain the C-terminal cytoplasmic region. Extensive expansion has also been observed for the Pho89 homologs in *H. werneckii*, as we identified six homologs of the Pho89 Na<sup>+</sup>, Pi symporter, with each homolog containing at least one PHO4 domain. In contrast with the abundant transporter families mentioned, only four homologs of three *S. cerevisiae* Ena Na<sup>+</sup> P-type ATPases have been identified in the *H. werneckii* genome (Lenassi et al., 2013). Previously, we identified and characterized two Ena-like P-ATPases (Gorjan and Plemenitaš, 2006). Analysis of the Ena Na<sup>+</sup> P-type ATPases identified in the genome of *H. werneckii* reveals that each homolog contains all four of the conserved domains found in the *S. cerevisiae* Ena proteins. Considering their multiplication, it appears that Nha transporters are more important than Ena. On the other hand, based on our previous data that demonstrated that *HwENA* genes are highly induced at alkaline pH (Gorjan and Plemenitaš,

2006), we speculate that they have complementary functions: Ena ATPases are more important at high pH, where the Nha antiporters cannot function correctly.

As well as the important role of plasma-membrane transport systems in ion homeostasis, in the cytosol, K<sup>+</sup> homeostasis and Na<sup>+</sup> detoxification are also connected to cation transport across the organelle membranes (Arino et al., 2010). In *S. cerevisiae*, endosomal Nhx1 (Nass and Rao, 1999) and Kha1 from the Golgi apparatus (Maresova and Sychrova, 2005) are Na<sup>+</sup>/H<sup>+</sup> exchangers, similar to Nha1 at the plasma membrane (Prior et al., 1996). The vacuolar Vnx1 (Cagnac et al., 2007) and the mitochondrial Mdm38 and Mrs7 (Nowikovsky et al., 2004; Zotova et al., 2010) have similar Na<sup>+</sup>/K<sup>+</sup>, H<sup>+</sup> exchanger functions, but different structures.

We found that homologs of Nhx1 and Kha1 are duplicated in the *H. werneckii* genome, all of which contain the domains that are typical for the Na<sup>+</sup>/H<sup>+</sup> exchanger family. The same has been observed for the Kha1 homologs. We also identified two homologs of transporters with high homology to the Mrs7 and Mdm38 transporters from *S. cerevisiae*. Of the intracellular cation transporters, only the homologs of the vacuolar Vnx1 are enriched in *H. werneckii* in comparison to *S. cerevisiae*. We identified eight homologs of the Vnx1 Na<sup>+</sup>/K<sup>+</sup>, H<sup>+</sup> antiporter, but

the homology of the HwVnx proteins compared to Vnx1 is low (Lenassi et al., 2013).

The activities of many transporters are closely connected to the proton gradients across the membranes, which are generated by the Pma1 P-type ATPase at the plasma membrane (Serrano et al., 1986; Ambesi et al., 2000) and the V-type ATPase at the vacuolar membrane (Graham et al., 2000). Different P-type ATPases use ATP hydrolysis as a source of energy for the transport of ions through the membrane, and they are structurally similar (Kuhlbrandt, 2004).

The enrichment of the transporters responsible for supplying the energy for the cation transporters in *H. werneckii* supports the importance of the complex cation transporter system for combating high environmental  $\text{Na}^+$ . We identified four homologs of Pma1 in *H. werneckii*, with each homolog containing three conserved domains that are also found in the *S. cerevisiae* Pma1 and Pma2 proteins. The importance of all four of the *H. werneckii* Pma homologs for cation homeostasis is also supported by expression analysis, as the expression profiles of the *PMA1* and *PMA2* homologs show different levels of response to saline conditions (Lenassi et al., 2013). Comparisons of the expression profiles of the *PMA* genes in *H. werneckii* with those described in *S. cerevisiae* have shown that in *S. cerevisiae*, *PMA1* is not induced by salt stress (Yale and Bohnert, 2001), while in *H. werneckii*, both *PMA1* and *PMA2* have salt-regulated transcription.

The yeast vacuolar ATPase does not only have a crucial role in the acidification of the vacuolar lumen, but it is also important for the correct functioning of other organelles (Arino et al., 2010). In the *H. werneckii* genome, we found homologs of all of the subunits of the *S. cerevisiae* V-ATPase complex. The *H. werneckii* vacuolar subunits in general share a lot of similarity with the *S. cerevisiae* subunits, which is not surprising, as their structures and function have been highly conserved through evolution (Graham et al., 2000). *S. cerevisiae* vacuolar ATPases are localized at different cellular locations; however, it remains to be determined where they are specifically localized in *H. werneckii*. In contrast to the transcription of the *HwPMAs*, which is salt regulated, no such trends have been seen for the expression of the VMA homologs under different salinities.

## WALLEMIA ICHTHYOPHAGA GENOME ANALYSIS

### CHARACTERISTICS OF THE GENOME AND THE TRANSCRIPTOMES

The genome of *W. ichthyophaga* has been deposited as a Whole Genome Shotgun project at DDBJ/EMBL/GenBank under the accession number APLC00000000 (Zajc et al., 2013). The genome of *W. ichthyophaga* is 9.6 Mb in size, and the sequence currently consists of 101 contigs and 82 scaffolds (Table 1). Most basidiomycetous haploid genomes are more than twice this size (and in some cases, larger by 40-fold or more; Gregory et al., 2007). The closely related species *W. sebi* also has a slightly larger genome (9.8 Mb) (Padamsee et al., 2012). Of the species investigated thus far, only the dandruff- and seborrhoeic-dermatitis-causing *Malassezia globosa* has a smaller genome (9.0 Mb; Gregory et al., 2007). The compactness of the genome of *W. ichthyophaga* is reflected in its low level of repetitive sequences (1.67%), and high density of genes (514 genes/Mb scaffold) (Zajc et al., 2013). This is only slightly lower than *W. sebi*

(538 genes/Mb), but more than in *M. globosa* (476 genes/Mb). This means that the coding DNA sequences in *W. ichthyophaga* cover almost three quarters of the genome. The GC content in *W. ichthyophaga* is 45.35%, while in *W. sebi* this is even lower, at 40.01% (Table 1). The absolute number of predicted proteins in *W. ichthyophaga* (4884; Zajc et al., 2013) is also unusually small for a basidiomycete (where more than 10,000 proteins are not uncommon), and is in the range observed for *Escherichia coli* (Lukjancenko et al., 2010). For comparison, the *W. sebi* genome codes for 5284 proteins, while *M. globosa* contains 4285 proteins. Interestingly, the reduction in genome size and gene number is not accompanied by a reduction in intron number, such as has been reported for some other fungi with small genomes (Kelkar and Ochman, 2012).

It has not been possible to assign functions for an unproportionally large number of the proteins that are found in *W. ichthyophaga* but not in *W. sebi*. With searches through the Pfam database, three quarters of these proteins could not be classified into any of the protein families (Zajc et al., 2013). Among those that could be identified, there were several proteins related to DNA processing and DNA damage.

The sequencing of the transcriptomes of *W. ichthyophaga* grown in 10% and 30% (w/v) NaCl has revealed that 13.1% of the genes are differentially expressed under these conditions (Zajc et al., 2013). Of these, two thirds are more expressed at lower salinity. Alternative splicing, which has been identified as intron retention, was detected for 15.0% of the genes, and in more than half of the cases (51.6%), alternative splicing was detected only at one of the two tested salinities (Zajc et al., 2013).

### COMPATIBLE SOLUTE MANAGEMENT

The strategy of osmo-adaptation of both the extremely halotolerant *H. werneckii* and the halophilic *W. ichthyophaga* is the accumulation of a mixture of polyols that act as compatible solutes (Figure 1). The main osmotically regulated polyol of both *H. werneckii* and *W. ichthyophaga* is glycerol, the levels of which are increased with increasing salinity and decreased after hyperosmotic shock (Kogej et al., 2007; Zajc et al., 2014). In addition to glycerol, we have reported erythritol, arabitol, and mannitol in *H. werneckii* (Plemenitaš et al., 2008), and smaller amounts of arabitol, and traces of mannitol in *W. ichthyophaga* (Zajc et al., 2014). The genes for the enzymes known to be involved in compatible solute management are found in the genome of *W. ichthyophaga* (Table 2). These are present in several copies, with the exception of the glycerol-3-phosphatase Gpp. *W. ichthyophaga* contains a homolog of *GPD1*, *WiGPD1*, the expression of which is salt-induced (Lenassi et al., 2011). A second homolog was also found by searching the genome. When compared with the homologs of *H. werneckii*, the expression level of *WiGPD1* is lower, and the response to hyperosmotic shock is slower (Lenassi et al., 2011). Expression of *WiGPD1* in *S. cerevisiae* boosted the osmotolerance of the *gpd1* and *gpd1gpd2* mutants. As was reported for homologs from *H. werneckii*, *WiGPD1* lacks the N-terminal peroxisomal targeting (PTS2) sequence (Lenassi et al., 2011), which is important for peroxisome localization of *WiGpd1* (Jung et al., 2010). This might mean that *WiGpd1* remains in the cytosol, which would be an advantage when living in extremely

saline environments, as it is this fraction that is important for the synthesis of the compatible solutes (Lenassi et al., 2011). During hyperosmotic shock, *S. cerevisiae* counteracts glycerol leakage by its active re-import using the protein Stl1, a glycerol/H<sup>+</sup> symporter in the plasma membrane (Ferreira et al., 2005). Similarly, the aquaglyceroporin channel Fps1 remains closed (while it opens during hypo-osmotic shock, to facilitate expulsion of excess glycerol) (Luyten et al., 1995). In *W. ichthyophaga*, four homologs of Stl1 have been found (Figure 1, Table 2), as well as three aquaglyceroporin-related proteins.

In the basidiomycete *Agaricus bisporus*, the solute D-mannitol is synthesized from fructose via a reduction step that is catalysed by two NADP-dependent mannitol dehydrogenases (Stoop and Mooibroek, 1998). *W. ichthyophaga* also contains two homologs of D-arabinitol-2-dehydrogenases, which are used in other fungi for the production of arabitol from an intermediate of the pentose phosphate pathway, D-ribulose-5-phosphate.

#### TRANSPORT OF ALKALI METAL IONS

In line with the strategy of compatible solutes, the intracellular levels of K<sup>+</sup> and Na<sup>+</sup> in *W. ichthyophaga* remain low at constant salinities (not above 30 nmol/mg dry biomass) (Zajc et al., 2013), even when compared to *H. werneckii* (not above 180 nmol/mg dry biomass) (Kogej et al., 2007). However, when under hyperosmotic shock, the levels of both cations increase significantly in *W. ichthyophaga*, indicating its poor capability to adjust to changing environments. The ratio between these cations decreases with increasing salinity, due to the rising levels of Na<sup>+</sup> and the lowering of K<sup>+</sup>. However, the intracellular K<sup>+</sup>/Na<sup>+</sup> ratio is higher across the whole salinity range in *W. ichthyophaga* compared to *H. werneckii* (Figure 1) and some other halotolerant fungi (e.g., *Aureobasidium pullulans*, *Debaryomyces hansenii*). In addition, the K<sup>+</sup>/Na<sup>+</sup> decrease over the salinity range is less steep in *W. ichthyophaga* compared to *H. werneckii*. The growth performance of *W. ichthyophaga* is greatest when the Na<sup>+</sup> content exceeds that of K<sup>+</sup> (Zajc et al., 2014). This indicates that these intracellular concentrations of Na<sup>+</sup> ions are not toxic to the cells.

Data from the genome show that there are only a low number of cation transporters, except for the enriched protein family of P-type ATPases (Figure 1, Table 2). Also, expression of the cation transporters is low and independent of salt, with only three minor exceptions (described below). This is probably associated with the life of *W. ichthyophaga* at relatively constant (although extremely high) salinities. Nevertheless, in its genome we observed a significant enrichment of the cation-transporting ATPases family. The identified proteins of this family are three H<sup>+</sup> and two Na<sup>+</sup> P-type ATPases (all of which are assumed to be located at the plasma membrane), two Ca<sup>2+</sup> P-type ATPases (vacuolar Pmc1 and Pmr1 from the Golgi apparatus), and a putative transporter of unknown specificity (Zajc et al., 2013). The *W. ichthyophaga* genome encodes three putative Pma proton pumps, while *W. sebi* contains only two (Zajc et al., 2013).

In environments with high concentrations of Na<sup>+</sup> salts, the cell must prevent the intracellular accumulation of the highly toxic Na<sup>+</sup>, without lowering the levels of K<sup>+</sup>. This is achieved by a variety of other secondary active transporters. *W. ichthyophaga* contains homologs of most known transporters from *S. cerevisiae*

(Arino et al., 2010) and unconventional yeast (Ramos et al., 2011), as either those located on intracellular membranes (Kha1, Mrs7/Mdm37, NhX1, Pmc1, Pmr1, Vnx1, Vma1) or at the plasma membrane (Ena, Nha1, Pho89, Pma, Trk1).

Judging by the genes that encode alkali metal cation transporters, the extremely halotolerant ascomycete *H. werneckii* and *W. ichthyophaga* use different salt-combating strategies. As described above, in *H. werneckii*, the numbers of most of the plasma-membrane alkali cation transporters are substantially increased. In *W. ichthyophaga* this is not the case (Figure 1, Table 2). *W. ichthyophaga* contains only one Trk homolog (inward K<sup>+</sup> transporter, with eight copies in *H. werneckii*) and no Tok homologs (outward K<sup>+</sup> channel, with four copies in *H. werneckii*). Similarly, *W. ichthyophaga* has only two Nha homologs (Na<sup>+</sup>/K<sup>+</sup> proton antiporters) and one Pho89 (Na<sup>+</sup>/Pi symporter), while *H. werneckii* contains eight and six, respectively (Lenassi et al., 2013; Zajc et al., 2013).

Active import of K<sup>+</sup> might contribute to ion homeostasis in hypersaline environments, and this would complement the action of passive K<sup>+</sup> channels. The known active transporters are K<sup>+</sup>-H<sup>+</sup> symporters (Hak symporters) and K<sup>+</sup>(Na<sup>+</sup>)-ATPase (Acu, alkali cation uptake transporters) (Benito et al., 2004; Ramos et al., 2011). While *H. werneckii* has no homologs of either of these transporter types (Lenassi et al., 2013), *W. ichthyophaga* contains two possible homologs of the otherwise rare Acu ATPases, of which only one contains a P-type ATPase domain (Zajc et al., 2013).

Several different membrane alkali metal transporters (mainly cation/H<sup>+</sup> antiporters) are also located on the organelle membranes: the Golgi apparatus (Kha1), mitochondria (Mdm38 or Mrs7), endosomes (NhX1), and vacuole (Vnx1) (Arino et al., 2010). In *W. ichthyophaga*, there are single-copy genes of all of these proteins (with the exception of the duplicated Kha1).

*W. ichthyophaga* can live at extremely high salinity; however, the above-described findings indicate that the salinity remains relatively constant and thus no rapid responses are needed. Under constant conditions, there is no need for a quick release of surplus K<sup>+</sup>, and thus the absence of the Tok outward K<sup>+</sup> channel might not be detrimental. On the other hand, *W. ichthyophaga* contains three homologs of aquaglyceroporins (Zajc et al., 2013), which fulfil the need for rapid expulsion of the accumulated compatible solute glycerol (Luyten et al., 1995). Two explanations are possible: (1) *W. ichthyophaga* might deal with hypo-osmotic shock-related K<sup>+</sup> expulsion in other ways than in other fungi; and (2) the aquaglyceroporin channels might serve some other functions than expulsion of glycerol during the shock.

In cells grown at 10 and 30% (w/v) NaCl, the large majority of the genes that encode metal-cation transporters are not differentially expressed. This is not as expected, since these transporters are believed to have crucial roles in adaptation to salt. The only exceptions are a homolog of the Pho89 Na<sup>+</sup>/Pi symporter, which shows elevated expression at high salinity, and a putative P-type Na<sup>+</sup> ATPase and a possible Acu K<sup>+</sup> importer, with the expression of both of these latter higher at low salinity. This is in stark contrast, for example, with other halotolerant fungi (e.g., *D. hansenii*, *H. werneckii*), where even at different constant salinities, differential expression of P-type H<sup>+</sup> and Na<sup>+</sup> ATPases has been observed

(Almagro et al., 2001; Gorjan and Plemenitaš, 2006; Lenassi et al., 2013). Furthermore, in *W. ichthyophaga* the expression of transporter coding genes is relatively low: of the total of 4884 genes, neither the  $\text{Na}^+$ -exporting P-type ATPases nor the two  $\text{Na}^+/\text{H}^+$  antiporters are among the 2000 most-expressed genes at high salinity. The possible post-transcriptional control of all of these genes remains to be investigated.

Continuous removal and/or compartmentalization of  $\text{Na}^+$  at a constant high salinity is extremely demanding energetically. The low numbers of transporters in *W. ichthyophaga* might reflect the relatively low adaptive potential of this species to changes in salt concentrations and/or its specialization with other mechanisms that are energetically more efficient. The apparent transcriptional non-responsiveness of transporters to salt can lead to similar conclusions. This would mean that the halophilic strategy of *W. ichthyophaga*, which is a unique example of a narrowly specialized fungal halophile (Gostinčar et al., 2010), is substantially different from that of *H. werneckii*, which contains a collection of  $\text{K}^+$ -channels and can adapt to a wide salinity range (Zajc et al., 2013).

## HYDROPHOBINS

The analysis of the *W. ichthyophaga* genome has revealed a significant expansion of seven protein families and contraction of 19. The most interesting of the expanded families are the hydrophobins, which are proteins that potentially have a role in the particular morphological adaptations of *W. ichthyophaga* (Figure 1; Zajc et al., 2013).

*W. ichthyophaga* has a characteristic morphology that can be seen in many stress-tolerant species: compact multicellular clumps that are similar to sarcinae (Zalar et al., 2005b). This morphology has been observed in, and it is believed to enhance survival in, high-stress environments (Wollenzien et al., 1995; Palkova and Vachova, 2006; Gostinčar et al., 2011). These cells have an abundant cover of extracellular polysaccharides (Kralj Kunčič et al., 2010) that can serve as protectants during desiccation (Selbmann et al., 2005), and possibly also when exposed to high concentrations of salt (Zajc et al., 2013). Additionally, at high salinity the morphology of *W. ichthyophaga* cells changes strikingly. The size of the meristematic cell clumps increases four-fold, and the cell walls become three-fold thicker, which also substantially decreases the intracellular volume (Kralj Kunčič et al., 2010).

The hydrophobins are proteins in the cell wall of filamentous fungi. These small ( $\leq 20$  kDa) and amphipathic molecules (Linder et al., 2005) are involved in a range of processes of cellular growth and development (Wosten, 2001). Possibly as a reflection of the many different roles they have, the hydrophobin genes are often present in multiple different copies. From an estimated 15 hydrophobin genes in the last common ancestor of *W. ichthyophaga* and *W. sebi*, these genes are enriched to 26 in *W. ichthyophaga*, while the number has fallen to 12 in *W. sebi*. This is the most significant protein family expansion in the genome of *W. ichthyophaga* (Zajc et al., 2013).

The hydrophobins that have been identified in *W. ichthyophaga* and *W. sebi* contain the characteristic hydrophobin pattern of the conserved spacing of eight cysteine residues (Zajc et al., 2013) that form four disulphide bridges (Hektor and Scholtmeijer, 2005;

Linder et al., 2005). These hydrophobins in both *W. ichthyophaga* and *W. sebi* contain a high proportion of acidic amino acids compared to homologs from other fungi (Zajc et al., 2013). This is similar to the archaeal halophilic proteins (Madern et al., 2000) and it might represent an adaptation to salt exposure. If acidic amino acids are exposed on a protein surface, they can bind salt and water, and thus help to avoid salt-induced changes in conformation that would lead to loss of activity (Sigliocco et al., 2011). This might also be the case for the hydrophobins of *W. ichthyophaga*, as these are among the few of its proteins that are actually exposed to high concentrations of NaCl and are not protected in the intracellular compatible-solute-rich environment. Interestingly, half of the genes encoding hydrophobins are differentially expressed during growth of *W. ichthyophaga* at different salinities, although their responses are not the same: at high salinity some of these hydrophobins have higher expression, and some have lower expression (Zajc et al., 2013).

The hydrophobins can self-assemble into amphipathic monolayers on hydrophobic–hydrophilic interfaces, which is crucial for their diverse functions. With their help, the cell can attach to hydrophobic surfaces, break through a water-air interface, and avoid water-logging without impeding the exchange of gasses. Hydrophobins can also strengthen the cell wall and make it more rigid, and they also impact on the movement of solutes (Wosten, 2001; Bayry et al., 2012). It is not difficult to imagine that these functions will be beneficial to cells exposed to high salinity (Zajc et al., 2013). Under these conditions, the leakage of compatible solutes and intrusion of toxic salt ions are among the greatest challenges to the cell. Increased strength and rigidity of the cell wall will help the cell to survive the structural stress that can be imposed by changes in the environmental osmolarity.

The hydrophobins might also have a role in the characteristic sarcina-like morphology of *W. ichthyophaga*, which is similar to *Fusarium verticillioides*, where the hydrophobins trigger microconidial chain formation (Fuchs et al., 2004). This might additionally be involved in the aggregation of *W. ichthyophaga* cells into compact cell clusters, as is characteristic of *W. ichthyophaga*, and which is probably a form of salt-stress response (Gostinčar et al., 2010). This characteristic sarcina-like morphology is considered to be among the main haloadaptations of *W. ichthyophaga*. Indeed, both the changes in the cell wall and the formation of multicellular structures have previously been suggested to be among the main adaptations of *W. ichthyophaga* to hypersaline environments (Kralj Kunčič et al., 2010).

## GENOMIC EVIDENCE OF ASEXUALITY

Traditional mycological approaches have not lead to any descriptions of mating behaviors for either *H. werneckii* or the *Wallemia* spp.. In both cases, no descriptions of fruiting bodies or reports of teleomorphs can be found in the literature. Basidiomycetous fungi have a tetrapolar *MAT* locus. Two additional unlinked loci encode the homeodomain-containing transcription factors and pheromone/pheromone receptors. In some cases, other mating architectures have evolved through the expansion and fusion of these *MAT* loci (Lee et al., 2010). The genome of *W. sebi* contains a single mating-type locus and lacks only a few meiosis-specific genes, which suggested that it can undergo sexual reproduction

(Padamsee et al., 2012). This is not the case for *W. ichthyophaga* (Zajc et al., 2013). Searches of the *W. ichthyophaga* genome for proteins that are similar to the gene products involved in mating in other basidiomycetes has resulted in the identification of only a few proteins, and even these are very dissimilar to proteins from other fungi. No discernible mating locus has been identified, and only three of eight meiosis-specific genes (as listed in Malik et al., 2008) have been found (Zajc et al., 2013).

Asexuality in fungi is not unusual. Asexual reproduction saves energy, as it eliminates the need to produce gametes and attractants, and this might be useful in extreme environments where careful management of energy is of key importance. In habitats that do not undergo major changes over longer time scales, this would also prevent the drowning of specific adaptations of local populations in the larger gene pool of the species (Gostinčar et al., 2010; Sun and Heitman, 2011). Therefore, in the case of these extremophilic fungal species, an asexual lifestyle might have evolutionary advantages (Gostinčar et al., 2010). This appears to be the case for *W. ichthyophaga*, and this might also be true for *H. werneckii*. While a putative mating type locus has been found in the genome of *H. werneckii*, no sexual stage of this species has been described to date. As mating genes can have roles outside of mating (Srikantha et al., 2012), their presence in itself is not confirmation of sexual reproduction in a species. The potential role of sexual reproduction in *H. werneckii* therefore needs to be further investigated.

## CONCLUSIONS

The extremely halotolerant and adaptable *H. werneckii* and the obligately halophilic *W. ichthyophaga* live in environments that are defined by low  $a_w$  and high concentrations of toxic inorganic ions. To withstand these harsh environmental conditions they use some common molecular mechanisms and also some specific molecular mechanisms. A model of the key adaptations in *H. werneckii* and *W. ichthyophaga* that have been discussed here is illustrated in **Figure 1**.

When a cell is exposed to a high osmolarity (salinity) environment, it has to react rapidly to the consequent loss of water. In *H. werneckii* and *W. ichthyophaga* this is achieved by the synthesis of glycerol and a few other compatible solutes (**Figure 1**). The expression of the key enzyme in glycerol synthesis, Gpd1, is under the control of the HOG signaling pathway, which is also important for other aspects of adaptation to high osmolarity environments. The key proteins of the HOG pathway are conserved in *H. werneckii* and *W. ichthyophaga*, although the HOG pathway architecture and regulation is different. While all of the HOG pathway components are present in at least two isoforms in *H. werneckii*, there are only two isoforms of the Hog1 kinase in *W. ichthyophaga* (**Figure 1**). In *H. werneckii*, HwHog1 is only phosphorylated under extracellular conditions of  $\geq 3$  M NaCl (Turk and Plemenitaš, 2002), while in *W. ichthyophaga*, the WiHog1 kinase is constitutively phosphorylated under optimal osmotic conditions and is dephosphorylated upon hyperosmotic or hypo-osmotic shock (Konte and Plemenitaš, 2013). In *H. werneckii*, HwHog1 promotes differential induction or repression of osmoreponsive genes, depending on the osmolarity, and also by physically interacting with chromatin and RNA polymerase

II (Vaupotič and Plemenitaš, 2007). While this specific architecture of the HOG pathway might represent the background for the extreme halotolerance of *H. werneckii*, the constitutive phosphorylation of the Hog1-like kinase in *W. ichthyophaga* might support its obligate halophilic nature.

In many natural hypersaline environments, the concentrations of toxic  $\text{Na}^+$  ions are far greater than those of  $\text{K}^+$  ions, and therefore the mechanisms that maintain the stable and high intracellular  $\text{K}^+/\text{Na}^+$  ratio are crucial for the survival in such environments. Both, *H. werneckii* and *W. ichthyophaga* can maintain high  $\text{K}^+/\text{Na}^+$  ratios over a wide range of environmental  $\text{Na}^+$  concentrations. In *H. werneckii*, this homeostasis is maintained by regulated transport of  $\text{K}^+$  and  $\text{Na}^+$  across the plasma membrane, as cation transporters are diverse and highly enriched in this fungus (**Figure 1**). *W. ichthyophaga* also regulates the entry and expulsion of cations; however, it mainly prevents their entry by dynamic cell-wall restructuring. An explanation for these observed differences in the way that these fungi combat the toxic environmental  $\text{Na}^+$  might be the need for *H. werneckii* to adapt rapidly to highly dynamic concentrations of NaCl (and other salts) that it typically encounters in its natural environment, while *W. ichthyophaga* thrives instead under continuously high salinity.

Large genetic redundancy is an important characteristic of *H. werneckii*, which has presumably resulted from the whole genome duplication. Although the whole genome duplication that is seen in *H. werneckii* is not uncommon for fungi, it is interesting that this duplication has not yet been followed by selective gene loss, as the large majority of genes in *H. werneckii* is still present in two copies. Such redundancy might be an excellent reservoir for cryptic genetic variability, which is of importance under stress environments that require good adaptability (Gostinčar et al., 2010). This might be especially the case, as the extent of sexual recombination as a means of generating genetic diversity in *H. werneckii* is still not known. Indeed, in *H. werneckii* there is a recognizable putative mating type locus, which indicates the possibility of (although so far not observed) a sexual reproduction cycle, while *W. ichthyophaga* lacks the cellular machinery for sexual reproduction altogether (**Figure 1**).

We believe that the recently published genome and transcriptome sequences of *H. werneckii* and *W. ichthyophaga* will accelerate research into the osmotic strategies of these fungi that can thrive in high salinity environments that allow the survival of only the most specialized minority of eukaryotes and prokaryotes.

## ACKNOWLEDGMENTS

The authors acknowledge financial support from the state budget of the Slovenian Research Agency (Infrastructural Centre Mycosmo, MRIC UL, Programme P1, Postdoctoral Project Z4-5531 to Cene Gostinčar, and Young Researcher Grants to Janja Zajc, Tilen Konte and Anja Kejžar). The study was also partly financed via the operation “Centre of excellence for integrated approaches in chemistry and biology of proteins,” number OP13.1.1.2.02.0005, financed by the European Regional Development Fund (85% share of financing) and the Slovenian Ministry of Higher Education, Science and Technology (15% share of financing).

## REFERENCES

- Almagro, A., Prista, C., Benito, B., Loureiro-Dias, M. C., and Ramos, J. (2001). Cloning and expression of two genes coding for sodium pumps in the salt-tolerant yeast *Debaromyces hansenii*. *J. Bacteriol.* 183, 3251–3255. doi: 10.1128/JB.183.10.3251-3255.2001
- Ambesi, A., Miranda, M., Petrov, V. V., and Slayman, C. W. (2000). Biogenesis and function of the yeast plasma-membrane H<sup>+</sup>-ATPase. *J. Exp. Biol.* 203(pt 1), 155–160.
- Arino, J., Ramos, J., and Sychrova, H. (2010). Alkali metal cation transport and homeostasis in yeasts. *Microbiol. Mol. Biol. Rev.* 74, 95–120. doi: 10.1128/MMBR.00042-09
- Bahn, Y. S., Geunes-Boyce, S., and Heitman, J. (2007). Ssk2 mitogen-activated protein kinase kinase kinase governs divergent patterns of the stress-activated Hog1 signaling pathway in *Cryptococcus neoformans*. *Eukaryot. Cell* 6, 2278–2289. doi: 10.1128/EC.00349-07
- Bayry, J., Aimanianda, V., Guijarro, J. I., Sunde, M., and Latge, J. P. (2012). Hydrophobins—unique fungal proteins. *PLoS Pathog.* 8:e1002700. doi: 10.1371/journal.ppat.1002700
- Benito, B., Garcialeblas, B., Schreier, P., and Rodriguez-Navarro, A. (2004). Novel P-type ATPases mediate high-affinity potassium or sodium uptake in fungi. *Eukaryot. Cell* 3, 359–368. doi: 10.1128/EC.3.2.359-368.2004
- Blomberg, A., and Adler, L. (1992). Physiology of osmotolerance in fungi. *Adv. Microb. Physiol.* 33, 145–212. doi: 10.1016/S0065-2911(08)60217-9
- Brewster, J. L., de Valoir, T., Dwyer, N. D., Winter, E., and Gustin, M. C. (1993). An osmosensing signal transduction pathway in yeast. *Science* 259, 1760–1763. doi: 10.1126/science.7681220
- Cagnac, O., Leterrier, M., Yeager, M., and Blumwald, E. (2007). Identification and characterization of Vnx1p, a novel type of vacuolar monovalent Cation/H<sup>+</sup> antiporter of *Saccharomyces cerevisiae*. *J. Biol. Chem.* 282, 24284–24293. doi: 10.1074/jbc.M703116200
- Cantrell, S. A., Casillas-Martinez, L., and Molina, M. (2006). Characterization of fungi from hypersaline environments of solar salterns using morphological and molecular techniques. *Mycol. Res.* 110, 962–970. doi: 10.1016/j.mycres.2006.06.005
- de Hoog, G. S., and Gerrits van den Ende, A. H. (1992). Nutritional pattern and eco-physiology of *Hortaea werneckii*, agent of human tinea nigra. *Anton. Leeuw.* 62, 321–329. doi: 10.1007/BF00572601
- de Hoog, G. S., and Guého, E. (1998). “Agents of white piedra, black piedra and tinea nigra,” in *Topley and Wilson's Microbiology and Microbial Infections*, Vol. 4, eds L. Ajello and R. J. Hay (London: Arnold Publications), 191–197.
- Ferreira, C., van Voorst, F., Martins, A., Neves, L., Oliveira, R., Kielland-Brandt, M. C., et al. (2005). A member of the sugar transporter family, Stl1p is the glycerol/H<sup>+</sup> symporter in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 16, 2068–2076. doi: 10.1091/mbc.E04-10-0884
- Fettich, M., Lenassi, M., Veranić, P., Gunde-Cimerman, N., and Plemenitaš, A. (2011). Identification and characterization of putative osmosensors, HwSho1A and HwSho1B, from the extremely halotolerant black yeast *Hortaea werneckii*. *Fungal Genet. Biol.* 48, 475–484. doi: 10.1016/j.fgb.2011.01.011
- Fuchs, U., Czymbek, K. J., and Sweigard, J. A. (2004). Five hydrophobin genes in *Fusarium verticillioides* include two required for microconidial chain formation. *Fungal Genet. Biol.* 41, 852–864. doi: 10.1016/j.fgb.2004.04.004
- Gorjan, A., and Plemenitaš, A. (2006). Identification and characterization of ENA ATPases HwENA1 and HwENA2 from the halophilic black yeast *Hortaea werneckii*. *FEMS Microbiol. Lett.* 265, 41–50. doi: 10.1111/j.1574-6968.2006.00473.x
- Gostinčar, C., Grube, M., de Hoog, S., Zalar, P., and Gunde-Cimerman, N. (2010). Extremotolerance in fungi: evolution on the edge. *FEMS Microbiol. Ecol.* 71, 2–11. doi: 10.1111/j.1574-6941.2009.00794.x
- Gostinčar, C., Lenassi, M., Gunde-Cimerman, N., and Plemenitaš, A. (2011). Fungal adaptation to extremely high salt concentrations. *Adv. Appl. Microbiol.* 77, 71–96. doi: 10.1016/B978-0-12-387044-5.00003-0
- Graham, L. A., Powell, B., and Stevens, T. H. (2000). Composition and assembly of the yeast vacuolar H<sup>+</sup>-ATPase complex. *J. Exp. Biol.* 203, 61–70.
- Gregory, T. R., Nicol, J. A., Tamm, H., Kullman, B., Kullman, K., Leitch, I. J., et al. (2007). Eukaryotic genome size databases. *Nucleic Acids Res.* 35, 332–338. doi: 10.1093/nar/gkl828
- Gunde-Cimerman, N., and Plemenitaš, A. (2006). Ecology and molecular adaptations of the halophilic black yeast *Hortaea werneckii*. *Rev. Environ. Sci. Biotechnol.* 5, 323–331. doi: 10.1007/s11157-006-9105-0
- Gunde-Cimerman, N., Zalar, P., de Hoog, S., and Plemenitaš, A. (2000). Hypersaline waters in salterns – natural ecological niches for halophilic black yeasts. *FEMS Microbiol. Ecol.* 32, 235–240. doi: 10.1016/S0168-6496(00)00032-5
- Hektor, H. J., and Scholtmeijer, K. (2005). Hydrophobins: proteins with potential. *Curr. Opin. Biotechnol.* 16, 434–439. doi: 10.1016/j.copbio.2005.05.004
- Hibbett, D. S., Binder, M., Bischoff, J. F., Blackwell, M., Cannon, P. F., Eriksson, O. E., et al. (2007). A higher-level phylogenetic classification of the Fungi. *Mycol. Res.* 111, 509–547. doi: 10.1016/j.mycres.2007.03.004
- Hohmann, S., Krantz, M., and Nordlander, B. (2007). Yeast osmoregulation. *Method Enzymol.* 428, 29–45. doi: 10.1016/S0076-6879(07)28002-4
- Iwatsu, T. U., and Udagawa, S. (1988). *Hortaea werneckii* isolated from sea-water. *Jpn. J. Med. Mycol.* 29, 142–145. doi: 10.3314/jjmm1960.29.142
- Jung, S., Marelli, M., Rachubinski, R. A., Goodlett, D. R., and Aitchison, J. D. (2010). Dynamic changes in the subcellular distribution of Gpd1p in response to cell stress. *J. Biol. Chem.* 285, 6739–6749. doi: 10.1074/jbc.M109.058552
- Kelkar, Y. D., and Ochman, H. (2012). Causes and consequences of genome expansion in fungi. *Genome Biol. Evol.* 4, 13–23. doi: 10.1093/gbe/evr124
- Kogej, T., Ramos, J., Plemenitaš, A., and Gunde-Cimerman, N. (2005). The halophilic fungus *Hortaea werneckii* and the halotolerant fungus *Aureobasidium pullulans* maintain low intracellular cation concentrations in hypersaline environments. *Appl. Environ. Microbiol.* 71, 6600–6605. doi: 10.1128/AEM.71.11.6600-6605.2005
- Kogej, T., Stein, M., Volkmann, M., Gorbushina, A. A., Galinski, E. A., and Gunde-Cimerman, N. (2007). Osmotic adaptation of the halophilic fungus *Hortaea werneckii*: role of osmolytes and melanization. *Microbiology* 153, 4261–4273. doi: 10.1099/mic.0.2007/010751-0
- Konte, T., and Plemenitaš, A. (2013). The HOG signal transduction pathway in the halophilic fungus *Wallemia ichthyophaga*: identification and characterisation of MAP kinases WiHog1A and WiHog1B. *Extremophiles* 17, 623–636. doi: 10.1007/s00792-013-0546-4
- Kralj Kunčič, M., Kogej, T., Drobne, D., and Gunde-Cimerman, N. (2010). Morphological response of the halophilic fungal genus *Wallemia* to high salinity. *Appl. Environ. Microbiol.* 76, 329–337. doi: 10.1128/AEM.02318-09
- Kralj Kunčič, M., Zajc, J., Drobne, D., Pipan Tkalec, Z., and Gunde-Cimerman, N. (2013). Morphological responses to high sugar concentrations differ from adaptation to high salt concentrations in the xerophilic fungi *Wallemia spp.* *Fungal Biol.* 117, 466–478. doi: 10.1016/j.funbio.2013.04.003
- Kuhlbrandt, W. (2004). Biology, structure and mechanism of P-type ATPases. *Nat. Rev. Mol. Cell Biol.* 5, 282–295. doi: 10.1038/nrm1354
- Lee, S. C., Ni, M., Li, W. J., Shertz, C., and Heitman, J. (2010). The evolution of sex: a perspective from the fungal kingdom. *Microbiol. Mol. Biol. R* 74, 298–340. doi: 10.1128/MMBR.00005-10
- Lenassi, M., Gostinčar, C., Jackman, S., Turk, M., Sadowski, I., Nislow, C., et al. (2013). Whole genome duplication and enrichment of cation transporters are associated with the halotolerant lifestyle of the black yeast *Hortaea werneckii*. *PLoS ONE* 30:e71328. doi: 10.1371/journal.pone.0071328
- Lenassi, M., and Plemenitas, A. (2007). Novel group VII histidine kinase HwHhk7B from the halophilic fungi *Hortaea werneckii* has a putative role in osmosensing. *Curr. Genet.* 51, 393–405. doi: 10.1007/s00294-007-0131-4
- Lenassi, M., Vaupotič, T., Gunde-Cimerman, N., and Plemenitaš, A. (2007). The MAP kinase HwHog1 from the halophilic black yeast *Hortaea werneckii*: coping with stresses in solar salterns. *Saline Systems* 3, 3. doi: 10.1186/1746-1448-3-3
- Lenassi, M., Zajc, J., Gostinčar, C., Gorjan, A., Gunde-Cimerman, N., and Plemenitaš, A. (2011). Adaptation of the glycerol-3-phosphate dehydrogenase Gpd1 to high salinities in the extremely halotolerant *Hortaea werneckii* and halophilic *Wallemia ichthyophaga*. *Fungal Biol.* 115, 959–970. doi: 10.1016/j.funbio.2011.04.001
- Linder, M. B., Szilvay, G. R., Nakari-Setala, T., and Penttila, M. E. (2005). Hydrophobins: the protein-amphiphiles of filamentous fungi. *FEMS Microbiol. Rev.* 29, 877–896. doi: 10.1016/j.femsre.2005.01.004
- Lukjancenko, O., Wassenaar, T. M., and Ussery, D. W. (2010). Comparison of 61 sequenced *Escherichia coli* genomes. *Microb. Ecol.* 60, 708–720. doi: 10.1007/s00248-010-9717-3
- Luyten, K., Albertyn, J., Skibbe, W. F., Prior, B. A., Ramos, J., Thevelein, J. M., et al. (1995). Fps1, a yeast member of the mip family of channel proteins, is a facilitator for glycerol uptake and efflux and is inactive under osmotic-stress. *EMBO J.* 14, 1360–1371.

- Madern, D., Ebel, C., and Zaccai, G. (2000). Halophilic adaptation of enzymes. *Extremophiles* 4, 91–98. doi: 10.1007/s007920050142
- Maeda, T., Wurgler Murphy, S. M., and Saito, H. (1994). A 2-component system that regulates an osmosensing Map kinase cascade in yeast. *Nature* 369, 242–245. doi: 10.1038/369242a0
- Malik, S. B., Pightling, A. W., Stefaniak, L. M., Schurko, A. M., and Logsdon, J. M. (2008). An expanded inventory of conserved meiotic genes provides evidence for sex in *Trichomonas vaginalis*. *PLoS ONE* 3:e2879. doi: 10.1371/journal.pone.0002879
- Maresova, L., and Sychrova, H. (2005). Physiological characterization of *Saccharomyces cerevisiae kha1* deletion mutants. *Mol. Microbiol.* 55, 588–600. doi: 10.1111/j.1365-2958.2004.04410.x
- Matheny, P. B., Gossmann, J. A., Zalar, P., Kumar, T. K. A., and Hibbett, D. S. (2006). Resolving the phylogenetic position of the Walleiomycetes: an enigmatic major lineage of Basidiomycota. *Can. J. Bot.* 84, 1794–1805. doi: 10.1139/b06-128
- Meena, N., Kaur, H., and Mondal, A. K. (2010). Interactions among HAMP domain repeats act as an osmosensing molecular switch in group III hybrid histidine kinases from Fungi. *J. Biol. Chem.* 285, 12121–12132. doi: 10.1074/jbc.M109.075721
- Mok, W. Y., Castelo, F. P., and Barreto Da Silva, M. S. (1981). Occurrence of *Exophiala werneckii* on salted freshwater fish *Osteoglossum bicirrhosum*. *Int J Food Sci Technol* 16, 505–512. doi: 10.1111/j.1365-2621.1981.tb01843.x
- Nass, R., and Rao, R. (1999). The yeast endosomal Na<sup>+</sup>/H<sup>+</sup> exchanger, Nhx1, confers osmotolerance following acute hypertonic shock. *Microbiology* 145, 3221–3228.
- Nowikovsky, K., Froschauer, E. M., Zsurka, G., Samaj, J., Reipert, S., Kolisek, M., et al. (2004). The LETM1/YOL027 gene family encodes a factor of the mitochondrial K<sup>+</sup> homeostasis with a potential role in the Wolf-Hirschhorn syndrome. *J. Biol. Chem.* 279, 30307–30315. doi: 10.1074/jbc.M403607200
- Ohm, R. A., Feau, N., Henrissat, B., Schuch, C. L., Horwitz, B. A., Barry, K. W., et al. (2012). Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. *PLoS Pathog.* 8:e1003037. doi: 10.1371/journal.ppat.1003037
- Padamsee, M., Kumar, T. K., Riley, R., Binder, M., Boyd, A., Calvo, A. M., et al. (2012). The genome of the xerotolerant mold *Wallemia sebi* reveals adaptations to osmotic stress and suggests cryptic sexual reproduction. *Fungal Genet. Biol.* 49, 217–226. doi: 10.1016/j.fgb.2012.01.007
- Palkova, Z., and Vachova, L. (2006). Life within a community: benefit to yeast long-term survival. *FEMS Microbiol. Rev.* 30, 806–824. doi: 10.1111/j.1574-6976.2006.00034.x
- Petelenz-Kurdziel, E., Eriksson, E., Smedh, M., Beck, C., Hohmann, S., and Goksor, M. (2011). Quantification of cell volume changes upon hyperosmotic stress in *Saccharomyces cerevisiae*. *Integr. Biol.* 3, 1120–1126. doi: 10.1039/c1ib00027f
- Petrović, U., Gunde-Cimerman, N., and Plemenitaš, A. (2002). Cellular responses to environmental salinity in the halophilic black yeast *Hortaea werneckii*. *Mol. Microbiol.* 45, 665–672. doi: 10.1046/j.1365-2958.2002.03021.x
- Plemenitaš, A., Vaupotič, T., Lenassi, M., Kogej, T., and Gunde-Cimerman, N. (2008). Adaptation of extremely halotolerant black yeast *Hortaea werneckii* to increased osmolarity: a molecular perspective at a glance. *Stud. Mycol.* 61, 67–75. doi: 10.3114/sim.2008.61.06
- Prior, C., Potier, S., Souciet, J. L., and Sychrova, H. (1996). Characterization of the NHA1 gene encoding a Na<sup>+</sup>/H<sup>+</sup>-antiporter of the yeast *Saccharomyces cerevisiae*. *FEBS Lett.* 387, 89–93. doi: 10.1016/0014-5793(96)00470-X
- Ramos, J., Arino, J., and Sychrova, H. (2011). Alkali-metal-cation influx and efflux systems in nonconventional yeast species. *FEMS Microbiol. Lett.* 317, 1–8. doi: 10.1111/j.1574-6968.2011.02214.x
- Saito, H., and Posas, F. (2012). Response to hyperosmotic stress. *Genetics* 192, 289–318. doi: 10.1534/genetics.112.140863
- Samson, R. A., Hoekstra, E. S., and Frisvad, J. C. (2004). *Introduction to Food- and Airborne Fungi*. Utrecht: Centraalbureau voor Schimmelcultures.
- Selbmann, L., de Hoog, G. S., Mazzaglia, A., Friedmann, E. I., and Onofri, S. (2005). Fungi at the edge of life: cryptoendolithic black fungi from Antarctic desert. *Stud. Mycol.* 1–32.
- Serrano, R., Kiellandbrandt, M. C., and Fink, G. R. (1986). Yeast plasma-membrane ATPase is essential for growth and has homology with (Na<sup>+</sup>/K<sup>+</sup>), K<sup>+</sup>- and Ca<sup>2+</sup>-ATPases. *Nature* 319, 689–693. doi: 10.1038/319689a0
- Sigliocco, A., Paiardini, A., Piscitelli, M., and Pasarella, S. (2011). Structural adaptation of extreme halophilic proteins through decrease of conserved hydrophobic contact surface. *BMC Struct. Biol.* 11:50. doi: 10.1186/1472-6807-11-50
- Srikantha, T., Daniels, K. J., Pujol, C., Sahni, N., Yi, S., and Soll, D. R. (2012). Nonsex genes in the mating type locus of *Candida albicans* play roles in a/alpha biofilm formation, including impermeability and fluconazole resistance. *PLoS Pathog.* 8:e1002476. doi: 10.1371/journal.ppat.1002476
- Stanley, J. T., Palmer, F., and Adams, J. B. (1982). Microcolonial fungi: common inhabitants on desert rocks? *Science* 215, 1093–1095. doi: 10.1126/science.215.4536.1093
- Stoop, J. M. H., and Mooibroek, H. (1998). Cloning and characterization of NADP mannitol dehydrogenase cDNA from the button mushroom, *Agaricus bisporus*, and its expression in response to NaCl stress. *Appl. Environ. Microbiol.* 64, 4689–4696.
- Sun, S., and Heitman, J. (2011). Is sex necessary? *BMC Biol.* 9:56. doi: 10.1186/1741-7007-9-56
- Turk, M., and Plemenitaš, A. (2002). The HOG pathway in the halophilic black yeast *Hortaea werneckii*: isolation of the HOG1 homolog gene and activation of HwHog1p. *FEMS Microbiol. Lett.* 216, 193–199. doi: 10.1111/j.1574-6968.2002.tb11435.x
- Vaupotič, T., and Plemenitaš, A. (2007). Differential gene expression and Hog1 interaction with osmoreponsive genes in the extremely halotolerant black yeast *Hortaea werneckii*. *BMC Genomics* 8:280. doi: 10.1186/1471-2164-8-280
- Widmann, C., Gibson, S., Jarpe, M. B., and Johnson, G. L. (1999). Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. *Physiol. Rev.* 79, 143–180.
- Wollenzien, U., Dehoog, G. S., Krumbein, W. E., and Urzi, C. (1995). On the isolation of microcolonial fungi occurring on and in marble and other calcareous rocks. *Sci. Total Environ.* 167, 287–294. doi: 10.1016/0048-9697(95)04589-S
- Wosten, H. A. (2001). Hydrophobins: multipurpose proteins. *Annu. Rev. Microbiol.* 55, 625–646. doi: 10.1146/annurev.micro.55.1.625
- Yale, J., and Bohnert, H. J. (2001). Transcript expression in *Saccharomyces cerevisiae* at high salinity. *J. Biol. Chem.* 276, 15996–16007. doi: 10.1074/jbc.M008209200
- Zajc, J., Kogej, T., Galinski, E. A., Ramos, J., and Gunde-Cimerman, N. (2014). Osmoadaptation strategy of the most halophilic fungus, *Wallemia ichthyophaga*, growing optimally at salinities above 15% NaCl. *Appl. Environ. Microbiol.* 80, 247–256. doi: 10.1128/AEM.02702-13
- Zajc, J., Liu, Y. F., Dai, W. K., Yang, Z. Y., Hu, J. Z., Gostinčar, C., et al. (2013). Genome and transcriptome sequencing of the halophilic fungus *Wallemia ichthyophaga*: haloadaptations present and absent. *BMC Genomics* 14:617. doi: 10.1186/1471-2164-14-617
- Zalar, P., Kocuvan, M. A., Plemenitaš, A., and Gunde-Cimerman, N. (2005a). Halophilic black yeasts colonize wood immersed in hypersaline water. *Bot. Mar.* 48, 323–326. doi: 10.1515/BOT.2005.042
- Zalar, P., Sybren de Hoog, G., Schroers, H. J., Frank, J. M., and Gunde-Cimerman, N. (2005b). Taxonomy and phylogeny of the xerophilic genus *Wallemia* (Walleiomycetes and Wallemiales, cl. et ord. nov.). *Anton. Leeuw.* 87, 311–328. doi: 10.1007/s10482-004-6783-x
- Zotova, L., Aleschko, M., Sponder, G., Baumgartner, R., Reipert, S., Prinz, M., et al. (2010). Novel components of an active mitochondrial K<sup>+</sup>/H<sup>+</sup> exchange. *J. Biol. Chem.* 285, 14399–14414. doi: 10.1074/jbc.M109.059956
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received:** 27 January 2014; **paper pending published:** 11 February 2014; **accepted:** 14 April 2014; **published online:** 05 May 2014.
- Citation:** Plemenitaš A, Lenassi M, Konte T, Kejžar A, Zajc J, Gostinčar C and Gunde-Cimerman N (2014) Adaptation to high salt concentrations in halotolerant/halophilic fungi: a molecular perspective. *Front. Microbiol.* 5:199. doi: 10.3389/fmicb.2014.00199
- This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.
- Copyright © 2014 Plemenitaš, Lenassi, Konte, Kejžar, Zajc, Gostinčar and Gunde-Cimerman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Polypliody in haloarchaea: advantages for growth and survival

**Karolin Zerulla and Jörg Soppa\***

Biocentre, Institute for Molecular Biosciences, Department of Biological Sciences, Goethe University Frankfurt, Frankfurt, Germany

**Edited by:**

R. Thane Papke, University of Connecticut, USA

**Reviewed by:**

Jonathan H. Badger, J. Craig Venter Institute, USA

James A. Coker, University of Maryland University College, USA

**\*Correspondence:**

Jörg Soppa, Biocentre, Institute for Molecular Biosciences, Department of Biological Sciences, Goethe University Frankfurt, Max-von-Laue-Strasse 9, D-60438 Frankfurt, Germany  
e-mail: soppa@bio.uni-frankfurt.de

The investigated haloarchaeal species, *Halobacterium salinarum*, *Haloferax mediterranei*, and *H. volcanii*, have all been shown to be polypliod. They contain several replicons that have independent copy number regulation, and most have a higher copy number during exponential growth phase than in stationary phase. The possible evolutionary advantages of polypliody for haloarchaea, most of which have experimental support for at least one species, are discussed. These advantages include a low mutation rate and high resistance toward X-ray irradiation and desiccation, which depend on homologous recombination. For *H. volcanii*, it has been shown that gene conversion operates in the absence of selection, which leads to the equalization of genome copies. On the other hand, selective forces might lead to heterozygous cells, which have been verified in the laboratory. Additional advantages of polypliody are survival over geological times in halite deposits as well as at extreme conditions on earth and at simulated Mars conditions. Recently, it was found that *H. volcanii* uses genomic DNA as genetic material and as a storage polymer for phosphate. In the absence of phosphate, *H. volcanii* dramatically decreases its genome copy number, thereby enabling cell multiplication, but diminishing the genetic advantages of polypliody. Stable storage of phosphate is proposed as an alternative driving force for the emergence of DNA in early evolution. Several additional potential advantages of polypliody are discussed that have not been addressed experimentally for haloarchaea. An outlook summarizes selected current trends and possible future developments.

**Keywords:** *Haloferax volcanii*, archaea, polypliody, gene conversion, desiccation, survival

## INTRODUCTION

Many species of eukaryotes are polypliod, and this is true for animals, plants, and lower unicellular eukaryotes. Last year, a special issue of Cytogenetic Genome Research focused on polypliody and assembled many excellent reviews focusing on different aspects of polypliody, including the mechanisms of its generation, the consequences for gene expression and genome biology, and the change of ploidy levels in both directions during evolution (e.g., Choleva and Janko, 2013; Wertheim et al., 2013; Hegarty et al., 2013; Weiss-Schneeweiss et al., 2013). Polypliody is even observed in human tissues, e.g., in the liver and in tumors (Duncan, 2013; Mayfield-Jones et al., 2013). While there has been tremendous advances in recent years, it is still not clear whether polypliody generally has a positive effect on the evolutionary success of a eukaryotic species (Madlung, 2013).

In stark contrast to eukaryotes, it has long been believed that prokaryotes (archaea and bacteria) are typically monoploid and contain a single copy of a circular chromosome, and this is still the current view of most reviews and textbooks (e.g., Madigan et al., 2012). Few exceptions were known and have been studied, e.g., the bacterium *Deinococcus radiodurans* that was isolated from irradiated meat and is highly resistant to X-ray irradiation and desiccation (Hansen, 1978). *D. radiodurans* has 5–8 copies of its chromosome and is thus “only” oligoploid (between 2 and 10 copies of the chromosome, more than 10 copies would be polypliod). Nevertheless,

*D. radiodurans* can quickly and efficiently regenerate complete chromosomes from overlapping fragments of severely scattered chromosomes, which is a process that involves DNA synthesis and homologous recombination (Zahradka et al., 2006; Slade et al., 2009). This would not be possible for a monoploid species, and thus, survival in DNA damaging conditions is an obvious evolutionary advantage of the oligoploid of *D. radiodurans*.

In recent years, results have accumulated showing that *D. radiodurans* and a few additional, long-known examples are by no means seldom and exotic exceptions, but that many species of archaea and bacteria are oligo- or polypliod (e.g., Komaki and Ishikawa, 2000; Breuert et al., 2006; Mendell et al., 2008; Michelsen et al., 2010; Tobiason and Seifert, 2010; Griesse et al., 2011; Hildenbrand et al., 2011; Pecoraro et al., 2011). Currently it seems that the opposite from the traditional view is correct, i.e., that monoploid species are a small minority among archaea and bacteria.

Several species of haloarchaea have also been shown to be polypliod in all growth phases and at various growth rates, e.g., *Halobacterium cutirubrum*, *Halobacterium salinarum*, *Haloferax volcanii* (review: Soppa, 2011), *H. mediterranei* (Zerulla and Soppa, unpublished results), and several new isolates (Zerulla and Soppa, unpublished results). Until now, no haloarchaeal species has been found to be monoploid; therefore, it might be that polypliody is a general trait of haloarchaea. Thus,

haloarchaea seem to be a suitable group of prokaryotes for studying polyploidy, its evolutionary advantages, and the molecular mechanisms of its regulation. Nine different possible evolutionary advantages of polyploidy for haloarchaea have recently been discussed (Soppa, 2013). For five of the advantages, experimental evidence had been published for haloarchaea at that time, and the remaining four were only theoretical considerations. In this contribution, we do not aim to reiterate the recent review, but to only shortly mention previously discussed angles and to focus on new results reported since this time, e.g., survival under extreme conditions on earth and perhaps elsewhere, the absence of an S-phase in the haloarchaeal cell cycle, the non-genetic role of genomic DNA as a phosphate storage polymer, and the possible driving force for the development of polyploidy in evolution.

### EVOLUTIONARY ADVANTAGES BASED ON HOMOLOGOUS RECOMBINATION

As mentioned above, it was shown long ago that the oligoploid bacterium *D. radiodurans* has a much higher resistance to DNA-damaging conditions than other bacteria with only one copy of their chromosome. Similarly, the polyploid haloarchaeon *H. salinarum* has been shown to be highly resistant to conditions that induce double strand breaks, i.e., X-ray irradiation and desiccation (Kottemann et al., 2005). Also, in this case, the chromosomes were scattered into many fragments, and complete chromosomes were re-generated by making use of overlapping fragments. In a further study, *H. salinarum* was challenged with increasing doses of X-ray irradiation, resulting in the selection of a mutant that had an even more enhanced irradiation resistance and was, in fact, believed to exhibit the highest resistance of all organisms on earth (DeVeaux et al., 2007). In the mutant, the expression of two genes encoding single strand DNA-binding proteins (*Ssb*) was highly up-regulated, indicating that the survival of DNA-damaging conditions involves homologous recombination in *H. salinarum*, like in *D. radiodurans*. In accordance with this view, overexpression of the homologous *rpaC* gene in *H. volcanii* resulted in an enhanced resistance against various DNA-damaging conditions (Skowyra and MacNeill, 2012). Based on these studies with haloarchaea, the role of the *Ssb* for the radiation resistance of *D. radiodurans* was recently investigated, and it was found that the *ssb* gene is not only essential, but furthermore, that the expression level is directly correlated with the degree of radiation resistance (Lockhart and DeVeaux, 2013).

It seems that polyploidy does not only confer resistance to DNA damaging agents but also results in a low rate of spontaneous mutations. The mutation rate of *H. volcanii* has been quantified in a genetic screen using the *pyrE* gene as a reporter, and it was found to be nearly one order of magnitude lower than that of other comparable mesophilic species (Mackwan et al., 2007). The authors proposed that the low mutation rate might be based on the polyploidy of *H. volcanii*, which enables the repair of mutated copies of the chromosome by making use of the information of wild-type copies that are simultaneously present in the cell.

The repair and the induction of mutations might be of evolutionary advantage. The presence of many copies of any given gene,

termed gene redundancy, allows for the mutation of some of the copies without losing the wild-type information of the remaining copies. This generates heterozygous cells, which might be able to grow under unfavorable conditions that inhibit growth of the homozygous wild-type. Several reports about heterozygous cells in specific laboratory settings are available. For example, the selection of heterozygous *H. volcanii* cells that can grow in the absence of leucine and tryptophan, which is impossible for homozygous mutants, has been described (Lange et al., 2011). The presence of heterozygous cells under specific selection conditions has also been described for a methanogenic archaeon (Hildenbrand et al., 2011) as well as for several cyanobacteria (e.g., Spence et al., 2004; Takahama et al., 2004; Nodop et al., 2008). The ease of selection of heterozygous cells in polyploid species of diverse phylogenetic groups suggests that such cells can also arise in natural populations under appropriate selection conditions.

An alternative potential mechanism for the formation of heterozygous cells is the fusion of two cells with non-identical genomes. For *H. volcanii*, it has been shown that genetic transfer between different auxotrophic cells is possible and involves cytoplasmic bridges and most likely the fusion of cells (Mevarech and Werczberger, 1985; Rosenshine et al., 1989). For a population of *Halorubrum* growing in a saltern, it was revealed that the cells exchanged genetic information promiscuously and that the linkage equilibrium was extremely low and, in fact, approached that of a sexual population (Papke et al., 2004). This indicates that heterozygous cells also form in nature, at least for a while. Recently, it was shown that even cells of two different species could fuse (Naor et al., 2012). The resulting heterozygous cells were not stable, but cells with identical genomes emerged that had integrated between 310 and 530 kbp of the genome of the other species into the main genome. Natural populations of the genus *Halorubrum* in two salterns and one salt lake have been characterized, and it has been found that clusters can be formed but that barriers to genetic exchange between the different “species” is leaky, indicating that cell fusion between cells of different “species” might also occur in nature (Papke et al., 2007). This is most likely not confined to haloarchaea. Recently, it has been discussed that lateral transfer between different species of prokaryotes has occurred so massively that a tree-like reconstruction of phylogeny does not adequately describe what has happened in evolution, and a network-like reconstruction is more appropriate (Dagan, 2011; Dagan and Martin, 2009). Of course genetic exchange is not only possible via cell fusion but also through other mechanisms that require the intermediate formation of cells that are heterozygous at least for some of their genes.

### INTERMOLECULAR GENE CONVERSION IN THE ABSENCE OF SELECTION

The repair of damaged or mutated copies of the chromosome using the wild-type information of other copies would require the un-reciprocal, intermolecular transfer of information from a donor to an acceptor molecule, a mechanism termed gene conversion. It has been shown that gene conversion indeed operates in *H. volcanii* and leads to the equalization of genome

copies in the absence of selection (Lange et al., 2011). Shortly, a heterozygous strain was constructed that simultaneously contained two different types of chromosomes that had either the *leuB* or *trpA* gene at the *leuB* locus. Selection for the presence of either of the two amino acids led to an equalization of genomes in the direction of the respective essential gene, while the genomes lost the information of the other gene. Most importantly, gene conversion also led to an equalization of genome copies in the absence of any selection. In addition, gene conversion occurred in the predicted direction, which required a smaller amount of DNA synthesis than the other direction. The experiment is schematically shown in **Figure 1**. Equalization of genome copies in the absence of selection has also been shown to operate in methanogenic Archaea (Hildenbrand et al., 2011). These are the only two studies that concentrated on intermolecular gene conversion in prokaryotes, and one additional study verified the existence of intermolecular gene conversion in chloroplasts (Khakhlova and Bock, 2006). Many more studies are available that concentrate on intramolecular gene conversion, which results in the concerted evolution of gene families and is a mechanism of antigenic variation or phase variation (Santoyo

and Romero, 2005; Palmer and Brayton, 2007). Even if intergenic gene conversion in polyploid prokaryotes is a neglected field of research, it can be predicted to occur in more bacterial and archaeal species than in *H. volcanii* and *Methanococcus maripaludis* and result in evolutionary advantages of polyploidy. In addition, it is an escape from “Muller’s ratchet” theory, which predicted that asexual polyploid species cannot exist because they would accumulate deleterious mutations (Muller, 1964; Lehman, 2003).

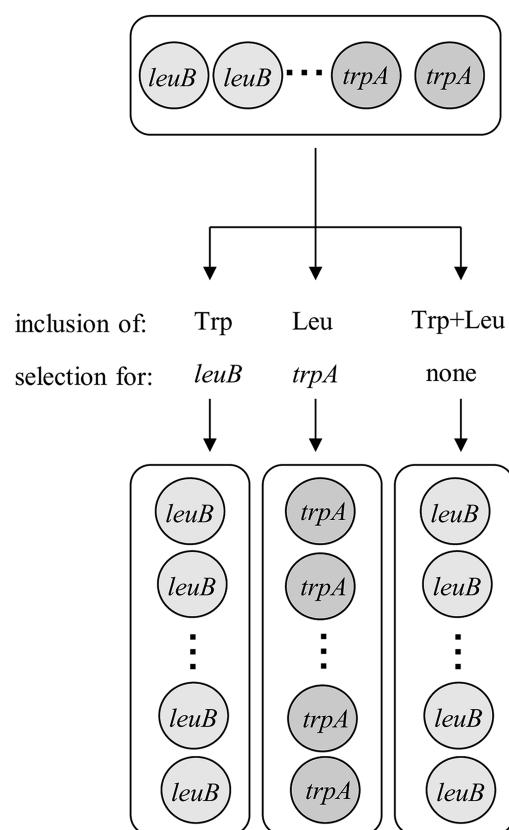
## LONG-TERM SURVIVAL AND SURVIVAL IN EXTREME ENVIRONMENTS

It has repeatedly been reported that haloarchaea have been isolated from ancient salt deposits that had remained undisturbed for geological times (e.g., Fendrihan et al., 2006; Park et al., 2009; Schubert et al., 2010; Gramain et al., 2011; Xiao et al., 2013). The salt deposits are in different parts of the world and have different ages, and the isolations were performed by several different research groups. Nevertheless, whether haloarchaea can survive for more than 100 000 years in salt deposits has been the matter of intense debate, and a counterargument has always been that the chemical stability of DNA is too small to allow its survival in an intact form over geological times (a good review of the different arguments is given by Grant et al., 1998). However, this counterargument does not hold true for polyploid species. As has been experimentally proven for *H. salinarum*, polyploid cells can regenerate intact chromosomes from scattered fragments (Kottemann et al., 2005). In liquid enclosures within halite crystals, it can be expected that the DNA damage will be small because irradiation and oxidative stress are absent. The maintenance energy for long-term survival, including the repair of double-strand breaks, might originate from lysis of some or even the majority of cells of the enclosed population. For *Escherichia coli*, it has indeed been shown that cell lysis promotes growth of the remaining population (Corchero et al., 2001). Recently, three haloarchaeal species have been freshly isolated from an ancient salt deposit, and their ploidy levels have been quantified. All three were shown to be polyploid, in agreement with the prediction that polyploidy enables long-term survival (Jaakkola et al., submitted for publication).

It has recently been observed that haloarchaea rapidly change their morphology upon exposure to conditions of low water activity and form 3–4 small spheres from one rod-like cell (Fendrihan et al., 2012). These spheres could outgrow to normal rods in favorable conditions. It could be that diminishing the surface to volume ratio is part of the strategy for long-term survival. Furthermore, roundish particles had indeed been observed in fluid inclusions of old halite crystals (Schubert et al., 2010). Notably, such a strategy would be impossible for monoploid species because only one of the 3–4 spheres would obtain a copy of the chromosome. The presence of a potassium pump has been discussed as another aspect of the long-term survival strategy of haloarchaea (Kixmüller and Greie, 2012).

Haloarchaea have not only been claimed to survive geological times enclosed in ancient salt deposits but also been shown to survive at extreme places on earth, e.g., places with extremely low water activity (e.g., Wierzchos et al., 2012). It has been discussed

*Haloferax volcanii*



**FIGURE 1 |** Schematic overview of an experiment verifying that gene conversion operates in the absence of selection in *Haloferax volcanii* (changed version of Figure 1 in Soppa, 2013).

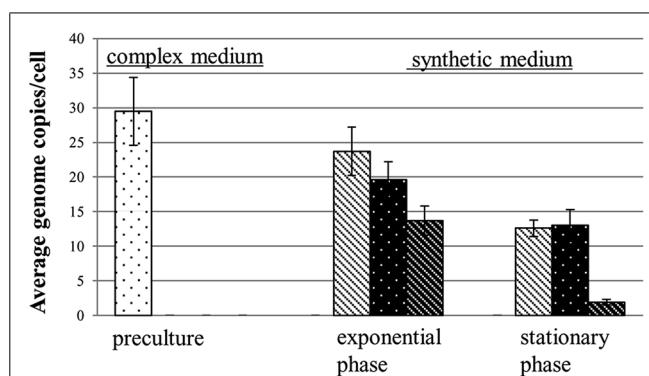
that some of these places are somewhat reminiscent of the conditions on Mars, and it has been proposed that haloarchaea could possibly survive on Mars. Therefore, *Halococcus dombrowskii* was exposed to simulated Martian UV irradiation, and the survival rate was found to be high when the cells were in fluid inclusions in halite crystals (Fendrihan et al., 2009). Notably, a considerable fraction of *Halococcus* sp. cells survived two weeks in space (Mancinelli and Klovstad, 2000), indicating that even extraterrestrial travel, e.g., on meteorites, might be possible for haloarchaea. Even if life on Mars is currently a fantastic idea, it induces experiments showing under which extreme conditions haloarchaea can survive – on earth.

### RELAXED REPLICATION CONTROL

Typically the cell cycle is highly regulated, and cell cycle checkpoints guarantee that its progression is stopped when problems occur, e.g., when a septum is not formed and the cell does not divide before replication is completed. DNA synthesis is usually confined to the so-called S-phase of the cell cycle, which is situated between the G1 and the G2 phases. Synchronized cultures of *H. salinarum* have been used to study the progression of cell cycle events, e.g., cell cycle-specific cyclic transcript level changes (Baumann et al., 2007). Unexpectedly, pulse labeling of newly synthesized DNA with 5-bromo-2'-deoxyuridine revealed that *H. salinarum* does not have an S-phase, but that DNA synthesis is constitutive (compare Figure 6 in Zerulla et al., 2014a). This loss of temporal replication regulation is not accompanied by a general relaxation of cell cycle control, e.g., the inhibition of replication results in a total stop of cell division, although the cell has approximately 30 genome copies and could easily divide several times without DNA synthesis (Herrmann and Soppa, 2002). In addition, DNA segregation control is intact, and the two daughter cells obtain equal amounts of DNA (Breuer et al., 2006), in contrast to other polyploid species (see below).

### GENOMIC DNA AS A PHOSPHATE STORAGE POLYMER

Most of the genetic advantages of polyploidy discussed above require the presence of homologous recombination. Recently, an additional evolutionary advantage of polyploidy has been described that is independent of homologous recombination, namely the usage of genomic DNA as a storage polymer for phosphate (Zerulla et al., 2014b). The study initially aimed at clarifying whether *H. volcanii* can use external environmental genomic DNA as “food.” It could indeed be shown that *H. volcanii* can use external genomic DNA as a source for carbon, nitrogen, and phosphate (Chimileski et al., 2014; Zerulla et al., 2014a). However, the negative control lacking any external source of phosphate revealed that *H. volcanii* could grow to a limited extent under this condition and, thus, must have an intracellular phosphate storage pool. Quantification of the chromosomes showed that their copy number was dramatically decreased from approximately 30 to only 2 during growth in the absence of an external phosphate source (Figure 2). The cell number increased “only” 8.4-fold in the absence of phosphate, thus the decrease in chromosome copy number was higher than the increase in cell number. This indicates that *H. volcanii* used genomic DNA as a phosphate storage polymer in two different ways: (1) the cells could divide approximately three times in the absence of external phosphate, which is

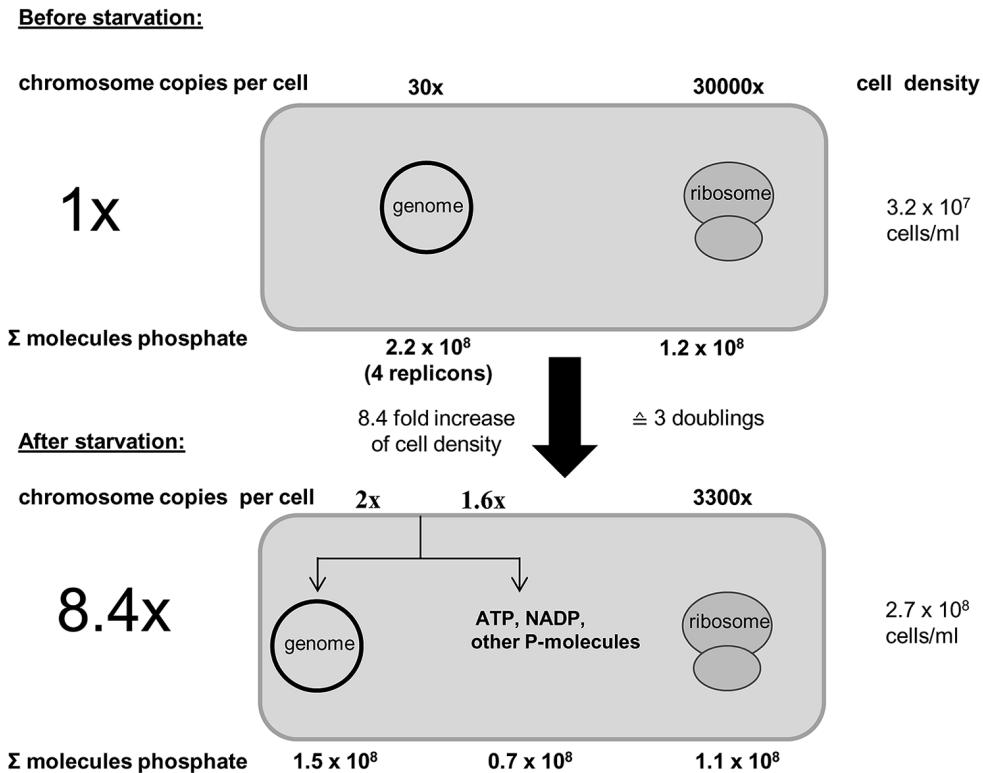


**FIGURE 2 | Chromosome copy numbers of *H. volcanii* cells grown in synthetic media with two different phosphate concentrations and in the absence of external phosphate.** The copy numbers were quantified during mid-exponential growth phase and at stationary phase, as indicated (taken from Zerulla et al., 2014b).

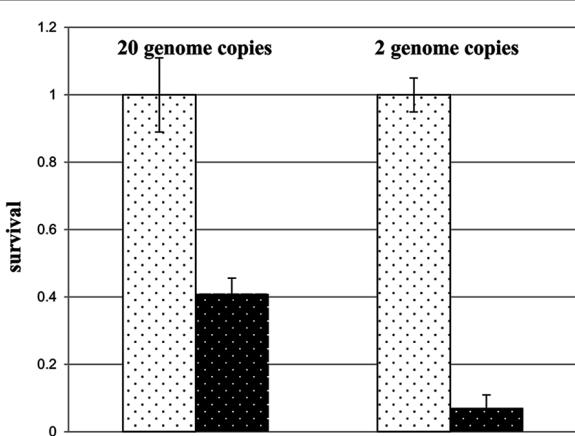
not possible for monoploid cells devoid of an alternative phosphate storage polymer such as polyphosphate, and (2) 1/3 of the genome copies were degraded to liberate phosphate for other phosphate-containing biomolecules lacking an intracellular storage pool, e.g., ATP, NADP<sup>+</sup>, phospholipids, phosphoproteins, and phosphosugars. The high copy number of *H. volcanii* implicates that one of the many biological functions of polyploidy is phosphate storage. The genetic advantages of polyploidy do not seem to require a copy number of 30, e.g., *D. radiodurans* has an extremely high resistance against irradiation and desiccation but harbors only eight copies of its chromosome. Re-addition of phosphate to phosphate-starved cells of *H. volcanii* led to fast and intense DNA synthesis and the copy number increased from 2 to 20 within 7 h.

Further analyses revealed that genomic DNA seems to be the only intracellular phosphate storage polymer for *H. volcanii*. The number of ribosomes decreased in accordance with the increase in cell number, and thus, ribosomes were distributed to offspring cells, but not degraded to liberate phosphate, and no indication for the presence of polyphosphate could be found. Figure 3 schematically summarizes the phosphate balance prior to and after growth of *H. volcanii* in the absence of phosphate.

In the first paragraphs of this contribution, a variety of genetic advantages of polyploidy were discussed, some or all of which might also apply to *H. volcanii*. The reduction of the chromosome copy number from 30 to 2 during growth in the absence of phosphate led to the prediction that concomitant genetic advantages should be diminished. This was indeed found to be the case. Comparison of the desiccation resistance of cells containing 20 and 2 copies of the chromosome revealed that the resistance of the former was fivefold higher than that of the latter (Figure 4). Therefore, if environmental conditions force *H. volcanii* to “choose” between several different advantages of polyploidy, genetic advantages are diminished in favor of cell multiplication. It will be interesting to reveal whether additional or all polyploid species also use this strategy, or whether other species abandon cell division in the absence of phosphate and retain the genetic advantages of polyploidy.



**FIGURE 3 | Schematic overview of the phosphate balance of *H. volcanii* prior to and after growth in the absence of external phosphate.** It is taken into account that the cell number increased 8.4-fold during growth in the absence of phosphate. The number of chromosomes and ribosomes were quantified (taken from Zerulla et al., 2014b).



**FIGURE 4 | Survival rates of *H. volcanii* cultures that entered a 12 day desiccation challenge with 20 chromosomes or with 2 chromosomes, as indicated (taken from Zerulla et al., 2014b).**

## THE POSSIBLE DRIVING FORCE FOR THE EMERGENCE OF DNA IN EVOLUTION

Several different and conflicting theories about the origin of life exist, including an autotrophic origin driven by natural gradients at geothermal vents and a heterotrophic origin in a “primordial

soup” (Miller et al., 1997; Wächtershäuser, 1997; Lane et al., 2010; Neveu et al., 2013). However, all theories are in agreement that RNA predated DNA and that, at some point, free-living cellular life forms existed that had RNA as their genetic material. Much later, DNA was introduced into cellular life and replaced RNA as genetic material. The current view is that the selective driving force for the emergence of DNA was its higher stability in comparison with RNA, leading to a reduction in mutation rates and enhanced survival of cells with DNA-based genomes, which would rapidly outgrow cells with RNA-based genomes. Recently, an alternative explanation has been proposed that is based on the observation that *H. volcanii* uses DNA as a phosphate storage polymer in addition to its role as genetic material (Zerulla et al., 2014b). The alternative hypothesis is that, during the pre-DNA era, cells could grow and live with their RNA-based genomes, but, during phosphate limitation, growth was impossible in the absence of a phosphate-storage polymer. The driving force for the emergence of DNA would, in this view, have been its much higher stability compared to alternative phosphate storage polymers, e.g., polyphosphate, which is rather unstable. Therefore, the first “polyploid” cells would have had a high content of DNA without using it as genetic material. Only later would DNA have evolved its additional function as a genetic information storage polymer and then be selected for by its higher stability compared to RNA.

## POSSIBLE ADDITIONAL ADVANTAGES NOT EXPERIMENTALLY VERIFIED FOR HALOARCHAEA

Haloarchaea typically contain several large replicons, which are either regarded as several different chromosomes or as one chromosome and one or more mega-plasmids. It has been argued that *H. volcanii* contains four different chromosomes and one small plasmid because the former replicons all contain a replication origin composed of a repeated set of a sequence motif that is bordered by a gene encoding an Origin Recognition Protein (Hartman et al., 2010). The copy numbers of these replicons are not identical; therefore, the gene dosage of a given gene depends on its localization on a specific replicon, and the gene dosages can vary by more than a factor of four (Zerulla et al., 2014a). In addition, the copy numbers of the different replicons is typically growth phase-dependent, and the numbers are smaller in stationary phase than in exponential growth phase. Furthermore, growth phase-dependent copy numbers are independently regulated for the different replicons, and this has been shown for *H. volcanii* (Zerulla et al., 2014a), *H. mediterranei* (Liu et al., 2013), and *H. salinarum* (Breuert et al., 2006). The independent differential regulation of gene dosages in response to growth phase or varying environmental conditions opens the possibility that haloarchaea apply the regulation of replicon copy number for the global regulation of gene expression. However, until now, it was unclear whether gene expression in haloarchaea was correlated in a systematic manner with gene dosage. Only for one gene, the dihydrofolate reductase (*dhfr*) gene, is it clear that a higher gene dosage results in a higher expression level (Zusman et al., 1989). It is readily possible to select mutants of *H. volcanii* that are resistant to trimethoprim, a competitive inhibitor of DHFR. Many of these mutants carry amplifications of the genome region including the *dhfr* gene, and this was initially used to clone the gene.

Polyplloid species might relax the strict control of DNA segregation and septum localization, without the danger of forming DNA-less daughter cells that waste the energy used for their production. It has in fact been shown that *Methanocaldococcus jannaschii* divides and generates daughter cells of different sizes with different DNA contents (Malandrin et al., 1999). However, relaxation of segregation control and cell division has not been reported for any haloarchaeal species.

A further attractive, yet untested, hypothesis is that the regulation of gene expression is different in monoploid and polyplloid species. Monoploid prokaryotes contain just a single copy of a promoter of a typical gene. The copy number of transcription factors is also usually low. Therefore, regulation of gene expression depends on a stochastic sequence of on- and off-states of genes over time that is based on the off-rate and on-rate of the respective transcription factor at the respective promoter. This leads to differential protein compositions of cells of identical genotypes, and, indeed, single-cell analyses have revealed that populations of prokaryotes exhibit phenotypic variations (e.g., de Jong et al., 2012). In contrast, polyplloid species similar to haloarchaea contain 20–30 copies of each promoter, and it is tempting to speculate that the copy numbers of transcription factors might be considerably higher than that in monoploid species. If this would be true, the regulation of gene expression in polyplloid species would

follow statistics rather than stochastics, and populations would be more uniform.

A further advantage that has not been verified for haloarchaea is the enlargement of cell size to the formation of “giant cells.” Giant cells escape predators that are specialized to feed on normal-sized prokaryotes. One example of a giant bacterium is *Epulopiscium*, which can reach a length of up to 600 μm. The cells are highly polyploid and have genome copy numbers of up to 10s of 1000s. It has been argued that cell enlargement depends on polyploidy because diffusion would be much too slow to distribute transcripts from a single genome throughout the volume of the cell (Mendell et al., 2008). Giant bacteria are found in several phylogenetic groups, and it has been described that all large bacterial species examined thus far are highly polyploid (Angert, 2012).

## OUTLOOK

It is now firmly established that several haloarchaeal species are polyploid and that polyploidy might even be a general trait of haloarchaea. However, genomes have been shown to be extremely flexible in evolution (Oliverio and Katz, 2014), and thus, many more haloarchaeal species must be tested before a generalization is possible. The molecular mechanisms of genome copy number regulation have not been unraveled. The major chromosome of *H. volcanii* contains four origins of replication, all of which can be deleted without loosing viability (Hawkins et al., 2013). This opens the possibility to construct small replicons as “haloarchaeal artificial chromosomes,” which carry a selected origin of replication that can be easily manipulated and used to study copy number regulation. In addition, it would be desirable to isolate isogenic strains with varying chromosome copy numbers, which would allow for systematically analyzing the advantages of polyploidy via the comparison of strains that are identical apart from their ploidy level. Furthermore, gene conversion has been primarily studied using eukaryotes (>90% of the literature) and only poorly studied with bacteria, and fewer than five studies have been performed with archaea (Lawson et al., 2009). Moreover, intermolecular gene conversion has only been verified for two prokaryotic species. Thus, haloarchaea might help in the initial understanding of the importance and mechanism of intermolecular gene conversion in prokaryotes. In addition, haloarchaea might tremendously contribute to unraveling the importance of polyploidy for survival mechanisms over geological times, at very extreme environments on earth, or even on other planets.

## ACKNOWLEDGMENT

The ploidy projects in my group were funded by the German Research Council (DFG) through grant So264/16.

## REFERENCES

- Angert, E. R. (2012). DNA replication and genomic architecture of very large bacteria. *Annu. Rev. Microbiol.* 66, 197–212. doi: 10.1146/annurev-micro-090110-102827
- Baumann, A., Lange, C., and Soppa, J. (2007). Transcriptome changes and cAMP oscillations in an archaeal cell cycle. *BMC Cell Biol.* 8:21. doi: 10.1186/1471-2121-8-21
- Breuert, S., Allers, T., Spohn, G., and Soppa, J. (2006). Regulated polyploidy in halophilic archaea. *PLoS ONE* 1:e92. doi: 10.1371/journal.pone.0000092

- Chimileski, S., Dolas, K., Naor, A., Gophna, U., and Papke, R. T. (2014). Extracellular DNA metabolism in *Haloferax volcanii*. *Front. Microbiol.* 5:57. doi: 10.3389/fmicb.2014.00057.eCollection 2014
- Choleva, L., and Janko, K. (2013). Rise and persistence of animal polyploidy: evolutionary constraints and potential. *Cytogenet. Genome Res.* 140, 151–170. doi: 10.1159/000353464
- Corchero, J. L., Cubarsi, R., Vila, P., Aris, A., and Villaverde, A. (2001). Cell lysis in *Escherichia coli* cultures stimulates growth and biosynthesis of recombinant proteins in surviving cells. *Microbiol. Res.* 156, 13–18. doi: 10.1078/0944-5013-00066
- Dagan, T. (2011). Phylogenomic networks. *Trends Microbiol.* 19, 483–491. doi: 10.1016/j.tim.2011.07.001
- Dagan, T., and Martin, W. (2009). Getting a better picture of microbial evolution en route to a network of genomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2187–2196. doi: 10.1098/rstb.2009.0040
- de Jong, I. G., Veening, J. W., and Kuipers, O. P. (2012). Single cell analysis of gene expression patterns during carbon starvation in *Bacillus subtilis* reveals large phenotypic variations. *Environ. Microbiol.* 14, 3110–3121. doi: 10.1111/j.1462-2920.2012.02892.x
- DeVeaux, L. C., Muller, J. A., Smith, J., Petrisko, J., Wells, D. P., and DasSarma, S. (2007). Extremely radiation-resistant mutants of a halophilic archaeon with increased single-stranded DNA-binding protein (RPA) gene expression. *Radiat. Res.* 168, 507–514. doi: 10.1667/RR0935.1
- Duncan, A. W. (2013). Aneuploidy, polyploidy and ploidy reversal in the liver. *Semin. Cell. Dev. Biol.* 24, 347–356. doi: 10.1016/j.semcdb.2013.01.003
- Fendrihan, S., Berces, A., Lammer, H., Musso, M., Ronto, G., Polacsek, T. K., et al. (2009). Inverstigating the effects of simulated Martian ultraviolet radiation on *Halococcus dombrowskii* and other extremely halophilic archaeabacteria. *Astrobiology* 9, 104–112. doi: 10.1089/ast.2007.0234
- Fendrihan, S., Dornmayr-Pfaffenhuemer, M., Gerbl, F. W., Holzinger, A., Grosbacher, M., Briza, P., et al. (2012). Spherical particles of halophilic archaea correlate with exposure to low water activity—implications for microbial survival in fluid inclusions of ancient halite. *Geobiology* 10, 424–433. doi: 10.1111/j.1472-4669.2012.00337.x
- Fendrihan, S., Legat, A., Pfaffenhuemer, M., Gruber, C., Weidler, G., Gerbl, F., et al. (2006). Extremely halophilic archaea and the issue of long-term microbial survival. *Rev. Environ. Sci. Biotechnol.* 5, 203–218. doi: 10.1007/s11157-006-0007-y
- Gramain, A., Diaz, G. C., Demergasso, C., Lowenstein, T. K., and McGenity, T. J. (2011). Archaeal diversity along a subterranean salt core from the Salar Grande (Chile). *Environ. Microbiol.* 13, 2105–2121. doi: 10.1111/j.1462-2920.2011.02435.x
- Grant, W. D., Gemmell, R. T., and McGenity, T. J. (1998). Halobacteria: the evidence for longevity. *Extremophiles* 2, 279–287. doi: 10.1007/s007920050070
- Griese, M., Lange, C., and Soppa, J. (2011). Ploidy in cyanobacteria. *FEMS Microbiol. Lett.* 323, 124–131. doi: 10.1111/j.1574-6968.2011.02368.x
- Hansen, M. T. (1978). Multiplicity of genome equivalents in the radiation-resistant bacterium *Micrococcus radiodurans*. *J. Bacteriol.* 134, 71–75.
- Hartman, A. L., Norais, C., Badger, J. H., Delmas, S., Haldenby, S., Madupu, R., et al. (2010). The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS ONE* 5:e9605. doi: 10.1371/journal.pone.0009605
- Hawkins, M., Malla, S., Blythe, M. J., Nieduszynski, C. A., and Allers, T. (2013). Accelerated growth in the absence of DNA replication origins. *Nature* 503, 544–547. doi: 10.1038/nature12650
- Hegarty, M., Coate, J., Sherman-Broyles, S., Abbott, R., Hiscock, S., and Doyle, J. (2013). Lessons from natural and artificial polyploids in higher plants. *Cytogenet. Genome Res.* 140, 204–225. doi: 10.1159/000353361
- Herrmann, U., and Soppa, J. (2002). Cell cycle-dependent expression of an essential SMC-like protein and dynamic chromosome localization in the archaeon *Halobacterium salinarum*. *Mol. Microbiol.* 46, 395–409. doi: 10.1046/j.1365-2958.2002.03181.x
- Hildenbrand, C., Stock, T., Lange, C., Rother, M., and Soppa, J. (2011). Genome copy numbers and gene conversion in methanogenic archaea. *J. Bacteriol.* 193, 734–743. doi: 10.1128/JB.01016-10
- Khakhlova, O., and Bock, R. (2006). Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J.* 46, 85–94. doi: 10.1111/j.1365-313X.2006.02673.x
- Kixmüller, D., and Greie, J. C. (2012). An ATP-driven potassium pump promotes long-term survival of *Halobacterium salinarum* within salt crystals. *Environ. Microbiol. Rep.* 4, 234–241. doi: 10.1111/j.1758-2229.2012.00326.x
- Komaki, K., and Ishikawa, H. (2000). Genomic copy number of intracellular bacterial symbionts of aphids varies in response to developmental stage and morph of their host. *Insect. Biochem. Mol. Biol.* 30, 253–258. doi: 10.1016/S0965-1748(99)00125-3
- Kottmann, M., Kish, A., Ilaoanusi, C., Bjork, S., and DiRuggiero, J. (2005). Physiological responses of the halophilic archaeon *Halobacterium sp. strain NRC1* to desiccation and gamma irradiation. *Extremophiles* 9, 219–227. doi: 10.1007/s00792-005-0437-4
- Lane, N., Allen, J. E., and Martin, W. (2010). How did LUCA make a living? Chemiosmosis in the origin of life. *Bioessays* 32, 271–280. doi: 10.1002/bies.200900131
- Lange, C., Zerulla, K., Breuert, S., and Soppa, J. (2011). Gene conversion results in the equalization of genome copies in the polyploid haloarchaeon *Haloferax volcanii*. *Mol. Microbiol.* 80, 666–677. doi: 10.1111/j.1365-2958.2011.07600.x
- Lawson, M. J., Jiao, J., Fan, W., and Zhang, L. (2009). A pattern analysis of gene conversion literature. *Comp. Funct. Genomics.* 2009, 1–11. doi: 10.1155/2009/761512
- Lehman, N. (2003). A case for the extreme antiquity of recombination. *J. Mol. Evol.* 56, 770–777. doi: 10.1007/s00239-003-2454-1
- Liu, X., Miao, D., Zhang, F., Wu, Z., Liu, J., and Xiang, H. (2013). Characterization of the minimal replicon of pHM300 and independent copy number control of major and minor chromosomes of *Haloferax mediterranei*. *FEMS Microbiol. Lett.* 339, 66–74. doi: 10.1111/1574-6968.12052
- Lockhart, J. S., and DeVeaux, L. C. (2013). The essential role of the Deinococcus radiodurans ssb gene in cell survival and radiation tolerance. *PLoS ONE* 8:e71651. doi: 10.1371/journal.pone.0071651
- Mackwan, R. R., Carver, G. T., Drake, J. W., and Grogan, D. W. (2007). An unusual pattern of spontaneous mutations recovered in the halophilic archaeon *Haloferax volcanii*. *Genetics* 176, 697–702. doi: 10.1534/genetics.106.069666
- Madigan, M., Martinko, J., Stahl, D., and Clark, D. (2012). *Brock Biology of Microorganisms*. San Francisco, CA: Pearson Education, Inc.
- Madlung, A. (2013). Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity (Edinb)* 110, 99–104. doi: 10.1038/hdy.2012.79
- Malandrin, L., Huber, H., and Bernander, R. (1999). Nucleoid structure and partition in *Methanococcus jannaschii*: an archaeon with multiple copies of the chromosome. *Genetics* 152, 1315–1323.
- Mancinelli, R. L., and Klovstad, M. (2000). Martian soil and UV irradiation: microbial viability assessment on spacecraft surfaces. *Plant. Space Sci.* 48, 1093–1097. doi: 10.1016/S0032-0633(00)00083-0
- Mayfield-Jones, D., Washburn, J. D., Arias, T., Edger, P. P., Pires, J. C., and Conant, G. C. (2013). Watching the grin fade: tracing the effects of polyploidy on different evolutionary time scales. *Semin. Cell Dev. Biol.* 24, 320–331. doi: 10.1016/j.semcdb.2013.02.002
- Mendell, J. E., Clements, K. D., Choat, J. H., and Angert, E. R. (2008). Extreme polyploidy in a large bacterium. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6730–6734. doi: 10.1073/pnas.0707522105
- Mevarech, M., and Werczberger, R. (1985). Genetic transfer in *Halobacterium volcanii*. *J. Bacteriol.* 162, 461–462.
- Michelsen, O., Hansen, F. G., Albrechtsen, B., and Jensen, P. R. (2010). The MG1363 and IL1403 laboratory strains of *Lactococcus lactis* and several dairy strains are diploid. *J. Bacteriol.* 192, 1058–1065. doi: 10.1128/JB.00900-09
- Miller, S. L., Schopf, J. W., and Lazcano, A. (1997). Oparin's "Origin of Life": sixty years later. *J. Mol. Evol.* 44, 351–353. doi: 10.1007/PL00006153
- Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutat. Res.* 106, 2–9. doi: 10.1016/0027-5107(64)90047-8
- Naor, A., Lapierre, P., Mevarech, M., Papke, R. T., and Gophna, U. (2012). Low species barriers in halophilic archaea and the formation of recombinant hybrids. *Curr. Biol.* 22, 1444–1448. doi: 10.1016/j.cub.2012.05.056
- Neveu, M., Kim, H. J., and Benner, S. A. (2013). The "strong" RNA world hypothesis: fifty years old. *Astrobiology* 13, 391–403. doi: 10.1089/ast.2012.0868
- Nodop, A., Pietsch, D., Hocker, R., Becker, A., Pistorius, E. K., Forchhammer, K., et al. (2008). Transcript profiling reveals new insights into the acclimation of the mesophilic fresh-water cyanobacterium *Synechococcus elongatus* PCC 7942 to iron starvation. *Plant Physiol.* 147, 747–763. doi: 10.1104/pp.107.114058

- Oliverio, A. M., and Katz, L. A. (2014). The dynamic nature of genomes across the tree of life. *Genome Biol. Evol.* 6, 482–488. doi: 10.1093/gbe/evu024
- Palmer, G. H., and Brayton, K. A. (2007). Gene conversion is a convergent strategy for pathogen antigenic variation. *Trends Parasitol.* 23, 408–413. doi: 10.1016/j.pt.2007.07.008
- Papke, R. T., Koenig, J. E., Rodriguez-Valera, F., and Doolittle, W. F. (2004). Frequent recombination in a saltern population of *Halorubrum*. *Science* 306, 1928–1929. doi: 10.1126/science.1103289
- Papke, R. T., Zhaxybayeva, O., Feil, E. J., Sommerfeld, K., Muise, D., and Doolittle, W. F. (2007). Searching for species in haloarchaea. *Proc. Natl. Acad. Sci. U.S.A.* 104, 14092–14097. doi: 10.1073/pnas.0706358104
- Park, J. S., Vreeland, R. H., Cho, B. C., Lowenstein, T. K., Timofeef, M. N., and Rosenzweig, W. D. (2009). Haloarchaeal diversity in 23, 121 and 419 MYA salts. *Geobiology* 7, 515–523. doi: 10.1111/j.1472-4669.2009.00218.x
- Pecoraro, V., Zerulla, K., Lange, C., and Soppa, J. (2011). Quantification of ploidy in proteobacteria revealed the existence of monoploid, (mero-)oligoploid and polyploid species. *PLoS ONE* 6:e16392. doi: 10.1371/journal.pone.0016392
- Rosenshine, I., Tchelet, R., and Mevarech, M. (1989). The mechanism of DNA transfer in the mating system of an archaeabacterium. *Science* 245, 1387–1389. doi: 10.1126/science.2818746
- Santoyo, G., and Romero, D. (2005). Gene conversion and concerted evolution in bacterial genomes. *FEMS Microbiol. Rev.* 29, 169–183. doi: 10.1016/j.femsre.2004.10.004
- Schubert, B. A., Lowenstein, T. K., Timofeef, M. N., and Parker, M. A. (2010). Halophilic Archaea cultured from ancient halite, Death Valley, California. *Environ. Microbiol.* 12, 440–454. doi: 10.1111/j.1462-2920.2009.02086.x
- Skowrya, A., and MacNeill, S. A. (2012). Identification of essential and non-essential single-stranded DNA-binding proteins in a model archaeal organism. *Nucleic Acids Res.* 40, 1077–1090. doi: 10.1093/nar/gkr838
- Slade, D., Lindner, A. B., Paul, G., and Radman, M. (2009). Recombination and replication in DNA repair of heavily irradiated *Deinococcus radiodurans*. *Cell* 136, 1044–1055. doi: 10.1016/j.cell.2009.01.018
- Soppa, J. (2011). Ploidy and gene conversion in Archaea. *Biochem. Soc. Trans.* 39, 150–154. doi: 10.1042/BST0390150
- Soppa, J. (2013). Evolutionary advantages of polyploidy in halophilic archaea. *Biochem. Soc. Trans.* 41, 339–343. doi: 10.1042/BST20120315
- Spence, E., Bailey, S., Nenninger, A., Moller, S. G., and Robinson, C. (2004). A homolog of Albino3/OxaI is essential for thylakoid biogenesis in the cyanobacterium *Synechocystis* sp. PCC6803. *J. Biol. Chem.* 279, 55792–55800. doi: 10.1074/jbc.M411041200
- Takahama, K., Matsuoka, M., Nagahama, K., and Ogawa, T. (2004). High-frequency gene replacement in cyanobacteria using a heterologous rps12 gene. *Plant Cell Physiol.* 45, 333–339. doi: 10.1093/pcp/pch041
- Tobiason, D. M., and Seifert, H. S. (2010). Genomic content of *Neisseria* species. *J. Bacteriol.* 192, 2160–2168. doi: 10.1128/JB.01593-09
- Wächtershäuser, G. (1997). The origin of life and its methodological challenge. *J. Theor. Biol.* 21, 483–494. doi: 10.1006/jtbi.1996.0383
- Weiss-Schneeweiss, H., Emadzade, K., Jang, T. S., and Schneeweiss, G. M. (2013). Evolutionary consequences, constraints and potential of polyploidy in plants. *Cytogenet. Genome Res.* 140, 137–150. doi: 10.1159/000351727
- Wertheim, B., Beukeboom, L. W., and van de Zande, L. (2013). Polyploidy in animals: effects of gene expression on sex determination, evolution and ecology. *Cytogenet. Genome Res.* 140, 256–269. doi: 10.1159/000351998
- Wierzchos, J., de los Rios, A., and Ascaso, C. (2012). Microorganisms in desert rocks: the edge of life on Earth. *Int. Microbiol.* 15, 173–183.
- Xiao, W., Wang, Z. G., Wang, Y. X., Schneegurt, M. A., Li, Z. Y., Lai, Y. H., et al. (2013). Comparative molecular analysis of the prokaryotic diversity of two salt mine soils in southwest China. *J. Basic Microbiol.* 53, 942–952. doi: 10.1002/jobm.201200200
- Zahradka, K., Slade, D., Bailone, A., Sommer, S., Averbeck, D., Petranovic, M., et al. (2006). Reassembly of shattered chromosomes in *Deinococcus radiodurans*. *Nature* 443, 569–573. doi: 10.1038/nature05160
- Zerulla, K., Baumann, A., and Soppa, J. (2014a). “Cell cycle and polyploidy in haloarchaea,” in *Genetics and Genomics of Halophiles*, eds R. T. Papke and A. Oren (Norfolk: Horizon Scientific Press), in press.
- Zerulla, K., Chimileski, S., Näther, D., Gophna, U., Papke, R. T., and Soppa, J. (2014b). DNA as a phosphate storage polymer and the alternative advantages of polyploidy for growth or survival. *PLoS ONE* 9:e94819. doi: 10.1371/journal.pone.0094819
- Zusman, T., Rosenshine, I., Boehm, G., Jaenicke, R., Leskiw, B., and Mevarech, J. (1989). Dihydrofolate reductase of the extremophily halophilic archaeabacterium *Halobacterium volcanii*. The enzyme and its coding gene. *J. Biol. Chem.* 264, 18878–18883.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 March 2014; accepted: 19 May 2014; published online: 13 June 2014.

Citation: Zerulla K and Soppa J (2014) Polyploidy in haloarchaea: advantages for growth and survival. *Front. Microbiol.* 5:274. doi: 10.3389/fmicb.2014.00274

This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.

Copyright © 2014 Zerulla and Soppa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Light-dependent expression of four cryptic archaeal circadian gene homologs

Michael Maniscalco, Jennifer Nannen<sup>†</sup>, Valerie Sodi<sup>†</sup>, Gillian Silver<sup>†</sup>, Phillip L. Lowrey and Kelly A. Bidle \*

Department of Biology, Rider University, Lawrenceville, NJ, USA

**Edited by:**

Jesse Dillon, California State University, Long Beach, USA

**Reviewed by:**

Scott Miller, The University of Montana, USA

Sonja-Verena Albers, Max Planck Institute for Terrestrial Microbiology, Germany

**\*Correspondence:**

Kelly A. Bidle, Department of Biology, Rider University, 2083 Lawrenceville Road, Lawrenceville, NJ 08648, USA  
e-mail: kbidle@rider.edu

**†Present address:**

Jennifer Nannen, NJ State Police Office of Forensic Sciences, DNA Lab, Hamilton, USA;  
Valerie Sodi, Department of Biochemistry and Molecular Biology, Drexel University College of Medicine, Philadelphia, USA;  
Gillian Silver, Department of Genetics, Rutgers University, New Brunswick, USA

Circadian rhythms are important biological signals that have been found in almost all major groups of life from bacteria to man, yet it remains unclear if any members of the second major prokaryotic domain of life, the Archaea, also possess a biological clock. As an initial investigation of this question, we examined the regulation of four cyanobacterial-like circadian gene homologs present in the genome of the haloarchaeon *Haloferax volcanii*. These genes, designated *cirA*, *cirB*, *cirC*, and *cirD*, display similarity to the KaiC-family of cyanobacterial clock proteins, which act to regulate rhythmic gene expression and to control the timing of cell division. Quantitative RT-PCR analysis was used to examine the expression of each of the four *cir* genes in response to 12 h light/12 h dark cycles (LD 12:12) in *H. volcanii* during balanced growth. Our data reveal that there is an approximately two to sixteen-fold increase in *cir* gene expression when cells are shifted from light to constant darkness, and this pattern of gene expression oscillates with the light conditions in a rhythmic manner. Targeted single- and double-gene knockouts in the *H. volcanii* *cir* genes result in disruption of light-dependent, rhythmic gene expression, although it does not lead to any significant effect on growth under these conditions. Restoration of light-dependent, rhythmic gene expression was demonstrated by introducing, *in trans*, a wild-type copy of individual *cir* genes into knockout strains. These results are noteworthy as this is the first attempt to characterize the transcriptional expression and regulation of the ubiquitous *kaiC* homologs found among archaeal genomes.

**Keywords:** archaea, halophiles, gene expression, mutants

## INTRODUCTION

Life on Earth is challenged by 24 h environmental oscillations, the most prevalent of which are the light/dark cycle and temperature fluctuations. To anticipate and respond appropriately to these recurrent environmental stimuli, organisms have evolved endogenous, cell-autonomous, self-sustained circadian clocks. These biological timekeepers drive circadian rhythms in biochemistry, gene expression, physiology and behavior and synchronize them to environmental time cues. The strong selective pressure for the precise temporal coordination of internal biological processes with the external day is evident in the diverse phylogenetic distribution of circadian clocks—from unicellular cyanobacteria, to fungi, plants, insects, and vertebrates, including humans (Dunlap et al., 2004).

To date, the only prokaryotes shown definitively to possess a circadian clock are the cyanobacteria (Johnson, 2007; Dong and Golden, 2008). This clock regulates many processes in the cyanobacterial cell, including global gene expression and cell division (Liu et al., 1995; Mori et al., 1996), and enhances reproductive fitness (Ouyang et al., 1998; Woelfle et al., 2004). The central clock of the cyanobacterium *Synechococcus elongatus* PCC 7942 is composed of just three proteins: KaiA, KaiB, and KaiC. Inactivation of any *kai* gene results in arrhythmicity (Ishiura et al., 1998). It was initially proposed that, similar to

the general model emerging in eukaryotes, the circadian mechanism in cyanobacteria was based on rhythmic *kaiBC* transcription and translation (Nakahira et al., 2004). The dramatic demonstration, however, that the self-sustained, temperature-compensated rhythm of KaiC phosphorylation could be reconstituted *in vitro* by mixing ATP with all three Kai proteins, strongly suggested that the primary pacemaker in the *Synechococcus* clock system is a posttranslational phosphorylation cycle rather than a transcriptional/translational feedback loop (Nakajima et al., 2005). Recent results, however, indicate that both a transcriptional/translational feedback loop and posttranslational phosphorylation cycle are necessary to maintain precise and robust circadian rhythms in cyanobacteria (Kitayama et al., 2008).

Analyses of sequenced archaeal genomes reveal an abundance of putative homologs of the cyanobacterial *kaiC* gene (Leipe et al., 2000; Dvornyk et al., 2003; Dvornyk and Knudsen, 2005; Ming et al., 2007) although no clear homologs of *kaiA* or *kaiB* are found among Archaea. Indeed, the crystal structure of a KaiC homolog from the hyperthermophilic archaeon, *Pyrococcus horikoshii* OT3, has been determined (Ming et al., 2007; Kang et al., 2009), yet no functional analyses have been reported to date. Given that many members of the Archaea and Bacteria inhabit overlapping ecological niches, and given the pervasiveness of horizontal gene transfer among prokaryotes

(Garcia-Vallve et al., 2000; Gogarten et al., 2002; Koonin and Wolf, 2008), it is not surprising that some Archaea harbor cyanobacterial-like circadian genes. What has yet to be determined, however, is whether these putative *kai* homologs are expressed, and if so, how they function in the Archaea.

In this report, we describe the characterization of four cryptic *kaiC*-like genes found in the halophilic archaeon, *Haloferax volcanii*, to determine how these genes are expressed and potentially regulated in response to cycles of light and dark. Here, we demonstrate the expression of these genes and their oscillating, light-dependent regulation using quantitative RT-PCR in an examination of RNA isolated from *H. volcanii* cultures grown in light/dark cycles. Targeted gene knockouts in three of the four genes reveal a disruption of this diurnal, light-dependent regulation in remaining wild-type genes, indicating that each of these genes is functionally required for this light-driven, regulatory pattern. The implications for these results are important, as this is the first functional demonstration, to our knowledge, of light/dark-driven transcriptional regulation in *kaiC*-like genes in the third major domain of life, the Archaea.

## MATERIALS AND METHODS

### STRAINS, PLASMIDS, AND CULTURE CONDITIONS

Strains and plasmids used for strain construction are found in **Table 1**. *Haloferax volcanii* strain DS70 (Wendoloski et al., 2001) was grown aerobically with shaking at 45°C in an incubator equipped with a programmable photosynthetic light bank (Innova 42R, Eppendorf). Cells were cultured in medium containing 125 g NaCl, 45 g MgCl<sub>2</sub>·6H<sub>2</sub>O, 10 g MgSO<sub>4</sub>·7H<sub>2</sub>O, 10 g KCl, 1.34 ml 10% CaCl<sub>2</sub>·2H<sub>2</sub>O, 3 g yeast extract, and 5 g tryptone, per liter (Robb et al., 1995). For constant light (LL) or constant dark (DD) conditions, liquid cultures were inoculated to a starting O.D.<sub>600</sub> of 0.1 from 48-h starter cultures, grown until mid-exponential phase, and then harvested by centrifugation for 5 min at 4°C, 6000 × g. For light and dark treatments (LD 12:12), cells were maintained in balanced growth conditions. To achieve balanced growth, an O.D.<sub>600</sub> reading was taken of an actively growing starter culture which was then diluted with fresh media to an O.D.<sub>600</sub> = 0.8 to initiate the experiments. This procedure was repeated every 12 h. Under these conditions, O.D.<sub>600</sub> values never varied more than 0.6–0.8 units from the end of one 12 h growth period before the addition of fresh media to the beginning of the next 12 h growth period after the addition of fresh media, ensuring cells remained in mid-exponential phase through the duration of the experiments. At 12 h time points, aliquots of cell cultures were centrifuged, supernatants were removed and pellets were stored at –80°C until processing for RNA. The amount of light was measured, in lux units, using a hand-held light meter (Reliability Direct, Inc., VWR). During dark growth conditions, lux units were equal to 0, while light growth conditions equaled 3000 lux.

### PHYLOGENETIC TREE CONSTRUCTION

Sequences used to construct a phylogenetic tree of cyanobacterial KaiC proteins and archaeal KaiC homologs were obtained by performing a protein BLAST search of microbial genomes within

**Table 1 | Table of strains and plasmids used in this study.**

Strains	Description	Source or reference
DS70	Wild-type isolate DS2 cured of plasmid pHV2	Wendoloski et al., 2001
H26	DS70 Δ <i>pyrE2</i>	Allers et al., 2004
JN1	H26 (Δ <i>cirB</i> )	This study
JN2	H26 (Δ <i>cirC</i> )	This study
JN3	H26 (Δ <i>cirD</i> )	This study
GL1	JN2 (Δ <i>cirBcirC</i> )	This study
GL2	JN3 (Δ <i>cirBcirD</i> )	This study
GL3	JN3 (Δ <i>cirCcirD</i> )	This study
<b>PLASMIDS</b>		
pTA131	Amp <sup>r</sup> ; pBluescript II containing P <sub>fdx-pyrE2</sub>	Allers et al., 2004
pTA354	<i>E. coli/H. volcanii</i> pTA131-based shuttle vector, with <i>pyrE2</i> marker and pHV1/4 replication origin	Norais et al., 2007
pMA7	pTA131 with <i>cirB</i> fragment containing a 250 bp deletion within its coding region	This study
pMA8	pTA131 with <i>cirC</i> fragment containing a 500 bp deletion within its coding region	This study
pAB6	pTA131 with <i>cirD</i> fragment containing a 122 bp deletion within its coding region	This study
pMM2	pTA354 with ~1.4 kb full-length <i>cirB</i> gene and 5' and 3' flanking regions	This study
pMM3	pTA354 with ~1.4 kb full-length <i>cirC</i> gene and 5' and 3' flanking regions	This study
pMM4	pTA354 with ~1.4 kb full-length <i>cirD</i> gene and 5' and 3' flanking regions	This study

GenBank on the NCBI website. Sequence alignments were performed using CLUSTALW in MEGA6 (Tamura et al., 2013). The Maximum Likelihood method based on the JTT matrix-based model (Jones et al., 1992) was used to create a tree in MEGA6 with the highest log likelihood (–5876.7) in MEGA6. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 19 amino acid sequences and all positions containing gaps and missing data were eliminated. There were a total of 201 positions in the final dataset.

### RNA ISOLATION AND cDNA SYNTHESIS

RNA isolation and cDNA synthesis were performed as previously described (Bidle et al., 2008). Briefly, *H. volcanii* cells were harvested, resuspended in Tri Reagent (MRC, Cincinnati, OH) and processed according to the manufacturer's protocol. RNA was resuspended in RNase-free water and treated with TURBO DNase (Ambion, Austin, TX). To ensure contaminating DNA was removed in the previous step, a 30-cycle PCR reaction was performed on purified RNA using the primer sets designed for qRT-PCR analyses, *cira* (F, 5'-GCCTGTATCT

CACCTTCGAAG; R, 5'-GTTCTTGATGCTCTGCTTGC), *cirB* (F, 5'-CGTCTACATCACGCTCGAAG; R, 5'-CCTCGTTCGTTGAGTCGTAC), *cirC* (F, 5'-CGAACCGAACGTACATGG; R, 5'-GAACCTCTCGATCGGGAC), or *cirD* (F, 5'-GGTTGACGAGCTCATTC; R, 5'-GCTCACCGAGGTTGATGAAG), as well as primers designed to amplify an ~400 bp region within the *H. volcanii* 16S rRNA gene (F, 5'-CGAACGGTTCATCGGGAAATCC; R, 5'-GTCATCACTGTAGTCGGAGC). DNA-free RNA samples were quantified and first-strand cDNA synthesis was initiated using 1 µg total RNA primed with random hexamers from the AffinityScript QPCR cDNA synthesis kit according to the manufacturer's protocol (Stratagene, La Jolla, CA). Negative controls contained no reverse transcriptase in the first strand cDNA reaction.

#### QUANTITATIVE RT-PCR

Transcript levels specific to each of the *cir* genes were analyzed using qRT-PCR for duplicate samples collected from replicate cultures of *H. volcanii* grown in several different conditions including LL, DD, or balanced growth in LD 12:12 cycles. Quantitative RT-PCR was performed using the RT SYBR Green qPCR Master Mix (SA Biosciences, Fredrick, MD) and initiated by adding 1 µl of the first-strand cDNA synthesis reaction to forward and reverse primers for either *cirA*, *cirB*, *cirC*, or *cirD*. Archaeal 16S rDNA amplification served as a normalizer. The reactions were performed using a RotorGene RG-3000 (Corbett Life Science, Qiagen) for 45 cycles at 95°C, 30 sec; 50°C, 1 min; 72°C, 1 min.

Standard curves for the reaction were created by amplifying each individual *cir* gene from genomic DNA and cloning into the pCR2.1-TOPO vector (Invitrogen). Serially diluted plasmid DNA served as a standard curve (avg.  $r^2$  value ~ 0.98) in subsequent qRT-PCR analyses. Amplification efficiency (efficiency =  $10^{(-1/\text{slope})}$ ) of all primer sets used was shown to be above 85%. *Cir* gene expression was normalized to 16 s rRNA expression and relative gene expression was calculated using the  $2^{-\Delta\Delta CT}$  method (Livak and Schmittgen, 2001).

#### CREATION OF KNOCKOUTS IN *cir* GENES

*H. volcanii* knockout strains were created using established gene disruption mutagenesis ("pop-in/pop-out" method) protocols that rely on transformant selection against a  $\Delta pyrE2$  background (Bitan-Banin et al., 2003) using constructs made in pTA131, a pBluescript-based suicide plasmid containing the *pyrE2* gene as a selectable marker (Allers et al., 2004). These constructs contain ~200–300 bp of the 5' region of the gene to be mutated fused to ~200–300 bp of the 3' end of the gene, creating a deletion within the middle of each gene. To facilitate the creation of these constructs, PCR was used to amplify both the 5' upstream and 3' downstream regions of interest, with the resulting PCR products subsequently being cloned into pCR2.1-TOPO (Invitrogen, Carlsbad, CA). Positive clones were then digested with the enzyme *Eco*RI, gel purified, and ligated together. Ligation products were screened using the 5' upstream gene's forward primer and 3' downstream gene's reverse primer to amplify a sequence that contained the 5' region concatenated to the 3' region in the proper orientation (i.e., only correctly ligated products, those that run 5' to 3' with a

truncation in the middle, will amplify). Following this, resulting PCR products were cloned into pCR2.1-TOPO, and positively identified clones were digested with the enzymes *Xba*I and *Bam*HI to facilitate cloning into the suicide vector pTA131. Simultaneously, each of the full-length *cir* genes, along with ~200 bp of flanking DNA on each 5' and 3' end, were cloned into the autonomously replicating shuttle vector pTA354 (Norais et al., 2007). These clones served both to monitor transformation efficiency during knockout strain construction, as well as to complement, *in trans*, the single gene knockouts in subsequent experiments. All constructs were subjected to DNA sequencing analysis prior to *H. volcanii* transformations to verify their DNA fidelity.

The *H. volcanii* uracil auxotrophic strain H26 [ $\Delta pyrE2$ ; (Allers et al., 2004)] was transformed with each construct using established methods (Cline et al., 1995). Positive transformants, displaying uracil prototrophy on selective medium, were subsequently patched onto Hv-CA plates (Hv-CA medium contains 0.5% (w/v) casamino acids, replacing the tryptone and yeast extract from the standard *H. volcanii* growth medium described above) containing 5-fluoroorotic acid (5-FOA) as uracil auxotrophs are unable to convert 5-FOA to its toxic analog, 5-fluorouracil. Thus, transformants that have lost the plasmid through homologous recombination events display 5-FOA resistance. *Ura*<sup>-</sup> 5-FOA<sup>R</sup> colonies were screened via PCR to determine if they harbored the wild-type or deleted version of each *cir* gene. The resulting knockout strains were confirmed via PCR and DNA sequencing analysis.

#### COMPLEMENTATION OF *cir* GENE KNOCKOUTS

Following knockout strain construction, full-length *cir* gene sequences were transformed into *H. volcanii* for complementation analyses. Each individual  $\Delta cir$  knockout strain was transformed with the autonomously replicating shuttle vector harboring the wild-type gene (Table 1). Following successful transformation, clones were grown in the aforementioned LD 12:12 diurnal conditions and gene expression was examined via qRT-PCR as previously described.

## RESULTS

#### IDENTIFICATION OF FOUR *H. volcanii* CIRCADIAN GENE HOMOLOGS

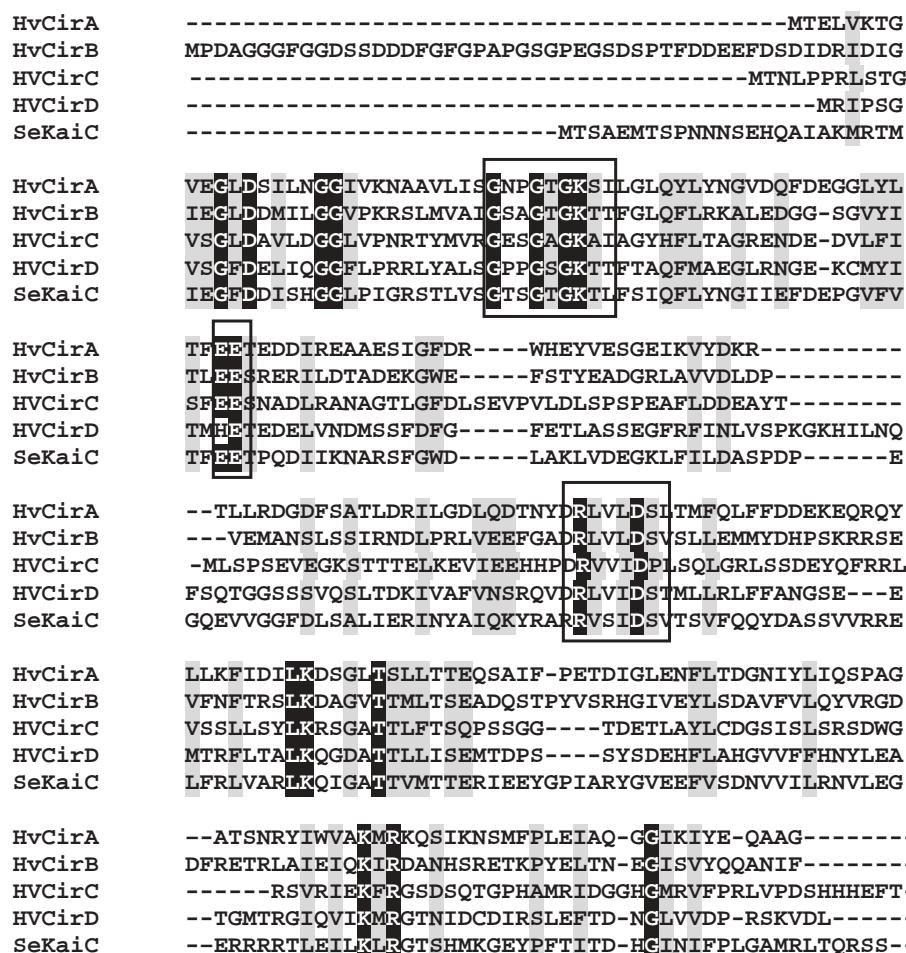
*H. volcanii* contains four cyanobacterial-like circadian rhythm gene homologs in its genome (Hartman et al., 2010), all displaying similarity to the KaiC-family of bacterial clock proteins. These genes were designated *cirA*, *cirB*, *cirC*, and *cirD* and they encode for predicted proteins of 27.5 kDa, 30.1 kDa, 52.7 kDa, and 29.7 kDa, respectively. These genes are not contained within an operon, but rather, are found distantly located from one another around the chromosome (Supplemental Figure 1). The *cirA* gene is located among a group of flagellar genes, while flanking the *cirB* gene are an NADH oxidase and ATP-NAD kinase; both *cirA* and *cirB* are divergently transcribed from their flanking genes. Interestingly, *cirC* is adjacent to *fixL*, an oxygen sensor protein; in cyanobacteria, *fixL* phosphorylates and dephosphorylates nitrogen fixation genes in response to environmental oxygen concentration (Monson et al., 1995). Directly upstream of *cirC* is the *orc2* gene that has been

shown to participate in DNA replication initiation (Robinson and Bell, 2005), although a recent study questions the role of replication origins in archaea, as deletion analyses indicate they are not solely required for cell division (Hawkins et al., 2013). Flanking *cirD* are an oxido-reductase and a hypothetical protein.

Each of these predicted protein sequences was aligned to *S. elongatus* KaiC, an ~58 kDa protein (Ishiura et al., 1998), and were shown to share between 28–33% identity and 48–55% similarity to this protein. More importantly, each of these predicted proteins contains a domain with a Walker A motif [(G/A)XXXXGK(T/S)] and Walker B [(RXXX(D/E)] motif where X = hydrophobic residue; (Walker et al., 1982), common to the KaiC family of proteins (Ishiura et al., 1998; Golden and Canales, 2003). These highly conserved regions are found within a diverse array of P-loop nucleoside triphosphate hydrolase protein families that, as well as KaiC, include RadA, RecA, and DnaB (Leipe et al., 2000). In addition to the Walker boxes found within *H. volcanii* Cir proteins, these predicted proteins also contained the

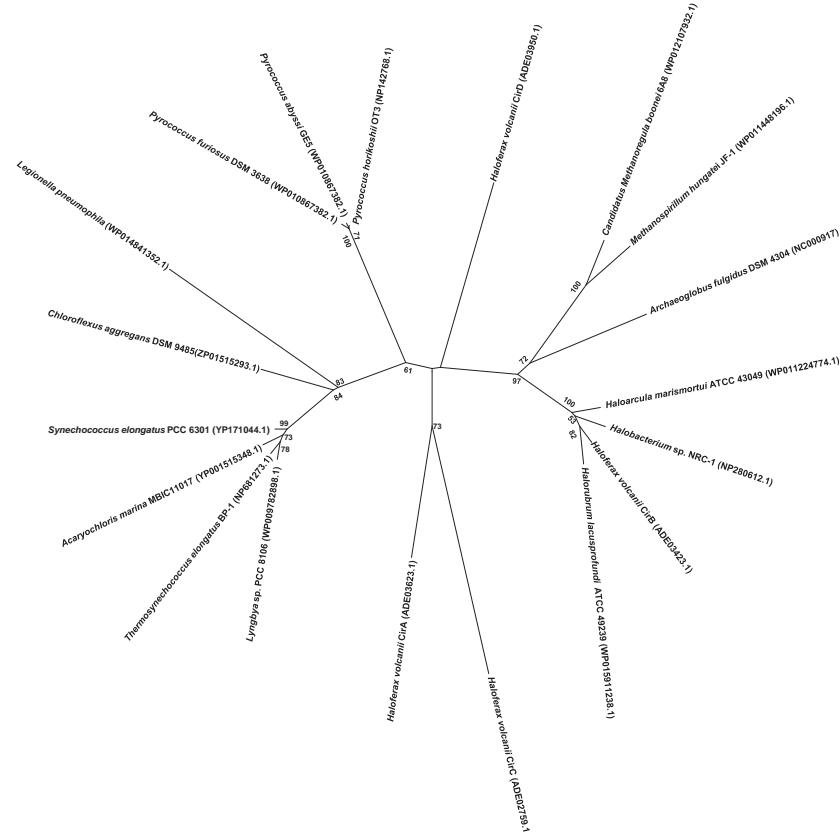
conserved pair of catalytic glutamate residues hypothesized to be involved in ATP binding (Yoshida and Amano, 1995). The Walker A and B motifs and conserved glutamate residues are highlighted within the *H. volcanii* Cir alignments with *S. elongatus* KaiC (**Figure 1**). It should be noted that only *CirC* contains the double Walker A and Walker B domain structure found in *S. elongatus* KaiC; the remaining homologs, *CirA*, *CirB*, and *CirD* only possess a single Walker A and Walker B domain. When compared to each other, the predicted *H. volcanii* Cir proteins share between 25–30% identity and 44–54% similarity at the amino acid level (data not shown).

A more extensive search of other annotated archaeal genomes indicates that KaiC homologs are widespread among all three archaeal kingdoms (i.e., *Euryarchaeota*, *Crenarchaeota*, and *Nanoarchaeota*) in this domain of life, consistent with other reports (Leipe et al., 2000; Dvornyk et al., 2003; Dvornyk and Knudsen, 2005; Ming et al., 2007). Indeed, greater than half of the ~150 annotated archaeal genomes in the GenBank database have homologs to *S. elongatus* KaiC, most of which have a single Walker



**FIGURE 1 |** Amino acid alignments of the N-terminal region of *S. elongatus* KaiC (SeKaiC) with *H. volcanii* (HV) *CirA*, *CirB*, *CirC*, and *CirD*. Predicted Walker A and B motifs, as well a

conserved pair of catalytic glutamate residues, are boxed. Identical residues are shaded in black; similar residues are shaded in gray.



**FIGURE 2 | Phylogenetic analysis of various known and predicted KaiC sequences from select cyanobacteria and archaea as compared with *H. volcanii* Cir sequences.** Sequence alignments and tree construction, using Maximum Likelihood, were performed using

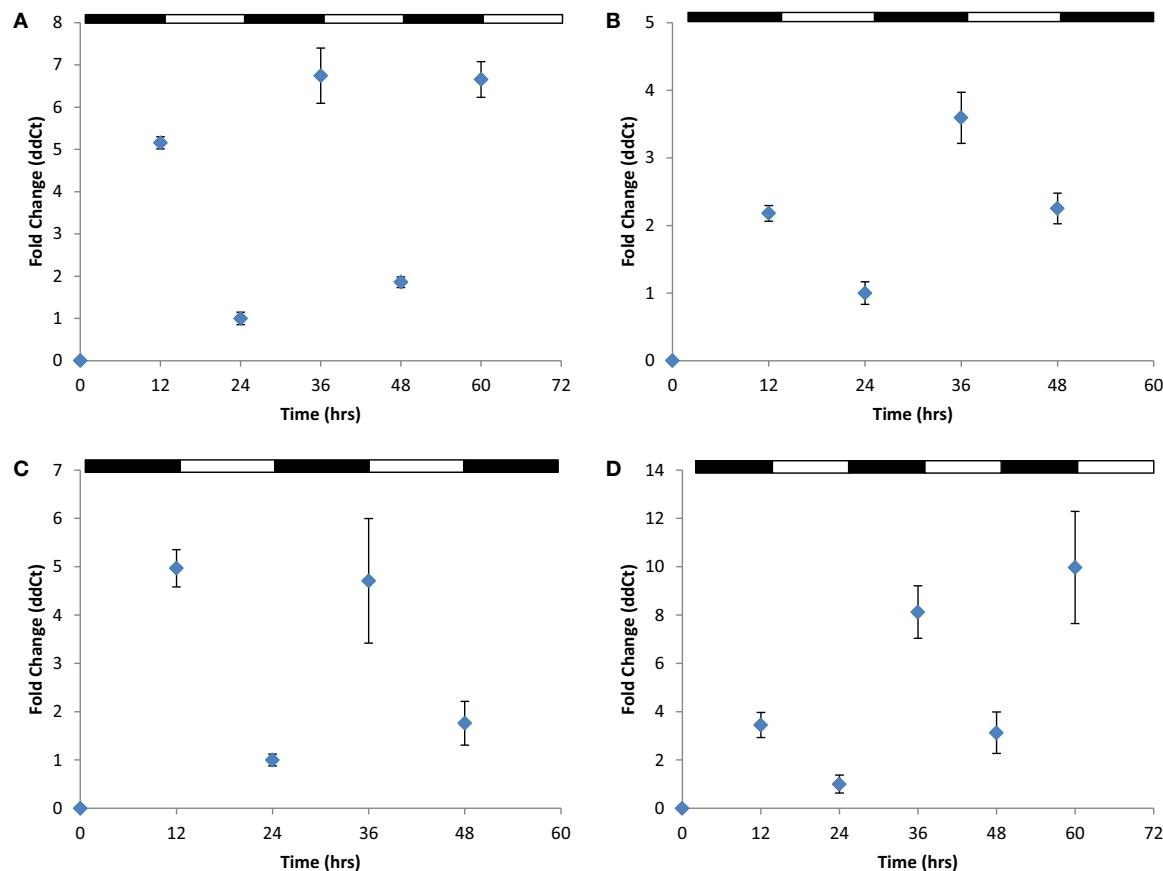
MEGA6 (Tamura et al., 2013). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Accession numbers for the sequences used in the trees are provided in the figure.

A and B domain structure (Dvornyk et al., 2003). Currently, their function remains unknown.

In order to visualize the relationship among the four Cir protein sequences found in *H. volcanii*, a phylogenetic tree was constructed to include other select annotated archaeal KaiC homologs, representative cyanobacterial KaiC proteins, including *S. elongatus*, and a recently characterized KaiC homolog from *Legionella pneumophila*. Interestingly, investigations of the *L. pneumophila* KaiC homolog revealed that it does not possess circadian properties, but rather appears to play a role in stress adaptation in the organism (Loza-Correa et al., 2014). As shown in Figure 2, the cyanobacterial KaiC proteins cluster within one group, distinct from *Legionella* and *Chloroflexus*, a member of the green-non-sulfur bacteria and an anoxygenic phototroph. The *H. volcanii* CirB homolog falls among several other haloarchaeal genera, while CirA, and CirC group together and distinct from the other sequences in the tree. (Figure 2). Given the fact that these proteins are only 25–30% identical to each other, these results are unsurprising and suggest that if they did arise from a gene duplication event, it was not recent in the organism's history.

#### GROWTH OF *H. volcanii* IN VARYING LIGHT CONDITIONS AND ESTABLISHMENT OF LIGHT-BASED DIFFERENTIAL *cir* GENE EXPRESSION

Before initiating qRT-PCR experiments, it was first established that growth in constant light (LL) or dark (DD) conditions had no discernable effect on, or advantage for, the growth of wild-type *H. volcanii*. There is very little difference in growth rate or abundance between the two conditions, thus ruling out the possibility that any detectable increased *cir* gene expression could be attributed to growth rate (Supplemental Figure 2). We next wanted to determine if any of the *cir* genes were regulated in a light- or dark-dependent manner; for this, a quantitative RT-PCR approach was used. Transcript levels specific for each *H. volcanii* *cir* gene were analyzed in triplicate using qRT-PCR with RNA isolated from duplicate mid-exponential cultures grown until mid-exponential phase under DD or LL conditions. Results of these initial experiments demonstrated that transcript levels of the four *H. volcanii* *cir* genes were greater during growth in DD (ranging from ~2.1-fold in *cirB* to ~16.2-fold in *cirD*; data not shown), as compared with expression during LL conditions. These results demonstrate that the *H. volcanii* *cir* genes are regulated in a light-dependent manner.



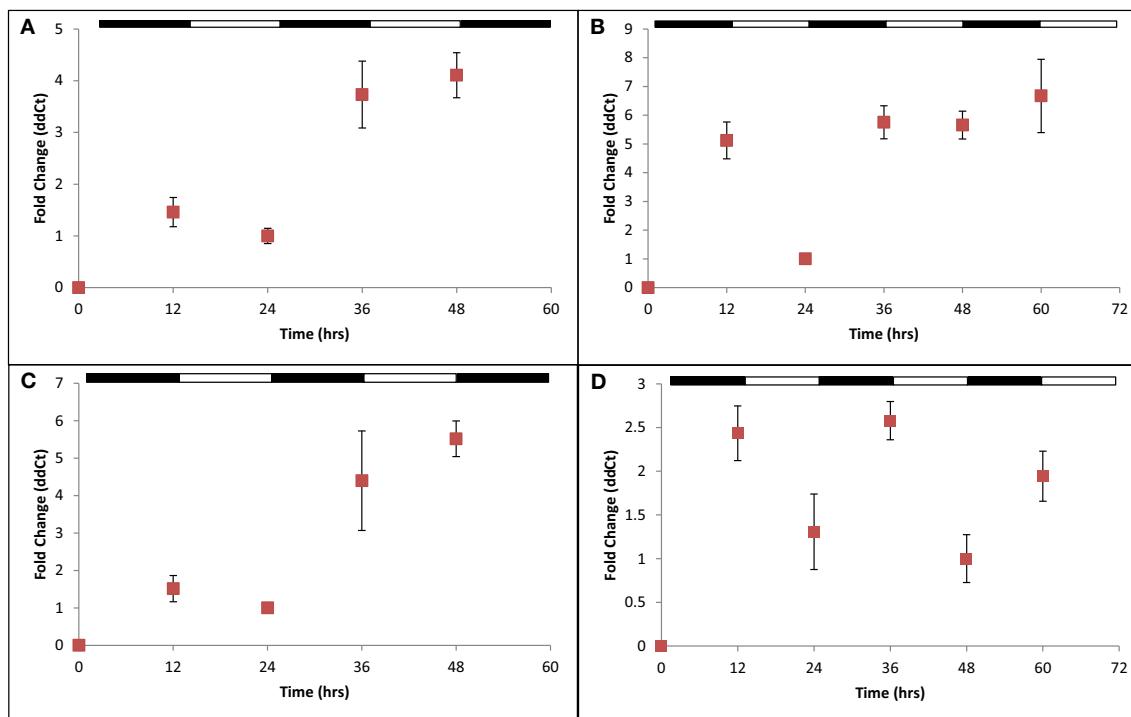
**FIGURE 3 | Quantitative RT-PCR analysis of *cir* gene expression.**  
Expression of (A) *cirA*, (B) *cirB*, (C) *cirC*, or (D) *cirD* during synchronous growth in LD 12:12 cycles. Relative gene expression, as indicated by

fold-change, was calculated using the  $2^{-\Delta\Delta CT}$  method (Livak and Schmittgen, 2001). Results given are pooled data collected from three independently-conducted experiments.

*H. volcanii* cells were next cultured in balanced growth conditions and grown under 12 h light/12 h dark conditions (LD 12:12; i.e., a typical diurnal cycle) for 72 h. These experiments demonstrated a consistent increase in gene expression for all four *cir* genes examined during each growth period in darkness (ranging from ~2.2-fold in *cirB* to ~10-fold in *cirD*) and decreased gene expression during growth in the light (**Figures 3A–D**). This pattern of regulation persisted regardless of whether the cycle was started in the light phase or the dark phase. To ensure that medium supplementation during the creation of balanced growth conditions was not altering the results of our studies, namely, affecting transcript abundance, we performed the following experiment. *H. volcanii* cells were cultured in constant LL or DD conditions and sampled every 3 h over a 12 h period. To maintain balanced growth, after every 3 h sampling period, cultures were supplemented with fresh growth medium to maintain the cells in mid-exponential phase. Results from these experiments revealed that gene expression remained constant over this sampling period (data not shown), indicating that the oscillations of gene expression seen in **Figure 3** are not in response to environmental manipulations (i.e., changing growth medium every 12 h).

#### HOW DOES A KNOCKOUT IN A *cir* GENE AFFECT RHYTHMIC GENE EXPRESSION IN THE REMAINING WILD-TYPE *cir* GENES?

To address this question, we created knockout strains using established gene disruption protocols (Bitan-Banin et al., 2003) in *cirB*, *cirC*, and *cirD* (JN1, JN2, and JN3, respectively; **Table 1**) and double knockout strains in *cirB/C*, *cirB/D*, and *cirC/D* (GL1, GL2, and GL3, respectively; **Table 1**). All strains were verified via PCR (**Supplemental Figure 3**) and DNA sequencing analysis to ensure the correct knockouts had been created. Despite repeated efforts, a knockout was unable to be generated in *cirA*. Nevertheless, the following results give us a clear picture that each of these genes appears to work together in *H. volcanii* to regulate gene expression in response to light conditions. We performed a comparative transcriptional analysis examining *cir* gene expression in strains JN1, JN2, and JN3 as compared with the *H. volcanii* wild-type parental strain H26. H26 and each of the  $\Delta$ *cir* strains were cultured as previously described in balanced growth in conditions of LD 12:12. Following this, qRT-PCR was used to examine *cir* gene expression in a  $\Delta$ *cir* background strain. For example, cDNA prepared from JN1 ( $\Delta$ *cirB*) RNA was subjected to qRT-PCR using *cirA*, *cirC*, and *cirD* primers. Using this strategy, we were able to ascertain whether gene



**FIGURE 4 | Quantitative RT-PCR analysis of *cir* gene expression in select  $\Delta$ *cir* mutants.** Expression of (A) *cirD* in JN1; (B) *cirB* in JN2; (C) *cirC* in JN3; and (D) *cirB* in JN1 (pMM2) during synchronous growth in LD 12:12 cycles. Loss of any single *cir* gene results in arrhythmic gene expression. Complementation of JN1 with an *in trans*

copy of wild-type *cirB* (pMM2) restores rhythmic gene expression. Relative gene expression, as indicated by fold-change, was calculated using the  $2^{-\Delta\Delta CT}$  method (Livak and Schmittgen, 2001). Results given are pooled data collected from two independently-conducted experiments.

expression in the three remaining viable *cir* genes was affected in this mutant background. In each knockout strain examined, wild-type *cir* gene expression was arrhythmic in a  $\Delta$ *cir* background, as compared with wild-type rhythmic expression. An example of arrhythmic gene expression for each mutant strain examined can be seen in Figures 4A–C, and can be compared against the rhythmic gene expression demonstrated in wild-type (Figures 3B–D). This pattern of arrhythmic gene expression was seen for every *cir* single knockout examined (data not shown due to number of combinations tested). This consistent pattern of arrhythmic gene expression leads us to speculate that expression of all four *H. volcanii* *cir* genes are coordinately regulated in response to light/dark conditions. This same loss of rhythmic gene expression was also seen in the double knockout strains GL1, GL2, and GL3 (data not shown). These results are similar to those of previous studies (Ishiura et al., 1998) that reported that loss of any single *kai* gene results in arrhythmicity in *S. elongatus*. Interestingly, growth studies on all six *H. volcanii* mutant strains revealed no significant change in growth rate or yield during LD 12:12 conditions (Supplemental Figure 4). Again, these results are reminiscent of work in cyanobacteria reporting that knockouts in any of the three *kai* genes do not affect growth of the organism (Ishiura et al., 1998).

## RESTORATION OF RHYTHMIC GENE EXPRESSION IN *cir* KNOCKOUT STRAINS VIA COMPLEMENTATION

After demonstrating that a single *cir* gene knockout disrupts rhythmic, light-dependent gene expression among the remaining wild-type *cir* genes, we next performed a series of complementation analyses. For this, we introduced, *in trans*, a copy of each wild-type *cir* gene on an autonomously replicating plasmid into its corresponding mutant strain. An example of this analysis can be seen in Figure 4D, whereby the full-length, wild-type *cirB* gene was introduced on plasmid pMM2 into the  $\Delta$ *cirB* strain JN1. Positive transformants were verified by PCR, cultured in LD 12:12 conditions, and examined for the restoration of rhythmic *cir* gene expression. Using this approach, all three *cir* deletion strains were complemented to wild-type levels of light-dependent, rhythmic gene expression (Figure 4D; complementation of JN2 and JN3 not shown due to space limitation).

## DISCUSSION

Circadian clocks confer on organisms the ability to anticipate and respond to recurrent 24 h environmental cycles. The selection pressures to establish an internal temporal order synchronized to the external environment were present from the evolution of the first life forms on earth (Paranjpe and Sharma, 2005). One of these pressures, mutagenic UV solar radiation,

which has given rise to the “escape from light” hypothesis for clock evolution (Pittendrigh, 1993; Rosato and Kyriacou, 2002), was significantly more intense during Earth’s early history than today (Karam, 2003). Unicellular cyanobacteria, among the most ancient life forms based on microfossil and biomarker records (Nisbet and Sleep, 2001), are the only prokaryotes in which circadian rhythms have been conclusively demonstrated (Johnson, 2007). Their role in the shift of the Earth’s atmosphere from a reductive to an oxidative one is well documented (Nisbet and Sleep, 2001). Thus, as photosynthesizers cyanobacteria must seek out light, yet also devise strategies to ameliorate the DNA-damaging effects of UV radiation. Furthermore, the development in non-heterocystous cyanobacteria of oxygenic photosynthesis and nitrogen fixation, two incompatible biochemical processes owing to the O<sub>2</sub>-sensitive nitrogenase complex, required the temporal separation of these activities (Huang et al., 1999; Berman-Frank et al., 2001; Church et al., 2005). Indeed, the *kaiC* clock gene seems to be ubiquitous among cyanobacterial species, suggesting that daily temporal control of cellular activity is not restricted to photosynthesis and N<sub>2</sub> fixation in these organisms (Lorne et al., 2000).

What then about other prokaryotes, and in particular the members of the Archaea, the third domain of life? Dating the origin of the Archaea is currently debated, yet some evidence suggests that archaeal methanogens may have been present by 3.4–2.6 billion years ago (Gribaldo and Brochier-Armanet, 2006). These dates coincide with the evolution of the cyanobacteria and thus suggest that the Archaea were exposed to similar selection pressures (e.g., intense UV radiation). Many halophilic archaea synthesize UV-protective pigments, gas vesicles for vertical migration, and/or flagella, all of which could conceivably be subject to circadian-regulated expression. The unique properties of the Archaea make it interesting to consider what molecular components and mechanisms might comprise a circadian system in these organisms. There is no reason to assume that an archaeal circadian system would necessarily be closely related to the cyanobacterial model, as the Archaea are no more closely related to the Bacteria than either domain is to the Eukarya (Woese, 1994). Indeed, archaeal transcriptional and translational mechanisms more closely resemble those of eukaryotes, while archaeal gene structure is more bacteria-like reviewed in (Allers and Mevarech, 2005). Thus, the components of an archaeal circadian system may resemble something between the bacterial and eukaryotic models. Even among the eukaryotic models studied, plants and mammals for example, there are significant clock differences at the molecular level.

To date, most of what we know about the phototactic response of Archaea comes from studies in the extreme haloarchaeon *Halobacterium* sp. NRC-1 (Schimz and Hildebrand, 1985; Hildebrand and Schimz, 1986; Rudolph and Oesterhelt, 1995; Nutsch et al., 2003; Oprian, 2003). Shortly after the publication of the NRC-1 genome (Ng et al., 2000), an *in silico* analysis was performed to identify a subset of genes involved in light sensing in this organism (DasSarma et al., 2001). Among the genes identified was a single *kaiC* homolog, leading the authors to speculate that circadian rhythms may be a property of some Archaea (DasSarma et al., 2001). An extension of these results came with

the publication of a microarray analysis by Whitehead et al., demonstrating the global regulation of diurnal gene expression in the transcriptome of NRC-1 (Whitehead et al., 2009). This study, while broadly focused in nature, is clearly an encouraging first step toward defining what a transcription/translation-based circadian response might look like in haloarchaea. Interestingly, the authors reported no change in gene expression in the single NRC-1 *kaiC* homolog in response to 12-h light/12-h dark (LD 12:12) conditions, nor was it directly addressed why this might be. This report is in direct contrast to results we have obtained demonstrating regulation of gene expression in the four *H. volcanii* *kaiC* homologs in response to LD 12:12 growth conditions. Finally, we must note that an intriguing study recently reported the presence of a circadian oscillation of peroxiredoxin oxidation in *H. salinarum* NRC-1, providing the first evidence of a posttranslational circadian mechanism that appears to be shared among the three domains of life (Edgar et al., 2012).

In conclusion, we have characterized four cyanobacterial, circadian clock *kaiC* homologs from the model haloarchaeon, *H. volcanii*, and have determined that they are transcriptionally regulated in a diurnal, light-dependent fashion. These results are noteworthy as this is the first attempt to directly characterize the gene expression of the ubiquitous *kaiC* homologs found among archaeal genomes. While we have demonstrated the transcriptional control of these four genes by environmental cues of light and darkness, much remains to be discovered as to their functional roles in *H. volcanii* as well as in other Archaea.

## ACKNOWLEDGMENTS

This research was supported by grants MCB10-51782 to Kelly A. Bidle and NIH 1R15GM086825-01 to Philip L. Lowrey, Jennifer Smith, Michael Anderson, and Alyssa Brown are thanked for help with plasmid construction and Nicole Ritzer is thanked for help with growth studies.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00079/abstract>

**Supplemental Figure 1 | Placement of the four *H. volcanii* cir genes in the context of the genome.** The Lasergene core suite (DNASTAR) was used to visualize the immediate genomic neighborhood of the four *H. volcanii* circadian genes based on data obtained from the NCBI GenBank database.

**Supplemental Figure 2 | Growth of wild-type *H. volcanii* at 45°C over a 72 h period in LL (■) or DD (○).** Values are averages taken from duplicate cultures.

**Supplemental Figure 3 | Verification of single (A) and double (B) knockout strains via PCR analysis.** Knockout strains (odd-numbered lanes) were compared against the H26 parental background (even-numbered lane) for confirmation of gene deletion. **(A)** Lane 1, molecular weight marker; lane 2, wild type *cirB* (1.2 kb); lane 3,  $\Delta$ *cirB* (JN1, 700 bp); lane 4, wild-type *cirC* (746 bp); lane 5,  $\Delta$ *cirC* (JN2, 500 bp); lane 6, wild-type *cirD* (550 bp); lane 7,  $\Delta$ *cirD* (JN3, 430 bp). **(B)** lane 1, 100 bp ladder; lane 2, wild-type *cirB* (750 bp) and *cirC* (1.2 kb); lane 3, GL1 harboring a deleted portion of *cirB* (500 bp) and *cirC* (700 bp); lane 4, wild-type *cirB* and *cirC* (323 bp); Lane 5,

GL2 harboring a deleted portion of *cirB* and *cirD* (201 bp); Lane 6, wild-type *cirC* and *cirD*; Lane 7, GL3 harboring a deleted portion of *cirC* and *cirD*. DNA sequencing analysis was used to confirm these results (data not shown).

**Supplemental Figure 4 | Growth curves of *H. volcanii* parental strain H26 as compared with various cir knockout strains. (A)** Growth of H26 as compared with JN1, JN2, and JN3. **(B)** Growth of H26 as compared with GL1, GL2, and GL3.

## REFERENCES

- Allers, T., and Mevarech, M. (2005). Archaeal genetics-the third way. *Nat. Rev. Gen.* 6, 58–73. doi: 10.1038/nrg1504
- Allers, T., Ngo, H.-P., Mevarech, M., and Lloyd, R. G. (2004). Development of additional selectable markers for the halophilic archaeon *Halofexax volcanii* based on the *leuB* and *trpA* genes. *Appl. Environ. Microbiol.* 70, 943–953. doi: 10.1128/AEM.70.2.943-953.2004
- Berman-Frank, I., Lundgren, P., Chen, Y. B., Kupper, H., Kolber, Z., Bergman, B., et al. (2001). Segregation of nitrogen fixation and oxygenic photosynthesis in the marine cyanobacterium *Trichodesmium*. *Science* 294, 1534–1537. doi: 10.1126/science.1064082
- Bidle, K. A., Kirkland, P. A., Nannen, J., and Maupin-Furlow, J. A. (2008). Proteomic analysis of *Halofexax volcanii* reveals salinity-mediated regulation of the stress response protein PspA. *Microbiology* 154, 1436–1443. doi: 10.1099/mic.0.2007/015586-0
- Bitan-Banin, G., Ortenberg, R., and Mevarech, M. (2003). Development of a gene knockout system for the halophilic archaeon *Halofexax volcanii* by use of the *pyrE* gene. *J. Bacteriol.* 185, 772–778. doi: 10.1128/JB.185.3.772-778.2003
- Church, M. J., Short, C. M., Jenkins, B. D., Karl, D. M., and Zehr, J. P. (2005). Temporal patterns of nitrogenase gene (*nifH*) expression in the oligotrophic North Pacific Ocean. *Appl. Environ. Microbiol.* 71, 5362–5370. doi: 10.1128/AEM.71.9.5362-5370.2005
- Cline, S. W., Pfeifer, F., and Doolittle, W. F. (1995). “Transformation of halophilic archaea,” in *Archaea: a Laboratory Manual*, eds F. T. Robb, A. R. Place, K. R. Sowers, H. J. Schreier, S. DasSarma, and E. M. Fleischmann (Cold Spring Harbor, NY: Cold Spring Harbor Press), 197–204.
- DasSarma, S., Kennedy, S. P., Berquist, B., Victor Ng, W., Baliga, N. S., Spudich, J. L., et al. (2001). Genomic perspective on the photobiology of *Halobacterium* species NRC-1, a phototrophic, phototactic, and UV-tolerant haloarchaeon. *Photosynth. Res.* 70, 3–17. doi: 10.1023/A:1013879706863
- Dong, G., and Golden, S. S. (2008). How a cyanobacterium tells time. *Curr. Opin. Microbiol.* 11, 541–546. doi: 10.1016/j.mib.2008.10.003
- Dunlap, J. C., Loros, J. J., and DeCoursey, P. J. (2004). *Chronobiology: Biological Timekeeping*. Sunderland, MA: Sinauer.
- Dvornyk, V., and Knudsen, B. (2005). Functional divergence of the circadian clock proteins in prokaryotes. *Genetica* 124, 247–254. doi: 10.1007/s10709-005-3146-0
- Dvornyk, V., Vinogradova, O., and Nevo, E. (2003). Origin and evolution of circadian clock genes in prokaryotes. *Proc. Natl. Acad. Sci. U.S.A* 100, 2495–2500. doi: 10.1073/pnas.0130099100
- Edgar, R., Green, E., Zhao, Y., van Ooijen, G., and Olmedo, M. (2012). Peroxiredoxins are conserved markers of circadian rhythms. *Nature* 485, 459–464. doi: 10.1038/nature11088
- Garcia-Vallve, S., Romeu, A., and Palau, J. (2000). Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Res.* 10, 1719–1725. doi: 10.1101/gr.130000
- Gogarten, J. P., Doolittle, W. F., and Lawrence, J. G. (2002). Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19, 2226–2238. doi: 10.1093/oxfordjournals.molbev.a004046
- Golden, S. S., and Canales, S. R. (2003). Cyanobacterial circadian clocks-timing is everything. *Nat. Rev. Microbiol.* 1, 191–199. doi: 10.1038/nrmicro774
- Gribaldo, S., and Brochier-Armanet, C. (2006). The origin and evolution of Archaea: a state of the art. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 1007–1022. doi: 10.1098/rstb.2006.1841
- Hartman, A. L., Norais, C., Badger, J. H., Delmas, S., Haldenby, S., Madupu, R., et al. (2010). The complete genome sequence of *Halofexax volcanii* DS2, a model archaeon. *PLoS ONE* 5:e9605. doi: 10.1371/journal.pone.0009605
- Hawkins, M., Malla, S., Blythe, M. J., Nieduszynski, C. A., and Allers, T. (2013). Accelerated growth in the absence of DNA replication origins. *Nature* 503, 544–547. doi: 10.1038/nature12650
- Hildebrand, E., and Schimz, A. (1986). Integration of photosensory signals in *Halobacterium halobium*. *J. Bacteriol.* 167, 305–311.
- Huang, T. C., Lin, R. F., Chu, M. K., and Chen, H. M. (1999). Organization and expression of nitrogen-fixation genes in the aerobic nitrogen-fixing unicellular cyanobacterium *Synechococcus* sp. strain RF-1. *Microbiology* 145, 743–753. doi: 10.1099/13500872-145-3-743
- Ishiura, M., Kutsuna, S., Aoki, S., Iwasaki, H., Andersson, C. R., Tanabe, A., et al. (1998). Expression of a gene cluster *kaiABC* as a circadian feedback process in cyanobacteria. *Science* 281, 1519–1523. doi: 10.1126/science.281.5382.1519
- Johnson, C. H. (2007). Bacterial circadian programs. *Cold Spring Harb. Symp. Quant. Biol.* 72, 395–404. doi: 10.1101/sqb.2007.72.027
- Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275–282. doi: 10.1093/bioinformatics/8.3.275
- Kang, H. J., Kubota, K., Ming, H., Miyazono, K. I., and Tanokura, M. (2009). Crystal structure of KaiC-like protein PH0186 from hyperthermophilic archaea *Pyrococcus horikoshii* OT3. *Proteins* 75, 1035–1039. doi: 10.1002/prot.22367
- Karam, P. A. (2003). Inconstant sun: how solar evolution has affected cosmic and ultraviolet radiation exposure over the history of life on Earth. *Health Phys.* 84, 322–333. doi: 10.1097/00004032-200303000-00005
- Kitayama, Y., Nishiwaki, T., Terauchi, K., and Kondo, T. (2008). Dual KaiC-based oscillations constitute the circadian system of cyanobacteria. *Genes Dev.* 22, 1513–1521. doi: 10.1101/gad.1661808
- Koonin, E. V., and Wolf, Y. I. (2008). Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* 36, 6688–6719. doi: 10.1093/nar/gkn668
- Leipe, D. D., Aravind, L., Grishin, N. V., and Koonin, E. V. (2000). The bacterial replicative helicase DnaB evolved from a RecA duplication. *Genome Res.* 10, 5–16. doi: 10.1006/jmbi.2001.5378
- Liu, Y., Tsiorremas, N. F., Johnson, C. H., Lebedeva, N. V., Golden, S. S., Ishiura, M., et al. (1995). Circadian orchestration of gene expression in cyanobacteria. *Genes Dev.* 9, 1469–1478. doi: 10.1101/gad.9.12.1469
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Lorne, J., Scheffer, J., Lee, A., Painter, M., and Miao, V. P. (2000). Genes controlling circadian rhythm are widely distributed in cyanobacteria. *FEMS Microbiol. Lett.* 189, 129–133. doi: 10.1111/j.1574-6968.2000.tb09218.x
- Loza-Correa, M., Sahr, T., Rolando, M., Daniels, C., Petit, P., Skarina, T., et al. (2014). The *Legionella pneumophila* *kai* operon is implicated in stress response and confers fitness in competitive environments. *Environ. Microbiol.* 16, 359–381. doi: 10.1111/1462-2920.12223
- Ming, H., Miyazono, K., and Tanokura, M. (2007). Cloning, expression, purification, crystallization and preliminary crystallographic analysis of selenomethionine-labelled KaiC-like protein PH0186 from *Pyrococcus horikoshii* OT3. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 63, 327–329. doi: 10.1107/S1744309107011402
- Monson, E. K., Ditta, G. S., and Helinski, D. R. (1995). The oxygen sensor protein, FixL, of *Rhizobium meliloti*: role of histidine residues in heme binding, phosphorylation, and signal transduction. *J. Biol. Chem.* 270, 5243–5250. doi: 10.1074/jbc.270.10.5243
- Mori, T., Binder, B., and Johnson, C. H. (1996). Circadian gating of cell division in cyanobacteria growing with average doubling times of less than 24 hours. *Proc. Natl. Acad. Sci. U.S.A.* 93, 10183–10188. doi: 10.1073/pnas.93.19.10183
- Nakahira, Y., Katayama, M., Miyashita, H., Kutsuna, S., Iwasaki, H., Oyama, T., et al. (2004). Global gene repression by KaiC as a master process of prokaryotic circadian system. *Proc. Natl. Acad. Sci. U.S.A.* 101, 881–885. doi: 10.1073/pnas.0307411100
- Nakajima, M., Imai, K., Ito, H., Nishiwaki, T., Murayama, Y., Iwasaki, H., et al. (2005). Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation *in vitro*. *Science* 308, 414–415. doi: 10.1126/science.1108451
- Ng, W. V., Kennedy, S. P., Mahairas, G. G., Berquist, B., Pan, M., Dutt Shukla, H., et al. (2000). Genome sequence of *Halobacterium* species NRC-1. *Proc. Natl. Acad. Sci. U.S.A.* 97, 12176–12181. doi: 10.1073/pnas.190337797
- Nisbet, E. G., and Sleep, N. H. (2001). The habitat and nature of early life. *Nature* 409, 1083–1091. doi: 10.1038/35059210

- Noraïs, C., Hawkins, M., Hartman, A., Eisen, J., Myllykallio, H., and Allers, T. (2007). Genetic and physical mapping of DNA replication origins in *Haloflexax volcanii*. *PLoS Genet.* 3, 729–743. doi: 10.1371/journal.pgen.0030077
- Nutsch, T., Marwan, W., Oesterhelt, D., and Gilles, E. D. (2003). Signal processing and flagellar motor switching during phototaxis of *Halobacterium salinarum*. *Genome Res.* 13, 2406–2412. doi: 10.1101/gr.1241903
- Oprian, D. D. (2003). Phototaxis, chemotaxis and the missing link. *Trends Biochem. Sci.* 28, 167–169. doi: 10.1016/S0968-0004(03)00056-2
- Ouyang, Y., Andersson, C. R., Kondo, T., Golden, S. S., and Johnson, C. H. (1998). Resonating circadian clocks enhance fitness in cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* 95, 8660–8664. doi: 10.1073/pnas.95.15.8660
- Paranjpe, D. A., and Sharma, V. K. (2005). Evolution of temporal order in living organisms. *J. Circadian Rhythms* 3:7. doi: 10.1186/1740-3391-3-7
- Pittendrigh, C. S. (1993). Temporal organization: reflections of a darwinian clock-watcher. *Annu. Rev. Physiol.* 55, 16–54. doi: 10.1146/annurev.ph.55.030193.000313
- Robb, F. T., Place, A. R., Sowers, K. R., Schreier, H. J., DasSarma, S., and Fleischmann, E. M. (eds.). (1995). *Archaea: A Laboratory Manual*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
- Robinson, N. P., and Bell, S. D. (2005). Origins of DNA replication in the three domains of life. *FEBS J.* 272, 3757–3766. doi: 10.1111/j.1742-4658.2005.04768.x
- Rosato, E., and Kyriacou, C. P. (2002). Origins of circadian rhythmicity. *J. Biol. Rhythms* 17, 506–511. doi: 10.1177/0748730402238232
- Rudolph, J., and Oesterhelt, D. (1995). Chemotaxis and phototaxis require CheA histidine kinase in the archaeon *Halobacterium salinarium*. *EMBO J.* 14, 667–673.
- Schimz, A., and Hildebrand, E. (1985). Response regulation and sensory control in *Halobacterium halobium* based on an oscillator. *Nature* 317, 641–643. doi: 10.1038/317641a0
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982). Distantly related sequences in the a- and b-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* 1, 945–951.
- Wendoloski, D., Ferrer, C., and Dyall-Smith, M. L. (2001). A new simvastatin (mevinolin)-resistance marker from *Haloarcula hispanica* and a new *Haloflexax volcanii* strain cured of plasmid pHV2. *Microbiology* 147, 959–964.
- Whitehead, K., Pan, M., Masumura, K., Bonneau, R., and Baliga, N. S. (2009). Diurnally entrained anticipatory behavior in archaea. *PLoS ONE* 4:e5485. doi: 10.1371/journal.pone.0005485
- Woelfle, M. A., Ouyang, Y., Phanvijhitsiri, K., and Johnson, C. H. (2004). The adaptive value of circadian clocks: an experimental assessment in cyanobacteria. *Curr. Biol.* 14, 1481–1486. doi: 10.1016/j.cub.2004.08.023
- Woese, C. R. (1994). Microbiology in transition. *Proc. Natl. Acad. Sci. U.S.A.* 91, 1601–1603. doi: 10.1073/pnas.91.5.1601
- Yoshida, M., and Amano, T. (1995). A common topology of proteins catalyzing ATP-triggered reactions. *FEBS Lett.* 359, 1–5. doi: 10.1016/0014-5793(94)01438-7

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Received: 20 December 2013; accepted: 12 February 2014; published online: 04 March 2014.*

*Citation: Maniscalco M, Nannen J, Sodi V, Silver G, Lowrey PL and Bidle KA (2014) Light-dependent expression of four cryptic archaeal circadian gene homologs. Front. Microbiol. 5:79. doi: 10.3389/fmicb.2014.00079*

*This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.*

*Copyright © 2014 Maniscalco, Nannen, Sodi, Silver, Lowrey and Bidle. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# Dihydroxyacetone metabolism in *Haloferax volcanii*

Matthew Ouellette, Andrea M. Makkay and R. Thane Papke \*

Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT, USA

**Edited by:**

Aharon Oren, The Hebrew University of Jerusalem, Israel

**Reviewed by:**

Jerry Eichler, Ben Gurion University of the Negev, Israel

Maria-Jose Bonete, University of Alicante, Spain

**\*Correspondence:**

R. Thane Papke, Department of Molecular and Cell Biology, University of Connecticut, 91 N. Eagleville Rd. Unit 3125, Storrs, CT 06268, USA  
e-mail: thane@uconn.edu

Dihydroxyacetone (DHA) is a ketose sugar that can be produced by oxidizing glycerol. DHA in the environment is taken up and phosphorylated to DHA-phosphate by glycerol kinase or DHA kinase. In hypersaline environments, it is hypothesized that DHA is produced as an overflow product from glycerol utilization by organisms such as *Salinibacter ruber*. Previous research has demonstrated that the halobacterial species *Haloquadratum walsbyi* can use DHA as a carbon source, and putative DHA kinase genes were hypothesized to be involved in this process. However, DHA metabolism has not been demonstrated in other halobacterial species, and the role of the DHA kinase genes was not confirmed. In this study, we examined the metabolism of DHA in *Haloferax volcanii* because putative DHA kinase genes were annotated in its genome, and it has an established genetic system to assay growth of mutant knockouts. Experiments in which *Hfx. volcanii* was grown on DHA as the sole carbon source demonstrated growth, and that it is concentration dependent. Three annotated DHA kinase genes (HVO\_1544, HVO\_1545, and HVO\_1546), which are homologous to the putative DHA kinase genes present in *Hqm. walsbyi*, as well as the glycerol kinase gene (HVO\_1541), were deleted to examine the effect of these genes on the growth of *Hfx. volcanii* on DHA. Experiments demonstrated that the DHA kinase deletion mutant exhibited diminished, but not absence of growth on DHA compared to the parent strain. Deletion of the glycerol kinase gene also reduced growth on DHA, and did so more than deletion of the DHA kinase. The results indicate that *Hfx. volcanii* can metabolize DHA and that DHA kinase plays a role in this metabolism. However, the glycerol kinase appears to be the primary enzyme involved in this process. BLASTp analyses demonstrate that the DHA kinase genes are patchily distributed among the Halobacteria, whereas the glycerol kinase gene is widely distributed, suggesting a widespread capability for DHA metabolism.

**Keywords:** dihydroxyacetone metabolism, dihydroxyacetone kinase, glycerol kinase, archaea, Halobacteria, Haloarchaea

## INTRODUCTION

Dihydroxyacetone (DHA) is a simple ketose sugar commonly used in sunless tanning lotions and sprays (Faurschou et al., 2004). DHA can be used as a carbon source by many different bacteria, yeast, and protists, and there are a number of different pathways in which it can be produced. In bacteria such as *Klebsiella pneumoniae*, DHA is produced anaerobically via glycerol oxidation by an NAD-dependent glycerol dehydrogenase (Forage and Lin, 1982). *Gluconobacter oxydans* and related bacteria also use glycerol oxidation to produce DHA, but they utilize a glycerol dehydrogenase that is pyrroloquinoline quinone (PQQ)-dependent and attached to the outer membrane. This pathway releases the DHA directly into the surrounding environment, which makes the *Gluconobacter* bacteria useful for industrial production of DHA (Deppenmeier et al., 2002). DHA can also be produced by methylotrophic yeast such as *Candida boidinii* by first oxidizing methanol to formaldehyde, after which a pyrophosphate-dependent transketolase transfers a two-carbon hydroxyethyl group to the formaldehyde to form DHA (Waites and Quayle, 1981).

Once DHA is obtained by a cell either via glycerol oxidation or uptake from the surrounding environment, it can then

be phosphorylated and subsequently metabolized. Two types of kinases phosphorylate DHA: glycerol kinase and DHA kinase. Glycerol kinase is considered less specific, and it is capable of phosphorylating both glycerol and DHA using ATP (Hayashi and Lin, 1967; Weinhouse and Benziman, 1976; Jin et al., 1982). DHA kinase is more specific, and it is only able to phosphorylate DHA and its isomer, D-glyceraldehyde (Erni et al., 2006). There are two major families of DHA kinases. The first consists of two subunits (DhaK and DhaL) and which are ATP-dependent. The DhaK subunit binds to the DHA substrate, and the DhaL subunit binds to ATP and transfers a phosphate group from ATP to DhaK-DHA (Daniel et al., 1995; Siebold et al., 2003). In the second family, the DHA kinases are made up of three subunits (DhaK, DhaL, and DhaM) and are phosphoenolpyruvate (PEP)-dependent. This family of DHA kinases uses the PEP:sugar phosphotransferase system (PTS) to transfer a phosphate group from PEP to the DhaM subunit, a multidomain protein with one domain predicted to be a member of the mannose (EIIMan) family of the PTS (Gutknecht et al., 2001; Zurbriggen et al., 2008). The DhaM then transfers the phosphate group to DhaL, which picks up the phosphate using an ADP cofactor bound to the subunit (Bachler et al., 2005). The phosphate is then transferred from DhaL to

the DhaK subunit, which phosphorylates the bound DHA substrate to DHA phosphate. The ATP-dependent family of DHA kinases is present in eukaryotes and some bacteria, whereas the PEP-dependent family of DHA kinases is present only in bacteria and archaea (Erni et al., 2006).

DHA has been hypothesized as a potential carbon source in hypersaline environments for heterotrophic halobacterial species (Elevi Bardavid et al., 2008). This hypothesis is supported by previous studies on glycerol oxidation in *Salinibacter ruber*, a halophilic bacterium common in hypersaline environments. In a study by Sher et al. (2004), which examined the oxidation of radio-labeled glycerol by *S. ruber*, an unknown soluble product consisting of 20% of the radioactivity from the added glycerol was observed to be excreted by the cells. This soluble product was later analyzed in a study by Elevi Bardavid and Oren (2008) using a colorimetric assay, and was identified as DHA; indicating that *S. ruber* could produce DHA in hypersaline environments as an overflow product via glycerol oxidation.

The ability of *Haloquadratum walsbyi*, a common halobacterial species, to metabolize DHA further supports the hypothesis that DHA is a carbon source in hypersaline environments. *Hqm. walsbyi* was first hypothesized to metabolize DHA after examination of the sequenced genome in a study Bolhuis et al. (2006) identified an uptake system for DHA involving three genes (HQ2672A, HQ2673A, and HQ2674A) encoding the subunits of a putative PEP-dependent DHA kinase. The DHA kinase encoded by these genes was hypothesized to use a phosphate group from the PTS system to phosphorylate DHA to DHA phosphate, which could then be incorporated into the metabolism of the cell. Elevi Bardavid and Oren (2008) tested DHA metabolism in *Hqm. walsbyi* by adding DHA to a cell culture of *Hqm. walsbyi* and measuring the change in DHA concentration over time. A decrease in DHA concentration was observed, indicating that the DHA was being taken up and metabolized by the *Hqm. walsbyi* cultures.

Overall, the current evidence supports a model where halobacterial species *Hqm. walsbyi* metabolizes DHA in hypersaline environments produced by *S. ruber*; however, there is still little known about DHA metabolism in Halobacteria. While DHA metabolism has been observed to occur in *Hqm. walsbyi*, no other halobacterial species has been shown to be able to metabolize DHA. Additionally, the putative DHA kinase genes in *Hqm. walsbyi* were never confirmed to be involved in DHA phosphorylation and metabolism. In this study, we sought to elucidate our understanding of halobacterial metabolism of DHA by examining DHA utilization in *Haloferax volcanii*, a halobacterial species isolated from Dead Sea sediment (Mullakhanbhai and Larsen, 1975). We used *Hfx. volcanii* because it has three putative PEP-dependent DHA kinase genes that are homologous to *Hqm. walsbyi* (Anderson et al., 2011), and it has an established genetic system that can be used to delete genes and test their function (Bitan-Banin et al., 2003; Allers et al., 2004; Blaby et al., 2010). We also used DHA metabolism genes in *Hfx. volcanii* to search the other sequenced halobacterial genomes to better understand the distribution of these genes among the Halobacteria. Our data provide important new insights into the metabolism of DHA in halobacterial organisms.

## MATERIALS AND METHODS

### STRAINS AND GROWTH CONDITIONS

Strains and plasmids used in this study are listed in **Table 1**. All *Hfx. volcanii* strains were grown in either Hv-YPG or Hv-CA medium at 42°C while shaking at 200 rpm. Hv-YPG and Hv-CA media were produced using the formulas outlined in *The Halohandbook* (Dyall-Smith, 2009). Hv-min medium used in growth experiments was modified from the formula in *The Halohandbook* to exclude a carbon source (Hv-min -C). Media were supplemented with uracil (50 µg/mL) and 5-fluoroorotic acid (50 µg/mL) as needed. For growth on Petri plates, 2% agar (w/v) was added to the media.

All *Escherichia coli* strains were grown in either S.O.C. media or LB-media at 37°C while shaking at 200 rpm. S.O.C. media was provided by Clontech (Cat. # 636763) and New England BioLabs (Cat. # B9020S). LB medium was produced by adding 5 g NaCl, 5 g tryptone, and 2.5 g of yeast extract to deionized water to a final volume of 500 mL and pH set to 7.0. LB was supplemented with ampicillin (100 µg/mL) as needed. When LB cell culture plates were produced, 1.5% agar (w/v) was added. LB plates were supplemented with 40 µL of X-gal (20 mg/mL) as needed.

### PCR AND DNA ISOLATION

All primers used in this study are listed in **Table 2**. DNA used for plasmid construction and screening was amplified via PCR. Reactions for PCR were assembled as 10 µL volumes and contained the following reagents: 5.9 µL of deionized water, 2 µL of 5x GC Phusion buffer (Thermo Scientific, Cat. # F-519), 1 µL of 100% DMSO (Thermo Scientific, Cat. # TS-20684), 0.4 µL of 10 mM dNTP (Promega, Cat. # U1511), 0.2 µL of 10 µM forward primer, 0.2 µL of 10 µM reverse primer, 0.2 µL of template DNA, and 0.1 µL of Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Cat. # F-530S). When needed, water was substituted with 20% acetamide. The reactions were performed in a Mastercycler EP Gradient (Eppendorf) with the following cycle: a DNA melting step at 94°C for 22 s, an annealing step at 58.1°C for 35 s, and an extension step at 72°C for 90 s. This cycle was repeated 40 times, after which a final annealing step at 72°C for 5 min was performed. Template DNA included *Hfx. volcanii* DS2 genomic DNA (20 ng/µL), plasmid DNA listed in **Table 1**, and DNA from *E. coli* and *Hfx. volcanii* colonies.

Gel electrophoresis was performed to separate and analyze the PCR products using 0.8% (w/v) agarose in 1 × TAE buffer (40 mM Tris acetate, 2 mM EDTA). After gel electrophoresis, PCR products were excised from the gel and purified using the Wizard SV Gel and PCR Clean-Up System (Promega). Plasmids from *E. coli* strains were extracted and purified using the PureYield Plasmid Miniprep System (Promega). Plasmids linearized via digestion with restriction enzymes (BamHI, HindIII, XbaI, or XbaI) were also purified using the Wizard SV Gel and PCR Clean-Up System.

### GENE DELETION IN *Hfx. volcanii*

Three *Hfx. volcanii* genes (*dhaKLM*; HVO\_1544, HVO\_1545, and HVO\_1546), which encode homologs to the putative DHA kinase genes in *Hqm. walsbyi*, and a glycerol kinase gene (*glpK*; HVO\_1541), were targeted for deletion in *Hfx. volcanii* strain H26

**Table 1 | List of plasmids and strains used in this study.**

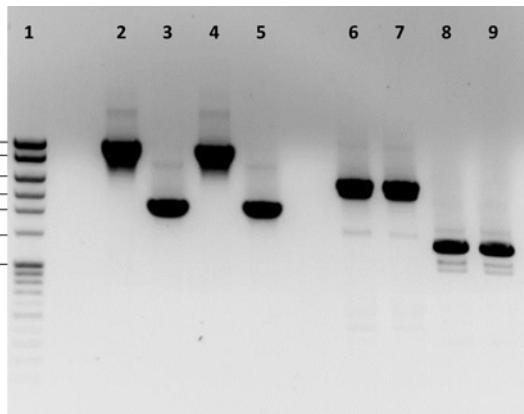
Plasmid or Strain	Description	References
pTA131	Cloning vector used for gene deletion in <i>Hfx. volcanii</i> . Contains lacZ cloning site, ampicillin resistance gene for screening in <i>E. coli</i> and pyrE2 gene for screening in <i>Hfx. volcanii</i> .	Allers et al., 2004
pTA409	Cloning vector used for gene complementation in <i>Hfx. volcanii</i> . Contains lacZ cloning site, ampicillin resistance gene for screening in <i>E. coli</i> and pyrE2 gene for screening in <i>Hfx. volcanii</i> .	Holzle et al., 2008
pΔdhaKLM	Derivative of pTA131 used to delete dhaKLM in <i>Hfx. volcanii</i> .	This study
pΔglpK	Derivative of pTA131 used to delete glpK in <i>Hfx. volcanii</i> .	This study
pdhaKLM	Derivative of pTA409 used to complement dhaKLM in ΔdhaKLM strain.	This study
pglpK	Derivative of pTA409 used to complement glpK in ΔglpK strain.	This study
HST08	An <i>E. coli</i> strain used for screening of constructed plasmids.	Clontech, Cat. # 636763
dam <sup>-</sup> /dcm <sup>-</sup>	An <i>E. coli</i> strain used to demethylate constructed plasmids.	New England BioLabs, Cat. # C2925H
H26	Uracil auxotrophic strain of <i>Hfx. volcanii</i> .	Allers et al., 2004
ΔdhaKLM	Derivative strain of H26 with dhaKLM operon deleted.	This study
ΔdhaKLM + pdhaKLM	Derivative strain of ΔdhaKLM with complementation of dhaKLM operon.	This study
ΔglpK	Derivative strain of H26 with glpK gene deleted.	This study
ΔglpK + pglpK	Derivative strain of ΔglpK with complementation of glpK gene.	This study
ΔdhaKLM ΔglpK	Derivative strain of H26 with dhaKLM operon and glpK gene deleted.	This study
ΔdhaKLM ΔglpK + pdhaKLM	Derivative strain of ΔdhaKLM ΔglpK with complementation of dhaKLM operon.	This study
ΔdhaKLM ΔglpK + pglpK	Derivative strain of ΔdhaKLM ΔglpK with complementation of glpK gene.	This study

**Table 2 | List of primers used in this study.**

Primer name	Description	Sequence
dhaKLM_FR1_F	Used to amplify flanking regions of dhaKLM for insertion into pTA131 digested with HindIII and BamHI to delete the operon.	5'- CGG TAT CGA TAA GCT GCC CTA CGC ACC CTA CAT G -3' 5'- TAG AAC TAG TGG ATC GCC TTC GGC TAC CCG CTC AT -3' 5'- GGA ATT CTA CCA GGC TCT GCG CTG AAC CGG CCG AA -3' 5'- GCC TGG TAG AAT TCC GAC TCA CCG TCC CTC ACG TT -3'
dhaKLM_R		
dhaKLMF	Used to amplify dhaKLM and native promoter for insertion into pTA409 digested with BamHI and Xhol to complement the operon.	5'- TAG AAC TAG TGG ATC AGG CGG TCG CGC GTT TCC GT -3' 5'- CGG GCC CCC CCT CGA ATC AGT TCA GCT TCC GGT AGT CGC G -3'
dhaKLMR		
glpK_FR1F	Used to amplify flanking regions of glpK for insertion into pTA131 digested with Xhol and XbaI to delete gene [external primers based on designs from Sherwood et al. (2009)].	5'- CGG GCC CCC CCT CGA TCG ACG ACC AGG CGT -3' 5'- TGG CGG CCG CTC TAG ACG ATG ACA ACG ATG T -3' 5'- GCC TGG GCA GAT CTC AAC ACG TGT TCG AAG -3' 5'- GAG ATC TGC CCA GGC TTC TAA CCA ACC TCG ATA CG -3'
glpK_FR1R		
glpK_FR2F		
glpK_FR2R		
glpKF	Used to amplify glpK and native promoter for insertion into pTA409 digested with BamHI and Xhol to complement gene.	5'- CGG GCC CCC CCT CGA CGC ACA ACT GAC GAA CGG GA -3' 5'- TAG AAC TAG TGG ATC TTA TTC CTC CCG TGC CCA GTC -3'
glpKR		

using the In-Fusion HD Cloning Kit (Clontech). The strategy for gene deletion was based on the methodology outlined in a study by Blaby et al. (2010) with a few modifications. Flanking regions of the targeted genes were developed to be between 800 and 1000 bp in length. The 15-bp linker used to combine the flanking regions was altered to so that EcoRI and BstOI sites were included for the dhaKLM deletion linker and BglII and BstOI sites were included for the glpK deletion linker. The pTA131 was linearized with HindIII and BamHI for the dhaKLM deletion and XbaI and XbaI for the glpK deletion. Constructed plasmids were transformed into Stellar Competent Cells (Clontech, Cat.

# 636763), according to the directions of the provider, and were plated on LB-amp plates with X-gal. White colonies were screened via colony PCR using the external primers of the target gene flanking regions. Confirmed deletion plasmids (listed in Table 1) were subcloned in dam<sup>-</sup>/dcm<sup>-</sup> Competent *E. coli* (New England BioLabs, Cat. # C2925H) to produce demethylated plasmids for transformation of *Hfx. volcanii*. *Hfx. volcanii* H26 colonies were screened for deleted genes via PCR using the external primers of the target gene flanking regions. The size of PCR products of screened cells were compared to those produced with wild-type DNA (Figure 1). Smaller product size indicated that the gene had



**FIGURE 1 | PCR analysis of H26,  $\Delta dhaKLM$ ,  $\Delta glpK$ , and  $\Delta dhaKLM \Delta glpK$ .** Analysis examined presence or absence of the *dhaKLM* operon (lanes 2–5) and *glpK* (lanes 6–9). Lane 1 contained exACTGene Mid Range Plus DNA Ladder (Fisher Scientific). Lanes 2 and 6 contained amplicons from H26. Lanes 3 and 7 contained amplicons from  $\Delta dhaKLM$ . Lanes 4 and 8 contained amplicons from  $\Delta glpK$ . Lanes 5 and 9 contained amplicons from  $\Delta dhaKLM \Delta glpK$ .

been deleted. The *Hfx. volcanii* H26 deletion strains produced by this process are listed in **Table 1**.

#### COMPLEMENTATION OF DELETED GENES

The *dhaKLM* and *glpK* genes deleted in *Hfx. volcanii* H26 were resuscitated by constructing complementation plasmids. Primers were designed which amplified the upstream native promoter and the coding region of the targeted genes in *Hfx. volcanii*. The primers were also designed to have 15 bp of homology with pTA409. Restriction digestion of pTA409 was performed using BamHI and XhoI to linearize the plasmid. After the linearized pTA409 and gene fragments were gel-purified, the DNA fragments were combined together using the In-Fusion HD Cloning Kit according to the instructions of the provider. The constructed plasmids were cloned, screened, and demethylated as described in the above gene deletion protocol. Purified constructed plasmids (listed in **Table 1**) were then transformed into the *Hfx. volcanii* H26 deletion strains using the PEG mediated transformation of Haloarchaea protocol from *The Halohandbook*. PCR was used to confirm transformation success. The *Hfx. volcanii* complementation strains produced by this process are listed in **Table 1**.

#### DHA GROWTH EXPERIMENTS

*Hfx. volcanii* strains listed in **Table 1** were grown to late-exponential phase ( $OD_{600} = \sim 0.6 - 0.8$ ) in Hv-YPC medium. The cell cultures were then centrifuged at 3220 RCF for 15 min and resuspended in Hv-min -C media supplemented with uracil. Centrifugation was repeated a total of three times to wash the cells of residual Hv-YPC media. During the final resuspension of the cells in Hv-min -C media, the cell cultures were diluted to  $OD_{600} \sim 0.01$ . Each cell culture was then distributed into the wells of a 96-well plate, with each well receiving 190  $\mu$ L of cell culture. Also, 200  $\mu$ L of Hv-min -C was added to the plate to be used as

a blank. Three wells of each culture were treated with 10  $\mu$ L of either 0.1 M DHA (final concentration of 5 mM DHA), 0.05 M DHA (final concentration of 2.5 mM DHA), 0.02 M DHA (final concentration of 1 mM DHA), or deionized water (negative control). The 96-well plate was then placed into a Multiscan FC plate reader (Fisher Scientific), which incubated the plate at 42°C while shaking it at low speed. The plate reader measured the  $OD_{620}$  of each well every hour for 72 h.

#### BIOINFORMATICS

The amino acid sequences of the *Hfx. volcanii* putative DHA kinase gene *dhaK* (HVO\_1546) and glycerol kinase gene *glpK* were used to perform BLASTp (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) searches of the NCBI database to determine other halobacterial species with DHA kinase and glycerol kinase genes. The amino acid sequences were retrieved from the NCBI database (*dhaK* GI number 292655696; *glpK* GI number 292655691). The search was restricted to the Halobacteriales (taxid 2235) with an *E*-value cut-off of 1e-20. Reciprocal BLASTp was performed to analyze only orthologous genes. The halobacterial genomes queried in this BLASTp search are listed in **Table 3**.

#### RESULTS

##### DHA KINASE IS PATCHILY DISTRIBUTED AMONG THE HALOBACTERIA

Three DHA kinase genes (HQ2672A, HQ2673A, and HQ2674A) have been annotated in the genome of *Hqm. walsbyi* (Bolhuis et al., 2006), a halobacterial species which is able to metabolize external DHA (Elevi Bardavid and Oren, 2008). Homologs of these three genes are also annotated in *Hfx. volcanii* (HVO\_1544, HVO\_1545, and HVO\_1546). In order to determine the prevalence of DHA kinase genes among the Halobacteria, the *Hfx. volcanii* *dhaK* gene (HVO\_1546) was used to perform a BLASTp search against the database of Halobacteria genomes available on NCBI. The search yielded significant hits among 31 different halobacterial species (**Table 4**). Except for *Haloferax larsenii* and *Haloferax elongans*, all queried *Haloferax* species yielded significant hits in the BLASTp search. Species from the *Halobiforma*, *Halococcus*, *Halorubrum*, and *Natronococcus* genera also yielded significant hits, but not all queried species from these genera produced results. All representatives from the genera *Haladaptatus*, *Halalkalicoccus*, *Halarchaeum*, *Haloquadratum*, *Halosarcina*, and *Salinarchaeum* yielded significant hits. Halobacteria genera that did not yield significant hits in the BLASTp search (*E*-value cut-off of 1e-20) include *Haloarcula*, *Halobacterium*, *Halobaculum*, *Halogeometricum*, *Halogramnum*, *Halomicrumbium*, *Halopiger*, *Haloplanus*, *Halorhabdus*, *Halosimplex*, *Halostagnicola*, *Haloterrigena*, *Halovivax*, *Natrialba*, *Natrinenema*, *Natronobacterium*, *Natronolimnobius*, *Natronomonas*, and *Natronorubrum*.

##### GROWTH ON DHA IN *Hfx. volcanii* IS CONCENTRATION DEPENDENT

Although putative DHA kinase genes are present in *Hfx. volcanii*, no previous research has demonstrated that *Hfx. volcanii* is able to grow on DHA as a carbon source. Therefore, experiments were performed to test the growth of *Hfx. volcanii* strain H26 on 5 mM, 2.5 mM, and 1 mM DHA. The results indicated that H26 was capable of growth on DHA as the sole carbon source. The cell

**Table 3 | List of halobacterial genomes queried in BLASTp search.****Queried halobacterial genomes**

<i>Haladaptatus paucihalophilus</i> DX253	<i>Haloferax larsenii</i> JCM 13917	<i>Halorubrum aidingense</i> JCM 13560	<i>Natrialba asiatica</i> DSM 12278
<i>Halalkalicoccus jeotgali</i> B3	<i>Haloferax lucentense</i> DSM 14919	<i>Halorubrum tebenquichense</i> DSM 14210	<i>Natrialba chahannaoensis</i> JCM 10990
<i>Halarachaeum acidiphilum</i> MH1-52-1	<i>Haloferax denitrificans</i> ATCC 35960	<i>Halorubrum terrestre</i> JCM 10247	<i>Natrialba hulunbeirensis</i> JCM 10989
<i>Haloarcula amylolytica</i> JCM 13557	<i>Haloferax elongans</i> ATCC BAA-1513	<i>Halorubrum arcis</i> JCM 13916	<i>Natrialba magadii</i> ATCC 43099
<i>Haloarcula argentinensis</i> DSM 12282	<i>Haloferax gibbonsii</i> ATCC 33959	<i>Halorubrum californiensis</i> DSM 19288	<i>Natrialba taiwanensis</i> DSM 12281
<i>Haloarcula californiae</i> ATCC 33799	<i>Haloferax mediterranei</i> ATCC 33500	<i>Halorubrum coriense</i> DSM 10284	<i>Natrinema altunense</i> JCM 12890
<i>Haloarcula hispanica</i> ATCC 33960	<i>Haloferax mucosum</i> ATCC BAA-1512	<i>Halorubrum distributum</i> JCM 9100	<i>Natrinema gari</i> JCM 14663
<i>Haloarcula japonica</i> DSM 6131	<i>Haloferax prahovense</i> DSM 18310	<i>Halorubrum ezzemoulense</i> DSM 17463	<i>Natrinema pallidum</i> DSM 3751
<i>Haloarcula marismortui</i> ATCC 43049	<i>Haloferax sulfurifontis</i> ATCC BAA-897	<i>Halorubrum hochstetnium</i> ATCC 700873	<i>Natrinema pellirubrum</i> DSM 15624
<i>Haloarcula sinaiensis</i> ATCC 33800	<i>Haloferax volcanii</i> DS2	<i>Halorubrum kocurii</i> JCM 14978	<i>Natrinema versiforme</i> JCM 10478
<i>Haloarcula vallismortis</i> ATCC 29715	<i>Haloferax</i> sp. ATCC BAA-644	<i>Halorubrum lacusprofundi</i> ATCC 49239	<i>Natrinema</i> sp. CX2021
<i>Haloarcula</i> sp. AS7094	<i>Haloferax</i> sp. ATCC BAA-645	<i>Halorubrum lipolyticum</i> DSM 21995	<i>Natrinema</i> sp. J7-1
<i>Halobacterium salinarum</i> NRC-1	<i>Haloferax</i> sp. ATCC BAA-646	<i>Halorubrum litoreum</i> JCM 13561	<i>Natrinema</i> sp. J7-2
<i>Halobacterium</i> sp. DL1	<i>Haloferax</i> sp. BAB2207	<i>Halorubrum saccharovorum</i> DSM 1137	<i>Natronobacterium gregoryi</i> SP2
<i>Halobacterium</i> sp. GN101	<i>Halogeometricum</i> borinquense DSM 11551	<i>Halorubrum</i> sp. T3	<i>Natronobacterium</i> sp. AS-7091
<i>Halobaculum gomorrense</i> JCM 9908	<i>Halogranum</i> salarium B-1	<i>Halosarcina</i> pallida JCM 14848	<i>Natronococcus amylolyticus</i> DSM 10524
<i>Halobiforma lacisalsi</i> AJ5	<i>Halomicrobiun</i> katesii DSM 19301	<i>Halosimplex carlsbadense</i> 2-9-1	<i>Natronococcus jeotgali</i> DSM 18795
<i>Halobiforma nitratireducens</i> JCM 10879	<i>Halomicrobiun</i> mukohataei DSM 12286	<i>Halostagnicola larsenii</i> XH-48	<i>Natronococcus occultus</i> SP4
<i>Halococcus hamelinensis</i> 100A6	<i>Halopiger</i> xanaduensis SH-6	<i>Haloterrigena limicola</i> JCM 13563	<i>Natronolimnobius innermongolicus</i> JCM 12255
<i>Halococcus morrhuae</i> DSM 1307	<i>Halopiger</i> sp. IIH2	<i>Haloterrigena</i> salina JCM 13891	<i>Natronomonas moolapensis</i> 8.8.11
<i>Halococcus saccharolyticus</i> DSM 5350	<i>Halopiger</i> sp. IIH3	<i>Haloterrigena</i> thermotolerans DSM 11522	<i>Natronomonas pharaonis</i> DSM 2160

(Continued)

**Table 3 | Continued****Queried halobacterial genomes**

<i>Halococcus salifodinae</i> DSM 8989	<i>Haloplanus natans</i> DSM 1798	<i>Haloterrigena turkmenica</i> DSM 5511	<i>Natronorubrum bangense</i> JCM 10635
<i>Halococcus thailandensis</i> JCM 13552	<i>Haloquadratum walsbyi</i> DSM 16790	<i>Halovivax asiaticus</i> JCM 14624	<i>Natronorubrum sulfidificiens</i> JCM 14089
<i>Halococcus</i> sp. 197A	<i>Halorhabdus tiamatea</i> SARL4B	<i>Halovivax ruber</i> XH-70	<i>Natronorubrum tibetense</i> GA33
<i>Haloferax alexandrinus</i> JCM 10717	<i>Halorhabdus utahensis</i> DSM 12940	<i>Natrialba aegyptia</i> DSM 13077	<i>Salinarchaeum</i> sp. <i>Harcht-Bsk1</i>

**Table 4 | Results of BLASTp search using *dhaK* (Performed on July 29, 2013).**

Species name	GI number	E-value	Species name	GI number	E-value
<i>Haloferax volcanii</i> DS2	292655696	0.0	<i>Natronococcus amylolyticus</i> DSM 10524	491710546	1e-169
<i>Haloferax</i> sp. BAB2207	493648700	0.0	<i>Halarchaeum acidiphilum</i> MH1-52-1	519064717	2e-169
<i>Haloferax alexandrinus</i> JCM 10717	445742333	0.0	<i>Halogramnum salarium</i> B-1	496767283	3e-165
<i>Haloferax sulfurifontis</i> ATCC BAA-897	494484188	0.0	<i>Halorubrum lipolyticum</i> DSM 21995	495278338	7e-165
<i>Haloferax lucentense</i> DSM 14919	490164612	0.0	<i>Halorubrum</i> sp. T3	515912844	2e-164
<i>Haloferax denitrificans</i> ATCC 35960	491112269	0.0	<i>Halorubrum kocurii</i> JCM 14978	496125287	3e-164
<i>Haloferax mediterranei</i> ATCC 33500	389847061	0.0	<i>Halococcus hamelinensis</i> 100A6	494968649	2e-162
<i>Haloferax</i> sp. ATCC BAA-644	445718309	0.0	<i>Halosarcina pallida</i> JCM 14848	495659148	2e-160
<i>Haloferax</i> sp. ATCC BAA-645	445712370	0.0	<i>Halorubrum lacusprofundi</i> ATCC 49239	222479879	6e-158
<i>Haloferax</i> sp. ATCC BAA-646	495849737	0.0	<i>Halococcus saccharolyticus</i> DSM 5350	492981238	3e-157
<i>Haloferax gibbonsii</i> ATCC 33959	491118466	0.0	<i>Halorubrum aidingense</i> JCM 13560	495274943	3e-157
<i>Haloferax prahovense</i> DSM 18310	445719493	0.0	<i>Haloquadratum walsbyi</i> DSM 16790	110668578	2e-154
<i>Haloferax mucosum</i> ATCC BAA-1512	495592772	0.0	<i>Halorubrum coriense</i> DSM 10284	493055434	6e-154
<i>Haladaptatus paucihalophilus</i> DX253	495255891	0.0	<i>Salinarchaeum</i> sp. <i>Harcht-Bsk1</i>	495690630	2e-150
<i>Halobiforma lacisalsi</i> AJ5	494236904	9e-180	<i>Halalkalicoccus jeotgali</i> B3	300710867	3e-145
<i>Natronococcus jeotgali</i> DSM 18795	495699224	2e-178			

density at which H26 reached stationary phase was also dependent on the initial concentration of DHA provided to the cells (**Figure 2**). H26 cells grown in medium supplemented with 1 mM DHA reached stationary phase at the lowest cell density, whereas cells grown with the highest tested concentration of 5 mM DHA reached stationary phase at the highest cell density. These data indicate that growth of *Hfx. volcanii* on DHA as a carbon source is concentration dependent.

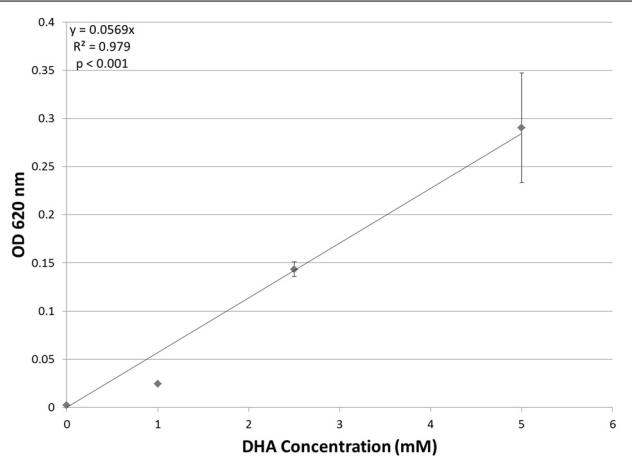
**DHA KINASE IS USED IN DHA METABOLISM IN *Hfx. volcanii***

Evidence indicates that *Hfx. volcanii*, like *Hqm. walsbyi*, can use DHA as a carbon source. Although both organisms have DHA kinase genes, no previous studies demonstrated these putative DHA kinase genes have a role in DHA metabolism. In order to determine that DHA metabolism in *Hfx. volcanii* utilizes the annotated DHA kinase, the operon *dhaKLM* (HVO\_1544–HVO\_1546) was deleted in *Hfx. volcanii* strain H26. The growth of this deletion strain ( $\Delta dhaKLM$ ) on 5 mM DHA was then tested in comparison to the parent strain H26 as well as a complementation strain ( $\Delta dhaKLM + pdhaKLM$ ). The results indicate that the deletion of *dhaKLM* causes a reduction in growth on DHA, and that complementation of the deleted genes negates

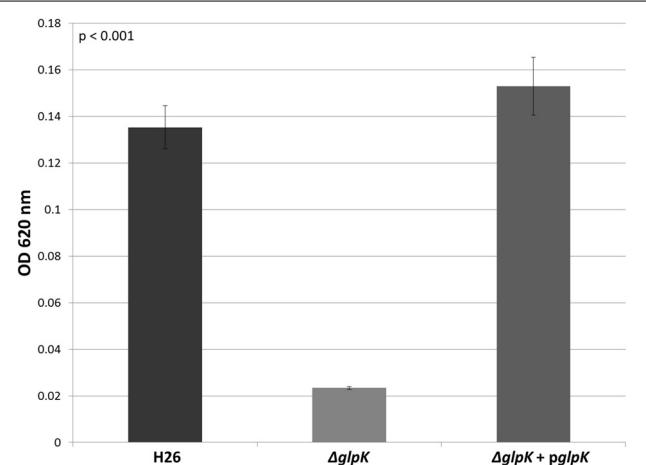
this growth deficiency (**Figure 3**). However, the  $\Delta dhaKLM$  was still capable of growth on DHA, exhibiting a 33% decrease in growth compared to H26. These results indicate that the *dhaKLM* genes are used by *Hfx. volcanii* in DHA metabolism, most likely for the phosphorylation of DHA to DHA phosphate, and that the genes are apparently not essential. Since it is still capable of growth on DHA there must be additional genes involved in the phosphorylation step.

**GLYCEROL KINASE IS MORE IMPORTANT THAN DHA KINASE**

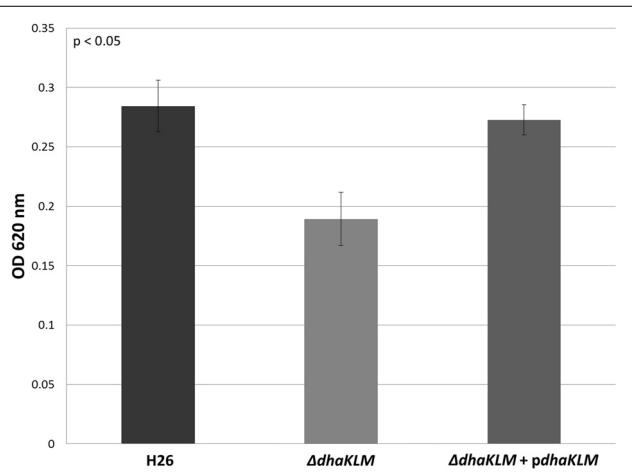
In other organisms, glycerol kinase is also capable of phosphorylating DHA (Hayashi and Lin, 1967; Weinhouse and Benziman, 1976; Jin et al., 1982). Therefore, the other gene involved DHA metabolism in *Hfx. volcanii* was hypothesized to be the glycerol kinase gene *glpK* (HVO\_1542). In order to test this hypothesis, the *glpK* gene was deleted in H26. The deletion strain ( $\Delta glpK$ ), and its complementation strain ( $\Delta glpK + pglpK$ ), were both grown on 5 mM DHA along with the parent strain H26. The results indicate that the deletion of *glpK* caused a reduction in growth on DHA even greater than deletion of *dhaKLM*, and that complementation of the *glpK* gene restores growth to normal levels (**Figure 4**). In comparison to the parent strain H26,  $\Delta glpK$



**FIGURE 2 | Cell density of *H. volcanii* H26 at stationary phase vs. concentration of DHA.** Cell density is represented by the average optical density ( $OD_{620}$ ) reading of three cell culture replicates. Error bars depict the standard deviation of the averages. The depicted line represents the line of best fit for the data. ANOVA single factor,  $p < 0.001$ .



**FIGURE 4 | Cell density of H26,  $\Delta glpK$ , and  $\Delta glpK + pglpK$  at stationary phase when grown on 5 mM DHA.** Cell density is represented by the average optical density ( $OD_{620}$ ) reading of three cell culture replicates. Error bars depict the standard deviation of the averages. ANOVA single factor,  $p < 0.001$ .



**FIGURE 3 | Cell density of H26,  $\Delta dhaKLM$ , and  $\Delta dhaKLM + pdhaKLM$  at stationary phase when grown on 5 mM DHA.** Cell density is represented by the average optical density ( $OD_{620}$ ) reading of three cell culture replicates. Error bars depict the standard deviation of the averages. ANOVA single factor,  $p < 0.05$ .

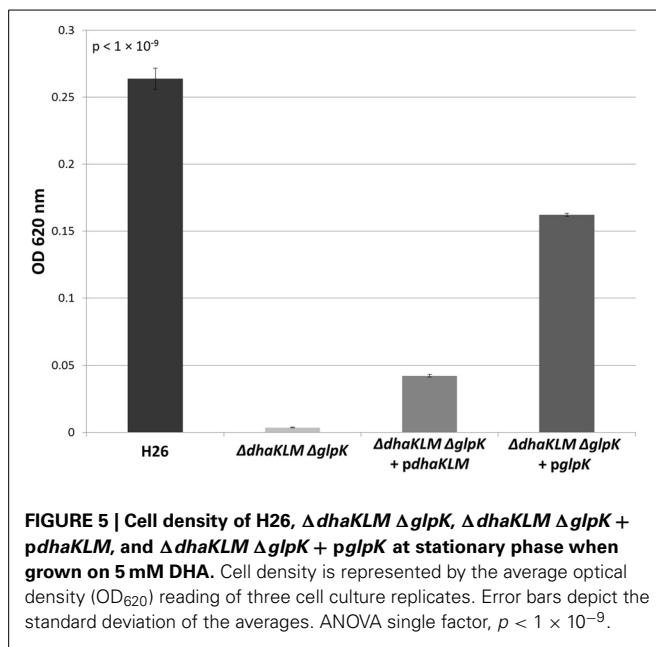
strain demonstrated an 83% decrease in growth. This decrease is far greater than the 33% decrease exhibited by the  $\Delta dhaKLM$  deletion mutant. These results indicate that the  $glpK$  gene is used by *Hfx. volcanii* in DHA metabolism, and that its role is potentially greater than that of the  $dhaKLM$  operon.

In order to further test the roles of the DHA kinase and glycerol kinase in DHA metabolism in *Hfx. volcanii*, the  $dhaKLM$  operon and  $glpK$  gene were both deleted in H26. This double deletion mutant ( $\Delta dhaKLM \Delta glpK$ ), along with a DHA kinase complementation strain ( $\Delta dhaKLM \Delta glpK + pdhaKLM$ ), a glycerol kinase complementation strain ( $\Delta dhaKLM \Delta glpK + pglpK$ ), and the parent strain H26, were then grown on 5 mM DHA.

The results indicate that the deletion of both kinases abolishes growth on DHA, and that complementation with glycerol kinase restores growth to a greater degree than complementation with DHA kinase (Figure 5). The  $\Delta dhaKLM \Delta glpK$  strain did not exhibit any growth, remaining at the initial  $OD_{620}$  of 0.0035. The  $\Delta dhaKLM \Delta glpK + pdhaKLM$  strain was able to grow on DHA, but demonstrated an 84% decrease compared to the H26 parent strain. The  $\Delta dhaKLM \Delta glpK + pglpK$  was also capable of limited growth on DHA, but demonstrated a 39% growth increase from H26 and a 390% growth increase compared with  $\Delta dhaKLM \Delta glpK + pdhaKLM$ . Overall, these data confirm that glycerol kinase is more important for DHA metabolism in *Hfx. volcanii* than DHA kinase.

#### GLYCEROL KINASE IS WIDELY DISTRIBUTED AMONG THE HALOBACTERIA

Since growth experiments indicated that glycerol kinase has a significant role in DHA metabolism, the presence of this gene in halobacterial species could potentially be a determinant of DHA metabolism in those species. Although the distribution of  $glpK$  homologs has been examined in previous studies (Sherwood et al., 2009; Anderson et al., 2011), a greater number of halobacterial genomes have become available since those studies. Therefore, the  $glpK$  gene in *Hfx. volcanii* was used to perform a BLASTp search against the halobacterial genomes available on NCBI. The search yielded 90 significant hits among 82 different species of Halobacteria (Table 5), indicating a much wider distribution of glycerol kinase compared to DHA kinase among the Halobacteria. Six species yielded more than one significant hit: *Halogeometricum borinquense* (3 hits), *Haladaptatus paucihalophilus* (3 hits), *Halofex prahovense* (2 hits), *Halofex mucosum* (2 hits), *Halofex gibbonsii* (2 hits), and *Natronomonas moolapensis* (2 hits). The multiple hits indicate the presence of  $glpK$  paralogs in these species. Only 18 of the 100 queried



halobacterial species did not yield significant hits: *Halococcus* sp. AS7094, *Halobacterium* sp. DL1, *Halobacterium* sp. GN101, *Halobaculum gomorrense*, *Halococcus* sp. 197A, *Halopiger* sp. IIH2, *Halopiger* sp. IIH3, *Haloplanus natans*, *Halorubrum ezzemoulense*, *Halosarcina pallida*, *Halostagnicola larsenii*, *Halovivax asiaticus*, *Halovivax ruber*, *Natrinema* sp. CX2021, *Natrinema* sp. J7-1, *Natronobacterium gregoryi*, *Natronobacterium* sp. AS-7091, and *Natronomonas pharaonis*. It should be noted, however, that only the genomes of *Halovivax ruber*, *Natronobacterium gregoryi*, and *Natronomonas pharaonis* are completely sequenced, whereas the other genomes without significant hits are incomplete, leaving open the possibility that these species might have *glpK* homologs. With the exception of *Halosarcina pallida*, which has an incompletely sequenced genome, all halobacterial species that yielded significant hits in the *dhaK* BLASTp search also yielded significant hits in the *glpK* BLASTp search.

## DISCUSSION

Previously, *Hqm. walsbyi* was the only halobacterial species known to be able to utilize DHA as a carbon source (Elevi Bardavid and Oren, 2008). In this study, we have identified *Hfx. volcanii* as the second halobacterial species known to be capable of metabolizing DHA. When DHA was added to growth medium as the sole carbon source, *Hfx. volcanii* was capable of growth. This growth was variable based on the concentration of DHA present in the growth medium. The ability of *Hfx. volcanii* to metabolize DHA suggests that the substrate could be an important carbon source in the Dead Sea environment where *Hfx. volcanii* naturally lives. Elevi Bardavid and Oren (2008) have suggested that *Salinibacter* might be a source of DHA in hypersaline environments, since it can produce DHA as an overflow product. However, *Salinibacter* has not been identified in the Dead Sea, making it an unlikely candidate for DHA producer. The DHA could potentially be produced as an overflow product from

*Dunaliella parva*, a halophilic alga that is the most prominent photosynthetic organism in the Dead Sea and is able to produce DHA (Ben-Amotz and Avron, 1974; Oren and Shilo, 1982). Elevi Bardavid and Oren (2008) hypothesized that the *Dunaliella* cell membrane could be permeable to DHA, allowing excess DHA produced by the cells to leak into the external environment. If *D. parva* produces a significant amount of DHA overflow, the substrate would be readily available for *Hfx. volcanii* to utilize as a source of carbon.

When Elevi Bardavid and Oren (2008) demonstrated that *Hqm. walsbyi* could utilize DHA as a carbon source, they hypothesized that the organism used a system involving a PEP-dependent DHA kinase to phosphorylate DHA to DHA kinase, based on genomic analysis from Bolhuis et al. (2006). However, their study did not demonstrate a direct connection between the putative DHA kinase and DHA metabolism. In our model halobacterial organism, *Hfx. volcanii*, we have demonstrated that DHA kinase is involved in metabolism of DHA. When the DHA kinase operon *dhaKLM* is deleted, growth of *Hfx. volcanii* on DHA is impeded, and complementation of the deleted genes with the *dhaKLM* operon restores growth. The growth of *Hfx. volcanii* is not completely abolished, however, and further analysis using a strain wherein the glycerol kinase gene *glpK* has been deleted indicates that *Hfx. volcanii* also uses glycerol kinase for DHA metabolism. Deletion of the *glpK* gene reduces growth on DHA more dramatically than the *dhaKLM* deletion, indicating that the role of glycerol kinase is more pronounced in DHA metabolism than that of DHA kinase for *Hfx. volcanii*. This enzyme primacy is further supported by the observation that, in the double deletion mutant  $\Delta dhaKLM \Delta glpK$ , complementation with *glpK* restores growth better than complementation with *dhaKLM*.

The primacy of the glycerol kinase in DHA metabolism is unexpected, since DHA kinase is usually the primary enzyme involved in DHA phosphorylation in other organisms due to the lower affinity of glycerol kinase for DHA. In *Klebsiella pneumoniae*, the glycerol kinase has a  $K_m$  of  $1 \times 10^{-3}$  M for DHA, whereas the DHA kinase has a  $K_m$  of  $1 \times 10^{-5}$  M (Jin et al., 1982). The glycerol kinase in *E. coli* has a  $K_m$  of  $5 \times 10^{-4}$  M for DHA (Hayashi and Lin, 1967), but the DHA kinase has a  $K_m$  of  $4.5 \times 10^{-7}$  M (Gutknecht et al., 2001). One possible explanation for the primacy of the glycerol kinase in *Hfx. volcanii* DHA metabolism is the glycerol kinase might have a higher affinity than DHA kinase for DHA. Another possible explanation might be differences in expression of the kinases. DHA kinase might be expressed at lower levels than glycerol kinase early in the *Hfx. volcanii* growth cycle, which would cause the glycerol kinase to be the primary DHA phosphorylating enzyme despite a possible lower affinity for DHA. Later in the growth cycle, however, *Hfx. volcanii* may increase expression of DHA kinase, leading to the higher affinity enzyme becoming the new primary enzyme for DHA phosphorylation. Growth experiments of  $\Delta dhaKLM \Delta glpK + pdhaKLM$ , in which the strain was grown beyond 72 h on 5 mM DHA, support this hypothesis, since growth of the strain on DHA increased significantly after 80 h, and actually surpassed  $\Delta dhaKLM \Delta glpK + pglpK$  after 96 h (data not shown). In-depth analysis into the enzymatic activity and kinetic constants of these enzymes toward DHA, as well as their expression levels,

**Table 5 | Results of BLASTp search using *glpK* (Performed on September 17, 2013).**

Species name	GI number	E-value	Species name	GI number	E-value
<i>Haloferax volcanii</i> DS2	292655691	0.0	<i>Haloferax mucosum</i> ATCC BAA-1512	445745541	0.0
<i>Haloferax</i> sp. BAB2207	432200129	0.0	<i>Natrialba hulunbeirensis</i> JCM 10989	445640226	0.0
<i>Haloferax lucentense</i> DSM 14919	445722906	0.0	<i>Halarchaeum acidiphilum</i> MH1-52-1	543417579	0.0
<i>Haloferax alexandrinus</i> JCM 10717	445742338	0.0	<i>Halorubrum californiensis</i> DSM 19288	445688091	0.0
<i>Haloferax</i> sp. ATCC BAA-646	445709004	0.0	<i>Halorubrum lipolyticum</i> DSM 21995	445813038	0.0
<i>Haloferax</i> sp. ATCC BAA-645	445712375	0.0	<i>Halorubrum lacusprofundi</i> ATCC 49239	222479549	0.0
<i>Haloferax</i> sp. ATCC BAA-644	445718304	0.0	<i>Salinarchaeum</i> sp. Harcht-Bsk1	510882182	0.0
<i>Haloferax sulfurifontis</i> ATCC BAA-897	445746251	0.0	<i>Natrialba chahannaoensis</i> JCM 10990	445643664	0.0
<i>Haloferax denitrificans</i> ATCC 35960	445749875	0.0	<i>Halorubrum hochstetnium</i> ATCC 700873	445701406	0.0
<i>Haloferax prahovense</i> DSM 18310	445719488	0.0	<i>Halomicrombium mukohataei</i> DSM 12286	257388556	0.0
<i>Haloferax elongans</i> ATCC BAA-1513	445734605	0.0	<i>Halorubrum tebenquichense</i> DSM 14210	445687222	0.0
<i>Haloferax larsenii</i> JCM 13917	445729767	0.0	<i>Haloarcula amylolytica</i> JCM 13557	445772086	0.0
<i>Haloferax gibbonsii</i> ATCC 33959	445726194	0.0	<i>Halomicrombium katesii</i> DSM 19301	517069632	0.0
<i>Haloferax mediterranei</i> ATCC 33500	389847056	0.0	<i>Halosimplex carlsbadense</i> 2-9-1	445671661	0.0
<i>Haloferax mucosum</i> ATCC BAA-1512	445747425	0.0	<i>Haloarcula vallismortis</i> ATCC 29715	445755712	0.0
<i>Halogeometricum borinquense</i> DSM 11551	313125210	0.0	<i>Haloarcula argentinensis</i> DSM 12282	445773756	0.0
<i>Halogeometricum borinquense</i> DSM 11551	313126426	0.0	<i>Halorubrum litoreum</i> JCM 13561	445813470	0.0
<i>Halobiforma nitratireducens</i> JCM 10879	445784518	0.0	<i>Haloarcula marismortui</i> ATCC 43049	55377424	0.0
<i>Natrinema pallidum</i> DSM 3751	445622526	0.0	<i>Haloarcula sinaiensis</i> ATCC 33800	445762583	0.0
<i>Haladaptatus paucihalophilus</i> DX253	320548735	0.0	<i>Haloarcula californiae</i> ATCC 33799	445763060	0.0
<i>Haloterrigena salina</i> JCM 13891	445666802	0.0	<i>Natronolimnobius innermongolicus</i> JCM 12255	445597617	0.0
<i>Haloterrigena thermotolerans</i> DSM 11522	445659630	0.0	<i>Natronorubrum tibetense</i> GA33	445585740	0.0
<i>Natrinema pellirubrum</i> DSM 15624	433590333	0.0	<i>Haloarcula japonica</i> DSM 6131	445778554	0.0
<i>Halococcus morrhuae</i> DSM 1307	445795889	0.0	<i>Natrialba aegyptia</i> DSM 13077	445651647	0.0
<i>Haloterrigena limicola</i> JCM 13563	445665007	0.0	<i>Natrialba taiwanensis</i> DSM 12281	445642534	0.0
<i>Halococcus salifodinae</i> DSM 8989	445798601	0.0	<i>Haloquadratum walsbyi</i> DSM 16790	110667688	0.0
<i>Halococcus hamelinensis</i> 100A6	445790305	0.0	<i>Natrialba magadii</i> ATCC 43099	289580614	0.0
<i>Natrinema</i> sp. J7-2	397773488	0.0	<i>Natrialba asiatica</i> DSM 12278	445650101	0.0
<i>Natrinema altunense</i> JCM 12890	445633695	0.0	<i>Natronomonas moolapensis</i> 8.8.11	452208319	0.0
<i>Natronorubrum sulfidificiens</i> JCM 14089	445594250	0.0	<i>Haloferax gibbonsii</i> ATCC 33959	445728401	0.0
<i>Natrinema gari</i> JCM 14663	445628815	0.0	<i>Natronococcus jeotgali</i> DSM 18795	445603927	0.0
<i>Halococcus thailandensis</i> JCM 13552	445801492	0.0	<i>Halorubrum saccharovorum</i> DSM 1137	445683831	0.0
<i>Natrinema versiforme</i> JCM 10478	445613765	0.0	<i>Halorhabdus utahensis</i> DSM 12940	257052548	0.0
<i>Halobiforma lacisalsi</i> AJ5	445778236	0.0	<i>Halorubrum</i> sp. B-1	399240308	0.0
<i>Haladaptatus paucihalophilus</i> DX253	320549923	0.0	<i>Halorubrum arcis</i> JCM 13916	445822264	0.0
<i>Haloterrigena turkmenica</i> DSM 5511	284166225	0.0	<i>Halorubrum terrestre</i> JCM 10247	445683460	0.0
<i>Haladaptatus paucihalophilus</i> DX253	516847391	0.0	<i>Halorubrum distributum</i> JCM 9100	445698917	0.0
<i>Halalkalicoccus jeotgali</i> B3	300711495	0.0	<i>Haloarcula hispanica</i> ATCC 33960	344211542	0.0
<i>Natronococcus occultus</i> SP4	435847946	0.0	<i>Halorubrum kocurii</i> JCM 14978	445806839	0.0
<i>Halopiger xanaduensis</i> SH-6	336253699	0.0	<i>Halorhabdus tiamatea</i> SARL4B	529078002	0.0
<i>Natronococcus amyloyticus</i> DSM 10524	445599450	0.0	<i>Halorubrum</i> sp. T3	515912305	0.0
<i>Halogeometricum borinquense</i> DSM 11551	445572938	0.0	<i>Halorubrum aidingense</i> JCM 13560	445818937	0.0
<i>Natronorubrum bangense</i> JCM 10635	445597786	0.0	<i>Halorubrum coriense</i> DSM 10284	445694991	0.0
<i>Haloferax prahovense</i> DSM 18310	445713901	0.0	<i>Natronomonas moolapensis</i> 8.8.11	452206238	0.0
<i>Halococcus saccharolyticus</i> DSM 5350	445793423	0.0	<i>Halobacterium salinarum</i> NRC-1	15790841	0.0

would enhance understanding on glycerol kinase primacy in *Hfx. volcanii* DHA metabolism.

DHA metabolism among the Halobacteria may extend beyond *Hfx. volcanii* and *Hqm. walsbyi*. Our BLASTp results for *dhaK* indicate that 29 other halobacterial species have a DHA kinase gene homologous to *dhaK* in *Hfx. volcanii* and *Hqm. walsbyi*.

Since our data indicate that the *dhaKLM* genes in *Hfx. volcanii* are involved in DHA metabolism, the homologs of these genes in other halobacterial species likely also have this function, allowing those species to utilize DHA. Halobacterial species without DHA kinase might also be capable of utilizing DHA if they possess a *glpK* gene, since our results indicate that glycerol kinase

also plays a role in DHA metabolism. BLASTp results for *glpK* indicate that 82 halobacterial species have homologs, and 51 of these species do not have *dhaKLM* homologs. We suspect that these species are also able to metabolize DHA. Eighteen halobacterial species are missing DHA and glycerol kinase genes, suggesting that they cannot metabolize DHA. However, only three of those genomes, *Halovivax ruber*, *Natronobacterium gregoryi*, and *Natronomonas pharaonis*, are not in draft form, leaving open the possibility for a near universal distribution of DHA metabolism in Halobacteria.

The broad taxonomic distribution of DHA and glycerol kinase genes among the Halobacteria suggests two interwoven hypotheses: (i) DHA is a common carbon source in hypersaline environments and (ii) DHA metabolism is widespread among the Halobacteria. A study by Elevi Bardavid and Oren (2008) detailed the conversion by the halophilic bacterium *S. ruber* of glycerol to DHA, which was then used as a growth substrate by *Hqm. walsbyi*. They speculated that DHA could be a common carbon source due to incomplete oxidation of glycerol, and from it being an intermediate of glycerol synthesis in *Dunaliella*. Our data demonstrating the extensive incidence of DHA and glycerol kinase genes provides support for their hypothesis that DHA is a common carbon source, and extends it to include that many if not most Halobacteria are capable of metabolizing it. However, future research on DHA production and turnover rates, and analysis on strains we predict to have DHA metabolism is necessary to elucidate the significance of this substrate to hypersaline ecosystems and Halobacteria.

## AUTHORS CONTRIBUTIONS

R. Thane Papke, Andrea M. Makkay, and Matthew Ouellette conceived the researched and wrote the manuscript. Andrea M. Makkay and Matthew Ouellette performed the research.

## ACKNOWLEDGMENTS

We would like to thank Dr. Thorsten Allers from the University of Nottingham for supplying us with *Haloflexax volcanii* strains and plasmids, and Dr. Aharon Oren from the Hebrew University of Jerusalem for supplying us with dihydroxyacetone. This research was supported by the National Science Foundation (award numbers, DEB0919290 and DEB0830024) and NASA Astrobiology: Exobiology and Evolutionary Biology Program Element (Grant Number NNX12AD70G).

## REFERENCES

- Allers, T., Ngo, H. P., Mevarech, M., and Lloyd, R. G. (2004). Development of additional selectable markers for the halophilic archaeon *Haloflexax volcanii* based on the *leuB* and *trpA* genes. *Appl. Environ. Microbiol.* 70, 943–953. doi: 10.1128/AEM.70.2.943-953.2004
- Anderson, I., Scheuner, C., Goker, M., Mavromatis, K., Hooper, S. D., Porat, I., et al. (2011). Novel insights into the diversity of catabolic metabolism from ten haloarchaeal genomes. *PLoS ONE* 6:e20237. doi: 10.1371/journal.pone.0020237
- Bachler, C., Flukiger-Bruhwiler, K., Schneider, P., Bahler, P., and Erni, B. (2005). From ATP as substrate to ADP as coenzyme: functional evolution of the nucleotide binding subunit of dihydroxyacetone kinases. *J. Biol. Chem.* 280, 18321–18325. doi: 10.1074/jbc.M500279200
- Ben-Amotz, A., and Avron, M. (1974). Isolation, characterization, and partial purification of a reduced nicotinamide adenine dinucleotide phosphate-dependent dihydroxyacetone reductase from the halophilic alga *Dunaliella parva*. *Plant Physiol.* 53, 628–631. doi: 10.1104/pp.53.4.628
- Bitan-Banin, G., Ortenberg, R., and Mevarech, M. (2003). Development of a gene knockout system for the halophilic archaeon *Haloflexax volcanii* by use of the *pyrE* gene. *J. Bacteriol.* 185, 772–778. doi: 10.1128/JB.185.3.772-778.2003
- Blaby, I. K., Phillips, G., Blaby-Haas, C. E., Gulig, K. S., El Yacoubi, B., and de Crecy-Lagard, V. (2010). Towards a systems approach in the genetic analysis of archaea: accelerating mutant construction and phenotypic analysis in *Haloflexax volcanii*. *Archaea* 2010, 426239. doi: 10.1155/2010/426239
- Bolhuis, H., Palm, P., Wende, A., Falb, M., Rampp, M., Rodriguez-Valera, F., et al. (2006). The genome of the square archaeon *Halococcus walsbyi*: life at the limits of water activity. *BMC Genomics* 7:169. doi: 10.1186/1471-2164-7-169
- Daniel, R., Stuertz, K., and Gottschalk, G. (1995). Biochemical and molecular characterization of the oxidative branch of glycerol utilization by *Citrobacter freundii*. *J. Bacteriol.* 177, 4392–4401.
- Deppenmeier, U., Hoffmeister, M., and Prust, C. (2002). Biochemistry and biotechnological applications of Gluconobacter strains. *Appl. Microbiol. Biotechnol.* 60, 233–242. doi: 10.1007/s00253-002-1114-5
- Dyall-Smith, M. (2009). *Protocols for Haloarchaeal Genetics* ©Ver. 7.1. Available online at: <http://www.haloarchaea.com/resources/halohandbook>
- Elevi Bardavid, R., Khristo, P., and Oren, A. (2008). Interrelationships between *Dunaliella* and halophilic prokaryotes in saltern crystallizer ponds. *Extremophiles* 12, 5–14. doi: 10.1007/s00792-006-0053-y
- Elevi Bardavid, R., and Oren, A. (2008). Dihydroxyacetone metabolism in *Salinibacter ruber* and in *Halococcus walsbyi*. *Extremophiles* 12, 125–131. doi: 10.1007/s00792-007-0114-x
- Erni, B., Siebold, C., Christen, S., Srinivas, A., Oberholzer, A., and Baumann, U. (2006). Small substrate, big surprise: fold, function and phylogeny of dihydroxyacetone kinases. *Cell. Mol. Life Sci.* 63, 890–900. doi: 10.1007/s00018-005-5518-0
- Faurschou, A., Janjua, N. R., and Wulf, H. C. (2004). Sun protection effect of dihydroxyacetone. *Arch. Dermatol.* 140, 886–887. doi: 10.1001/archderm.140.7.886
- Forage, R. G., and Lin, E. C. (1982). DHA system mediating aerobic and anaerobic dissimilation of glycerol in *Klebsiella pneumoniae* NCIB 418. *J. Bacteriol.* 151, 591–599.
- Gutknecht, R., Beutler, R., Garcia-Alles, L. F., Baumann, U., and Erni, B. (2001). The dihydroxyacetone kinase of *Escherichia coli* utilizes a phosphoprotein instead of ATP as phosphoryl donor. *EMBO J.* 20, 2480–2486. doi: 10.1093/emboj/20.10.2480
- Hayashi, S. I., and Lin, E. C. (1967). Purification and properties of glycerol kinase from *Escherichia coli*. *J. Biol. Chem.* 242, 1030–1035.
- Holzle, A., Fischer, S., Heyer, R., Schutz, S., Zacharias, M., Walther, P., et al. (2008). Maturation of the 5S rRNA 5' end is catalyzed *in vitro* by the endonuclease tRNase Z in the archaeon *H. volcanii*. *RNA* 14, 928–937. doi: 10.1261/rna.933208
- Jin, R. Z., Forage, R. G., and Lin, E. C. (1982). Glycerol kinase as a substitute for dihydroxyacetone kinase in a mutant of *Klebsiella pneumoniae*. *J. Bacteriol.* 152, 1303–1307.
- Mullakhanbhai, M. F., and Larsen, H. (1975). *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Arch. Microbiol.* 104, 207–214. doi: 10.1007/BF00447326
- Oren, A., and Shilo, M. (1982). Population dynamics of *Dunaliella parva* in the Dead Sea. *Limnol. Oceanogr.* 27, 201–211. doi: 10.4319/lo.1982.27.2.0201
- Sher, J., Elevi, R., Mana, L., and Oren, A. (2004). Glycerol metabolism in the extremely halophilic bacterium *Salinibacter ruber*. *FEMS Microbiol. Lett.* 232, 211–215. doi: 10.1016/S0378-1097(04)00077-1
- Sherwood, K. E., Cano, D. J., and Maupin-Furrow, J. A. (2009). Glycerol-mediated repression of glucose metabolism and glycerol kinase as the sole route of glycerol catabolism in the haloarchaeon *Haloflexax volcanii*. *J. Bacteriol.* 191, 4307–4315. doi: 10.1128/JB.00131-09
- Siebold, C., Arnold, I., Garcia-Alles, L. F., Baumann, U., and Erni, B. (2003). Crystal structure of the *Citrobacter freundii* dihydroxyacetone kinase reveals an eight-stranded alpha-helical barrel ATP-binding domain. *J. Biol. Chem.* 278, 48236–48244. doi: 10.1074/jbc.M305942200
- Waites, M. J., and Quayle, J. R. (1981). The interrelation transketolase and dihydroxyacetone synthase activities in the methylotrophic yeast *Candida*

- boidinii*. *J. Gen. Microbiol.* 124, 309–316. doi: 10.1099/00221287-124-2-309
- Weinhause, H., and Benziman, M. (1976). Phosphorylation of glycerol and dihydroxyacetone in *Acetobacter xylinum* and its possible regulatory role. *J. Bacteriol.* 127, 747–754.
- Zurbriggen, A., Jeckelmann, J. M., Christen, S., Bienossek, C., Baumann, U., and Erni, B. (2008). X-ray structures of the three *Lactococcus lactis* dihydroxyacetone kinase subunits and of a transient intersubunit complex. *J. Biol. Chem.* 283, 35789–35796. doi: 10.1074/jbc.M804893200

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 October 2013; paper pending published: 01 November 2013; accepted: 21 November 2013; published online: 16 December 2013.

Citation: Ouellette M, Makkay AM and Papke RT (2013) Dihydroxyacetone metabolism in *Haloflexax volcanii*. *Front. Microbiol.* 4:376. doi: 10.3389/fmicb.2013.00376

This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.

Copyright © 2013 Ouellette, Makkay and Papke. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Protective role of salt in catalysis and maintaining structure of halophilic proteins against denaturation

Rajeshwari Sinha and Sunil K. Khare \*

Department of Chemistry, Indian Institute of Technology Delhi, Delhi, India

**Edited by:**

Aharon Oren, The Hebrew University of Jerusalem, Israel

**Reviewed by:**

Aharon Oren, The Hebrew University of Jerusalem, Israel  
Dominique Madern, Institut de Biologie Structurale, France

**\*Correspondence:**

Sunil K. Khare, Enzyme and Microbial Biochemistry Laboratory, Department of Chemistry, Indian Institute of Technology, Delhi, MS 801 (A), Hauz Khas, New Delhi 110016, India  
e-mail: skhare@rocketmail.com; skkhare@chemistry.iitd.ac.in

Search for new industrial enzymes having novel properties continues to be a desirable pursuit in enzyme research. The halophilic organisms inhabiting under saline/ hypersaline conditions are considered as promising source of useful enzymes. Their enzymes are structurally adapted to perform efficient catalysis under saline environment wherein non-halophilic enzymes often lose their structure and activity. Haloenzymes have been documented to be polyextremophilic and withstand high temperature, pH, organic solvents, and chaotropic agents. However, this stability is modulated by salt. Although vast amount of information have been generated on salt mediated protection and structure function relationship in halophilic proteins, their clear understanding and correct perspective still remain incoherent. Furthermore, understanding their protein architecture may give better clue for engineering stable enzymes which can withstand harsh industrial conditions. The article encompasses the current level of understanding about haloadaptations and analyzes structural basis of their enzyme stability against classical denaturants.

**Keywords:** halophiles, haloadaptations, structure, denaturants, secondary structure, tertiary structure

## INTRODUCTION

Halophiles are the class of extremophiles which inhabit saline/hypersaline habitats. Halophilic proteins retain their structural and functional integrity under such high salt conditions (Oren, 2008). Certain unique structural features enable them to sustain their structure and physiological activities at high salt. These proteins, thus offer a unique model system to decipher structure function modulation under saline environment.

A perfect model that accurately explains how salts stabilize a protein is still debatable. Earlier studies on some of the extreme halophilic and haloarchaeal enzymes like *Haloarcula marismortui* malate dehydrogenases (Hm MDH) (Mevarech et al., 1977) and *Halobacterium salinarum* ferredoxins (Gafni and Werber, 1979) indicated that their enzymatic properties are fully expressed only in the presence of salt and that the gradual withdrawal of salt leads to the unfolding of protein. However, current understanding has emerged that requirements of high salt for activity and stability would rather be a restrictive definition of the halophilic proteins isolated from extreme halophiles. In many cases such as MDH from *Salinibacter ruber* (Madern and Zaccai, 2004),  $\alpha$ -amylase from *Har. hispanica* (Hutcheon et al., 2005), the enzyme is not completely inactivated in the absence of salt. Hm MDH from different studies have shown the enzyme to be stable at millimolar concentration of salts in the presence of divalent cations or coenzyme (Bonneté et al., 1994; Madern and Zaccai, 1997).

Haloadaptations of proteins viz. presence of increased number of acidic amino acid residues on protein surface, smaller hydrophobic patches as well as salt bridges between acidic and strategically positioned basic residues have been previously defined (Lanyi, 1974; Eisenberg et al., 1992; Danson and Hough, 1997; Madern et al., 2000). Since then, structural analyses have

revealed two significant differences in the characteristics of the surface of the halophilic enzymes. The first of these is that the excess of acidic residues are predominantly located on the enzyme surface forming a hydration shell that protects the enzyme from aggregation under high salinity. Secondly, the surface also displays a significant reduction in exposed hydrophobic character, which arises from a reduction in surface-exposed lysine residues. Oren (2013) recently reviewed the occurrence of acidic proteomes in halophiles for a better understanding of the modes of haloadaptation at the cellular level.

The halophilic proteins remain highly soluble in high salt milieu whereas their non-halophilic counterparts precipitate. The most appropriate model that explains the changes in solvent properties of halophilic proteins is the “solvation-stabilization model” (Ebel et al., 1999; Costenaro et al., 2002; Ebel et al., 2002). The thermodynamic basis in such cases has been well explained by Zaccai (2013). The model reflects that solubility, and changes in stability are intrinsically coupled. Extensive characterization of orthologous enzymes from extreme halophilic microorganisms, have shown that the unique adaptive feature which is shared by stable and unstable halophilic proteins is their high solubility at high salt concentration (Coquelle et al., 2010). Crystallographic analysis on halophilic and non-halophilic MDH from *Salinibacter ruber* and *Chloroflexus aurantiacus* respectively successfully established that acidic amino acids in the former were involved in disruption of pentagonally arranged network of water molecules within the hydration shell of halophilic protein (Talon et al., 2014). The acidic enrichment was attributed to an “evolutionary innovation” which enabled the protein to adapt under saline stress and suitably alter inherent protein-solvent interactions. A correlation between an increase of acidic amino acid and a

favorable change of solubility in halophilic proteins has also been established by Tadeo et al. (2009).

A good amount of data has emerged in recent years which indicate a protective role of salt in stabilizing these proteins against classical denaturants. The present article encompasses the major haloadaptations directed toward understanding the basis of stability against classical denaturants. A critical understanding of how halophilic proteins, under the influence of salt, retain structural and functional integrity amidst denaturing milieu will provide guidelines and templates for engineering stable proteins/enzymes for industrial applications.

## BASIC ASPECTS IN PROTEIN STABILITY

The effects of salt on structure and function of non-halophilic proteins have been well worked out. The presence of high salt in a protein solution will have the following implications: disturbance in local water structure around the protein; decreased propensity for intermolecular hydrogen bonds, affecting protein solubility, binding, stability and crystallization; increased surface tension of water, striping off the essential water layer from the protein surface and increased hydrophobic interactions, causing protein aggregation and precipitation. To sum up, the protein structure and consequent functions are adversely affected by high salt concentrations.

## PROTEIN DENATURATION

Denaturation of a protein refers to the loss of biological activity due to structural changes in the protein brought about by physical or chemical factors such as pH, temperature, salt, detergents, organic solvents or chaotropic agents. The secondary, tertiary or quaternary structures are largely affected upon denaturation. Some of the important mechanisms of protein denaturation are (<http://class.fst.ohio-state.edu/FST822/lectures/Denat.htm>):

- High temperature weakens the inherent bonds in protein. Further heating disrupts the hydrogen bonds within protein and new hydrogen bonds are established with surrounding water molecules thus breaking the helical structure.
- Water miscible solvent (less polar than water) lowers the dielectric constant of the system thereby strengthening electrostatic interactions among molecules. The protein largely unfolds exposing hydrophobic groups to the solvent causing aggregation and precipitation.
- Proteins are usually more soluble in dilute salt solutions because the salts in their ionic forms associate with opposite charges within the protein moiety, leading to increased hydration of the surface. However, at very high salt concentration, the increased surface tension of water generates a competition between protein and salt ions for hydration. Salts strip off the essential layer of water molecules from the protein surface eventually denaturing the protein.
- Protein denaturation by urea may occur by direct or indirect mechanisms (Lindgren and Westlund, 2010). Urea may directly interact with proteins by hydrogen bonding with the polarized areas on protein surface, weakening intermolecular bonds and protein structure. Guanidium hydrochloride (GdmCl) has a similar mechanism but is a more effective denaturant than

urea. In an indirect mechanism, urea may disturb the water structure causing destabilization of the protein.

- Acids and bases alter the pH of the solution as well as disrupt the salt bridges which are primarily stabilizing ionic interactions between opposite charged amino acid residues on protein surface. Heavy metal salts *viz.*  $Hg^{2+}$ ,  $Cd^{2+}$ ,  $Pb^{2+}$ ,  $Ag^+$  similarly disrupt salt bridges or disulfide linkages in proteins leading to an insoluble metal protein complex. SDS induced protein denaturation involved unfolding of tertiary structure and “chain expansion” (Bhuyan, 2010).

## SALT IS ESSENTIAL FOR MAINTAINING STRUCTURE AND FUNCTION OF HALOPHILIC PROTEINS

Evidently, presence of salt is a prerequisite for functioning of halophilic proteins (Mevarech et al., 2000). Large number of studies have been undertaken to investigate the role of salt in regulating structures and function of extremely/ moderately halophilic enzymes. Some of these are summarized in Table 1. The data suggests at the precise salt dependence of protein structures in halophiles. Majority of the studies indicate loss of enzymatic activity upon salt removal. These observations have been well supported by structural data. While the protein remained predominantly unfolded or randomly coiled in salt free medium, salt promoted increase in negative ellipticity and subsequent refolding. The effect of other metal ions on the activity and stability of halophilic enzymes have also been investigated. Differential roles of divalent  $Ca^{2+}$  and monovalent  $Na^+$  in preventing unfolding and regulation of catalytic activity respectively was also reported recently for *Bacillus* sp. EMB9 protease (Sinha and Khare, 2013a). Salt and divalent metal ions were reported to independently stabilize and regulate catalysis and folding of RNase H1 from *Hbt.* sp. NRC-1 (Tannous et al., 2012). In a different study, changes in tertiary structure of ferredoxin were associated with removal of  $Fe^{3+}$  (Gafni and Werber, 1979).

However, anomalies to the above generic trend cannot be ruled out. *S. ruber* MDH and *Har. hispanica* amylase remained completely active and structured even in absence of salt (Madern and Zaccari, 2004; Hutcheon et al., 2005). Likewise, glutamate dehydrogenase (GDH) from *Hbt. salinarum* was catalytically active under both low and high salt (Ishibashi et al., 2002). The exact reason for such stability is not known but it is plausible that the folded protein is structurally rigid enough to remain folded in the correct conformation and withstand non-saline environment.

## EFFECT OF DENATURANTS ON HALOPHILIC PROTEINS: PROTECTIVE ROLE OF SALTS

Haloadaptations are perceived to impart stability to halophilic proteins against denaturants. Increasing evidences have gathered now to indicate the role of salt in protecting proteins against denaturants. Studies show that in the presence of salt, secondary and tertiary structure are maintained rigidly against denaturants. However, the question remains the precise mechanism which enables this stability in halophilic enzymes. Few of the important instances have been discussed below.

**Table 1 | Effect of salt on activity and structure of halophilic enzymes.**

Enzyme	Microorganism	Salt range investigated	Structure-function correlation	Reference
Malate dehydrogenase	extreme halophilic bacteria from the Dead Sea	0–4.0 M NaCl	Salt withdrawal caused loss of enzymatic activity Distorted ellipticity with complete loss of secondary structure	Pundak and Eisenberg, 1981
NADP-glutamate dehydrogenase	<i>Haloferax mediterranei</i>	33 mM	Fluorescence spectra remained unaltered until KCl concentration was lowered to 33 mM At 33 mM KCl, time dependent decay of activity and complete fluorescence quenching	Ferrer et al., 1998
Thermolysin	<i>Bacillus thermoproteolyticus</i>	1–5 M NaCl	Activity enhanced by 15 fold in presence of 4 M NaCl Heat stability doubled in presence of 1.5 M NaCl supported by CD spectral data	Inouye et al., 1998
Transducers HtrX and HtrX1	<i>Halobacterium salinarum</i>	0.2–3.6 M NaCl or KCl	Under low-ionic-strength conditions (~0.2 M NaCl or KCl) HtrXI assumed a random coil structure while HtrX retained 85% $\alpha$ -helicity Addition of NaCl or KCl led to increase in $\alpha$ -helical characters for both	Larsen et al., 1999
Fe2-S2 ferredoxin	<i>Halobacterium salinarum</i>	0.1–4.5 M	Unfolding at low salt; time dependent loss of secondary structure Destabilization of the Fe2-S2 center of ferredoxin at low salt	Bandyopadhyay and Sonawat, 2000
Dihydrofolate reductase (DHFR)	<i>Haloferax volcanii</i>	0.12 M and 3 M KCl	<i>Hv</i> DHFR inactivation below 0.5 M KCl Secondary and tertiary structures showed analogy with the activity trend	Wright et al., 2002
Nucleoside diphosphate kinase (NDK)	<i>Natrialba magadii</i>	90 mM–3.5 M NaCl	Red shift of $\lambda_{\text{max}}$ from 333 nm (at 1.75 M NaCl) to 340 nm (without NaCl) in fluorescence spectra	Polosina et al., 2002
Isocitrate dehydrogenase (ICDH)	<i>Haloferax volcanii</i>	0–5.0 M NaCl	Dissociation of the protein at low NaCl High $\alpha$ -helical content at 4 M NaCl was indicative of a fully folded active enzyme while partially folding observed at low concentration Irreversible denaturation below a threshold concentration	Madern et al., 2004
$\alpha$ -amylase AmyH	halophilic archaeon <i>Haloarcula hispanica</i>	0–4.0 M NaCl	Retained structural and functional integrity in the absence of salt Intrinsic fluorescence suggest that it did not unfold at low salt but may get slightly more loosely folded Very stable in high-salt buffer	Hutcheon et al., 2005
Esterolytic enzyme LipC	archaeon <i>Haloarcula marismortui</i> and overexpressed in <i>E. coli</i> BL21	0–5.1 M NaCl	Far UV-CD showed maximum negative ellipticity at 3.4 M NaCl. Considerable loss of secondary structure, as salt concentration was varied away from the optimal value	Rao et al., 2009
Recombinant esterase	<i>Haloarcula marismortui</i> , (Hm EST), cloned and overexpressed in <i>E. coli</i>	0–500 mM KCl	Unfolded protein in salt-free medium Pronounced negative ellipticity, increase in $\alpha$ -helical content upon addition of KCl	Müller-Santos et al., 2009

(Continued)

**Table 1 | Continued**

Enzyme	Microorganism	Salt range investigated	Structure-function correlation	Reference
Protease	<i>Geomicrobium</i> sp. EMB2	0–10% (w/v) NaCl	Secondary structure of the protease unfolded in salt-free medium Structure regained by inclusion of 2–5% NaCl	Karan and Khare, 2011
Nep extracellular protease	<i>Natrialba magadii</i>	1.0–3.0 M NaCl	Irreversible denaturation, aggregation and loss of activity of Nep in the absence of salt; random coil structure in CD spectra Secondary conformation sufficiently folded in salinity profile	Souza et al., 2012

### EFFECT OF TEMPERATURE

Thermal stability in proteins is attributed to the combination of factors like improved core packing, increased ionic interactions, decreased hydrophobic surface area, helix stabilization and reduced conformational strain (Sinha and Khare, 2013b). Stability at high temperatures in halophilic proteins has been found to be regulated by presence of salt.  $\beta$ -lactamase from halophilic *Chromohalobacter* sp. retained ~82% of its activity after heat treatment at 100°C for 5 min (Tokunaga et al., 2004). This was indicative of a “reversible renaturation” of the lactamase induced by salt. *Hbt. salinarum* NDK also refolded back post heat treatment in a similar manner at higher concentrations of salt (Ishibashi et al., 2002). “Irreversible aggregation” may have possibly been averted due to the presence of acidic amino acid residues on the protein surface. In another study, higher thermal stability was imparted to *Hbt.* sp. SP1(1) by 4 M NaCl than 2 M (Akolkar and Desai, 2010).

### EFFECT OF CHAOTROPIC AGENTS

Investigations on the effect of urea or GdmCl on some halophilic proteins revealed that these are relatively more stable toward denaturation compared to their non-halophilic analog (Dodia et al., 2008; Karan and Khare, 2011). Protective role of salt against urea induced denaturation has been evidenced in case of NDK from *Nab. magadii* (Polosina et al., 2002). At 3.5 M NaCl, it retained complete activity even at 6 M urea. This unique stability was attributed to the possible formation of strong intersubunit contacts within the quaternary structure of the halophilic protein which may have imparted stability to the subdomains and prevented denaturation. Fluorescence spectral analysis established that *Har. hispanica*  $\alpha$ -amylase, at 4 M NaCl remained fully folded and conformationally active even in the presence of 6 M urea (Hutcheon et al., 2005). The protein was less structured in absence of salt and gradually lost its overall structure upon increasing urea concentrations. It is likely that charge screening effect imposed by salt ions may have prevented the urea from coming in close vicinity of the polar patches on halophilic protein thereby averting their interaction and subsequent denaturation.

### EFFECT OF ORGANIC SOLVENTS

Organic solvents behave as mild chaotropic agents. They disrupt hydrogen bonds between protein subunits and reduce the catalytic efficiency by affecting the critical water concentration at the

active site. Solvent stability is increasingly being evidenced as a generic trait among halophilic enzymes (Gupta and Khare, 2009). The presence of high salt reduces the water activity significantly. Halophilic enzymes are thus uniquely adapted to function in low water/ non-aqueous media (Kumar and Khare, 2012). Significant solvent stability among halophilic enzymes has been reported by different groups (Shafiei et al., 2011; Li and Yu, 2012; Li et al., 2012; Sinha and Khare, 2013a). However, very little has been investigated about the structure of halophilic enzymes in organic solvents.

Recently, the solvent induced conformational changes have been assessed. Fluorescence investigations of halophilic alcohol dehydrogenase (ADH2) from *Hfx. volcanii* affirmed that salt influences the correct folding of proteins in organic solvents (Alsafadi and Paradisi, 2013). The  $\alpha$ -helical content of *Geomicrobium* sp. protease remained unaffected in 50% (v/v) n-hexane and n-decane in presence of 5% (w/v) NaCl (Karan and Khare, 2011). Withdrawal of salt caused loss of  $\alpha$ -helical structure.

### EFFECT OF MUTATIONS

Preliminary understanding about the mechanism of salt supported structural protection has also emerged from site directed mutagenesis (SDM) experiments. The importance of acidic peptide motif in halophilic enzymes to withstand saline stress was shown by Evilla and Hou (2006). Presence of insertion peptide in extreme halophile *Hbt.* sp. NRC-1 cysteinyl tRNA synthetase (CysRS) showed strong salt dependence, and enhanced enzyme stability at low salt. Deletion of the motif reduced aminoacylation efficiency.

Extensive SDM on halophilic MDH from *Har. marismortui* produced a mutant which was more “halophilic” than the wild type enzyme (Madern et al., 1995). Its structure was solved allowing highlighting the important role of protein solvent interactions in the stabilization of a halophilic protein (Richard et al., 2000). Others SDM studies on Hm MalDH have also demonstrated the important role of anion binding site in the stabilization process (Madern et al., 2000; Irimia et al., 2003; Madern and Ebel, 2007). Mutation of several solvent exposed acidic amino acid residues with lysine in *Hfx. mediterranei* glucose dehydrogenase resulted in proteins displaying a slightly less halophilic behavior than the wild type enzyme (Esclapez et al., 2007). Mutation studies on glucose dehydrogenase and isocitrate dehydrogenase from extremely halophilic Archaea *Hfx. mediterranei* and *Hfx. volcanii* have also

recently contributed to understanding of the molecular basis of salt tolerance for halophilic adaptation (Esclapez et al., 2013).

Replacement of 7 Ser residues in thermolysin by Asp improved stability and activity of its mutants (Takita et al., 2008). In the presence of 4 M NaCl, hydrolytic activity of all mutants increased 17–19 folds. Halophilic characteristics were imparted to *Pseudomonas* NDK by replacing 2 Ala residues at its C-terminal end with acidic residues (Glu-Glu) (Tokunaga et al., 2008). Conversely, introduction of Ala-Ala into *Halomonas* sp. 593 NDK (Ha NDK) in place of Glu-Glu caused loss of enzyme halophilicity and critically affected the enzyme properties as well as secondary structure. While the wild type Ha NDK remained stable against dilution induced inactivation, the mutant form was easily irreversibly destabilized by dilution, regardless of presence of salt.

Above studies highlight that protein halophilicity bears a strong correlation with (i) the necessity and importance of acidic amino acid residues on its surface and (ii) the presence of salt which serves important role in restoring/preserving functionality in a halophilic enzyme. On the basis of available data, we may infer that there possibly exists a subset among halophilic proteins which successfully retain structure and activity even under unfavorable conditions. Although, this cannot be generalized, the ability of moderately halophilic enzymes to withstand denaturing environments has not been explored much and provides ample scope for future research.

## CONCLUSION

The article reviews the present understanding about the responses of halophilic enzymes in solution toward chaotropic reagents and denaturants. Salt is essential in maintaining native structure of halophilic proteins. Sufficient experimental evidences conclude that salt significantly contributes in the modulation of the protein structure/activity toward different chaotropic conditions. The salt induced protective effect on the structure against chaotropic agents, temperature and solvents as well the corresponding structural transitions establish a unique structure-function correlation in halophilic enzymes. Comprehending their differential behavior and stability under harsher conditions could lead to better understanding about the biochemical and biophysical characteristics of these proteins and their exploitation for applications in biocatalysis or biotransformation under saline or low water conditions.

## ACKNOWLEDGMENTS

The financial support to the study by the Department of Biotechnology and research fellowship to Rajeshwari Sinha by the Council of Scientific and Industrial Research (Govt. of India) is gratefully acknowledged.

## REFERENCES

- Akolkar, A. V., and Desai, A. J. (2010). Catalytic and thermodynamic characterization of protease from *Halobacterium* sp. SP1(1). *Res. Microbiol.* 161, 355–362. doi: 10.1016/j.resmic.2010.04.005
- Alsaifadi, D., and Paradisi, F. (2013). Effect of organic solvents on the activity and stability of halophilic alcohol dehydrogenase (ADH2) from *Haloflexax volcanii*. *Extremophiles* 17, 115–122. doi: 10.1007/s00792-012-0498-0
- Bandyopadhyay, A. K., and Sonawat, H. M. (2000). Salt dependent stability and unfolding of [Fe2-S2] ferredoxin of *Halobacterium salinarum*: spectroscopic investigations. *Biophys. J.* 79, 501–510. doi: 10.1016/S0006-3495(00)76312-0
- Bhuyan, A. K. (2010). On the mechanism of SDS-induced protein denaturation. *Biopolymers* 93, 186–199. doi: 10.1002/bip.21318
- Bonneté, F., Madern, D., and Zaccai, G. (1994). Stability against denaturation mechanisms in halophilic malate dehydrogenase “adapt” to solvent conditions. *J. Mol. Biol.* 244, 436–447. doi: 10.1006/jmbi.1994.1741
- Coquelle, N., Talon, R., Juers, D. H., Girard, É., Kahn, R., and Madern, D. (2010). Gradual adaptive changes of a protein facing high salt concentrations. *J. Mol. Biol.* 404, 493–505. doi: 10.1016/j.jmb.2010.09.055
- Costenaro, L., Zaccai, G., and Ebel, C. (2002). Link between protein-solvent and weak protein-protein interactions gives insight into halophilic adaptation. *Biochemistry* 41, 13245–13252. doi: 10.1021/bi025830z
- Danson, M. J., and Hough, D. W. (1997). The structural basis of protein halophilicity. *Comp. Biochem. Physiol. A Physiol.* 117, 307–312. doi: 10.1016/S0300-9629(96)00268-X
- Dodia, M. S., Bhimani, H. G., Rawal, C. M., Joshi, R. H., and Singh, S. P. (2008). Salt dependent resistance against chemical denaturation of alkaline protease from a newly isolated haloalkaliphilic *Bacillus* sp. *Bioresour. Technol.* 99, 6223–6227. doi: 10.1016/j.biortech.2007.12.020
- Ebel, C., Costenaro, L., Pascu, M., Faou, P., Kernel, B., Proust-De Martin, F., et al. (2002). Solvent interactions of halophilic malate dehydrogenase. *Biochemistry* 41, 13234–13244. doi: 10.1021/bi0258290
- Ebel, C., Faou, P., Kernel, B., and Zaccai, G. (1999). Relative role of anions and cations in the stabilization of halophilic malate dehydrogenase. *Biochemistry* 38, 9039–9047. doi: 10.1021/bi9900774
- Eisenberg, H., Mevarech, M., and Zaccai, G. (1992). Biochemical, structural, and molecular genetic aspects of halophilism. *Adv. Protein. Chem.* 43, 1–62. doi: 10.1016/S0065-3233(08)60553-7
- Esclapez, J., Camacho, M., Pire, C., and Bonete, M. J. (2013). “Biochemical analysis of halophilic dehydrogenases altered by site-directed mutagenesis,” in *Genetic Manipulation of DNA and Protein - Examples From Current Research*, ed D. Figurski (InTech). doi: 10.5772/34565. Available online at: [http://cdn.intechopen.com/pdfs/42535/InTech-Biochemical\\_analysis\\_of\\_halophilic\\_dehydrogenases\\_altered\\_by\\_site\\_directed\\_mutagenesis.pdf](http://cdn.intechopen.com/pdfs/42535/InTech-Biochemical_analysis_of_halophilic_dehydrogenases_altered_by_site_directed_mutagenesis.pdf)
- Esclapez, J., Pire, C., Bautista, V., Martínez-Espínosa, R. M., Ferrer, J., and Bonete, M. J. (2007). Analysis of acidic surface of *Haloflexax mediterranei* glucose dehydrogenase by site-directed mutagenesis. *FEBS Lett.* 581, 837–842. doi: 10.1016/j.febslet.2007.01.054
- Evilla, C., and Hou, Y. (2006). Acquisition of an insertion peptide for efficient aminoacylation by a halophile tRNA synthetase. *Biochemistry* 45, 6835–6845. doi: 10.1021/bi0521386
- Ferrer, J., Cremades, R., Pire, C., and Bonete, M. J. (1998). Fluorescence and quenching comparative studies of halophilic and bovine glutamate dehydrogenase. *J. Photochem. Photobiol. B.* 47, 148–154. doi: 10.1016/S1011-1344(98)00214-0
- Gaffni, A., and Werber, M. M. (1979). Ferredoxin from *Halobacterium* of the Dead Sea. Structural properties revealed by fluorescence techniques. *Arch. Biochem. Biophys.* 196, 363–370. doi: 10.1016/0003-9861(79)90588-5
- Gupta, A., and Khare, S. K. (2009). Enzymes from solvent tolerant microbes: useful biocatalysts for non-aqueous enzymology. *Crit. Rev. Biotechnol.* 29, 44–54. doi: 10.1080/07388550802688797
- Hutcheon, G. W., Vasisht, N., and Bolhuis, A. (2005). Characterisation of a highly stable  $\alpha$ -amylase from the halophilic archaeon *Halocarcula hispanica*. *Extremophiles* 9, 487–495. doi: 10.1007/s00792-005-0471-2
- Inouye, K., Kuzuya, K., and Tonomura, B. (1998). Sodium chloride enhances markedly the thermal stability of thermolysin as well as its catalytic activity. *Biochim. Biophys. Acta* 1388, 209–214. doi: 10.1016/S0167-4838(98)00189-7
- Irimia, A., Ebel, C., Madern, D., Richard, S. B., Cosenza, L. W., Zaccai, G., et al. (2003). The oligomeric states of *Halocarcula marismortui* malate dehydrogenase are modulated by solvent components as shown by crystallographic and biochemical studies. *J. Mol. Biol.* 326, 859–873. doi: 10.1016/S0022-2836(02)01450-X
- Ishibashi, M., Arakawa, T., Philo, J. S., Sakashita, K., Yonezawa, Y., Tokunaga, H., et al. (2002). Secondary and quaternary structural transition of the halophilic archaeon nucleoside diphosphate kinase under high- and low-salt conditions. *FEMS Microbiol. Lett.* 216, 235–241. doi: 10.1111/j.1574-6968.2002.tb11441.x
- Karan, R., and Khare, S. K. (2011). Stability of haloalkaliphilic *Geomicrombium* sp. protease modulated by salt. *Biochemistry (Mosc.)* 76, 686–693. doi: 10.1134/S0006297911060095

- Kumar, S., and Khare, S. K. (2012). Purification and characterization of maltooligosaccharide-forming alpha-amylase from moderately halophilic *Marinobacter* sp. EMB8. *Bioresour. Technol.* 116, 247–251. doi: 10.1016/j.biortech.2011.11.109
- Lanyi, J. K. (1974). Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriol. Rev.* 38, 272–290.
- Larsen, R. W., Yang, J., Hou, S., Helms, M. K., Jameson, D. M., and Alam, M. (1999). Spectroscopic characterization of two soluble transducers from the archaeon *Halobacterium salinarum*. *J. Protein Chem.* 18, 269–275. doi: 10.1023/A:1021031227197
- Li, X., Wang, H., Li, T., and Yu, H. (2012). Purification and characterization of an organic solvent-tolerant alkaline cellulase from a halophilic isolate of *Thalassobacillus*. *Biotechnol. Lett.* 34, 1531–1536. doi: 10.1007/s10529-012-0938-z
- Li, X., and Yu, H. Y. (2012). Purification and characterization of novel organic-solvent-tolerant  $\beta$ -amylase and serine protease from a newly isolated *Salimicrobium halophilum* strain LY20. *FEMS Microbiol. Lett.* 329, 204–211. doi: 10.1111/j.1574-6968.2012.02522.x
- Lindgren, M., and Westlund, P. O. (2010). On the stability of chymotrypsin inhibitor 2 in a 10 M urea solution. The role of interaction energies for urea-induced protein denaturation. *Phys. Chem. Chem. Phys.* 12, 9358–9366. doi: 10.1039/b925726h
- Madern, D., Camacho, M., Rodríguez-Arnedo, A., Bonete, M. J., and Zaccai, G. (2004). Salt-dependent studies of NADP-dependent isocitrate dehydrogenase from the halophilic archaeon *Haloferax volcanii*. *Extremophiles* 8, 377–384. doi: 10.1007/s00792-004-0398-z
- Madern, D., and Ebel, C. (2007). Influence of an anion-binding site in the stabilization of halophilic malate dehydrogenase from *Haloarcula marismortui*. *Biochimie* 89, 981–987. doi: 10.1016/j.biochi.2007.03.008
- Madern, D., Ebel, C., and Zaccai, G. (2000). Halophilic adaptation of enzymes. *Extremophiles* 4, 91–98. doi: 10.1007/s007920050142
- Madern, D., Pfister, C., and Zaccai, G. (1995). Mutation at a single acidic amino acid enhances the halophilic behaviour of malate dehydrogenase from *Haloarcula marismortui* in physiological salts. *Eur. J. Biochem.* 230, 1088–1095. doi: 10.1111/j.1432-1033.1995.tb20659.x
- Madern, D., and Zaccai, G. (1997). Stabilisation of halophilic malate dehydrogenase from *Haloarcula marismortui* by divalent cations. *Eur. J. Biochem.* 249, 607–611. doi: 10.1111/j.1432-1033.1997.00607.x
- Madern, D., and Zaccai, G. (2004). Molecular adaptation: the malate dehydrogenase from the extreme halophilic bacterium *Salinibacter ruber* behaves like a non-halophilic protein. *Biochimie* 86, 295–303. doi: 10.1016/j.biochi.2004.04.004
- Mevarech, M., Eisenberg, H., and Neumann, E. (1977). Malate dehydrogenase isolated from extremely halophilic bacteria of the Dead Sea. 1. Purification and molecular characteristics. *Biochemistry* 16, 3781–3785. doi: 10.1021/bi00636a009
- Mevarech, M., Frolow, F., and Gloss, L. M. (2000). Halophilic enzymes: proteins with a grain of salt. *Biophys. Chem.* 86, 155–164. doi: 10.1016/S0301-4622(00)00126-5
- Müller-Santos, M., de Souza, E. M., Pedrosa, F. O., Mitchell, D. A., Longhi, S., Carrière, F., et al. (2009). First evidence for the salt-dependent folding and activity of an esterase from the halophilic archaea *Haloarcula marismortui*. *Biochim. Biophys. Acta* 1791, 719–729. doi: 10.1016/j.bbaplip.2009.03.006
- Oren, A. (2008). Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Syst.* 4, 2. doi: 10.1186/1746-1448-4-2
- Oren, A. (2013). Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. *Front. Microbiol.* 4:315. doi: 10.3389/fmicb.2013.00315
- Polosina, Y. Y., Zamyatkin, D. F., Kostyukova, A. S., Filimonov, V. V., and Fedorov, O. V. (2002). Stability of *Natrialba magadii* NDP kinase? comparisons with other halophilic proteins. *Extremophiles* 6, 135–142. doi: 10.1007/s007920100232
- Pundak, S., and Eisenberg, H. (1981). Structure and activity of malate dehydrogenase of the extreme halophilic bacteria of the Dead Sea. 1. Conformation and interaction with water and salt between 5 M and 1 M NaCl concentration. *Eur. J. Biochem.* 118, 463–470. doi: 10.1111/j.1432-1033.1981.tb05542.x
- Rao, L., Zhao, X., Pan, F., Li, Y., Xue, Y., Ma, Y., et al. (2009). Solution behavior and activity of a halophilic esterase under high salt concentration. *PLoS ONE* 4:e6980. doi: 10.1371/journal.pone.0006980
- Richard, S. B., Madern, D., Garcin, E., and Zaccai, G. (2000). Halophilic adaptation: novel solvent protein interactions observed in the 2.9 and 2.6 Å resolution structures of the wild type and a mutant of malate dehydrogenase from *Haloarcula marismortui*. *Biochemistry* 39, 992–1000. doi: 10.1021/bi991001a
- Shafiei, M., Ziae, A. A., and Amoozegar, M. A. (2011). Purification and characterization of an organic-solvent-tolerant halophilic  $\alpha$ -amylase from the moderately halophilic *Nesterenkonia* sp. strain F. *J. Ind. Microbiol. Biotechnol.* 38, 275–281. doi: 10.1007/s10295-010-0770-1
- Sinha, R., and Khare, S. K. (2013a). Characterization of detergent compatible protease of a halophilic *Bacillus* sp. EMB9?: differential role of metal ions in stability and activity. *Bioresour. Technol.* 145, 357–361. doi: 10.1016/j.biortech.2012.11.024
- Sinha, R., and Khare, S. K. (2013b). “Thermostable proteases,” in *Thermophilic Microbes in Environmental and Industrial Biotechnology*, eds T. Satyanarayana, J. Littlechild, and Y. Kawarabayasi (Dordrecht: Springer), 859–880. doi: 10.1007/978-94-007-5899-5\_32
- Souza, T. A., Okamoto, D. N., Ruiz, D. M., Oliveira, L. C., Kondo, M. Y., Tersario, I. L., et al. (2012). Correlation between catalysis and tertiary structure arrangement in an archaeal halophilic subtilase. *Biochimie* 94, 798–805. doi: 10.1016/j.biochi.2011.11.011
- Tadeo, X., López-Méndez, B., Trigueros, T., Laín, A., Castaño, D., and Millet, O. (2009). Structural basis for the aminoacid composition of proteins from halophilic archaea. *PLoS Biol.* 7:e1000257. doi: 10.1371/journal.pbio.1000257
- Takita, T., Aono, T., Sakurama, H., Itoh, T., Wada, T., Minoda, M., et al. (2008). Effects of introducing negative charges into the molecular surface of thermolysin by site-directed mutagenesis on its activity and stability. *Biochim. Biophys. Acta* 1784, 481–488. doi: 10.1016/j.bbapap.2007.12.004
- Talon, R., Coquelle, N., Madern, D., and Girard, E. (2014). An eperimental point of view on hydration/ solvation in halophilic proteins. *Front. Microbiol.* 5:66. doi: 10.3389/fmicb.2014.00066
- Tannous, E., Yokoyama, K., You, D. J., Koga, Y., and Kanaya, S. (2012). A dual role of divalent metal ions in catalysis and folding of RNase H1 from extreme halophilic archaeon *Halobacterium* sp. NRC-1. *FEBS Open Bio.* 2, 345–352. doi: 10.1016/j.fob.2012.10.003
- Tokunaga, H., Arakwa, T., and Tokunaga, M. (2008). Engineering of halophilic enzymes: two acidic amino acid residues at the carboxy-terminal region confer halophilic characteristics to *Halomonas* and *Pseudomonas* nucleoside diphosphate kinases. *Protein Sci.* 17, 1603–1610. doi: 10.1110/ps.035725.108
- Tokunaga, H., Ishibashi, M., Arakawa, T., and Tokunaga, M. (2004). Highly efficient renaturation of  $\beta$ -lactamase isolated from moderately halophilic bacteria. *FEBS Lett.* 558, 7–12. doi: 10.1016/S0014-5793(03)01508-4
- Wright, D. B., Banks, D. D., Lohman, J. R., Hilsenbeck, J. L., and Gloss, L. M. (2002). The effect of salts on the activity and stability of *Escherichia coli* and *Haloferax volcanii* dihydrofolate reductases. *J. Mol. Biol.* 323, 327–344. doi: 10.1016/S0022-2836(02)00916-6
- Zaccai, G. (2013). Hydration shells with a pinch of salt. *Biopolymers* 99, 233–238. doi: 10.1002/bip.22154

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Received: 30 December 2013; accepted: 27 March 2014; published online: 09 April 2014.*

*Citation: Sinha R and Khare SK (2014) Protective role of salt in catalysis and maintaining structure of halophilic proteins against denaturation. Front. Microbiol. 5:165. doi: 10.3389/fmicb.2014.00165*

*This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.*

*Copyright © 2014 Sinha and Khare. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# An experimental point of view on hydration/solvation in halophilic proteins

Romain Talon<sup>1,2,3†</sup>, Nicolas Coquelle<sup>1,2,3†</sup>, Dominique Madern<sup>1,2,3\*</sup> and Eric Girard<sup>1,2,3\*</sup>

<sup>1</sup> Institut de Biologie Structurale, Université Grenoble Alpes, Grenoble, France

<sup>2</sup> CEA, DSV, Institut de Biologie Structurale, Grenoble, France

<sup>3</sup> Institut de Biologie Structurale, Centre National de la Recherche Scientifique, Grenoble, France

**Edited by:**

R. Thane Papke, University of Connecticut, USA

**Reviewed by:**

Aharon Oren, The Hebrew University of Jerusalem, Israel  
Ida Helene Steen, University of Bergen, Norway

**\*Correspondence:**

Dominique Madern and Eric Girard,  
Equipe ELMA, Institut de Biologie Structurale, 6 Rue Jules Horowitz,  
38000 Grenoble, France  
e-mail: dominique.madern@ibs.fr;  
eric.girard@ibs.fr

<sup>†</sup>These authors have contributed equally to this work.

Protein-solvent interactions govern the behaviors of proteins isolated from extreme halophiles. In this work, we compared the solvent envelopes of two orthologous tetrameric malate dehydrogenases (MalDHs) from halophilic and non-halophilic bacteria. The crystal structure of the MalDH from the non-halophilic bacterium *Chloroflexus aurantiacus* (*Ca* MalDH) solved, *de novo*, at 1.7 Å resolution exhibits numerous water molecules in its solvation shell. We observed that a large number of these water molecules are arranged in pentagonal polygons in the first hydration shell of *Ca* MalDH. Some of them are clustered in large networks, which cover non-polar amino acid surface. The crystal structure of MalDH from the extreme halophilic bacterium *Salinibacter ruber* (*Sr*) solved at 1.55 Å resolution shows that its surface is strongly enriched in acidic amino acids. The structural comparison of these two models is the first direct observation of the relative impact of acidic surface enrichment on the water structure organization between a halophilic protein and its non-adapted counterpart. The data show that surface acidic amino acids disrupt pentagonal water networks in the hydration shell. These crystallographic observations are discussed with respect to halophilic protein behaviors in solution

**Keywords:** halophilic, solvation, hydration, water pentagon, malate dehydrogenase, acidic proteins, adaptation, *salinibacter*

## INTRODUCTION

*Salinibacter ruber* (*Sr*) is a halophilic bacterium that was isolated from saltern crystallizer ponds in Spain (Antón et al., 2002). In contrast to most bacterial species that equilibrate osmotic pressure with compatible solute, *S. ruber* accumulates high KCl concentration within its cytoplasm, an adaptive strategy similar to that of haloarchaea (*Halobacteriaceae*) (Oren, 2002). *S. ruber* genome sequence has revealed some interesting characteristics related to haloadaption: numerous lateral gene transfers from haloarchaea and a mean *pI*-value of 5.2 of its whole proteome (Mongodin et al., 2005). This proteomic *pI* shift toward low values, which is typical in haloarchaea, is the consequence of an enrichment of Asp and Glu residues and is considered an adaptive signature of proteins facing high salt concentration (Oren, 2013). However this explanation has been recently challenged by the characterization of a bacterium (*Halorodospira*) that does not accumulate high KCl concentration in its cytoplasm and has nonetheless a high acidic proteome (Deole et al., 2013). Among the few cytoplasmic enzymes isolated from *S. ruber* (Bonete et al., 2003; Madern and Zaccai, 2004), the tetrameric malate dehydrogenase (MalDH) remains the most extensively characterized, at the biochemical and structural level (Coquelle et al., 2010). As observed for non-halophilic counterparts, this halophilic enzyme does not require salt to maintain its conformational stability. However, the *Sr* MalDH structure revealed an acidic amino acids enriched surface, typical to that observed for a halophilic enzyme, which is responsible for a favorable change of solubility in high concentration of salts (Coquelle et al., 2010).

According to the solvation-stabilization model for halophilic protein (Madern et al., 2000; reviewed in Zaccai, 2013), high salt concentrations exert a major selective pressure through a strong impact on protein solubility. In order to compete against this deleterious effect of salts, halophilic proteins stay highly soluble by maintaining a solvation envelope composition as close as possible as the composition of the bulk. This model is based on biophysical measurements that have shown that a halophilic protein recruits a solvation envelope of high ionic concentration (Costenaro et al., 2002; Ebel et al., 2002). In the solvation-stabilization model, surface acidic amino acids are suggested to be responsible for this particular solvent organization. Even if several structures of halophilic protein have been solved (Frolow et al., 1996; Richard et al., 2000; Bieger et al., 2003; Irimia et al., 2003; Zeth et al., 2004; Besir et al., 2005; Britton et al., 2006; Winter et al., 2009; Yamamura et al., 2009; Wende et al., 2010; Bracken et al., 2011), attempts to describe how the solvation shell of a halophilic protein interacts with acidic residues using X-rays crystallography is still a challenge.

In our follow-up crystallographic study on *Sr* MalDH (Coquelle et al., 2010); we determined the direct effect on solvent organization due to its acidic surface, by using a comparison with a non-halophilic counterpart. For this purpose, we solved *de novo* the crystal structure of the non-halophilic *Chloroflexus aurantiacus* (*Ca*) MalDH at 1.7 Å resolution. It allowed the determination of a hydration shell consisting in 945 water molecules, which cluster themselves in large networks of structured water through pentameric/hexameric polygons. Direct and indirect effects of acidic

amino acids substitutions, avoiding the formation of structured water in *Sr* MalDH are described here through the comparison with *Ca* MalDH. The data are analyzed with respect to the solvation-stabilization model for halophilic protein. In particular, we underline that difference in hydration-solvation characteristics should always be kept in mind while analyzing the solvation layer of a halophilic protein, using X-ray crystallography, or any other techniques.

## MATERIALS AND METHODS

### PROTEIN PRODUCTION AND PURIFICATION

*Ca* MalDH overexpression was done accordingly to Dalhus et al. (2002). The cells were lysed by sonication in a 50 mM Tris-HCl buffered at pH 7. The crude extract was incubated for half an hour at 70°C and centrifuged for 15 min at 17,000 g. The soluble portion of the extract was loaded on a Q sepharose column equilibrated in 50 mM Tris-HCl buffer at pH 7. The protein was eluted using a linear gradient of 0–1 M NaCl. Fractions containing *Ca* MalDH were extensively dialyzed against 50 mM potassium phosphate buffer (pH 7) and deposited on a hydroxyapatite column equilibrated with the same buffer. The enzyme was eluted with a linear gradient of 50–1000 mM ammonium phosphate. The active fractions were pooled and concentrated by centrifugation using an Amicon PM30. They were deposited on a Sephadryl S300 gel filtration column (1 × 100 cm) and then eluted using an isocratic buffer of 50 mM Tris-HCl buffered at pH 7. The purified fractions were concentrated at 20 mg/ml and stored at 4°C.

### CRYSTALLIZATION

Crystallization was performed by vapor diffusion using the hanging-drop method at 293 K. Native *Ca* MalDH crystals ( $\approx 500 \times 400 \times 400 \mu\text{m}^3$ ) were grown within 2 days by mixing 1.5  $\mu\text{L}$  of 20 mg· $\text{mL}^{-1}$  protein solution and 1.5  $\mu\text{L}$  of 4–14% PEG 400, 100 mM sodium acetate buffer at pH 4.6 and 40 mM cadmium acetate reservoir solution. *Ca* MalDH derivative crystals were obtained by a 10 s soaking of a native crystal in a 2.0  $\mu\text{L}$  solution equivalent to the mother liquor containing 100 mM of GdHPDO3A lanthanide complex (Girard et al., 2003). Then the crystal was quickly back-soaked in 2.0  $\mu\text{L}$  of the corresponding reservoir solution without the lanthanide complex.

Prior to data collection, native and derivative crystals were cryo-cooled in liquid nitrogen using mother liquor containing 25% PEG 400 as cryo-protectant.

### DATA COLLECTION AND DATA PROCESSING

Gd-derivative data were collected on a Nonius FR591 X-Ray home source (1.541 Å). Native data were collected on the FIP-BM30A beamline at the ESRF (Grenoble, France) with the X-ray beam wavelength set to 0.979 Å. Diffraction frames were integrated using the program XDS (Kabsch, 2010) and the integrated intensities were scaled and merged using the CCP4 programs SCALA and TRUNCATE (Winn et al., 2011) respectively. A summary of the processing statistics is given in Table 1.

*Ca* MalDH crystals belong to the P3<sub>1</sub>21 space group with one A-D dimer per asymmetric unit leading to a solvent content of 49.5%.

**Table 1 | Data collection and processing statistics.**

	Data set	
	GdHPDO3A derivative	Native
$\lambda$ (Å)	1.541	0.979
Space group	P3121	
Cell parameter (Å)	$a = 106.77, c = 103.53$	$a = 106.23, c = 102.57$
Resolution (Å)	19.63–1.90 (2.00–1.90)	19.70–1.70 (1.79–1.70)
Unique reflexions	54120 (7772)	68673 (10178)
$R_{\text{merge}}$ (%) <sup>a</sup>	5.0 (23.7)	8.4 (35.6)
$R_{\text{pim}}$ (%) <sup>b</sup>	1.9 (8.5)	4.4 (20.3)
$R_{\text{ano}}$ (%) <sup>c</sup>	3.6 (9.6)	
$I/I(\sigma)$ (d)	12.4 (3.3)	6.3 (2.2)
Completeness (%)	98.8 (99.1)	93.7 (95.8)
Multiplicity	10.6 (9.8)	4.2 (3.9)

<sup>a</sup>  $R_{\text{merge}} = \sum_h \sum_i |\bar{I}(h) - I_i(h)| / \sum_h \sum_i |I_i(h)|$  where  $I_i(h)$  is the *i*th measurement of reflection  $h$  and  $\bar{I}(h)$  is the mean measurement of reflection  $h$ .

<sup>b</sup>  $R_{\text{p.i.m.}} = \sum_h \left( \frac{1}{(N-1)} \right)^{1/2} \sum_i |I_i(h) - \bar{I}(h)| / \sum_h \sum_i |I_i(h)|$ . This indicator, which describes the precision of the averaged measurement, is most relevant. (Weiss, 2001).

<sup>c</sup>  $R_{\text{ano}} = \sum_h |\bar{I}^+(h) - \bar{I}^-(h)| / \sum_h |\bar{I}^+(h) + \bar{I}^-(h)|$  where  $\bar{I}^+(h)$  and  $\bar{I}^-(h)$  are the mean intensities of a Friedel mate.

<sup>d</sup>  $I/\sigma(I)$  is the signal-to-noise ratio for merged intensities.

### EXPERIMENTAL SIRAS PHASING

*Ca* MalDH structure was determined *de novo* by the SIRAS (Single Isomorphous Replacement with Anomalous Scattering) method. As shown in Table 1, the high value of Rano clearly indicated the presence of GdHPDO3A complex binding sites, which was then confirmed by inspection of the anomalous Patterson map. Gadolinium positions were determined within the asymmetric unit using the program SHELXD (Sheldrick, 2010). Heavy-atom refinement and initial phasing were performed using the program SHARP (Bricogne et al., 2003). Phases from SHARP were improved by density modification using the CCP4 program DM (Cowtan and Main, 1996) leading to figures of merit of 0.235 and 0.793 after SHARP and density modification respectively. Automatic model building was performed with the program BUCCANEER (Cowtan, 2006) leading to an initial model consisting in 552 over the expected 618 A-D dimer residues.

### REFINEMENT AND WATER MOLECULES BUILDING

The model was manually completed and improved in COOT (Emsley et al., 2010) prior to refinement with PHENIX (Adams et al., 2010). This model was then optimized through iterative rounds of refinement and model building. At the end stages of the refinement, TLS was used with TLS-groups determined with the TLSMD server (Brünger, 1992; Painter and Merritt, 2006a,b). The 1.7 Å resolution *Ca* MalDH final model consists in the complete (N-terminus, C-terminus and catalytic loop) residues sequence for each monomer of the *Ca* MalDH A-D dimer. The analysis of this final model (Table 2) showed no residues in disallowed regions of the Ramachandran plot (99.7% in preferred regions).

**Table 2 | Refinement statistics and model quality of the *Ca* MalDH structure.**

PDB Code	4CL3
Resolution (Å)	19.70–1.70
Rwork (%) <sup>a</sup>	17.02
Rfree (%)	21.01
Number of reflexion used	67,827
<b>ATOMIC COMPOSITION</b>	
Protein	4694
Water	945
Ions	12
Ligands	55
Res.out of Ramachandran (%)	0.30
<b>GLOBAL STANDARD DEVIATION</b>	
Bond length (Å)	0.010
Bond angle (°)	1.292
<b>BFACTOR VALUES</b>	
Mean protein Bfactor (Å <sup>2</sup> )	21.50
Min protein Bfactor (Å <sup>2</sup> )	9.17
Max protein Bfactor (Å <sup>2</sup> )	109.92
Mean water Bfactor (Å <sup>2</sup> )	36.04
Mean ions Bfactor (Å <sup>2</sup> )	34.55
Mean ligand Bfactor (Å <sup>2</sup> )	32.47

<sup>a</sup> $R = \sum_h |F_o - F_c| / \sum_h |F_o|$  where  $F_o$  and  $F_c$  are the observed and calculated structure factor amplitudes of reflection  $h$  respectively. Rfree (Brünger, 1992) is the  $R$  for the test reflection data set for cross validation (5% of excluded reflections). Rwork is the  $R$  for the working reflection data set.

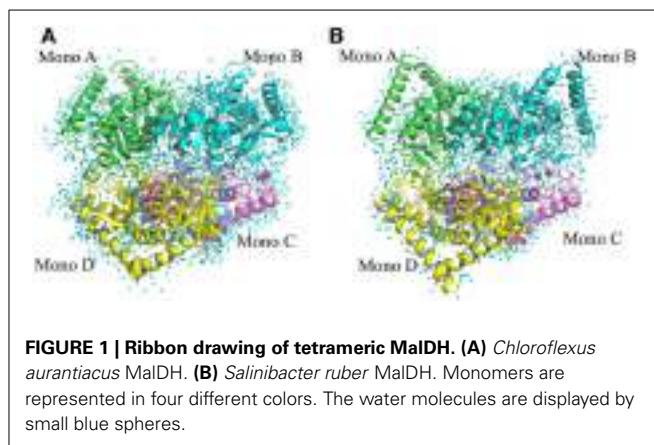
In order to precisely assign the 945 water molecules in the model, we allowed the PHENIX program to automatically build solvent molecule up to 5.0 Å above the protein surface, with a distance of 1.7–3.0 Å between two water molecules or between a water molecules and the coordinated residue and only if the 1.0 σ contoured  $2F_o - F_c$  electron density map was interpretable. At the end, each water molecule was manually verified in COOT.

All the figures were made by using the Pymol program: The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC. All electrostatic calculations were performed using the Pymol plugin for APBS (Baker et al., 2001).

## RESULTS

### QUALITY OF *Ca* AND *Sr* MALDH MODELS

The structure of *Ca* MalDH enzyme was determined at 1.7 Å resolution using SIRAS phasing. The asymmetric unit contains a dimer, the physiological tetramer being generated by the crystal symmetry operators of the P3<sub>1</sub>21 space group (Figure 1). This dimer delineated A-D will serve as the reference for all comparisons through this study. Our *Ca* MalDH model (4BGT) does not present major fold difference compared to the previously deposited (PDB accession code: 1GUY) structure (Dalhus et al., 2002), as confirmed by a root-mean-square deviation (RMSD) value of 0.42 Å for 594 A-D dimer superimposed residues. Moreover, the mobile loop (residues 83–89, following the linear numbering of 4BGT) covering the catalytic site, as well as the residues of the N- and C- termini have been modeled in each



**FIGURE 1 | Ribbon drawing of tetrameric MalDH. (A)** *Chloroflexus aurantiacus* MalDH. **(B)** *Salinibacter ruber* MalDH. Monomers are represented in four different colors. The water molecules are displayed by small blue spheres.

monomer of this new *Ca* MalDH structure. The detailed analysis of *Ca* MalDH fold and stabilization mechanism based on the 1GUY model has previously been published (Dalhus et al., 2002), and thus will not be further described in this study. The striking new feature in our model is the incredibly large number of modeled water molecules, i.e., 945 for the dimer A-D, which allows a detailed analysis of water organization.

*Sr* MalDH shares more than 72% of sequence similarity with its non-halophilic counterpart *Ca* MalDH. The *Sr* MalDH model was obtained at a resolution of 1.55 Å, and also contains a large number of water molecules: 680 for the equivalent *Ca* MalDH A-D dimer (Coquelle et al., 2010). The overall structural similarity between one monomer of *Sr* and *Ca* MalDHs led to a RMSD of about 0.6 Å for 258 superimposed Cα.

Therefore, these two structures of excellent resolution, with a large number of water molecules in their solvent layers, provide a unique combination to finely compare the water organization at their surface.

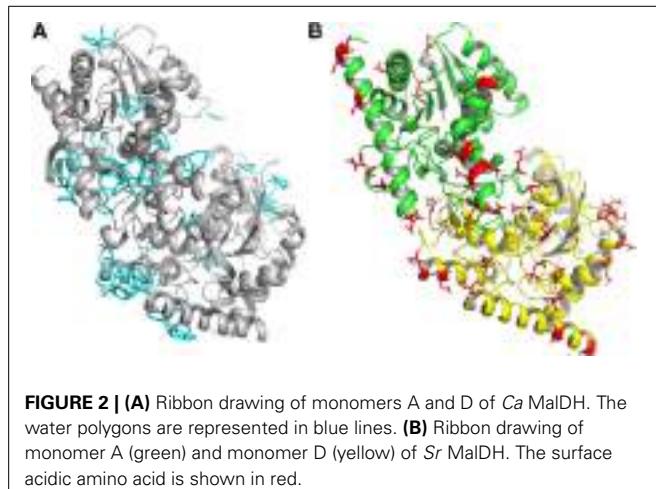
### COMPARISON OF HALOPHILIC AND NON-HALOPHILIC HYDRATION PATTERNS

A detailed analysis of the geometry of the 945 water molecules surrounding the dimeric *Ca* MalDH model (distance and angle) was performed and is presented in Table 3. It is outside the scope of this study to describe in great details both the geometry and interactions with the protein of all these water molecules. The role of water molecules in the folding process and stabilization of proteins has been well described in a work based on a larger set of proteins (Matsuoka and Nakasako, 2009). The most interesting feature of the water molecules in *Ca* MalDH structure is that 28% of them are organized in polygons (pentagons or hexagons), which can form extended clusters (Figure 2). These polygons are only observed at the surface of apolar residues. Geometrical properties of these polygons (Table 3) are in good agreement with those determined from a large statistical study using high-resolution structures (Lee and Kim, 2009).

Based on *Ca* MalDH water analysis, a careful inspection of the halophilic *Sr* MalDH hydration layer at the surface of the protein was performed to detect any water polygon. Even though 43% of *Sr* MalDH water molecules were considered to be superimposable with those from *Ca* MalDH (using a cut off distance of

**Table 3 | Water statistics for dimer AD of *Ca* MalDH.**

Number of water molecules: 945		
Water per residues: 1.57		
Water molecules involved in polygons: 28%		
76 polygons: 10 hexagons and 66 pentagons		
Size of clustered polygons: Up to 15		
Planar polygons: 64%		
Distance between surface residues and polygons (in Å)		
Minimal 2.58	Maximal 4.02	Average 3.23
Distance between two water molecules forming pentagons (in Å)		
Minimal 2.11	Maximal 4.19	Average 2.86
Angle between three water molecules forming pentagons (in °)		
Minimal 73.86	Maximal 138.74	Average 107.56
Distance between two water molecules forming hexagons (in Å)		
Minimal 2.39	Maximal 3.32	Average 2.66
Angle between three water molecules forming hexagons (in °)		
Minimal 81.45	Maximal 142.52	Average 111.72

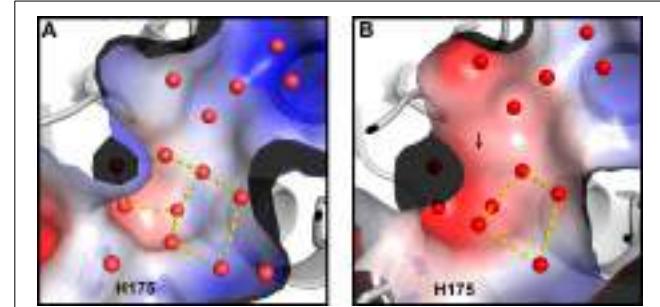


1.5 Å), no polygons were observed at the surface of the halophilic MalDH. However, 14 water molecules lie in the catalytic pocket of *Sr* MalDH, all of which are conserved in *Ca* MalDH. Five are organized as a pentagon, the only one observed in *Sr* MalDH (**Figure 3**). In *Ca* MalDH, the same water pentagon is present, but the catalytic pocket of *Ca* MalDH contains an extra water molecule, which closes a second pentagon in the catalytic pocket, adjacent to the first one (**Figure 3A**). A black arrow indicates the missing water molecule in *Sr* MalDH (**Figure 3B**).

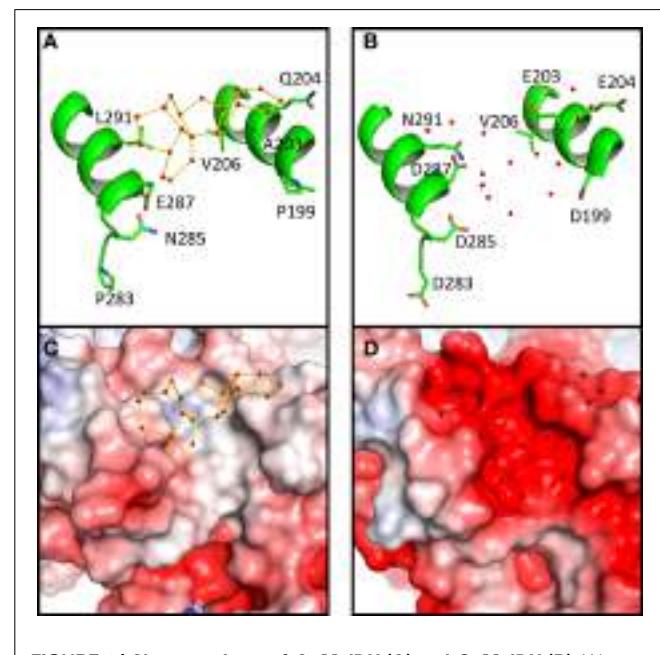
We therefore decided to have a closer look at surface regions where polygons are present in *Ca* MalDH to figure out the reasons why none are observed in *Sr* MalDH.

#### ACIDIC *Sr* MALDH SURFACE PREVENTS THE FORMATION OF STRUCTURED WATER

As mentioned, large networks of connected water polygons are present in *Ca* MalDH (**Figure 2A**). An example of such network is shown in **Figure 4A**. This network is anchored between helices  $\alpha$ 1G- $\alpha$ 1G and  $\alpha$ H and is made up of five pentagons



**FIGURE 3 | Close up views of the catalytic pocket.** Electrostatic surface representation of *Ca* MalDH **(A)** and *Sr* MalDH **(B)**. Water molecules are shown in small red spheres. Dashed lines colored in yellow delineates the polygons. The catalytic histidine (H175) is indicated. Numbering of amino acids corresponds to linear numbering of *Ca* MalDH.



**FIGURE 4 | Close up views of *Ca* MalDH **(A)** and *Sr* MalDH **(B)**.** Water molecules are shown in small red spheres. Dashed lines colored in yellow delineates the polygons. Important amino acids are represented in sticks. Electrostatic surface representation of *Ca* MalDH **(C)** and *Sr* MalDH **(D)**.

and one hexagon. In the same protein region, no water polygon is observed in *Sr* MalDH (**Figure 4B**), which possesses four extra negative charges compared to *Ca* MalDH, due to substitutions at positions 199, 203, 283, and 285. These substitutions led to important electrostatic surface changes, with a highly negative one for *Sr* MalDH compared to the apolar surface of *Ca* MalDH (**Figures 4C,D**). The lateral chain of acidic residues D287 in *Sr* MalDH is orientated in such a conformation that the *Sr* MalDH hydration pattern is modified when compared to that of *Ca* MalDH. The data suggest that the replacement of non-polar amino acid residues by acidic amino acid in a halophilic protein modifies properties of the hydration shell. Around apolar surfaces of the non-halophilic MalDH, water molecules cannot form direct hydrogen bonds with the protein, and thus organize

themselves as polygons with their nearest stable water neighbors. Acidic amino acids enrichment in these regions of *Sr* MalDH surface favors direct hydrogen bonding with water and therefore prevents polygons formation.

We also observe that water polygons formation is hampered in halophilic *Sr* MalDH, not only by direct acidic amino acid substitution but also by the side chain reorganization of conserved residues, as illustrated in **Figure 5**. In *Sr* MalDH compared to *Ca* MalDH, two acidic amino acids are observed at position 158 and 204. Glutamate at position 158 induces a direct perturbation of water pentagon P1, as previously observed. But Glutamate 204 promotes an interaction with R201 side chain, which moved to a new position that hinders appropriate hydrogen bonding geometries requested for the formation of water polygon P2 (**Figure 5**).

These two examples clearly illustrate the key influence of acidic amino acid enrichment in halophilic protein on the water organization at their surface; either through direct impacts or *via* conformational rearrangements of surrounding residues. This leads to the destabilization of almost all water polygons observed in the non-halophilic protein structure.

## DISCUSSION

This study presents for the first time a detailed analysis of the water organization at the surface of a halophilic protein and its non-halophilic counterpart. Both crystal structures were obtained at high resolution (better than 1.7 Å) and display similar crystallographic quality. The comparison of these hydration envelopes shows the effect of surface composition changes on the hydration shell structure. In the structure of *Ca* MalDH, we observed a large amount of stable water polygons. These specific water arrangements were first observed in the crystal structure of Crambin (Teeter, 1984). It has been analyzed that these water are not the results of crystallization process and are likely due to intrinsic interaction mode with the local hydrophobic surface of proteins (Nakasako, 1999, 2004). Water organization observed in *Ca* MalDH is in good agreement as apolar surface prevent direct hydrogen bonding of water molecules with the protein and favors polygonal structures. Acidic residues substitutions at the surface of *Sr* MalDH promote hydrogen bonding between the solvent and the protein. In particular, we observed that the changes in water structure organization in *Sr* MalDH are not only due to direct

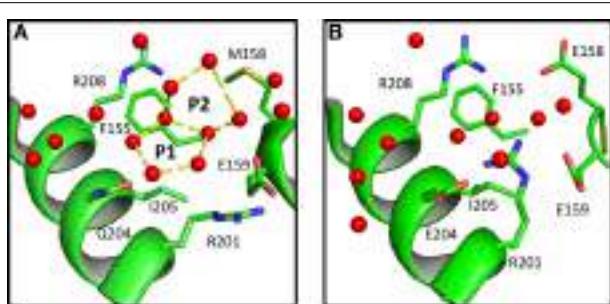
effects but also to long-range effects of amino acid substitutions. The latter is an indirect consequence of amino acids substitutions, selected to increase the *Sr* MalDH enzymatic activity at high salt concentration as analyzed in our previous work (Coquelle et al., 2010). Indeed, these changes modify the local dynamics of the protein surface, which should impact the dynamical properties of the nearest hydration water molecules, as previously observed (Nakasako et al., 2001).

At this stage, it is important to remind the concept of solvation/hydration of proteins.

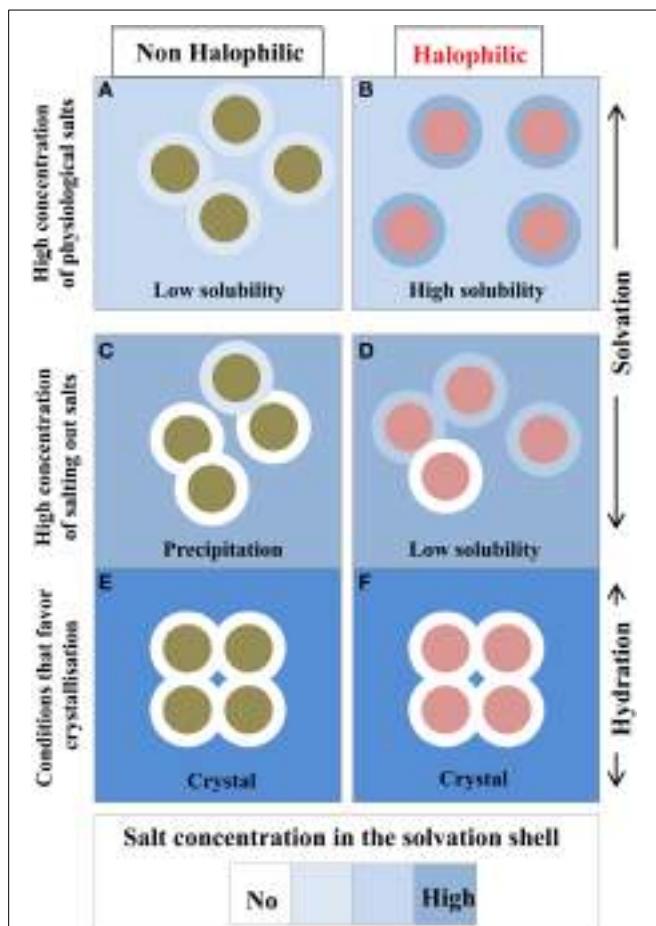
## ARE HALOPHILIC PROTEINS SOLVATED OR HYDRATED?

This is an important issue that should be discussed. Because of the chemical properties of the protein surface, the solvent composition at the vicinity of a given protein surface is different from the bulk. In a simple binary system containing water and protein without any cosolvents, such as salt or other macromolecular solutes, a hydration shell surrounds the protein. In the presence of high concentration of additional compounds such as salts, sugars, precipitating agents etc., the protein solution should be described as a ternary system in which the protein is enveloped by a solvation shell. The thermodynamics of proteins in the three-component system is well understood in terms of preferential binding parameters (Von Hippel and Schleich, 1969; Inoue and Timasheff, 1972; Arakawa and Timasheff, 1982; Zaccai and Eisenberg, 1990; Timasheff, 1991; reviewed in Zaccai, 2013). In conditions that maintain protein solubility, the chemical potential of the solvation shell and the bulk are equilibrated (**Figure 6**). In salting-out conditions that favor protein aggregation and crystallization, the equilibrium is strongly perturbed because the small solutes are excluded from the solvation shell (Tardieu et al., 2002). In this case the solvation shell looks like a hydration shell.

Cytoplasmic protein isolated from extreme halophilic prokaryotes that use the KCl-in adaptive strategy, such as *S. ruber* or the *Halobacteriaceae*, maintain a high solubility at molar concentration of various salt (Coquelle et al., 2010). In the case of the tetrameric MalDH from *Haloarcula marismortui*, the measurements of the preferential binding parameters have shown that the enzyme obey the general thermodynamics rules of the three components system (Costenaro et al., 2002; Ebel et al., 2002): In salting out conditions, the solvation envelope of *Hm* MalDH is strongly depleted in salt and it looks like a hydration shell; such behaviors is equivalent to the situation encountered with a non-halophilic protein. However, in high concentration of various physiological salts, it has been measured that *Hm* MalDH preferential binding parameters depend on salt type, demonstrating that the composition of its solvation shell varies (Costenaro et al., 2002; Ebel et al., 2002). In these physiological salts, *Hm* MalDH solvation envelope is enriched in salt, reflecting its halophilic adaptation. Consequently, as the chemical potential of the solvation layer and the bulk solvent are close, *Hm* MalDH remains highly soluble at high salt concentration. We determined that *Sr* MalDH remains highly soluble in high concentration of physiological salts (Coquelle et al., 2010). Based on the observation made on *Hm* MalDH, this suggests that *Sr* MalDH solvation layer should also be enriched with salts.



**FIGURE 5 |** Close up view of *Ca* MalDH (A) and *Sr* MalDH (B). Water molecules are in red spheres. Linear numbering as in *Ca* MalDH.



**FIGURE 6 | Compositional changes of the solvation layer between non-halophilic and halophilic proteins.** Filled circles either represent non-halophilic (Green) or halophilic (Red) proteins with their solvation shell (external circle). Solubility measurements are taken from Coquelle et al. (2010). With the halophilic protein, due to its acidic enriched surface, the dominant inter particular effect is repulsive and its solvation shell composition is similar to the bulk solvent (**B**). These two effects strongly favor high solubility even in high physiological salt concentration. In equivalent conditions the non-halophilic protein solubility is reduced (**A**). In non-physiological conditions, i.e., in the presence of salting out salts (**C,D**) or additives that promote crystallization (**E,F**), the solvation shell of each enzyme starts to be depleted in salt. This situation impacts the solubility of each enzyme to diverse extent and could promote precipitation. In crystal conditions, the solvation envelope is a hydration shell, which contains none or very few ions.

#### SOLUBILITY AND ACIDIC AMINO ACID SURFACE ENRICHMENT

The relationship between an increase in protein solubility and the shift toward negatively charged protein surfaces is not restricted to halophilic protein (Trevino et al., 2007). The favorable effect of acidic residues on protein solubility has been highlighted by an elegant thermodynamical work based on seven non-halophilic proteins, which displayed pIs ranging from 3.5 to 8 (Kramer et al., 2012). In the case of halophilic proteins, it has been demonstrated that their high negative charge density maintains a weak repulsive protein-protein interactions in high salt concentration (Costenaro et al., 2002; Ebel et al., 2002). Theoretically, this repulsive effect between macromolecules of

same net charge could also be induced by positively charged amino acids. However, calculation of the solvent-accessible areas of the side-chain components between negatively and positively charged residues unravel their relative efficiency on solubility. Compared to positively charged residues, the favorable effect of acidic residues is due to the lower hydrophobic solvent exposed surface of their side chain (Britton et al., 1998).

#### APPLICATION TO HALOPHILIC PROTEIN STRUCTURES

Our comparative study of the hydration shell of *Ca* and *Sr* MalDHs sheds light on the close relation between solubility, acidic residue enrichment and solvation.

First, conclusions from the solvent properties analysis using X-ray crystallography should be drawn with cautiousness. Our observation confirms that salting-out conditions deplete the solvation layer in salts. Indeed, the additives used for crystal growth shift the conditions toward salt depletion in the solvation envelope. Consequently, even if several halophilic proteins have been crystallized in the presence of high concentrations of physiological salts, one should take precaution when discussing the role of the solvent layer as obtained from X-ray structures.

Second, the favorable change in solubility of halophilic proteins is driven by their protein surface enrichment in acidic residues, which plays a dual role. Indeed, our study shows that acidic residues, through their carboxyl groups that are known to form strong hydrogen bonds, can organize the solvation shell by direct as well as indirect interactions. They are therefore good candidates for interactions with hydrated salt ions as proposed by Zaccai (2013). Moreover, they promote slightly repulsive inter-particular interactions between each protein molecule, favoring solubility.

#### CONCLUSION

Recent data have suggested that acidic enrichment, considered as an adaptive signature of halophilic proteins, could also be due to genetic drift (Deole et al., 2013). Whatever the precise evolutionary mechanism responsible for the peculiar composition of protein isolated from halophilic microorganisms, our work helps to understand that acidic acid enrichment was an appropriate evolutionary innovation in the case of microorganisms that accumulates high concentration of KCl in their cytoplasm to maintain their turgor pressure in highly salted environment. Such enrichment allows halophilic proteins to compete against aggregation via their ability to reorganize protein-solvent interactions.

The role of acidic amino acids substitution on the solvent organization, highlighted in the present work, has to be completed by further studies involving enzymes from halophilic organisms that used different strategies to cope with high concentration of salts.

#### ACKNOWLEDGMENTS

This work was supported in part by the Agence Nationale de la Recherche Grants “Ln23” ANR-13-BS07-0007-02. Romain Talon and Eric Girard also thank scientists of the FIP-BM30A beamline at the European Synchrotron radiation Facility (ESRF) for their help.

#### AUTHOR CONTRIBUTIONS

Dominique Madern and Eric Girard designed research, Romain Talon, Nicolas Coquelle, Dominique Madern, and Eric Girard

performed research, Romain Talon, Nicolas Coquelle, Dominique Madern, and Eric Girard were involved in data analysis. Romain Talon, Nicolas Coquelle, Dominique Madern, and Eric Girard wrote the paper.

## REFERENCES

- Adams, P. D., Afonine, P. V., Bunkóczki, G., Chen, V. B., Davis, I. W., Echols, N., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221. doi: 10.1107/S0907444909052925
- Antón, J., Oren, A., Benloch, S., Rodríguez-Valera, F., Amann, R., and Rosselló-Mora, R. (2002). *Salinibacter ruber* gen. nov., sp. nov., a novel, extremely halophilic member of the Bacteria from saltern crystallizer ponds. *Int. J. Syst. Evol. Microbiol.* 52, 485–491. doi: 10.1186/1746-1448-4-15
- Arakawa, T., and Timasheff, S. N. (1982). Preferential interactions of proteins with salts in concentrated solutions. *Biochemistry* 21, 6545–6552. doi: 10.1021/bi00268a034
- Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* 98, 10037–10041. doi: 10.1073/pnas.181342398
- Besir, H., Zeth, K., Bracher, A., Heider, U., Ishibashi, M., Tokunaga, M., et al. (2005). Structure of a halophilic nucleoside diphosphate kinase from *Halobacterium salinarum*. *FEBS Lett.* 579, 6595–6600. doi: 10.1016/j.febslet.2005.10.052
- Bieger, B., Essen, L. O., and Oesterhelt, D. (2003). Crystal structure of halophilic dodecin. A novel dodecameric flavin binding protein from *Halobacterium salinarum*. *Structure* 4, 375–385. doi: 10.1016/S0969-2126(03)00048-0
- Bonete, M. J., Pérez-Pomares, F., Diaz, S., Ferrer, J., and Oren, A. (2003). Occurrence of two different glutamate dehydrogenase activities in the halophilic bacterium *Salinibacter ruber*. *FEMS Microbiol. Lett.* 226, 181–186. doi: 10.1016/S0378-1097(03)00592-5
- Bracken, C. D., Neighbor, A. M., Lamenn, K. K., Thomas, G. C., Schubert, H. L., Whiting, F. G., et al. (2011). Crystal structures of a halophilic archaeal malate synthase from *Haloferax volcanii* and comparisons with isoforms A and G. *BMC Struct. Biol.* 11:23. doi: 10.1186/1472-6807-11-23
- Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D Biol. Crystallogr.* 59, 2023–2030. doi: 10.1107/S0907444903017694
- Britton, K. L., Baker, P. J., Fisher, M., Ruzheinikov, S., Gilmour, D. J., Bonete, M. J., et al. (2006). Analysis of protein solvent interactions in glucose dehydrogenase from the extreme halophile *Haloferax mediterranei*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 4846–4851. doi: 10.1073/pnas.0508854103
- Britton, K. L., Stillman, T. J., Yip, K. S., Forterre, P., Engel, P. C., and Rice, D. W. (1998). Insights into the molecular basis of salt tolerance from the study of glutamate dehydrogenase from *Halobacterium salinarum*. *J. Biol. Chem.* 273, 9023–9030. doi: 10.1074/jbc.273.15.9023
- Brünger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472–475. doi: 10.1038/355472a0
- Coquelle, N., Talon, R., Juers, D. H., Girard, E., Kahn, R., and Madern, D. (2010). Gradual adaptive changes of a protein facing high salt concentrations. *J. Mol. Biol.* 404, 493–505. doi: 10.1016/j.jmb.2010.09.055
- Costenaro, L., Zaccai, G., and Ebel, C. (2002). Link between protein–solvent and weak protein–protein interactions gives insight into halophilic adaptation. *Biochemistry* 41, 13245–13252. doi: 10.1021/bi025830z
- Cowtan, K. D. (2006). The Buccaneer software for automated model building. *Acta Crystallogr. D Biol. Crystallogr.* 62, 1002–1011. doi: 10.1107/S0907444906022116
- Cowtan, K. D., and Main, P. (1996). Phase combination and cross validation in iterated density modification calculations. *Acta Crystallogr. D Biol. Crystallogr.* 52, 43–48. doi: 10.1107/S090744499500761X
- Dalhus, B., Saarinen, M., Sauer, U. H., Eklund, P., Johansson, K., Karlsson, A., et al. (2002). Structural basis for thermophilic protein stability: structures of thermophilic and mesophilic malate dehydrogenases. *J. Mol. Biol.* 318, 707–721. doi: 10.1016/S0022-2836(02)00050-5
- Deole, R., Challacombe, J., Raiford, D. W., and Hoff, W. D. (2013). An extremely halophilic proteobacterium combines a highly acidic proteome with a low cytoplasmic potassium content. *J. Biol. Chem.* 288, 581–588. doi: 10.1074/jbc.M112.420505
- Ebel, C., Costenaro, L., Pascu, M., Faou, P., Kernel, B., Proust-De Martin, F., et al. (2002). Solvent interactions of halophilic malate dehydrogenase. *Biochemistry* 41, 13234–13244. doi: 10.1021/bi0258290
- Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. D. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501. doi: 10.1107/S0907444910007493
- Frolow, F., Harel, M., Sussman, J. L., Mevarech, M., and Shoham, M. (1996). Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe–2S ferredoxin. *Nat. Struct. Biol.* 3, 452–458. doi: 10.1038/nsb0596-452
- Girard, E., Stelter, M., Anelli, P. L., Vicat, J., and Kahn, R. (2003). A new class of gadolinium complexes employed to obtain high-phasing-power heavy-atom derivatives: results from SAD experiments with hen egg-white lysozyme and urate oxidase from *Aspergillus flavus*. *Acta Crystallogr. D Biol. Crystallogr.* 59, 118–126. doi: 10.1107/S0907444902020176
- Inoue, H., and Timasheff, S. N. (1972). Preferential and absolute interactions of solvent components with proteins in mixed solvent systems. *Biopolymers* 11, 737–743. doi: 10.1002/bip.1972.360110402
- Irimia, A., Ebel, C., Madern, D., Richard, S. B., Cosenza, L. W., Zaccai, G., et al. (2003). The Oligomeric states of Haloarcula marismortui malate dehydrogenase are modulated by solvent components as shown by crystallographic and biochemical studies. *J. Mol. Biol.* 325, 859–873. doi: 10.1016/S0022-2836(02)01450-X
- Kabsch, E. (2010). XDS. *Acta Crystallogr. D. Biol. Crystallogr.* 66, 125–132. doi: 10.1107/S0907444909047337
- Kramer, R. M., Shende, V. R., Motl, N., Pace, C. N., and Scholtz, J. M. (2012). Toward a molecular understanding of protein solubility: increased negative surface charge correlates with increased solubility. *Biophys J.* 102, 1907–1915. doi: 10.1016/j.bpj.2012.01.060
- Lee, J., and Kim, S. H. (2009). Water polygons in high-resolution protein crystal structures. *Protein Sci.* 18, 1370–1376. doi: 10.1002/pro.162
- Madern, D., Ebel, C., and Zaccai, G. (2000). Halophilic adaptation of enzymes. *Extremophiles* 4, 91–98. doi: 10.1007/s007920050142
- Madern, D., and Zaccai, G. (2004). Molecular adaptation: the malate dehydrogenase from the extreme halophilic bacterium *Salinibacter ruber* behaves like a non-halophilic protein. *Biochimie* 86, 295–303. doi: 10.1016/j.biochi.2004.04.004
- Matsuoka, D., and Nakasako, M. (2009). Probability distributions of hydration water molecules around polar protein atoms obtained by a database analysis. *J. Phys. Chem. B* 113, 11274–11292. doi: 10.1021/jp902459n
- Mongodin, E. F., Nelson, K. E., Daugherty, S., DeBoy, R. T., Wister, J., Khouri, H., et al. (2005). The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18147–18152. doi: 10.1073/pnas.0509073102
- Nakasako, M. (1999). Large-scale networks of hydration water molecules around bovine beta-trypsin revealed by cryogenic X-ray crystal structure analysis. *J. Mol. Biol.* 289, 547–564. doi: 10.1006/jmbi.1999.2795
- Nakasako, M. (2004). Water–protein interactions from high-resolution protein crystallography. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 1191–1204. doi: 10.1098/rstb.2004.1498
- Nakasako, M., Fujisawa, T., Adachi, S., Kudo, T., and Higuchi, S. (2001). Large-scale domain movements and hydration structure changes in the active-site cleft of unligated glutamate dehydrogenase from *Thermococcus profundus* studied by cryogenic X-ray crystal structure analysis and small-angle X-ray scattering. *Biochemistry* 40, 3069–3079. doi: 10.1021/bi002482x
- Oren, A. (2002). *Halophilic Microorganisms and their Environments*. Dordrecht: Kluwer Scientific Publishers. doi: 10.1007/0-306-48053-0
- Oren, A. (2013). Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. *Front. Microbiol.* 4:315. doi: 10.3389/fmicb.2013.00315
- Painter, J., and Merritt, E. A. (2006a). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D. Biol. Crystallogr.* 62, 439–450. doi: 10.1107/S0907444906005270
- Painter, J., and Merritt, E. A. (2006b). TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* 39, 109–111. doi: 10.1107/s0021889805038987
- Richard, S. B., Madern, D., Garcin, E., and Zaccai, G. (2000). Halophilic adaptation: novel solvent protein interactions observed in the 2.9 and 2.6 Å resolution

- structures of the wild type and a mutant of malate dehydrogenase from *Haloarcula marismortui*. *Biochemistry* 39, 992–1000. doi: 10.1021/bi991001a
- Sheldrick, G. M. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* 66, 479–485. doi: 10.1107/S0907444909038360
- Tardieu, A., Bonneté, F., Finet, D., and Vivares, D. (2002). Understanding salt or PEG induced attractive interactions to crystallize biological macromolecules. *Acta Crystallogr. D Biol. Crystallogr.* 58, 1549–1553. doi: 10.1107/S0907444902014439
- Teeter, M. M. (1984). Water structure of a hydrophobic protein at atomic resolution: pentagon rings of water molecules in crystals of crambin. *Proc. Natl. Acad. Sci. U.S.A.* 81, 6014–6018. doi: 10.1073/pnas.81.19.6014
- Timasheff, S. N. (1991). Solvent effect on protein stability. *Curr. Opin. Struct. Biol.* 2, 35–39. doi: 10.1016/0959-440X(92)90173-5
- Trevino, S. R., Scholtz, J. M., and Pace, C. N. (2007). Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favourably than the other hydrophilic amino acids in RNase Sa. *J. Mol. Biol.* 366, 449–446. doi: 10.1016/j.jmb.2006.10.026
- von Hippel, P. H., and Schleich, T. W. (1969). “The effects of neutral salts on the structure and conformational stability of macromolecules in solution,” in *Biological Macromolecules*, Vol. II, eds G. D. Fasman and S. N. Timasheff (New York, NY: Marcel Dekker Inc.), 417–584.
- Weiss, M. S. (2001). Global indicators of X-ray data quality. *J. Appl. Crystallogr.* 34, 130–135. doi: 10.1107/S0021889800018227
- Wende, A., Johansson, P., Vollrath, R., Dyall-Smith, M., Oesterhelt, D., and Grininger, M. (2010). Structural and biochemical characterization of a halophilic archaeal alkaline phosphatase. *J. Mol. Biol.* 400, 52–62. doi: 10.1016/j.jmb.2010.04.057
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* 67, 235–242. doi: 10.1107/S0907444910045749
- Winter, J. A., Christofi, P., Morroll, S., and Bunting, K. A. (2009). The crystal structure of *Haloflexax volcanii* proliferating cell nuclear antigen reveals unique surface charge characteristics due to halophilic adaptation. *BMC Struct. Biol.* 9:55 doi: 10.1186/1472-6807-9-55
- Yamamura, A., Ichimura, T., Kamekura, M., Mizuki, T., Usami, R., Makino, T., et al. (2009). Molecular mechanism of distinct salt-dependent enzyme activity of two halophilic nucleoside diphosphate kinases. *Biophys. J.* 96, 4692–700. doi: 10.1016/j.bpj.2009.03.012
- Zaccaï, G. (2013). Hydration shells with a pinch of salt. *Biopolymers* 4, 233–238. doi: 10.1002/bip.22154
- Zaccaï, G., and Eisenberg, H. (1990). Halophilic proteins and the influence of solvent on protein stabilization. *Trends Biochem. Sci.* 15, 333–337. doi: 10.1016/0968-0004(90)90068-M
- Zeth, K., Offermann, S., Essen, L. O., and Oesterhelt, D. (2004). Iron-oxo clusters biomineralizing on protein surfaces: structural analysis of *Halobacterium salinarum* DpsA in its low- and high-iron states. *Proc. Natl. Acad. Sci. U.S.A.* 101, 13780–13785. doi: 10.1073/pnas.0401821101
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 30 December 2013; accepted: 04 February 2014; published online: 21 February 2014.*
- Citation:* Talon R, Coquelle N, Madern D and Girard E (2014) An experimental point of view on hydration/solvation in halophilic proteins. *Front. Microbiol.* 5:66. doi: 10.3389/fmicb.2014.00066
- This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.*
- Copyright © 2014 Talon, Coquelle, Madern and Girard. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# The haloarchaeal MCM proteins: bioinformatic analysis and targeted mutagenesis of the $\beta$ 7- $\beta$ 8 and $\beta$ 9- $\beta$ 10 hairpin loops and conserved zinc binding domain cysteines

Tatjana P. Kristensen<sup>1†‡</sup>, Reeja Maria Cherian<sup>1†‡</sup>, Fiona C. Gray<sup>1</sup> and Stuart A. MacNeill<sup>1,2\*</sup>

<sup>1</sup> Department of Biology, University of Copenhagen, Københavns Biocenter, Copenhagen N, Denmark

<sup>2</sup> School of Biology, University of St. Andrews, North Haugh, St. Andrews, Fife, UK

**Edited by:**

R. Thane Papke, University of Connecticut, USA

**Reviewed by:**

Nils-Kåre Birkeland, University of Bergen, Norway

Antonio Ventosa, University of Sevilla, Spain

**\*Correspondence:**

Stuart A. MacNeill, School of Biology, University of St. Andrews, North Haugh, St. Andrews, Fife, KY16 9ST, UK

e-mail: stuart.macneill@st-andrews.ac.uk

**†Present address:**

Tatjana P. Kristensen, Department of Veterinary Disease Biology, University of Copenhagen, Frederiksberg C, Denmark;

Reeja Maria Cherian, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Göteborg, Sweden

<sup>‡</sup>These authors have contributed equally to this work.

The hexameric MCM complex is the catalytic core of the replicative helicase in eukaryotic and archaeal cells. Here we describe the first *in vivo* analysis of archaeal MCM protein structure and function relationships using the genetically tractable haloarchaeon *Haloferax volcanii* as a model system. *Hfx. volcanii* encodes a single MCM protein that is part of the previously identified core group of haloarchaeal MCM proteins. Three structural features of the N-terminal domain of the *Hfx. volcanii* MCM protein were targeted for mutagenesis: the  $\beta$ 7- $\beta$ 8 and  $\beta$ 9- $\beta$ 10  $\beta$ -hairpin loops and putative zinc binding domain. Five strains carrying single point mutations in the  $\beta$ 7- $\beta$ 8  $\beta$ -hairpin loop were constructed, none of which displayed impaired cell growth under normal conditions or when treated with the DNA damaging agent mitomycin C. However, short sequence deletions within the  $\beta$ 7- $\beta$ 8  $\beta$ -hairpin were not tolerated and neither was replacement of the highly conserved residue glutamate 187 with alanine. Six strains carrying paired alanine substitutions within the  $\beta$ 9- $\beta$ 10  $\beta$ -hairpin loop were constructed, leading to the conclusion that no individual amino acid within that hairpin loop is absolutely required for MCM function, although one of the mutant strains displays greatly enhanced sensitivity to mitomycin C. Deletions of two or four amino acids from the  $\beta$ 9- $\beta$ 10  $\beta$ -hairpin were tolerated but mutants carrying larger deletions were inviable. Similarly, it was not possible to construct mutants in which any of the conserved zinc binding cysteines was replaced with alanine, underlining the likely importance of zinc binding for MCM function. The results of these studies demonstrate the feasibility of using *Hfx. volcanii* as a model system for reverse genetic analysis of archaeal MCM protein function and provide important confirmation of the *in vivo* importance of conserved structural features identified by previous bioinformatic, biochemical and structural studies.

**Keywords:** *Haloferax volcanii*, archaea, Haloarchaea, MCM helicase, DNA replication, reverse genetics, zinc binding domain

## INTRODUCTION

In all forms of life, successful chromosomal DNA replication requires efficient unwinding of the DNA double helix at the replication forks, a reaction catalyzed by the replicative DNA helicase. In eukaryotes the replicative helicase is the CMG complex, a tripartite molecular machine composed of Cdc45, MCM and GINS (reviewed by Onesti and MacNeill, 2013). The MCM (mini-chromosome maintenance) complex is the catalytic core of this machine (Vijayraghavan and Schwacha, 2012). MCM is a ring-shaped hexamer composed of six related but non-identical subunits, each of which is a member of the AAA+ (ATPases associated with diverse cellular activities) protein superfamily (Duderstadt and Berger, 2008). MCM is loaded onto chromosomal replication origins as a head-to-head double hexamer in the G1 phase of the cell cycle in a reaction known as replication licensing (Ervin et al., 2009; Remus et al., 2009; Gambus et al., 2011). Cdc45 and GINS then assemble at the G1-S boundary to form the CMG, activation of which involves MCM subunit

phosphorylation by DDK (Dbf4-dependent protein kinase). Once activated, individual CMG complexes move with the replication forks from origin to inter-origin sequences.

Consistent with their shared evolutionary history, homologs of the major eukaryotic replication factors have been identified and characterized in the archaea, including homologs of the three components of the CMG. Owing to their high sequence similarity to their eukaryotic counterparts, archaeal MCM proteins were the first to be identified and biochemically characterized. Many archaea encode single MCM proteins that have been shown to form—or are presumed to form—homohexameric helicase complexes (reviewed by Slaymaker and Chen, 2012). The best studied examples of this type of MCM are from the euryarchaeon *Methanothermobacter thermoautotrophicum* and the crenarchaeon *Sulfolobus solfataricus*. However, a number of species encode multiple MCM proteins, such as *Thermococcus kodakarensis* and *Methanococcus maripaludis*, which encode three and four, respectively (Walters and Chong, 2010; Ishino et al.,

2011; Pan et al., 2011). It is not impossible that in some species these proteins could form eukaryotic-like heterohexameric complexes *in vivo*. Interestingly, only one of the three *T. kodakarensis* MCM proteins is essential for cell viability (Ishino et al., 2011; Pan et al., 2011)

Structural information is available for a number of archaeal MCM proteins and has been used to guide biochemical investigations of protein structure-function relationships (reviewed by Slaymaker and Chen, 2012). At a structural level, individual archaeal MCM proteins comprise a non-catalytic N-terminal domain followed by the catalytic AAA+ domain and, at the extreme C-terminus, a short winged helix-turned-helix (wHTH) domain. The most extensive crystal structure is that of near full-length *S. solfataricus* MCM which spans the N-terminal and catalytic AAA+ domains but not the wHTH domain (Brewster et al., 2008). Efforts to determine the structure of the wHTH by NMR are ongoing (Wiedemann et al., 2013). Additional crystal structures include the N-terminal domains of *S. solfataricus*, *M. thermoautotrophicum*, and *Thermoplasma acidophilum* MCM proteins, with the latter forming a right-handed spiral filament (Fletcher et al., 2003; Liu et al., 2008; Fu et al., 2014). A left-handed filament structure for near full-length *S. solfataricus* MCM has also been determined (Slaymaker et al., 2013), as well as a full-length structure of a catalytically inactive MCM homolog from *Methanopyrus kandleri* (Bae et al., 2009). The biological significance, if any, of the filamentous forms remains to be determined.

Unlike the MCM proteins, archaeal GINS and Cdc45 homologs share only very limited sequence similarity with their eukaryotic counterparts (Marinsek et al., 2006). The eukaryotic GINS complex is a heterotetramer, comprising the related Sld5, Psf1, Psf2, and Psf3 proteins (reviewed by Kamada, 2012). Both homotetrameric and heterotetrameric (i.e., dimer of dimer or A<sub>2</sub>B<sub>2</sub>) complexes have been identified in archaea and the structure of the *T. kodakarensis* A<sub>2</sub>B<sub>2</sub> heterotetrameric GINS has been solved (Oyama et al., 2011). Archaeal Cdc45 homologs have only very recently been positively identified as such (Sanchez-Pulido and Ponting, 2011; Krastanova et al., 2012; Makarova et al., 2012). These proteins belong to the RecJ nuclease branch of the DHH hydrolase superfamily. Unlike eukaryotic Cdc45 proteins, at least some archaeal RecJ/Cdc45 proteins possess nuclease activity (Li et al., 2011; Yuan et al., 2013), the precise function of which is unclear. The existence of all three CMG components in archaea suggest that these organisms may have a valuable role to play as models for dissecting the function of the individual CMG components.

Using multiple sequence alignments and crystal structures as a guide, a number of laboratories have reported detailed mutagenesis studies of MCM structure-function relationships (reviewed by Slaymaker and Chen, 2012). In all cases, the effects of the mutations on MCM function were determined *in vitro*, using purified recombinant proteins in various biochemical assays. To our knowledge, there has been no *in vivo* reverse genetic analysis of the effects of mutations of MCM function, largely due to the difficulty or impossibility of conducting such studies in species where genetic tools are either

rudimentary or unavailable. In this report we describe the first results of reverse genetic analysis of archaeal MCM function *in vivo*, using the haloarchaeal organism *Haloferax volcanii* as a model system. The haloarchaea present a particularly attractive model to study archaeal chromosome replication owing to the ease with which representative species can be manipulated genetically (reviewed by Farkas et al., 2013). *Hfx. volcanii* in particular has proved a highly successful model, with a number of components of the *Hfx. volcanii* replication machinery already characterized, including multiple origins of replication (Hawkins et al., 2013), origin binding proteins (Norais et al., 2007), single-stranded DNA binding proteins (Skowyra and MacNeill, 2012; Stroud et al., 2012), the sliding clamp PCNA (Morgunova et al., 2009; Winter et al., 2009) and both ATP- and NAD-dependent DNA ligases (Poidevin and MacNeill, 2006; Zhao et al., 2006). In addition to their genetic tractability, over 100 haloarchaeal genomes have been now sequenced, offering a wealth of information for comparative protein sequence analysis. *Hfx. volcanii* encodes a single MCM protein, a member of the previously defined core group of haloarchaeal MCM proteins discussed further below (MacNeill, 2009). In the work presented here, three conserved features of the protein are targeted for mutagenesis: the β7-β8 and β9-β10 β-hairpin loops and the four conserved cysteines of the putative zinc binding domain. The results of these studies establish *Hfx. volcanii* as a valuable model for detailed structure-function analysis of MCM helicase and provide confirmation of the importance of these conserved features for MCM function *in vivo*.

## MATERIALS AND METHODS

### DATABASE SEARCHING AND SEQUENCE HANDLING

Protein sequences were obtained from the UniProt Knowledgebase (UniProtKB) database (Magrane and UniProt Consortium, 2011): primary accession numbers are listed in **Table 1**. Multiple sequence alignments and phylogenetic trees were generated using ClustalX 2.1 (Larkin et al., 2007) and njplot (Perriere and Gouy, 1996), respectively. Intein sequences were initially identified by visual inspection as large sequence insertions in comparative sequence analysis. Inteins boundaries were defined by the presence of the N-terminal (block A) and C-terminal (block G) intein splicing motifs as defined at the InBase intein database (Perler, 2002).

### STRAINS AND GROWTH CONDITIONS

*Hfx. volcanii* strains used in this study are listed in **Table 2**. All strains were grown in either Hv-YPC or Hv-CA medium at 45°C as described in the Halohandbook v7.2 ([www.haloarchaea.com/resources/halohandbook](http://www.haloarchaea.com/resources/halohandbook)). For selection procedures, tryptophan was added to Hv-CA medium at a final concentration of 50 µg/ml. For counter-selection using 5-fluoroorotic acid (5-FOA), Hv-CA was supplemented with uracil and 5-FOA at final concentrations of 10 µg/ml and 50 µg/ml, respectively. For mitomycin C sensitivity assays, wild-type (H26) and mutant cells were grown at 45°C in Hv-YPC medium to an OD<sub>650 nm</sub> of 0.2–0.32, before being serially diluted in 18% SW. 5 µl aliquots were then spotted on Hv-YPC plates containing 0, 10, 20, or 30 ng/ml mitomycin C and incubated at 45°C for 5 days.

**Table 1 | List of MCM proteins analyzed in this study.**

Key to Figure 1	Species	UniProtKB accession number	Length (amino acids)	Inteins			
				A	B	C	D
1	<i>Haladaptatus paucihalophilus</i> DX253	E7QNU9	698				
2	<i>Halalkalicoccus jeotgali</i> B3	D8J3U5	700				
3A	<i>Haloarcula marismortui</i> ATCC 43049	Q5UYX8	1175	+			
3B		Q5V011	681				
3C		Q5V814	649				
4	<i>Halobacterium salinarum</i> R1	B0R796	879	+			
5A	<i>Halobiforma lacisalsi</i> AJ5	M0LR18	1342				+
5B		M0LZ47	312				
6	<i>Halococcus thailandensis</i> JCM 13552	M0N8X7	698				
7	<i>Haloferax volcanii</i> DS2	D4GZG5	702				
8	<i>Halogeometricum borinquense</i> DSM 11551	E4NRK9	1818	+			+
9	<i>Halogramnum salarium</i> B-1	J2ZGA0	700				
10A	<i>Halomicrombium mukohataei</i> DSM 12286	C7P1F9	873	+			
10B		C7NZ81	458				
11A	<i>Halopiger xanaduensis</i> SH-6	F8D3Z0	702				
11B		F8DEM3	698				
11C		F8DET6	315				
12	<i>Haloquadratum walsbyi</i> HBSQ001	Q18E84	2216	+	+	+	+
13	<i>Halorhabdus utahensis</i> AX-2	C7NUH7	1412	+			+
14A	<i>Halorubrum lacusprofundi</i> DSM 5036	B9LTB1	700				
14B		B9LUI3	717				
15	<i>Halosarcina pallida</i> JCM 14848	M0D6S0	1172	+			
16A	<i>Halosimplex carlsbadense</i> 2-9-1	M0D2C8	698				
16B		M0CE93	712				
17A	<i>Haloterrigena turkmenica</i> DSM 5511	D2RUS4	700				
17B		D2S3H9	314				
18	<i>Halovivax asiaticus</i> JCM 14624	M0BKP0	876	+			
19A	<i>Natrialba asiatica</i> DSM 12278	M0AKM2	700				
19B		M0AIF1	315				
20	<i>Natrinema pallidum</i> DSM 3751	L9Z845	700				
21A	<i>Natronobacterium gregoryi</i> SP2	L0AHL1	1172	+			
21B		L0AN78	709				
21C		L0ANL5	662				
22	<i>Natronococcus amylolyticus</i> DSM 10524	L9XCC3	1404	+	+		
23	<i>Natronolimnobius innermongolicus</i> JCM 12255	L9X2E8	700				
24A	<i>Natronomonas pharaonis</i> DSM 2160	Q3IML4	1037	+			
24B		Q3IPB6	676				
25	<i>Natronorubrum tibetense</i> GA33	L9VHE4	1814	+			+
26	<i>Salinarchaeum</i> sp. Harcht-Bsk1	R4W989	697				
	<i>Methanosaerina acetivorans</i> C2A	Q8TSW4	701				
	<i>Methanothermobacter thermautotrophicus</i> ΔH	O27798	666				
	<i>Sulfolobus solfataricus</i> P2	Q9UXG1	686				
	<i>Cenarchaeum symbiosum</i> A	A0RYB8	697				
	<i>Korarchaeum cryptofilum</i> OPF8	B1L6L9	703				
	<i>Thermoplasma acidophilum</i> DSM 1728	Q9HK10	698				
	<i>Archaeoglobus fulgidus</i> DSM 4304	O29733	586				
	<i>Homo sapiens</i> Mcm2	P49736	904				
	<i>Homo sapiens</i> Mcm3	P25205	808				
	<i>Homo sapiens</i> Mcm4	P33991	863				
	<i>Homo sapiens</i> Mcm5	P33992	734				
	<i>Homo sapiens</i> Mcm6	Q14566	821				
	<i>Homo sapiens</i> Mcm7	P33993	719				

The numbers shown in the first column provide a key for the proteins represented in the phylogenetic tree shown in **Figure 1**. The plus signs in the final four columns indicate the present of intein insertions at each of the four conserved positions A–D shown in **Figure 2**. Protein 3A, for example, Q5UYX8 from *Haloarcula marismortui*, has a single intein at position A.

**Table 2 |** *Haloferax volcanii* strains used in this study.

Strain no.	Genotype	Notes	References
SMH693	–	Wild-type strain DS70	Wendoloski et al., 2001
SMH630	$\Delta pyrE2$	Strain H26	Allers et al., 2004
SMH656	<i>mcm-rM1</i> $\Delta pyrE2$	Single alanine substitution Q186A	This study
SMH654	<i>mcm-rM3</i> $\Delta pyrE2$	Single alanine substitution E190A	This study
SMH658	<i>mcm-rM4</i> $\Delta pyrE2$	Single alanine substitution R193A	This study
SMH660	<i>mcm-rM5</i> $\Delta pyrE2$	Single alanine substitution E196A	This study
SMH662	<i>mcm-rM6</i> $\Delta pyrE2$	Single alanine substitution Q199A	This study
SMH638	<i>mcm-bH1</i> $\Delta pyrE2$	Paired alanine substitution H226A/I227A	This study
SMH640	<i>mcm-bH2</i> $\Delta pyrE2$	Paired alanine substitution E228A/Q229A	This study
SMH642	<i>mcm-bH3</i> $\Delta pyrE2$	Paired alanine substitution Q230A/T231A	This study
SMH644	<i>mcm-bH4</i> $\Delta pyrE2$	Paired alanine substitution S232A/G233A	This study
SMH646	<i>mcm-bH5</i> $\Delta pyrE2$	Paired alanine substitution N234A/E235A	This study
SMH648	<i>mcm-bH6</i> $\Delta pyrE2$	Paired alanine substitution K236A/T237A	This study
SMH649	<i>mcm-hD1</i> $\Delta pyrE2$	Two amino acid deletion T231/S232	This study
SMH650	<i>mcm-hD2</i> $\Delta pyrE2$	Four amino acid deletion Q230/T231/S232/G233	This study
SMH652	<i>mcm-S1</i> $\Delta pyrE2$	Silent Acc65I restriction site replacing codons 141 and 142	This study

## MOLECULAR CLONING REAGENTS

Enzymes for molecular cloning were purchased from New England Biolabs (NEB), Promega or Fermentas. Oligonucleotides were synthesized by DNA Technology A/S (Risskov, Denmark). DNA sequencing was performed by Eurofins MWG Operon (Ebersberg, Germany). DNA purification kits were from Qiagen. PCR amplification was performed using the GC-rich PCR system (Roche) with *Taq* polymerase (NEB) substituting for the GC-rich enzyme as necessary. For routine cloning purposes, *E.coli* DH5α (*fhuA2*  $\Delta$  (*argF-lacZ*)*U169* *phoA* *glnV44*  $\Phi$ 80  $\Delta$  (*lacZ*)*M15* *gyrA96* *recA1* *relA1* *endA1* *thi-1* *hsdR17*) was used (Invitrogen). To prepare unmethylated plasmid DNA for *Hfx. volcanii* transformation, *E.coli* GM121 (F- *dam-3* *dcm-6* *ara-14* *fhuA31* *galK2* *galT22* *hsdR3* *lacY1* *leu-6* *thi-1* *thr-1* *tsx-78*) was used.

## CONSTRUCTION OF MUTANT *Hfx. Volcanii* STRAINS

Mutant strains were constructed using the pop-in/pop-out method (Bitan-Banin et al., 2003) in *Hfx. volcanii*  $\Delta$ *pyrE2* strain H26 as follows. First, plasmid pTA131-HfxMCM-HXba was constructed by using oligonucleotide primers HfxMCM-5H and HfxMCM-3X (designed to include *HindIII* and *XbaI* sites respectively, see Table 3) to amplify a 1.0 kb fragment of *Hfx. volcanii* genomic DNA spanning the region from 100 nucleotides upstream of the *mcm* ORF to 900 nucleotides inside the ORF. The PCR product was digested with *HindIII* and *XbaI*, cloned into plasmid pTA131 digested with the same two enzymes (Allers et al., 2004) and sequenced to confirm the absence of PCR errors.

The resulting plasmid (pTA131-HfxMCM-HXba) was then used as a template for PCR overlap extension mutagenesis (OEM) to create 25 mutant derivatives. The sequences of the mutagenic primers are given in Table 3. OEM was performed using the GC-Rich PCR system buffer and GC-Rich enzyme mix (Roche) for the first round of PCR and the GC-Rich PCR system buffer and *Taq* polymerase (NEB) for the second round. Oligonucleotides HfxMCM-5H and HfxMCM-3X were used as

the flanking primers throughout. The resulting PCR products were then restricted with *Bsu36I* and *BstEII* (for  $\beta$ 7- $\beta$ 8  $\beta$ -hairpin and zinc binding domain mutants) or *BstEII* and *XbaI* (for  $\beta$ 9- $\beta$ 10  $\beta$ -hairpin mutants) and the 329 bp *Bsu36I-BstEII* or 271 bp *BstEII-XbaI* pieces carrying the mutations cloned back into pTA131-HfxMCM-HXba from which the corresponding wild-type restriction fragments had been removed. Plasmids were again sequenced to confirm the absence of unwanted sequence changes before being passaged through *E.coli* GM121 to generate unmethylated DNA for transformation into *Hfx. volcanii*  $\Delta$ *pyrE2* strain H26 (Table 2).

Transformation of *Hfx.volcanii* was accomplished as described in the Halohandbook v7.2 (www.haloarchaea.com/resources/halohandbook). Transformants obtained on Hv-CA medium lacking uracil were grown for 30 generations at 45°C in non-selective Hv-YPC medium to allow loss of the plasmid before being plated on Hv-CA plates containing 50  $\mu$ g/ml 5-FOA and a 10  $\mu$ g/ml uracil. Colonies formed on these plates were then inoculated into 500  $\mu$ l of Hv-YPC liquid medium and grown overnight at 45°C. Genomic DNA from these was prepared by taking 10  $\mu$ l of overnight culture, adding to 500  $\mu$ l of sterile water and heating to 70°C for 10 min to lyse the cells. For the  $\beta$ -hairpin and zinc binding domain mutants, 1  $\mu$ l of the resulting mix was used in PCR reactions to screen for the presence of the mutants in the chromosome. To discriminate between wild-type and mutant sequences, oligonucleotide primers with mismatched 3' sequences were used in PCR reactions in conjunction with primer HFXMCM-R1150, which lies 1150 nucleotides into the *mcm* ORF and which is therefore not present in pTA131-HfxMCM-HXba (see Table 3 for oligonucleotide sequences). For the silent restriction site mutant, the partial *mcm* ORF was amplified using primers HfxMCM-5H and HfxMCM-3X and the PCR products digested with Acc65I to identify mutants. In all cases, putative positive clones were then re-streaked twice to single colonies on Hv-YPC agar at

**Table 3 | Oligonucleotides used in this study.**

Primer name	Sequence
<b>A. FOR CONSTRUCTION OF pTA131-HfxMCM-HXba</b>	
HfxMCM-5H	5'-GTGTTGTTAAG <u>CCTCCCGCAGGGCAGCGA</u> -3'
HfxMCM-3X	5'-GGTGGTGGT <u>TCTAGACATGGCAATCTTCCTG</u> -3'
<b>B. FOR CONSTRUCTION OF <math>\beta</math>7-<math>\beta</math>8 <math>\beta</math>-HAIRPIN MUTANTS BY OEM</b>	
HFXMCM-rM1-F	5'-AAACTGCGCTCGCCGAGTCCCCGAGGGC-3'
HFXMCM-rM1-R	5'-GCCCTCGGGGGACTCGGCACGCCAGTT-3'
HFXMCM-rM2-F	5'-AAACTGCGCTCAGGCC <u>CTCCCCGAGGGCCTGCG</u> -3'
HFXMCM-rM2-R	5'-CGCAGGCC <u>CTCGGGGAGGCC</u> TCGACGCCAGTT-3'
HFXMCM-rM3-F	5'-CAGGAGTCCC <u>GCGGCC</u> CTGCGGGGG-3'
HFXMCM-rM3-R	5'-GCCCCCGCGCAGGCC <u>GGGGGACTCTG</u> -3'
HFXMCM-rM4-F	5'-CCGAGGGC <u>CTGGCGGGGCGAGCCG</u> -3'
HFXMCM-rM4-R	5'-CGCGT <u>CTGCCCGCC</u> CAAGCCCTCGGG-3'
HFXMCM-rM5-F	5'-CTGCGGGG <u>CGCACGCC</u> CAGACCATC-3'
HFXMCM-rM5-R	5'-GAT <u>GCTCTGGCGTGGCGCCCCGCG</u> AG-3'
HFXMCM-rM6-F	5'-GGCAGACGCC <u>CCAGCATGCACATCAAC</u> -3'
HFXMCM-rM6-R	5'-GTTGAT <u>GTCATGGCCGCTCTCG</u> GG-3'
HFXMCM-rD1-F	5'-CAGGAGTCCC <u>GAGGGCAGACGCC</u> CAG-3'
HFXMCM-rD1-R	5'-CTGCG <u>CGCTCGCC</u> CTGGGGGACTCTG-3'
HFXMCM-rD2-F	5'-GTCAGGA <u>GATCCCC</u> GAGACGCCAGAGC-3'
HFXMCM-rD2-R	5'-GCT <u>CTCGGGCTCT</u> GGGGGACTCTGGAC-3'
HFXMCM-rD3-F	5'-CGCGT <u>CCAGGATCC</u> -ACGCCGAGACATC-3'
HFXMCM-rD3-R	5'-GAT <u>GCTCTGGCGTGGACTCC</u> GGACGCG-3'
HFXMCM-rD4-F	5'-CTGCG <u>CTCAGGAG</u> -CCGAGACATCGAC-3'
HFXMCM-rD4-R	5'-GTC <u>ATGTCCTGCGG</u> CTCTGGACGCCAG-3'
<b>C. FOR CONSTRUCTION OF <math>\beta</math>9-<math>\beta</math>10 <math>\beta</math>-HAIRPIN MUTANTS BY OEM</b>	
HFXMCM-bH1-F	5'-GTCGGCGT <u>CTCGCAGCGAACAGCAGACATCG</u> -3'
HFXMCM-bH1-R	5'-CGAT <u>GTCGCTGTTCGCTGAGGACGCC</u> GAC-3'
HFXMCM-bH2-F	5'-GTC <u>CTCACATCGCAGCGAACATCGGGCAAC</u> -3'
HFXMCM-bH2-R	5'-GTT <u>CCCGATGTCTCGCTGCGATGTGGAGAC</u> -3'
HFXMCM-bH3-F	5'-CACAT <u>GAACAGGCAGCGTGGCAACGAGAAG</u> -3'
HFXMCM-bH3-R	5'-CTT <u>CTGTTGCCGACGCTGCTGATGTG</u> -3'
HFXMCM-bH4-F	5'-GAACAG <u>CGACAGCAGCGAACGAGAACGCC</u> -3'
HFXMCM-bH4-R	5'-GGCGT <u>CTTCCTCGTTCGCTGCTGTGTT</u> -3'
HFXMCM-bH5-F	5'-CAGAC <u>ATGGCGCAGCGAACGCCCC</u> TTC-3'
HFXMCM-bH5-R	5'-GAAGAC <u>GGGCCTTCGCTGCGCCGATGTCTG</u> -3'
HFXMCM-bH6-F	5'-TCGGG <u>CAACAGGCAGGCCGCTTCACTAC</u> -3'
HFXMCM-bH6-R	5'-TAGTC <u>GAAGACGGCGCTGCCCTGCCCCG</u> A-3'
HFXMCM-hD1-F	5'-CACAT <u>GAACAGCAG-GGCAACGAGAACG</u> -3'
HFXMCM-hD1-R	5'-CGT <u>CTCTCGTTGCC-CTGCTGTTGATGTG</u> -3'
HFXMCM-hD2-F	5'-CTCC <u>CACATCGAACAG-AACGAGAACGCC</u> -3'
HFXMCM-hD2-R	5'-GGCGT <u>CTTCCTCGTTCTGATGTGGAG</u> -3'
HFXMCM-hD3-F	5'-GTC <u>CTCACATCGAA-GAGAACGCCCC</u> TG-3'
HFXMCM-hD3-R	5'-GACGGGCG <u>TCTCTCTGATGTGGAGGAC</u> -3'
HFXMCM-hD4-F	5'-GGCGT <u>CTCCACATCGAACATCGAACGCC</u> -3'
HFXMCM-hD4-R	5'-GAAGAC <u>GGGCCTTCGTT-GATGTGGAGGACGCC</u> -3'
<b>D. FOR CONSTRUCTION OF ZINC BINDING DOMAIN MUTANTS BY OEM</b>	
HFXMCM-C137A-F	5'-CCGC <u>CTTCGAGGCGCAGCGCTCGG</u> -3'
HFXMCM-C137A-R	5'-CCG <u>AGCGCTGCCCTGAAGCGGG</u> -3'
HFXMCM-C140A-F	5'-AGT <u>GCCAGCGCGGGGACGATGAG</u> -3'
HFXMCM-C140A-R	5'-CT <u>CATCGCCCCGCGCTGGCACT</u> -3'
HFXMCM-C159A-F	5'-AAC <u>CCCACGAGGCGCAGGGATGCGA</u> -3'
HFXMCM-C159A-R	TCG <u>CATCCCCGCGCTGTTGAGG</u> -3'
HFXMCM-C162A-F	5'-AGT <u>GTCAGGGAGCGGAGGCCAGGG</u> -3'
HFXMCM-C162A-R	5'-CC <u>CTGGCGCTCGCTCCCTGACACT</u> -3'
<b>E. FOR CONSTRUCTION OF SILENT Acc65I mutation by OEM</b>	
HFXMCM-S1-F	5'-TG <u>CCAGCGCTCGGTACCATGAGTACATC</u> -3'
HFXMCM-S1-R	5'-GAT <u>GTAGCTATGGTACCGCAGCGCTGGCA</u> -3'

(Continued)

**Table 3 | Continued**

Primer name	Sequence
<b>F. FOR DETECTION OF <math>\beta</math>7-<math>\beta</math>8 <math>\beta</math>-HAIRPIN MUTANTS</b>	
HFXMCM-rM1-WT	5'-CAGAA <u>ACTGCGCTCCAG</u> -3'
HFXMCM-rM1-MUT	5'-CAGAA <u>ACTGCGCTCGCC</u> -3'
HFXMCM-rM2-WT	5'-AA <u>ACTGCGCTCCAGGAG</u> -3'
HFXMCM-rM2-MUT	5'-AA <u>ACTGCGCTCCAGGCC</u> -3'
HFXMCM-rM3-WT	5'-GTC <u>CAAGGAGTCCCCGAG</u> -3'
HFXMCM-rM3-MUT	5'-GTC <u>CAAGGAGTCCCCGCC</u> -3'
HFXMCM-rM4-WT	5'-T <u>CCCCCGAGGGCCTGCG</u> -3'
HFXMCM-rM4-MUT	5'-T <u>CCCCCGAGGGCCTGGCG</u> -3'
HFXMCM-rM5-WT	5'-G <u>GCCTGCGGGGGCGAG</u> -3'
HFXMCM-rM5-MUT	5'-G <u>GCCTGCGGGGGCGCC</u> -3'
HFXMCM-rM6-WT	5'-GGGG <u>CGAGACGCCGAG</u> -3'
HFXMCM-rM6-MUT	5'-GGGG <u>CGAGACGCCGCC</u> -3'
HFXMCM-rD1-MUT	5'-GGAG <u>TCCCCGAGGGCGAG</u> -3'
HFXMCM-rD2-MUT	5'-CCAGG <u>AGTCCCCGAGAC</u> -3'
HFXMCM-rD3-MUT	5'-CGT <u>CCAGGAGTCCACGCC</u> -3'
HFXMCM-rD4-MUT	5'-CTG <u>CGCTCCAGGAGGCC</u> -3'
<b>G. FOR DETECTION OF <math>\beta</math>9-<math>\beta</math>10 <math>\beta</math>-HAIRPIN MUTANTS</b>	
HFXMCM-bH1-WT	5'-GTC <u>GGCGTCTCCACATC</u> -3'
HFXMCM-bH1-MUT	5'-GTC <u>GGCGTCTCGCAGCG</u> -3'
HFXMCM-bH2-WT	5'-GTC <u>CTCCACATCGAACAG</u> -3'
HFXMCM-bH2-MUT	5'-GTC <u>CTCCACATCGCAGCG</u> -3'
HFXMCM-bH3-WT	5'-CACAT <u>CGAACAGCAGACA</u> -3'
HFXMCM-bH3-MUT	5'-CACAT <u>CGAACAGGCCAGCG</u> -3'
HFXMCM-bH4-WT	5'-GAAC <u>AGCAGACATCGGG</u> -3'
HFXMCM-bH4-MUT	5'-GAAC <u>AGCAGACAGCAGCG</u> -3'
HFXMCM-bH5-WT	5'-CAGAC <u>ATCGGGCAACGAG</u> -3'
HFXMCM-bH5-MUT	5'-CAGAC <u>ATCGGGCGCAGCG</u> -3'
HFXMCM-bH6-WT	5'-ACAT <u>CGGGCAACGAGAAGA</u> -3'
HFXMCM-bH6-MUT	5'-ACAT <u>CGGGCAACGAGGGCAG</u> -3'
HFXMCM-hD1-WT	5'-CACAT <u>CGAACAGCAGACA</u> -3'
HFXMCM-hD1-MUT	5'-CACAT <u>CGAACAGCAGGGC</u> -3'
HFXMCM-hD2-WT	5'-CTCC <u>ACATCGAACAGCAG</u> -3'
HFXMCM-hD2-MUT	5'-CTCC <u>ACATCGAACAGAAC</u> -3'
HFXMCM-hD3-WT	5'-TC <u>CTCCACATCGAACAGC</u> -3'
HFXMCM-hD3-MUT	5'-TC <u>CTCCACATCGAAAGAGA</u> -3'
HFXMCM-hD4-WT	5'-GG <u>CGTCTCCACATCGAA</u> -3'
HFXMCM-hD4-MUT	5'-GG <u>CGTCTCCACATCAAG</u> -3'
<b>H. FOR DETECTION OF ZINC BINDING DOMAIN MUTANTS</b>	
HFXMCM-C137A-WT	5'-GAAG <u>CCGCTTCGAGTGC</u> -3'
HFXMCM-C137A-MUT	5'-GAAG <u>CCGCTTCGAGGGCG</u> -3'
HFXMCM-C140A-WT	5'-TTC <u>GAAGTGCCAGCGCTGC</u> -3'
HFXMCM-C140A-MUT	5'-TTC <u>GAAGTGCCAGCGCGC</u> -3'
HFXMCM-C159A-WT	5'-CAGGA <u>ACCCACGAGGT</u> -3'
HFXMCM-C159A-MUT	5'-CAGGA <u>ACCCACGAGGCC</u> -3'
HFXMCM-C162A-WT	5'-CAC <u>GAAGTGTCAAGGGATGC</u> -3'
HFXMCM-C162A-MUT	5'-CAC <u>GAAGTGTCAAGGGAGCG</u> -3'
<b>I. REVERSE PRIMER FOR MUTANT DETECTION</b>	
HFXMCM-R1150	5'-G <u>CGAATCCCGAGCCGTC</u> -3'

Restriction sites in oligonucleotide primers used for pTA131-HfxMCM-HXba construction underlined. Oligonucleotide primers for overlap extension mutagenesis (OEM) are shown in top strand-bottom strand pairs with the mutated bases underlined in the top strand primer (for amino acid substitutions) or with deletion boundaries indicated with a dash. Oligonucleotide primers for mutant detection are shown with 3' mismatched bases underlined.

45°C before being re-tested by PCR using wild-type and mutant primers. Finally, the presence of the mutation was confirmed by sequencing of the relevant part of the *mcm* ORF amplified by PCR from genomic DNA (see Supplementary Information for DNA sequence traces).

## RESULTS

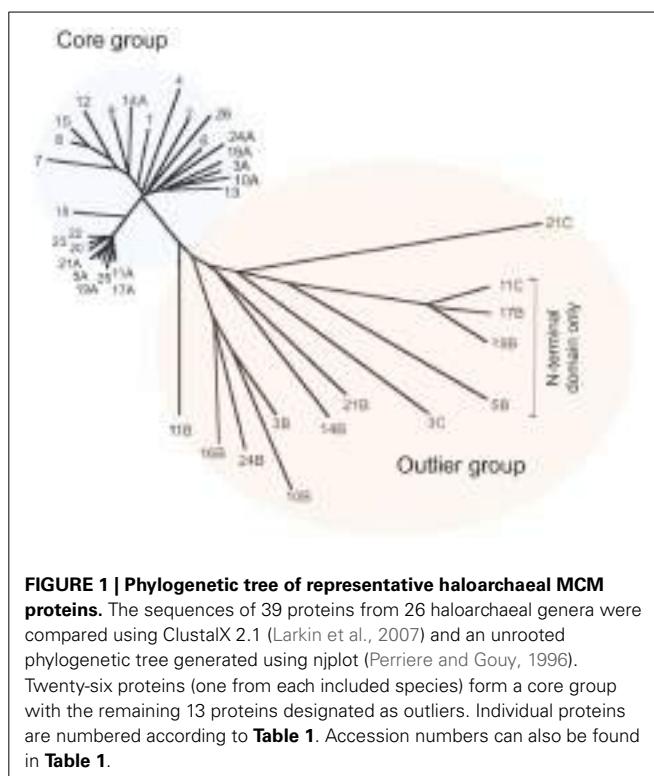
HALOARCHAEAL MCM PROTEINS

Prior to embarking on reverse genetic analysis of *Hfx. volcanii* MCM protein function, we undertook a detailed bioinformatic analysis of MCM protein distribution and conservation across the haloarchaea. To identify haloarchaeal MCM proteins, we searched the UniProtKB sequence database using BLAST with the *Hfx. volcanii* MCM protein (HVO\_0220, UniProtKB accession number D4GZG5) as the query sequence. Almost 200 proteins were identified in ~120 species belonging to 26 different genera of the *Halobacteriaceae*. To simplify further analysis, we selected a single species as a representative of each genus. These 26 species encoded a total of 39 MCM proteins. A complete list of the species under investigation, together with accession numbers for the proteins we identified, can be found in **Table 1**.

To investigate the relationship between the proteins in greater detail, we constructed multiple protein sequence alignments using ClustalX 2.1 (Larkin et al., 2007). Previously, when analysing a significantly smaller number of genomes (five), we identified a core group of five MCM proteins and a further three proteins that we designated as outliers (MacNeill, 2009). We performed a similar analysis of the larger dataset, using ClustalX to generate multiple sequence alignments of protein sequences (from which inteins were first removed—see below) and njplot (Perriere and Gouy, 1996) to generate unrooted phylogenetic trees. **Figure 1** shows the phylogenetic tree for the 39 proteins. As before, two groups are apparent, corresponding to the core and outlier groups defined previously (MacNeill, 2009). Each of the 26 species encodes a single member of the core group that includes the *Hfx. volcanii* MCM protein. The core group MCM proteins range in length from 697 to 702 amino acids (after inteins are removed—see below) and display a minimum pairwise protein sequence identity of 68% (range 68–95%). Three hundred and twenty four residues (46% of the protein sequence) are absolutely conserved across all 26 core group proteins. This includes the key catalytic residues that make up the Walker A (P-loop) motif (GDPGTGKS in all 26 proteins compared to the consensus GX<sub>4</sub>GKS/T), the Walker B motif (DELD in all 26 proteins) and arginine finger motif (SRF in all 26 proteins).

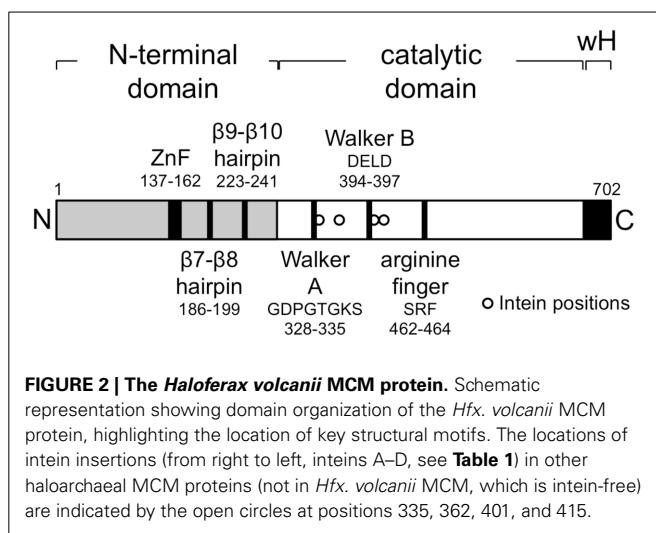
The outlier group proteins display much greater variety in length (312–717 amino acids) and sequence similarity (24–48% identity with *Hfx. volcanii* MCM). The four shortest proteins (312–315 amino acids) are made up of sequences related to the non-catalytic N-terminal domain of MCM only and cannot possess helicase activity. The remaining nine outliers all possess conserved Walker B (DEL/ID) and arginine finger (SRF) motifs, as well as the key lysine in the Walker A (P-loop) motif, suggesting that these proteins may well have ATPase and/or helicase activities. To date, none has been characterized biochemically.

Inteins are parasitic genetic elements capable of efficient self-splicing at the protein level (Gogarten and Hilario, 2006) and are



a common feature of the haloarchaeal MCM proteins. Amongst the 39 proteins, inteins were found in 13, all members of the core group defined above, and at four different locations within the C-terminal catalytic domain of the protein, with the largest number of inteins in an individual MCM protein being four (**Table 1**). **Figure 2** shows the position of the four inteins relative to conserved sequence features found in archaeal MCM proteins. As noted previously, inteins are frequently located at or very near to highly conserved and functionally important sequence regions (Gogarten and Hilario, 2006) and the haloarchaeal MCM inteins are no exception: intein insertion site A is located immediately C-terminal to the essential lysine in the Walker A sequence GDPGTGKS mentioned above, while intein C lies just four amino acids C-terminal to the Walker B sequence DELD (**Table 1**, **Figure 2**). Intein splicing is therefore likely to be essential for protein function.

In a number of archaeal organisms, including representatives of the crenarchaea, euryarchaea, thaumarchaea and korarchaea, the gene encoding MCM is found adjacent to that encoding a GINS subunit (MacNeill, 2010). To ask whether this arrangement is also found in any of the species under investigation in this report, we examined the chromosomal context of each of the 26 core MCM genes and also the 13 outlier proteins. None of the genes encoding core group proteins is located adjacent to a gene encoding GINS or indeed, to any known replication gene (data not shown). A similar situation is seen with genes encoding 12 of the 13 outlier proteins, the sole exception being the pNG3053 protein (UniProtKB accession number Q5V814, labeled 3C in **Figure 1**) encoded by plasmid pNG300 of *Haloarcula marismortui* which is located immediately 3' to an ORF encoding



the C-terminally truncated GINS homolog pNG3052 (data not shown).

#### REVERSE GENETIC ANALYSIS OF MCM FUNCTION

*Hfx. volcanii* encodes a single intein-free MCM protein of 702 amino acids in length (Table 1) that comprises an N-terminal domain that spans residues 1–283, an AAA+ catalytic core domain spanning residues 283–632 and a C-terminal winged helix-turn-helix (wH) domain spanning residues 633–702. All the conserved sequence features characteristic of archaeal MCM proteins are found in the *Hfx. volcanii* protein (Figure 2). In order to test whether *Hfx. volcanii* was a workable model for reverse genetic analysis of archaeal MCM function, three of these conserved features were targeted for mutagenesis: the β7-β8 β-hairpin loop (also known as the allosteric communication loop, ACL), the β9-β10 β-hairpin loop and the putative zinc binding domain (Figure 2).

To introduce mutations into the *mcm* gene, we used the pop-in/pop-out method (Bitan-Banin et al., 2003). To achieve this, a plasmid was constructed carrying a 1.0 kb region of the *Hfx. volcanii* genome spanning the first 900 nucleotides of the *mcm* open reading frame together with 100 nucleotides of 5' flanking region. The plasmid also carries the *pyrE2* selectable marker (*pyrE2* function is required for uracil prototrophy) but does not possess a replication origin capable of promoting autonomous replication in *Hfx. volcanii*. Stable maintenance therefore requires that the plasmid integrates (pops-in) into the *Hfx. volcanii* genome by means of homologous recombination between the *mcm* sequences on the plasmid and the *mcm* gene in the chromosome. PCR overlap extension mutagenesis was used to generate a series of mutated forms of the plasmid in which the targeted amino acids were either replaced, singly or in pairs, with one or two alanine residues respectively, or deleted altogether. Next, the plasmids were introduced into *Hfx. volcanii* Δ*pyrE2* strain H26 by standard methods and transformant (pop-in) colonies obtained on Hv-CA plates lacking uracil. Multiple independent colonies were then individually picked and grown in non-selective (uracil-containing) Hv-YPC medium before

being plated onto Hv-CA plates containing 5-fluoroorotic acid to select for (pop-out) clones that had lost the plasmid (see Materials and Methods). Pop-out colonies were then screened by PCR using oligonucleotide primers specific for either the wild-type or mutant sequences (see Table 3 for primer sequences). Candidate mutant strains were sequentially re-streaked twice on non-selective medium (Hv-YPC) before the presence of the mutation was confirmed by PCR using wild-type- and mutant-specific primers and by sequencing of amplified genomic DNA (see Materials and Methods). Table 2 lists the strains constructed by this method.

#### MUTAGENESIS OF THE β7-β8 β-HAIRPIN LOOP

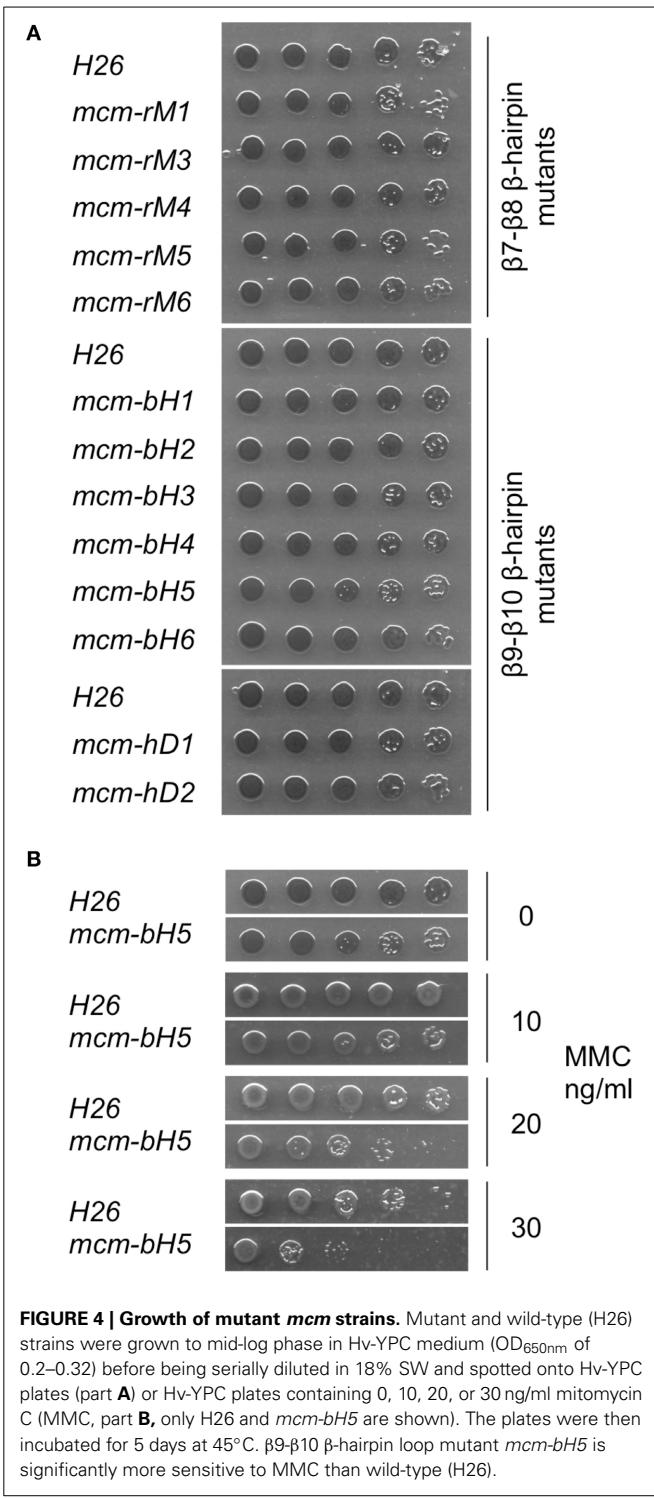
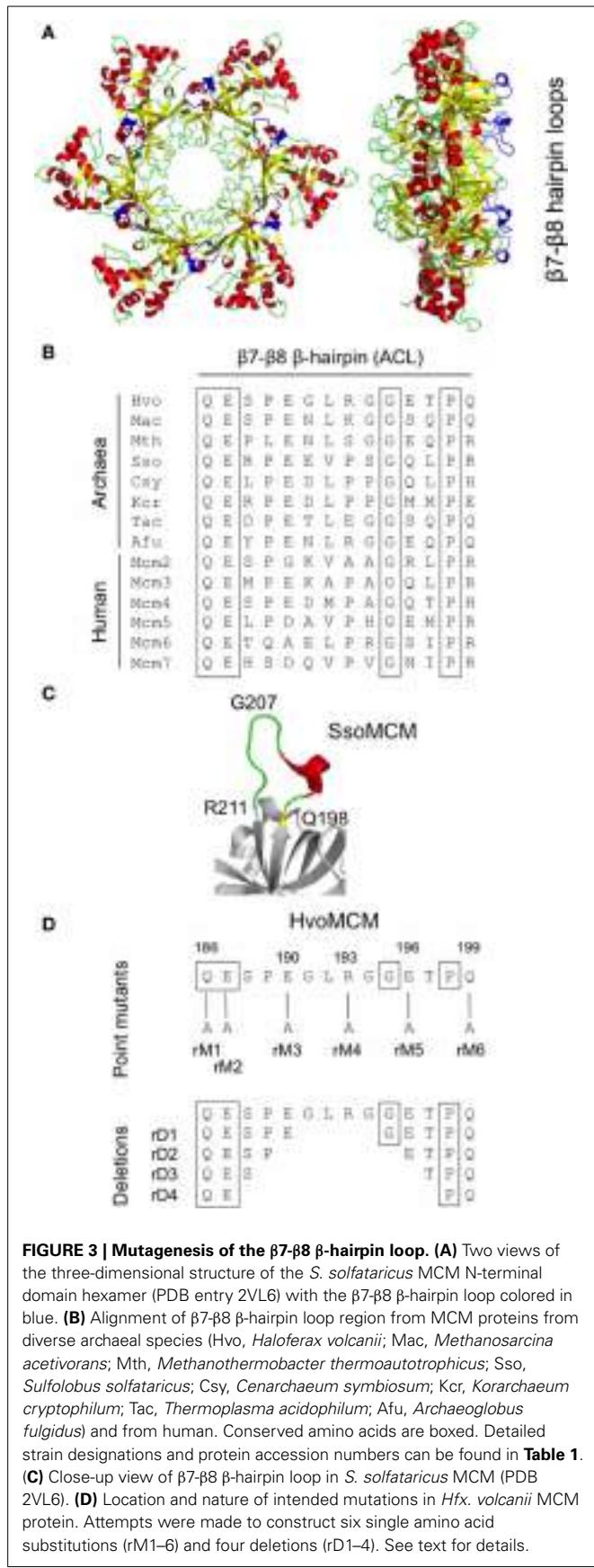
First identified on the basis of its high degree of sequence conservation across species, the β7-β8 β-hairpin loop (also known as the allosteric communication loop, ACL) is located at the interface between the N-terminal and AAA+ catalytic domains of the MCM protein (Figure 3). A number of mutants in β7-β8 loop have previously been analyzed biochemically, leading to the conclusion that the loop plays a role in coupling the activities of these two domains (Sakakibara et al., 2008; Barry et al., 2009). The mutations tested include six single point mutants in recombinant *M. thermoautotrophicum* MCM (including four at the boxed conserved residues shown in Figure 3B) as well as a triple point mutant and a complete replacement of the loop by the tripeptide serine-asparagine-glycine in *S. solfataricus* MCM (Sakakibara et al., 2008; Barry et al., 2009).

In order to test whether the function of the β7-β8 loop function was essential *in vivo* in *Hfx. volcanii*, we attempted to construct 10 different alleles (Figure 3): six single point mutants in which individual charged or bulky polar amino acids within the loop were replaced with alanine (mutants *mcm-rM1*–*mcm-rM6*) and four deletions, of two, four, six, and eight amino acids (mutants *mcm-rD1*–*mcm-rD4*, respectively). Two of the individual amino acids targeted (Q186 and E187) are conserved across species (Figure 3B); the equivalent residues were mutated in *M. thermoautotrophicum* MCM (Sakakibara et al., 2008).

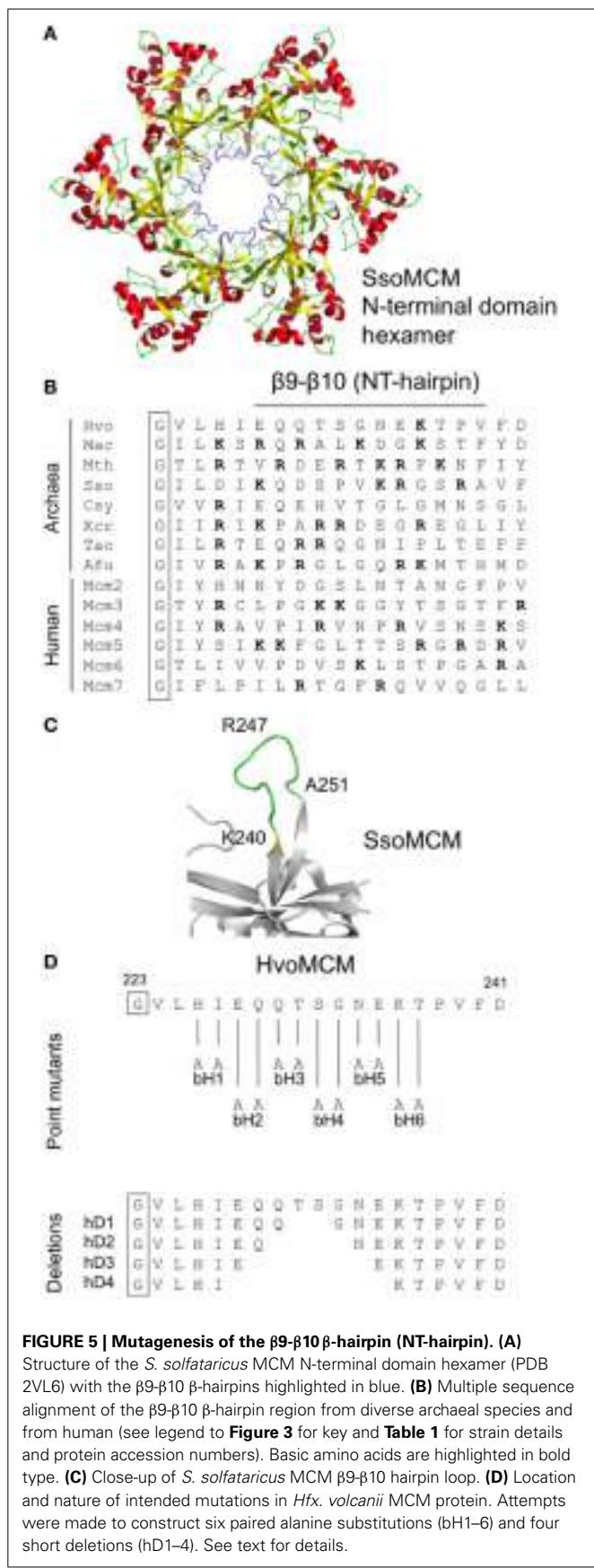
Of the 10 mutants, five (*mcm-rM1*, *mcm-rM3*–*mcm-rM6*) were readily identified by PCR screening using genomic DNA templates and oligonucleotide primers specific for the wild-type and mutant sequences (see Table 3) and confirmed by DNA sequencing (Supplementary Figure S1). Despite extensive screening (see Materials and methods), the remaining four strains (*mcm-rM2*, *mcm-rD1*–*mcm-rD4*) could not be isolated, suggesting that these mutations either inactivate or significantly impair the function of the *Hfx. volcanii* MCM protein in cells grown at 45°C. Growth of the five viable mutants was indistinguishable from the parental wild-type strain H26 (Figure 4A, upper panel). These results indicate that the β7-β8 loop is indeed essential for *Hfx. volcanii* MCM function although it is also clear that some point mutations (including replacement of conserved amino acid glutamine 186 with alanine) can be tolerated.

#### MUTAGENESIS OF THE β9-β10 β-HAIRPIN LOOP

Structural analysis of archaeal MCM proteins identifies four β-hairpins in each monomer, three of which protrude, to a greater or lesser extent, into the central channel through which single-



or double-stranded DNA is thought to pass (Slaymaker and Chen, 2012). One of these hairpins, the positively charged  $\beta$ 9- $\beta$ 10  $\beta$ -hairpin (also known as the NT-hairpin) is located in the N-terminal domain of the protein (see **Figure 5A**). Unlike the  $\beta$ 7- $\beta$ 8 loop described above, the sequence of this part of the MCM protein is not well-conserved across evolution (**Figure 5B**).



However, it appears that the positively charged nature of the hairpin is important for function: mutation of arginine 226 and lysine 228 in the *M. thermoautotrophicum* MCM abolishes that protein's ability to bind to DNA (Fletcher et al., 2003).

To probe the *in vivo* function of the  $\beta$ 9– $\beta$ 10  $\beta$ -hairpin loop in *Hfx. volcanii*, and in the absence of highly conserved amino acids presenting themselves as obvious targets for mutagenesis, we initially attempted to construct six mutants, *mcm-bH1*–*mcm-bH6*, in which adjacent amino acids were replaced with paired alanines (Figure 5D). All six mutants were recovered following PCR screening (and confirmed by sequencing, see Supplementary Figure S2), implying that the  $\beta$ 9– $\beta$ 10  $\beta$ -hairpin loop is readily mutable. None of six mutants exhibited obvious growth defects at 45°C (Figure 4A, middle panel).

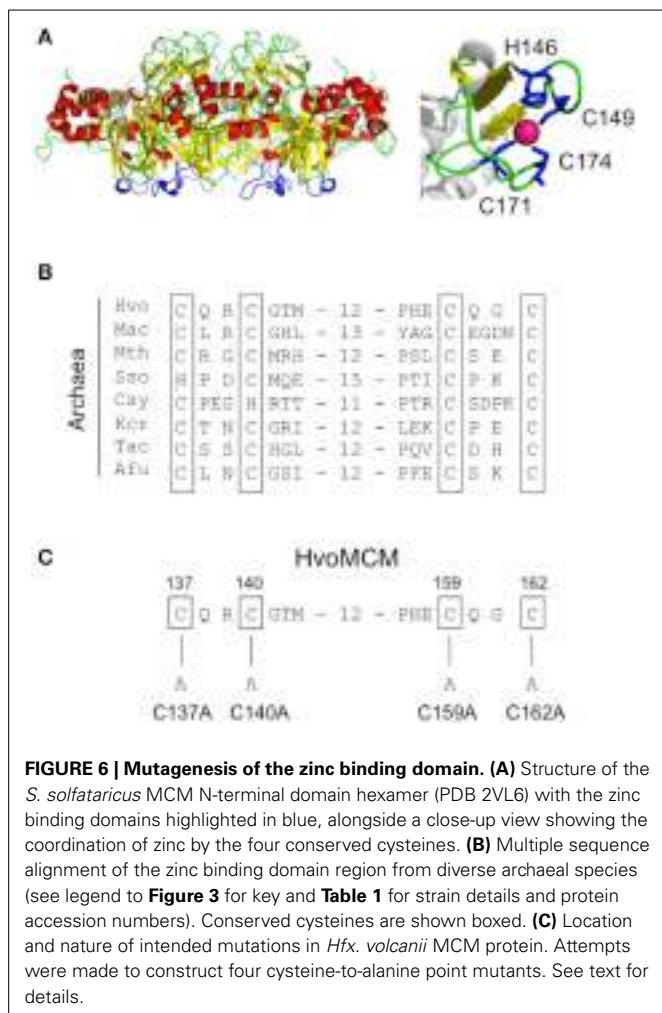
We therefore extended this analysis by attempting to create strains carrying deletions of increasing size in the  $\beta$ 9– $\beta$ 10 loop (mutants *mcm-hD1*–*mcm-hD4*, see Figure 5D). However, despite extensive screening, only two of the four could be isolated: *mcm-hD1* and *mcm-hD2*. Growth of these strains, like *mcm-bH1*–*mcm-bH6*, was indistinguishable from the parental wild-type H26 (Figure 4A, lower panel). That we were unable to isolate *mcm-hD3* and *mcm-hD4* strains strongly implies that these deletions either inactivate or significantly impair the function of the *Hfx. volcanii* MCM protein. We conclude from this analysis that while the precise sequence of the  $\beta$ 9– $\beta$ 10  $\beta$ -hairpin loop is not absolutely required for the function of the protein, the loop itself does have a crucial role.

#### MUTAGENESIS OF THE ZINC BINDING DOMAIN

The N-terminal domain of the archaeal MCM proteins contains cysteine and histidine residues that fold into a zinc binding domain (Figure 6) (Slaymaker and Chen, 2012). A similar domain appears also to be present in eukaryotic MCM proteins but its precise function in either kingdom is unknown. We attempted to construct mutants in which each of the four cysteines was individually replaced with alanine (Figure 6C). However, none of these four mutants that we attempted to construct (*mcm-C137A*, *mcm-C140A*, *mcm-C159A*, and *mcm-C162A*, see Figure 6C) could be isolated, implying that all four cysteines—and presumably zinc binding—is essential for MCM function *in vivo*. As a control, we tested whether it was possible to introduce a silent mutation into the vicinity of the cysteine 140 codon by changing the sequence GGGACG encoding glycine 141 and threonine 142 to an *Acc65I* restriction site, GGTACC (mutation *mcm-S1*). Sixty colonies were screened by *Acc65I* digestion of the PCR amplified *mcm* gene: 26 contained the *Acc65I* site (data not shown), indicating that the region of the gene encoding the zinc binding domain can be mutated in this manner.

#### SENSITIVITY TO MITOMYCIN C

In total, we isolated 13 mutant strains carrying either single alanine substitutions, paired alanine substitutions and short sequence deletions in two conserved sequence elements, the  $\beta$ 7– $\beta$ 8 and  $\beta$ 9– $\beta$ 10  $\beta$ -hairpin loops. None of these strains displayed obvious growth deficiencies: none of the strains grew slowly (Figure 4A) nor were any of the strains cold-sensitive or



temperature-sensitive (data not shown). We also tested whether the strains might be sensitive to DNA damage induced by the DNA modifying agent mitomycin C (MMC). MMC forms three types of MMC-DNA adducts: monoadducts, intrastrand biadducts and interstrand crosslinks (Tomasz, 1995; Bargonetti et al., 2010). Of the 13 tested strains, one strain, *mcm-bH5*, showed increased sensitivity to MMC treatment (Figure 4B). This mutant has two adjacent residues in the  $\beta$ 7- $\beta$ 10  $\beta$ -hairpin loop, asparagine 234 and glutamate 235, replaced with alanine.

## DISCUSSION

The MCM helicase is the key catalytic engine of DNA unwinding during chromosome replication in eukaryotes and most likely in archaea also. Three factors make the archaeal MCM proteins excellent models for their human counterparts. First, the relatively high level of sequence similarity between the eukaryotic and archaeal MCM proteins throughout the non-catalytic N-terminal and catalytic AAA+ domains. Second, the relative simplicity of the homohexameric archaeal MCM complexes compared to the heterohexameric eukaryotic complexes. Third, the relative ease with which certain archaeal MCM proteins can be purified in recombinant form, assayed for various activities and crystallized

for structure determination. It is striking that although more than 10 years have passed since publication of the first partial crystal structure of an archaeal MCM (Fletcher et al., 2003), no high-resolution eukaryotic MCM structures have been solved. Until this occurs, the archaea will remain an important model allowing detailed dissection of MCM protein function.

Using multiple protein sequence alignments and crystal structures as a guide, a number of groups have identified regions of the MCM protein with potentially important roles in MCM function. These have then been mutated and the consequences for MCM activity determined by a variety of *in vitro* biochemical assays. The  $\beta$ 7- $\beta$ 8  $\beta$ -hairpin loop was first identified in this way, for example (Sakakibara et al., 2008; Barry et al., 2009). However, despite the growing number of similar studies in the literature, no attempt has been made to examine the *in vivo* consequences of such mutations.

Here we describe the results of probing the *in vivo* function of specific amino acids within an archaeal MCM protein using the haloarchaeon *Hfx. volcanii* as a model. *Hfx. volcanii* encodes a single MCM protein of 702 amino acids. This protein is part of the core group of haloarchaeal MCM proteins defined in a previous study (MacNeill, 2009) and expanded upon here. Each of 26 haloarchaeal species investigated encodes a single member of this group (Table 1, Figure 1). Ten species encode additional MCM family proteins classed as outliers. Given that a single core protein is found in all species examined and that 16 of the species encode only a single MCM, it is highly likely that these proteins act as replicative helicases. In support of this, the single MCM protein encoded by *Hbt. salinarum* NRC-1 (equivalent to the *Hbt. salinarum* R1 protein listed in Table 1 and labeled as number 4 in Figure 1) has previously been shown to be essential for cell survival (Berquist et al., 2007). While we have not attempted to delete the *Hfx. volcanii mcm* gene in its entirety, our inability to isolate certain mutant *mcm* alleles in this study strongly points to the *Hfx. volcanii mcm* being essential also.

The cellular functions of the outlying haloarchaeal MCM proteins are unknown. With the exception of the group of four C-terminally truncated MCM proteins highlighted in Figure 1 (proteins labeled 5B, 11C, 17B, and 19B), all the outliers possess intact Walker A, Walker B and arginine finger motifs, suggesting that they may be active as ATPase and/or DNA helicases. Interestingly, *T. kodakarensis* encodes three MCM proteins, all three of which have helicase activity, although only one is essential for cell viability and which therefore is the likely replicative helicase (Ishino et al., 2011; Pan et al., 2011).

In order to determine whether *Hfx. volcanii* presented a workable model for reverse genetic analysis of archaeal MCM function, we targeted three regions of the proteins for investigation. Initially, we focused on the  $\beta$ 7- $\beta$ 8  $\beta$ -hairpin loop located at the interface between the N-terminal and catalytic domains (Figure 3A) and previously identified as having a role in communicating conformational changes from the DNA binding N-terminal domain to the catalytic AAA+ domain (Sakakibara et al., 2008; Barry et al., 2009). The  $\beta$ 7- $\beta$ 8 loop is well conserved across eukaryotic and archaeal evolution (Figure 3B). In this study, we constructed five strains (*mcm-rM1*, *mcm-rM3*—*mcm-rM6*)

carrying single point mutations in the *Hfx. volcanii* MCM  $\beta$ 7- $\beta$ 8  $\beta$ -hairpin loop (**Table 2**, **Figures 3, 4**). The mutated residues included glutamine 186, which is conserved in both archaeal and eukaryotic MCM proteins (**Figure 3B**) but which can be replaced by alanine without significantly affecting cell growth, as well as three charged residues (glutamates 190 and 196, and arginine 193) and glutamine 199, all of which were also replaced by alanine without markedly affecting growth rates. In addition, none of the five  $\beta$ 7- $\beta$ 8 loop mutants led to increased sensitivity to MMC exposure in spotting assays (**Figure 4**).

In contrast, we were unable to isolate mutant *mcm-rM2* encoding a protein in which conserved residue glutamate 187 was replaced by alanine, nor any of the four deletion alleles *mcm-rD1*—*mcm-rD4*. Replacement of glutamate 182 in the *M. thermoautotrophicum* MCM protein, the residue corresponding to glutamate 187 in *Hfx. volcanii*, with arginine led to marked reductions in ATPase and helicase activities *in vitro*, without affecting either ATP or DNA binding (Sakakibara et al., 2008), whereas mutation of conserved glutamine 181 (equivalent to glutamine 186 in *Hfx. volcanii* MCM, the residue mutated in *mcm-rM1*) to alanine had little impact. Clearly, conservation is no predictor of *in vivo* essentiality.

We turned next to the  $\beta$ 7- $\beta$ 8  $\beta$ -hairpin loop, also known as the NT-hairpin. This loop extends into the central channel of the MCM hexamer and may have a role in tracking DNA through the channel based on the fact that mutating two positively charged residues at the tip of the loop in the *M. thermoautotrophicum* MCM protein abolishes DNA binding *in vitro* (Fletcher et al., 2003). In *H. volcanii*, only a single basic residue is present in the  $\beta$ 9- $\beta$ 10 loop region, arginine 236. This is likely to lie at or near the tip of the  $\beta$ -hairpin (**Figures 5B,C**). By comparison with *M. thermoautotrophicum*, it would be reasonable to predict that this amino acid would be essential for N-terminal domain DNA binding and thus for MCM function *in vivo*. In the absence of more widespread sequence conservation, we chose to probe the *in vivo* function of the  $\beta$ 7- $\beta$ 8  $\beta$ -hairpin loop initially by attempting to construct a series of six mutants in which adjacent amino acids were replaced with pairs of alanines (**Figure 5D**). All six mutant strains (**Table 2**, **Figures 3, 5**) were viable, including *mcm-bH6* in which arginine 236 was replaced with alanine. Next, we attempted to construct four different  $\beta$ 9- $\beta$ 10 loop deletion mutants but only two were viable: *mcm-hD1* and *mcm-hD2* (**Table 2**, **Figures 3, 5**). The former removes amino acids 231 and 232, and the latter amino acids 230–233. In contrast, strains carrying two larger deletions, *mcm-hD3* and *mcm-hD4*, removing amino acids 229–234 and 228–235, respectively, could not be isolated. Thus, these results clearly demonstrate that while no individual amino acid within the  $\beta$ 9- $\beta$ 10 hairpin loop is essential for *Hfx. volcanii* MCM function *in vivo*, the loop does have a key role to play.

In addition to mutagenizing the  $\beta$ 7- $\beta$ 8 and  $\beta$ 9- $\beta$ 10 loops, we also attempted to individually replace with alanine each of the four cysteine residues that make up the zinc binding domain (**Figure 6**). However, we were unable to recover any of four desired mutants, *mcm-C137A*, *mcm-C140A*, *mcm-C159A*, or *mcm-C162A*, and conclude that the zinc binding domain is therefore likely to have an essential function *in vivo*. Consistent with

this, replacing the equivalent of *Hfx. volcanii* MCM cysteine 162 with serine produces an *M. thermoautotrophicum* MCM protein with impaired ATPase and single-stranded DNA binding activities and no helicase activity (Poplawski et al., 2001).

Finally, the mutant *Hfx. volcanii* strains generated in this study were tested for increased sensitivity to the DNA modifying drug mitomycin C (MMC). MMC is a potent DNA inter-strand crosslinker and is widely used as replication fork blocking agent, as replication cannot continue past such crosslinks. We tested all 13 viable *mcm* alleles for MMC sensitivity by spotting serially diluted cultures onto medium containing increasing concentrations of MMC and found one strain, *mcm-bH5*, that displayed significantly enhanced sensitivity compared with the parental wild-type H26 (**Figure 4B**). On medium lacking MMC, growth of *mcm-bH5*, as with the other viable  $\beta$ 9- $\beta$ 10 loop mutants, is indistinguishable from wild-type (**Figure 4A**). The *mcm-bH5* protein carries a double substitution in  $\beta$ 9- $\beta$ 10 loop with asparagine 234 and glutamate 235 being replaced with a pair of significantly less bulky alanines. In the absence of a crystal structure of the *Hfx. volcanii* MCM protein, it is difficult to predict whether the impact of the *mcm-bH5* mutations is confined to the  $\beta$ 9- $\beta$ 10 loop alone or whether these amino acid changes will have a wider impact on N-terminal domain structure. It is also difficult to envisage how the *mcm-bH5* mutations would cause cells to become supersensitive to MMC, particularly as interpreting the effect of MMC on cells is complicated by the different modes of action of this compound. MMC forms at least four different types of DNA adduct: two species of MMC-mono-dG-adduct, intra-strand dG-MMC-dG biadducts and inter-strand dG-MMC-dG crosslinks, with the latter being assumed to be the replication fork blocking lesion (Tomasz, 1995; Bargonetti et al., 2010). Given the key role of the MCM helicase at the fork it is tempting to speculate that the supersensitivity of *mcm-bH5* is a result of difficulties that arise when the mutant helicase encounters an inter-strand MMC crosslink. However, confirmation of this clearly requires additional experimentation.

In conclusion, the results presented here demonstrate the feasibility of using *Hfx. volcanii* as a model system for reverse genetic analysis of archaeal MCM protein function, provide important confirmation of the *in vivo* importance of conserved structural features identified by previous bioinformatic, biochemical and structural studies, and offer the prospect of more extensive mutational analysis, not only of MCM but of other key replication factors, in the future.

## AUTHOR CONTRIBUTIONS

Tatjana P. Kristensen, Rejea Maria Cherian, and Fiona C. Gray performed the bulk of the experimental work. Stuart A. MacNeill designed the study, carried out some of the experimental work and wrote the manuscript.

## ACKNOWLEDGMENTS

We would like to thank our colleagues in both Copenhagen and St. Andrews for their help with this study. This work was funded by Forskningsrådet for Natur og Univers (FNU sagsnr. 272-05-0446).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00123/abstract>

### Supplementary Figure S1 | Sequence analysis of $\beta7$ - $\beta8$ $\beta$ -hairpin mutant strains.

Codons mutated to encode alanine are boxed.

**Supplementary Figure S2 | Sequence analysis of  $\beta9$ - $\beta10$   $\beta$ -hairpin mutant strains.** Codons mutated to encode alanine pairs (mutants bH1–6) are boxed. Vertical lines in hD1 and hD2 indicate location of deleted sequences.

## REFERENCES

- Allers, T., Ngo, H. P., Mevarech, M., and Lloyd, R. G. (2004). Development of additional selectable markers for the halophilic archaeon *Halofexax volcanii* based on the *leuB* and *trpA* genes. *Appl. Environ. Microbiol.* 70, 943–953. doi: 10.1128/AEM.70.2.943-953.2004
- Bae, B., Chen, Y. H., Costa, A., Onesti, S., Brunzelle, J. S., Lin, Y., et al. (2009). Insights into the architecture of the replicative helicase from the structure of an archaeal MCM homolog. *Structure* 17, 211–222. doi: 10.1016/j.str.2008.11.010
- Bargonetti, J., Champeil, E., and Tomasz, M. (2010). Differential toxicity of DNA adducts of mitomycin C. *J. Nucleic Acids* 2010:698960. doi: 10.4061/2010/698960
- Barry, E. R., Lovett, J. E., Costa, A., Lea, S. M., and Bell, S. D. (2009). Intersubunit allosteric communication mediated by a conserved loop in the MCM helicase. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1051–1056. doi: 10.1073/pnas.0809192106
- Berquist, B. R., Dassarma, P., and Dassarma, S. (2007). Essential and non-essential DNA replication genes in the model halophilic Archaeon, *Halobacterium* sp. NRC-1. *BMC Genet.* 8:31. doi: 10.1186/1471-2156-8-31
- Bitan-Banin, G., Ortenberg, R., and Mevarech, M. (2003). Development of a gene knockout system for the halophilic archaeon *Halofexax volcanii* by use of the *pyrE* gene. *J. Bacteriol.* 185, 772–778. doi: 10.1128/JB.185.3.772-778.2003
- Brewster, A. S., Wang, G., Yu, X., Greenleaf, W. B., Carazo, J. M., Tjajadi, M., et al. (2008). Crystal structure of a near-full-length archaeal MCM: functional insights for an AAA+ hexameric helicase. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20191–20196. doi: 10.1073/pnas.0808037105
- Duderstadt, K. E., and Berger, J. M. (2008). AAA+ ATPases in the initiation of DNA replication. *Crit. Rev. Biochem. Mol. Biol.* 43, 163–187. doi: 10.1080/10409230802058296
- Ervin, C., Clarke, P., Zech, J., Lurz, R., Sun, J., Uhle, S., et al. (2009). A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20240–20245. doi: 10.1073/pnas.0911500106
- Farkas, J. A., Picking, J. W., and Santangelo, T. J. (2013). Genetic techniques for the archaea. *Annu. Rev. Genet.* 47, 539–561. doi: 10.1146/annurev-genet-111212-133225
- Fletcher, R. J., Bishop, B. E., Leon, R. P., Scalfani, R. A., Ogata, C. M., and Chen, X. S. (2003). The structure and function of MCM from archaeal *M. thermoautotrophicum*. *Nat. Struct. Biol.* 10, 160–167. doi: 10.1038/nsb893
- Fu, Y., Slaymaker, I. M., Wang, J., Wang, G., and Chen, X. S. (2014). The 1.8- $\text{\AA}$  crystal structure of the N-terminal domain of an archaeal MCM as a right-handed filament. *J. Mol. Biol.* 426, 1512–1523. doi: 10.1016/j.jmb.2013.12.025
- Gambus, A., Khoudoli, G. A., Jones, R. C., and Blow, J. J. (2011). MCM2-7 form double hexamers at licensed origins in *Xenopus* egg extract. *J. Biol. Chem.* 286, 11855–11864. doi: 10.1074/jbc.M110.199521
- Gogarten, J. P., and Hilaro, E. (2006). Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evol. Biol.* 6:94. doi: 10.1186/1471-2148-6-94
- Hawkins, M., Malla, S., Blythe, M. J., Nieduszynski, C. A., and Allers, T. (2013). Accelerated growth in the absence of DNA replication origins. *Nature* 503, 544–547. doi: 10.1038/nature12650
- Ishino, S., Fujino, S., Tomita, H., Ogino, H., Takao, K., Daiyasu, H., et al. (2011). Biochemical and genetical analyses of the three mcm genes from the hyperthermophilic archaeon, *Thermococcus kodakarensis*. *Genes Cells* 16, 1176–1189. doi: 10.1111/j.1365-2443.2011.01562.x
- Kamada, K. (2012). The GINS complex: structure and function. *Subcell. Biochem.* 62, 135–156. doi: 10.1007/978-94-007-4572-8\_8
- Krastanova, I., Sannino, V., Amenitsch, H., Gileadi, O., Pisani, F. M., and Onesti, S. (2012). Structural and functional insights into the DNA replication factor Cdc45 reveal an evolutionary relationship to the DHH family of phospho-esterases. *J. Biol. Chem.* 287, 4121–4128. doi: 10.1074/jbc.M111.285395
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404
- Li, Z., Pan, M., Santangelo, T. J., Chemnitz, W., Yuan, W., Edwards, J. L., et al. (2011). A novel DNA nuclease is stimulated by association with the GINS complex. *Nucleic Acids Res.* 39, 6114–6123. doi: 10.1093/nar/gkr181
- Liu, W., Pucci, B., Rossi, M., Pisani, F. M., and Ladenstein, R. (2008). Structural analysis of the *Sulfolobus solfataricus* MCM protein N-terminal domain. *Nucleic Acids Res.* 36, 3235–3243. doi: 10.1093/nar/gkn183
- MacNeill, S. A. (2009). The haloarchaeal chromosome replication machinery. *Biochem. Soc. Trans.* 37, 108–113. doi: 10.1042/BST0370108
- MacNeill, S. A. (2010). Structure and function of the GINS complex, a key component of the eukaryotic replisome. *Biochem. J.* 425, 489–500. doi: 10.1042/BJ20091531
- Magrane, M., and UniProt Consortium. (2011). UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* 2011:bar009. doi: 10.1093/database/bar009
- Makarova, K. S., Koonin, E. V., and Kelman, Z. (2012). The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. *Biol. Direct* 7:7. doi: 10.1186/1745-6150-7-7
- Marinsek, N., Barry, E. R., Makarova, K. S., Dionne, I., Koonin, E. V., and Bell, S. D. (2006). GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep.* 7, 539–545. doi: 10.1038/sj.embo.7400649
- Morgunova, E., Gray, F. C., MacNeill, S. A., and Ladenstein, R. (2009). Structural insights into the adaptation of proliferating cell nuclear antigen (PCNA) from *Halofexax volcanii* to a high-salt environment. *Acta Crystallogr. D Biol. Crystallogr.* 65, 1081–1088. doi: 10.1107/S0907444909029321
- Norais, C., Hawkins, M., Hartman, A. L., Eisen, J. A., Myllykallio, H., and Allers, T. (2007). Genetic and physical mapping of DNA replication origins in *Halofexax volcanii*. *PLoS Genet.* 3:e77. doi: 10.1371/journal.pgen.0030077
- Onesti, S., and MacNeill, S. A. (2013). Structure and evolutionary origins of the CMG complex. *Chromosoma* 122, 47–53. doi: 10.1007/s00412-013-0397-x
- Oyama, T., Ishino, S., Fujino, S., Ogino, H., Shirai, T., Mayanagi, K., et al. (2011). Architectures of archaeal GINS complexes, essential DNA replication initiation factors. *BMC Biol.* 9:28. doi: 10.1186/1741-7007-9-28
- Pan, M., Santangelo, T. J., Li, Z., Reeve, J. N., and Kelman, Z. (2011). *Thermococcus kodakarensis* encodes three MCM homologs but only one is essential. *Nucleic Acids Res.* 39, 9671–9680. doi: 10.1093/nar/gkr624
- Perler, F. B. (2002). InBase: the intein database. *Nucleic Acids Res.* 30, 383–384. doi: 10.1093/nar/30.1.383
- Perriere, G., and Gouy, M. (1996). WWW-query: an on-line retrieval system for biological sequence banks. *Biochimie* 78, 364–369. doi: 10.1016/0300-9084(96)84768-7
- Poidevin, L., and MacNeill, S. A. (2006). Biochemical characterisation of LigN, an NAD<sup>+</sup>-dependent DNA ligase from the halophilic euryarchaeon *Halofexax volcanii* that displays maximal *in vitro* activity at high salt concentrations. *BMC Mol. Biol.* 7:44. doi: 10.1186/1471-2199-7-44
- Poplawski, A., Grabowski, B., Long, S. E., and Kelman, Z. (2001). The zinc finger domain of the archaeal minichromosome maintenance protein is required for helicase activity. *J. Biol. Chem.* 276, 49371–49377. doi: 10.1074/jbc.M108519200
- Remus, D., Beuron, F., Tolun, G., Griffith, J. D., Morris, E. P., and Diffley, J. F. (2009). Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* 139, 719–730. doi: 10.1016/j.cell.2009.10.015
- Sakakibara, N., Kasiviswanathan, R., Melamud, E., Han, M., Schwarz, F. P., and Kelman, Z. (2008). Coupling of DNA binding and helicase activity is mediated by a conserved loop in the MCM protein. *Nucleic Acids Res.* 36, 1309–1320. doi: 10.1093/nar/gkm1160
- Sanchez-Pulido, L., and Ponting, C. P. (2011). Cdc45: the missing RecJ ortholog in eukaryotes? *Bioinformatics* 27, 1885–1888. doi: 10.1093/bioinformatics/btr332
- Skowyra, A., and MacNeill, S. A. (2012). Identification of essential and non-essential single-stranded DNA-binding proteins in a model archaeal organism. *Nucleic Acids Res.* 40, 1077–1090. doi: 10.1093/nar/gkr838

- Slaymaker, I. M., and Chen, X. S. (2012). MCM structure and mechanics: what we have learned from archaeal MCM. *Subcell. Biochem.* 62, 89–111. doi: 10.1007/978-94-007-4572-8\_6
- Slaymaker, I. M., Fu, Y., Toso, D. B., Ranatunga, N., Brewster, A., Forsburg, S. L., et al. (2013). Mini-chromosome maintenance complexes form a filament to remodel DNA structure and topology. *Nucleic Acids Res.* 41, 3446–3456. doi: 10.1093/nar/gkt022
- Stroud, A., Liddell, S., and Allers, T. (2012). Genetic and biochemical identification of a novel single-stranded DNA-binding complex in *Haloferax volcanii*. *Front. Microbiol.* 3:224. doi: 10.3389/fmicb.2012.00224
- Tomasz, M. (1995). Mitomycin C: small, fast and deadly (but very selective). *Chem. Biol.* 2, 575–579. doi: 10.1016/1074-5521(95)90120-5
- Vijayraghavan, S., and Schwacha, A. (2012). The eukaryotic Mcm2-7 replicative helicase. *Subcell. Biochem.* 62, 113–134. doi: 10.1007/978-94-007-4572-8\_7
- Walters, A. D., and Chong, J. P. (2010). An archaeal order with multiple minichromosome maintenance genes. *Microbiology* 156, 1405–1414. doi: 10.1099/mic.0.036707-0
- Wendoloski, D., Ferrer, C., and Dyall-Smith, M. L. (2001). A new simvastatin (mevinolin)-resistance marker from *Haloarcula hispanica* and a new *Haloferax volcanii* strain cured of plasmid pHV2. *Microbiology* 147, 959–964.
- Wiedemann, C., Ohlenschläger, O., Medagli, B., Onesti, S., and Görlich, M. (2013).  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  chemical shift assignments for the winged helix domains of two archaeal MCM C-termini. *Biomol. NMR Assign.* doi: 10.1007/s12104-013-9516-0. [Epub ahead of print].
- Winter, J. A., Christofi, P., Morroll, S., and Bunting, K. A. (2009). The crystal structure of *Haloferax volcanii* proliferating cell nuclear antigen reveals unique surface charge characteristics due to halophilic adaptation. *BMC Struct. Biol.* 9:55. doi: 10.1186/1472-6807-9-55
- Yuan, H., Liu, X.-P., Han, Z., Allers, T., Hou, J.-L., and Liu, J.-H. (2013). RecJ-like protein from *Pyrococcus furiosus* has 3'-5' exonuclease activity on RNA: implications for proofreading of 3'-mismatched RNA primers in DNA replication. *Nucleic Acid Res.* 41, 5817–5826. doi: 10.1093/nar/gkt275
- Zhao, A., Gray, F. C., and MacNeill, S. A. (2006). ATP- and NAD $^+$ -dependent DNA ligases share an essential function in the halophilic archaeon *Haloferax volcanii*. *Mol. Microbiol.* 59, 743–752. doi: 10.1111/j.1365-2958.2005.04975.x

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Received: 09 February 2014; accepted: 10 March 2014; published online: 26 March 2014.*

*Citation: Kristensen TP, Maria Cherian R, Gray FC and MacNeill SA (2014) The haloarchaeal MCM proteins: bioinformatic analysis and targeted mutagenesis of the  $\beta_7$ - $\beta_8$  and  $\beta_9$ - $\beta_{10}$  hairpin loops and conserved zinc binding domain cysteines. *Front. Microbiol.* 5:123. doi: 10.3389/fmicb.2014.00123*

*This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.*

*Copyright © 2014 Kristensen, Maria Cherian, Gray and MacNeill. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# Identification of carotenoids from the extremely halophilic archaeon *Haloarcula japonica*

Rie Yatsunami<sup>1\*</sup>, Ai Ando<sup>1</sup>, Ying Yang<sup>1</sup>, Shinichi Takaichi<sup>2</sup>, Masahiro Kohno<sup>1</sup>, Yuriko Matsumura<sup>1</sup>, Hiroshi Ikeda<sup>1</sup>, Toshiaki Fukui<sup>1</sup>, Kaoru Nakasone<sup>3</sup>, Nobuyuki Fujita<sup>4</sup>, Mitsuo Sekine<sup>4</sup>, Tomonori Takashina<sup>5</sup> and Satoshi Nakamura<sup>1</sup>

<sup>1</sup> Department of Bioengineering, Tokyo Institute of Technology, Yokohama, Japan

<sup>2</sup> Department of Biology, Nippon Medical School, Kawasaki, Japan

<sup>3</sup> Department of Biotechnology and Chemistry, Kinki University, Hiroshima, Japan

<sup>4</sup> Biotechnology Field, National Institute of Technology and Evaluation, Tokyo, Japan

<sup>5</sup> Department of Applied Bioscience, Toyo University, Gunma, Japan

## Edited by:

R. Thane Papke, University of Connecticut, USA

## Reviewed by:

Jonathan L. Klassen, University of Wisconsin-Madison, USA

Bonnie K. Baxter, Westminster College, USA

## \*Correspondence:

Rie Yatsunami, Department of Bioengineering, Tokyo Institute of Technology, 4259-J2-13,

Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan  
e-mail: yatsunami.r.aa@m.titech.ac.jp

The carotenoids produced by extremely halophilic archaeon *Haloarcula japonica* were extracted and identified by their chemical, chromatographic, and spectroscopic characteristics (UV-Vis and mass spectrometry). The composition (mol%) was 68.1% bacterioruberin, 22.5% monoanhydrobacterioruberin, 9.3% bisanhydrobacterioruberin, <0.1% isopentenyldehydrorhodopin, and trace amounts of lycopene and phytoene. The *in vitro* scavenging capacity of a carotenoid, bacterioruberin, extracted from *Haloarcula japonica* cells against 1,1-diphenyl-2-picrylhydrazyl (DPPH) radicals was evaluated. The antioxidant capacity of bacterioruberin was much higher than that of β-carotene.

**Keywords:** extremely halophilic archaeon, *Haloarcula japonica*, C<sub>50</sub> carotenoid, bacterioruberin, antioxidant capacity

## INTRODUCTION

Carotenoids are yellow to red pigments, which originate from the terpenoid biosynthetic pathway. They are synthesized by plants, algae, some fungi, bacteria, and archaea. They are involved in photosynthesis as accessory pigments, and function as antioxidants, light protection pigments, and membrane stabilizers. Their antioxidant properties are closely related to their chemical structure, including aspects such as the number of conjugated double bonds (CDB), the type of structural end-group, and oxygen-containing substituents (Albrecht et al., 2000). Carotenoids are efficient scavengers of reactive nitrogen species, reactive oxygen species (ROS), especially singlet oxygen species, and non-biological radicals (Burton, 1989; Di Mascio et al., 1989; Miller et al., 1996; Chisté et al., 2011).

Extremely halophilic archaea generating red-colored colonies produce phytoene, lycopene, β-carotene, acyclic C<sub>50</sub> bacterioruberin (BR), and its precursors, such as isopentenyldehydrorhodopin (IDR), bisanhydrobacterioruberin (BABR), and monoanhydrobacterioruberin (MABR) (Goodwin, 1980). A halophilic archaeon, *Halobacterium salinarum*, grows chemoorganotrophically in the dark. In the light, they can utilize light energy, even though they still depend on organic nutrients as a carbon source. The molecule responsible for their light utilization is bacteriorhodopsin, which functions as a proton pump to generate ATP through *cis-trans* isomerization of the chromophore retinal, an end product of carotenoid biosynthesis. In the early steps of the carotenoid and retinal biosynthetic pathways, two geranylgeranyl pyrophosphate (GGPP) molecules are condensed to form a C<sub>40</sub> carotenoid, phytoene, which

undergoes a series of desaturation reactions to form the red carotenoid lycopene (Kushwaha et al., 1976). For retinal synthesis, lycopene is cyclized to β-carotene, and then cleaved to a C<sub>20</sub> retinal cofactor (Peck et al., 2002). Alternatively, lycopene may be used as a precursor for BR, which is a C<sub>50</sub>-xanthophyll functioning to increase membrane rigidity and provide protection against UV light (Lazrak et al., 1998; Shahmohammadi et al., 1998).

*Haloarcula japonica*, the extremely halophilic archaeon, has flat red cells that are predominantly triangular in shape (Takashina et al., 1990; Otozai et al., 1991), suggesting this organism might produce carotenoids. This organism, which requires 2.6–4.3 M NaCl for growth, has a large amount of glycoprotein (CSG) on its cell surface (Nakamura et al., 1992; Nishiyama et al., 1992; Horikoshi et al., 1993). By using flash-induced fluorescence spectroscopic analysis, a bacteriorhodopsin-like retinal protein was identified on the cell envelope vesicles of *Haloarcula japonica* (Yatsunami et al., 1997). These results suggest that *Haloarcula japonica* has both carotenoids and retinal biosynthetic pathways. Recently, the draft genome sequence of *Haloarcula japonica* has been determined (Nakamura et al., 2011). However, the carotenoid composition and both carotenoids and retinal biosynthetic pathways of *Haloarcula japonica* have not been identified yet.

Here, we present the carotenoid composition of *Haloarcula japonica* and evaluate the antioxidant potential of an extracted carotenoid using the 1,1-diphenyl-2-picrylhydrazyl (DPPH) method and compare its activity with that of β-carotene.

## MATERIALS AND METHODS

### STRAIN AND CULTIVATION CONDITIONS

Extremely halophilic archaeon *Haloarcula japonica* strain TR-1 (JCM 7785<sup>T</sup>) was pre-cultured at 37°C in the dark with a complex medium as described previously (Das Sarma and Fleischmann, 1995). 4 ml of pre-inoculum was transferred to a 2 L Erlenmeyer flask containing 400 mL of the liquid medium and cultured to a stationary phase for 10 days under the same conditions. The cells were harvested by centrifugation at 4,400 × g for 20 min, washed with a basal salt solution [20% (w/v) NaCl and 4% (w/v) MgSO<sub>4</sub>·7H<sub>2</sub>O], and stored at -80°C until used.

### EXTRACTION OF CAROTENOIDS

The extraction of the carotenoids was performed under dim light as follows. A frozen cell pellet was thawed, and 10 times volume of acetone/methanol (7:3, v/v) was added. The suspension was sonicated with a sonic oscillator (VP-5S, Taitec, Koshigaya, Japan) for several seconds and centrifuged. The supernatant was collected and evaporated. The carotenoids were dissolved in a small volume of *n*-hexane/acetone and loaded onto a DEAE-Toyopearl 650 M (Tosoh, Tokyo, Japan) column to remove the polar lipids (Takaichi and Ishidzu, 1992). The non-adsorption fraction including the carotenoids was recovered and evaporated. The carotenoids were also dissolved in a small volume of *n*-hexane/acetone and loaded on a column of silica gel 60 (Merck, Darmstadt, Germany). Separation was achieved by binary graduation elution using an initial composition of 90% *n*-hexane and 10% acetone, which was decreased stepwise to 50% *n*-hexane and 50% acetone. All fractions were recovered, evaporated, and further analyzed by HPLC with a μBondapack C<sub>18</sub> column (8 × 100 mm, RCM type, Waters, Milford, MA, USA), as described previously (Takaichi and Ishidzu, 1992). The elution was performed with 100% methanol at 1.8 ml min<sup>-1</sup>. The absorption spectra were recorded with a photodiode-array detector (250–580 nm, 1.3-nm intervals, MCPD-3600, Otsuka Electronics, Osaka, Japan) attached to the High-performance liquid chromatography (HPLC) apparatus as described previously (Takaichi and Shimada, 1992). The peaks of lycopene and phytoene were collected again, further separated by HPLC with Novapack C<sub>18</sub> column (8 × 100 mm, RCM type, Waters) and eluted with a mixture of acetonitrile, methanol, and tetrahydrofuran (58:35:7, v/v) (2.0 ml min<sup>-1</sup>) as described previously (Takaichi, 2000). The lycopene and phytoene were identified by a combination of the HPLC retention times and the absorption spectra. Other carotenoids, including IDR, BABR, MABR, and BR were detected with absorbance at 490 nm. To identify the each elution peak, the relative molecular masses of the purified carotenoids were measured by field-desorption mass spectrometry using a double-focusing gas chromatograph/mass spectrometer equipped with a field-desorption apparatus (M-2500, Hitachi, Tokyo, Japan) according to the method of Takaichi (1993). The 500 MHz <sup>1</sup>H NMR spectrum of the BR was recorded in CDCl<sub>3</sub> at 25°C on a Varian VXR-500S spectrometer (Varian Medical Systems, Palo Alto, CA, USA).

### QUANTIFICATION OF CAROTENOIDS

For the quantification of carotenoids, ethyl β-apo-8'-carotenoate (Wako Pure Chemical, Osaka, Japan) was used as an internal

standard. 10 μl of 0.5 mM ethyl β-apo-8'-carotenoate in ethanol was added to the samples upon extraction. The suspension was disrupted by sonication with a Ultrasonic disruptor UD-201 (Tomy Seiko, Tokyo, Japan) for several seconds and centrifuged. The supernatant was collected and evaporated. The carotenoids were dissolved in a small volume of *n*-hexane/acetone, analyzed by HPLC with a μBondapack C<sub>18</sub> column (3.9 × 300 mm, 125 Å, 10 μm, Waters). The HPLC system consisted of a SCL-10A chromatograph fitted with a photodiode-array detector (SPD-M20A, Shimadzu, Kyoto, Japan) and controlled with an LC solution (Shimadzu). The carotenoids were eluted with methanol/water (9:1) for the first 10 min and then with 100% methanol (1.5 ml min<sup>-1</sup>). Detection was performed at 490 nm, and the online spectra were acquired in the 190–800 nm wavelength range with 1.2 nm resolution. Each carotenoid was identified by the retention time on the HPLC and the absorption spectrum in the eluent by a photodiode-array detector. The absorption coefficients of BR and its derivatives at 490 nm were assumed to be 167 mM<sup>-1</sup> cm<sup>-1</sup> (Kelly et al., 1970). That of the ethyl β-apo-8'-carotenoate at 445 nm was assumed to be 100 mM<sup>-1</sup> cm<sup>-1</sup>, which is the same as that of the β-apo-8'-carotenoic acid (Isler et al., 1959).

### DPPH RADICAL SCAVENGING ASSAY

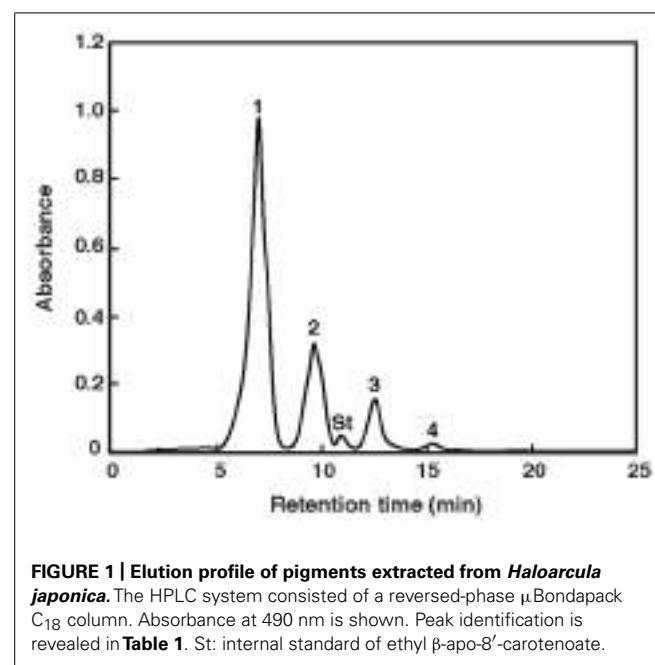
DPPH radical scavenging activity was measured using an ESR spectrometer (JES-FA-100, JEOL, Tokyo, Japan). The stable free radical, DPPH was dissolved in acetone (200 μM). β-Carotene, a standard antioxidant, was used as a positive control. The BR and β-carotene were diluted in 100 μl of acetone, yielding concentrations of 0–200 and 0–800 μM, respectively. The 100 μl DPPH solution and the 100 μl carotenoid solution were mixed, and the DPPH radical was measured after 60 s. The spin adduct was detected by ESR spectrometer exactly 2 min later. The ESR measurement conditions were as follows: field sweep, 330.500–340.500 mT; field modulation frequency, 100 kHz; field modulation width, 0.25 mT; sweep time, 2 min; time constant, 0.1 s; microwave frequency, 9.427 GHz; and microwave power, 4 mW. All the scavenging activities in the present study were calculated using the following equation, in which H and H<sub>0</sub> were the peak areas of the radical signals with and without a sample, respectively:

$$\text{Radical scavenging activity (\%)} = [1 - (H/H_0)] \times 100.$$

## RESULTS

### CHARACTERIZATION OF CAROTENOID PROFILES OF *Haloarcula japonica*

The carotenoids occurring in *Haloarcula japonica* were extracted and identified based on their chemical, chromatographic, and spectroscopic characteristics (UV-Vis and mass spectrometry). **Figure 1** shows an elution profile on HPLC system that corresponds to the carotenoids obtained from *Haloarcula japonica*. **Table 1** summarizes the identification for each chromatographic peak, and **Figure 2** shows the corresponding chemical structures. Peak 1, which was the major carotenoid, was assigned as all-trans-BR. The UV-Vis spectrum, mass spectrum, CD spectrum (data not shown), and NMR spectra (data not shown) were compatible with those of BR from *Haloferax volcanii*.



**Table 1 | Characteristics of carotenoids produced by *Haloarcula japonica*.**

Peak <sup>a</sup>	Carotenoid	Retention time (min)	$\lambda_{\text{max}}$ (nm) in methanol	[M] <sup>+</sup> (m/z)
1	BR	7.0	387, 466, 488, 525	741.0
2	MABR	9.6	369, 385, 465, 492, 524	722.8
3	BABR	12.5	370, 386, 460, 492, 524	704.5
4	IDR	15.3	375, 455, 480, 511	620.7
-	Lycopene	19.4	296, 363, 445, 472, 501 <sup>b</sup>	not done
-	Phytoene	19.4	277, 288, 298 <sup>b</sup>	not done

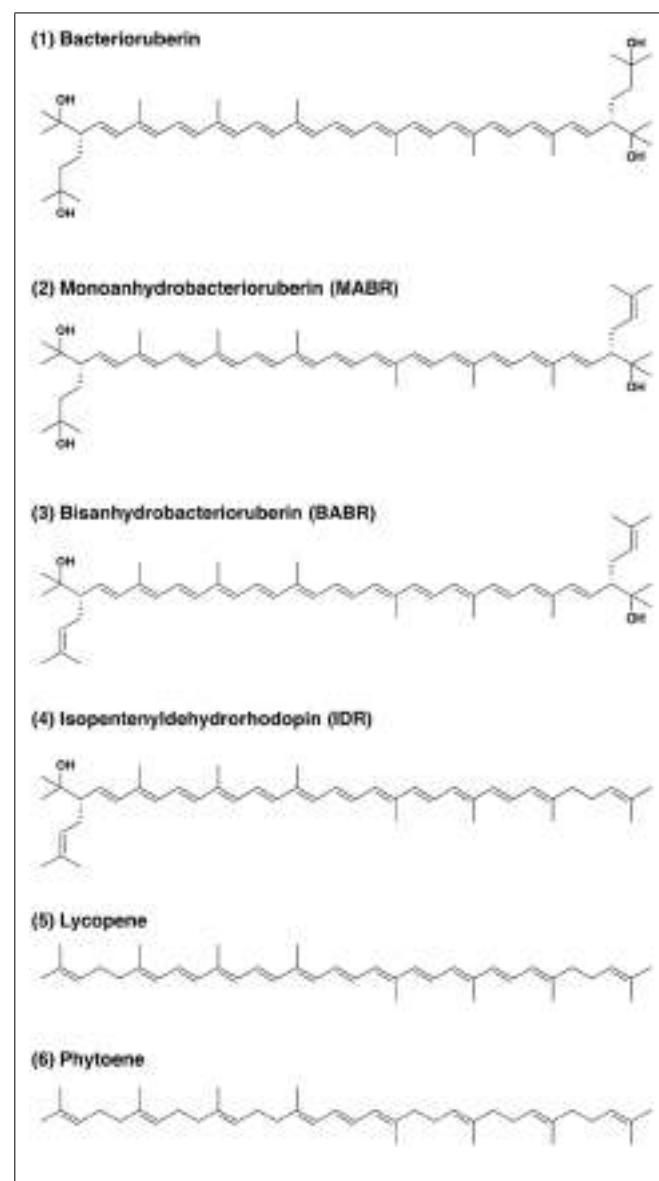
<sup>a</sup> Peak numbers are based on Figure 1 using μBondapack C<sub>18</sub> column.

<sup>b</sup>  $\lambda_{\text{max}}$  in acetonitrile/MeOH/THF (58:35:7) using Novapack C<sub>18</sub> column.

(Rønneklev and Liaaen-Jensen, 1995). Peaks 2 and 3 were similarly identified as MABR and BABR, respectively. Peak 4 was a minor C<sub>45</sub>-carotenoid, IDR. In addition to these carotenoids, two minor C<sub>40</sub> carotenoids were also eluted at 19.4 min. They were identified as phytoene and lycopene using another HPLC system (date not shown). These carotenoids were also found in *Halobacterium salinarum* (Kelly et al., 1970).

#### QUANTITATIVE ANALYSIS OF CAROTENOID

The total carotenoid content was 335  $\mu\text{g g}^{-1}$  of dry mass, although the contents in *Halobacterium salinarum* and *Halococcus morrhuae* were 89 and 45  $\mu\text{g g}^{-1}$ , respectively (Mandelli et al., 2012). These contents were about four and seven times compared to that of *Halobacterium salinarum* and *Halococcus morrhuae*, respectively. BR was the major pigment, accounting for up to 68.1% of the total carotenoids (mol%). Therefore, it was the main one responsible for the red color of this organism. The BR content in *Haloarcula*



**FIGURE 2 | Structure of carotenoids produced by *Haloarcula japonica*.**

*japonica* was similar to those in other halophilic archaea (Mandelli et al., 2012). Other major pigments were MABR (22.5%) and BABR (9.3%), and IDR was found at a lower level (<0.1%). These results suggest that BR is produced as a final product in *Haloarcula japonica* and is synthesized from other C<sub>50</sub> carotenoids, such as IDR, BABR, and MABR as well as other halophilic archaea.

#### ANTIOXIDANT CAPACITY

An ROS formed under photo-oxidation stress can react with macromolecules like lipids and proteins and cause a cellular damage. Antioxidants are substances that have the ability to reduce ROS and prevent macromolecules from oxidation (Klein et al., 2012). DPPH method was carried out to evaluate the antioxidant capacity of the carotenoid extracted. ESR spin trapping provides a sensitive, direct, and accurate means of monitoring reactive

species (Guo et al., 1999). DPPH is a stable free radical donor, which is widely used to test the free radical scavenging effect of natural antioxidants. DPPH method involves the scavenging of a performed stable radical by an electron transfer mechanism from the carotenoid to the radical, generating a carotenoid radical cation (Huang et al., 2005). The scavenging capacity of the reactive species was dependent on the carotenoid concentration. **Figure 3** shows that the scavenging capacity of BR was much higher than that of  $\beta$ -carotene. The oxidant capacity seems to relate to both the length of the CDB and the presence of the function group. The antioxidant capacity of the carotenoids increases with increased extension and maximum overlap of the CDB molecular orbitals (Albrecht et al., 2000; Tian et al., 2007). BR molecule contains 13 CDB, which are much more than the nine CDB of the  $\beta$ -carotene. Therefore, the BR would be a good DPPH radical scavenger.

## DISCUSSION

In the present study, we extracted the carotenoid pigments occurring in *Haloarcula japonica* and identified them. This is the first report concerned with the carotenoids produced by the extremely halophilic archaeon *Haloarcula japonica*. The production of carotenoids from *Haloarcula japonica* is very attractive. Previous data suggested that the C<sub>50</sub>-carotenoids found in halophilic archaea may be incorporated in membranes due to their length, and that their two polar end-groups may facilitate the adjustment to such membranes; moreover, it has been reported that bacterioruberin reinforces the lipid membrane of *Halobacterium* spp. (Ourisson and Nakatani, 1989).

In addition, the carotenoids synthesized by these microorganisms have a function to protect their cells against the lethal actions of ionizing radiation, UV radiation, and hydrogen peroxide (Shah-mohammadi et al., 1998). Saito et al. (1997) have extracted BR

from *Rubrobacter radiotolerans*. They studied the OH scavenging effect using a system of thymine degradation and compared with that of  $\beta$ -carotene. These results have shown that the OH radical scavenging ability of BR was much higher than that of  $\beta$ -carotene. In this work, the scavenging capacity of BR extracted from *Haloarcula japonica* toward DPPH free radicals was measured. BR exhibited higher DPPH free radical scavenging than  $\beta$ -carotene. The present result was consistent with the previous study.

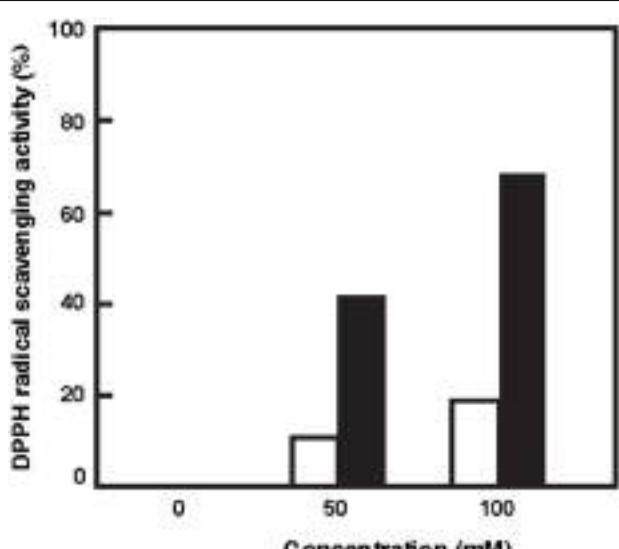
Since the carotenoids produced by halophilic archaea can play both roles of membrane stabilization and protection against oxidizing agents, these compounds are essential for the survival of such microorganisms. In order to clarify the function of BR *in vivo*, further studies using of BR-deficient *Haloarcula japonica* mutant were needed.

## ACKNOWLEDGMENTS

This study was partially supported by Grant-in-Aid for Scientific Research (C) from Japan Society for the Promotion of Science and Takahashi Industrial and Economic Research Foundation to Rie Yatsunami.

## REFERENCES

- Albrecht, M., Takaichi, S., Steiger, S., Wang, Z., and Sandmann, G. (2000). Novel hydroxycarotenoids with improved antioxidative properties produced by gene combination in *Escherichia coli*. *Nat. Biotechnol.* 18, 843–846. doi: 10.1038/78443
- Burton, G. W. (1989). Antioxidant action of carotenoids. *J. Nutr.* 119, 109–111.
- Chisté, R. C., Mercadante, A. Z., Gomes, A., Fernandes, E., Lima, S. F. D., and Bragagnolo, N. (2011). In vitro scavenging capacity of annatto seed extracts against reactive oxygen and nitrogen species. *Food Chem.* 127, 419–426. doi: 10.1016/j.foodchem.2010.12.139
- Das Sarma, S., and Fleischmann, E. M. (1995). *Archaea, a Laboratory Manual*. Cold Spring Harbor: Cold Springer Harbor Laboratory Press.
- Di Mascio, P., Kaiser, S., and Sies, H. (1989). Lycopene as the most efficient biological carotenoid singlet oxygen quencher. *Arch. Biochem. Biophys.* 274, 532–538. doi: 10.1016/0003-9861(89)90467-0
- Goodwin, T. W. (1980). “Non-photosynthetic bacteria,” in *The Biochemistry of the Carotenoids*, Vol.1, ed. T. W. Goodwin (London and New York: Chapman and Hall), 290–317.
- Guo, Q., Zhao, B., Shen, S., Hou, J., and Xin, W. (1999). ESR study on the structure-antioxidant activity relationship of tea catechins and their epimers. *Biochim. Biophys. Acta.* 1427, 13–23. doi: 10.1016/S0304-4165(98)00168-8
- Horikoshi, K., Aono, R., and Nakamura, S. (1993). The triangular halophilic archaeabacterium *Haloarcula japonica* strain TR-1. *Experientia* 49, 497–502. doi: 10.1007/BF01955151
- Huang, D., Ou, B., and Prior, R. L. (2005). The chemistry behind antioxidant capacity assays. *J. Agric. Food Chem.* 53, 1841–1856. doi: 10.1021/jf030723c
- Isler, O., Guex, W., Rüegg, R. G., Saucy, G., Schwietter, U., Walter, M., et al. (1959). Synthesen in der carotenoid-reihle 16. Mittellung carotenoide vom typus des torularhodins. *Helv. Chim. Acta.* 42, 864–871. doi: 10.1002/hlca.19590420329
- Kelly, M., Norgård, S., and Liaasen-Jensen, S. (1970). Bacterial carotenoids. XXXI. C<sub>50</sub>-carotenoids 5. Carotenoids of *Halobacterium salinarum*, especially, bacterioruberin. *Acta Chem. Scand.* 24, 2169–2182. doi: 10.3891/acta.chem.scand.24-2169
- Klein, B. C., Walter, C., Lange, H. A., and Buchholz, R. (2012). Macroalgae as natural sources for antioxidative compounds. *J. Appl. Phycol.* 24, 1133–1139. doi: 10.1007/s10811-011-9743-7
- Kushwaha, S. C., Kates, M., and Porter, J. W. (1976). Enzymatic synthesis of C<sub>40</sub> carotenoids by cell-free preparation from *Halobacterium cutirubrum*. *Can. J. Biochem.* 54, 816–823. doi: 10.1139/o76-117
- Lazrak, T., Wolff, G., Albrecht, A.-M., Nakatani, Y., Ourisson, G., and Kates, M. (1998). Bacterioruberins reinforce reconstituted *Halobacterium* lipid membranes. *Biochim. Biophys. Acta.* 939, 160–162. doi: 10.1016/0005-2736(88)90057-0



**FIGURE 3 | Radical scavenging activities of BR and  $\beta$ -carotene against DPPH radicals.** White bars,  $\beta$ -carotene; Black bars, BR.

- Mandelli, F., Miranda, S. V., Rodrigus, E., and Marcadante, A. A. (2012). Identification of carotenoids with high antioxidant capacity produced by extremophile microorganisms. *World J. Microbiol. Biotechnol.* 28, 1781–1790. doi: 10.1007/s11274-011-0993-y
- Miller, N. J., Sampson, J., Candeias, L. P., Bramley, P. M., and Rice-Evans, C. A. (1996). Antioxidant activities of carotenoids and xanthophylls. *FEBS Lett.* 384, 240–242. doi: 10.1016/0014-5793(96)00323-7
- Nakamura, S., Aono, R., Muzutani, S., Takashina, T., Grant, W. D., and Horikoshi, K. (1992). The cell surface glycoprotein of *Haloarcula japonica* TR-1. *Biosci. Biotechnol. Biochem.* 56, 996–998. doi: 10.1271/bbb.56.996
- Nakamura, S., Nakasone, K., and Takashina, T. (2011). “Genetics and genomics of triangular disc-shaped halophilic archaeon *Haloarcula japonica* TR-1,” in *Extremophiles Handbook*, eds K. Horikoshi, G. Antranikian, A. T. Bull, F. T. Robb, and K. O. Stetter (Heidelberg: Springer), 364–381.
- Nishiyama, Y., Takashina, T., Grant, W. D., and Horikoshi, K. (1992). Ultrastructure of cell wall of the triangular halophilic archaeabacterium *Haloarcula japonica* strain TR-1. *FEMS Microbiol. Lett.* 99, 43–48. doi: 10.1111/j.1574-6968.1992.tb05539.x
- Otozai, K., Takashina, T., and Grant, W. D. (1991). “A novel triangular archaeabacterium, *Haloarcula japonica*,” in *Superbugs: Microorganisms in Extreme Environments*, eds K. Horikoshi and W. D. Grant (Tokyo: Scientific Societies Press), 63–75.
- Ourisson, G., and Nakatani, Y. (1989). “Bacterial carotenoids as membrane reinforcers. A general role of polyterpenoids: membrane stabilization.” in *Carotenoids, Chemistry and Biology*, eds N. I. Krinsky, M. N. Mathews-Roth, and R. T. Taylor (New York: Plenum), 237–246.
- Peck, R. F., Johnson, E. A., and Krebs, M. P. (2002). Identification of a lycopene  $\beta$ -cyclase required for bacteriorhodopsin biogenesis in the archaeon *Halobacterium salinarum*. *J. Bacteriol.* 184, 2889–2897. doi: 10.1128/JB.184.11.2889-2897.2002
- Rønneklev, M., and Liaaen-Jensen, S. (1995). Bacterial carotenoids 53, C50-carotenoids 23; carotenoids of *Haloflexax volcanii* versus other halophilic bacteria. *Biochem. Syst. Ecol.* 23, 627–634. doi: 10.1016/0305-1978(95)00047-X
- Saito, T., Miyabe, Y., Ide, H., and Yamamoto, O. (1997). Hydroxyl radical scavenging ability of bacterioruberin. *Radiat. Phys. Chem.* 50, 267–269. doi: 10.1016/S0969-806X(97)00036-4
- Shahmohammadi, H. R., Asgarani, E., Terato, H., Saito, T., Ohyama, Y., Gekko, K., et al. (1998). Protective roles of bacterioruberin and intracellular KCl in the resistance of *Halobacterium salinarum* against DNA-damaging agents. *J. Radiat. Res.* 39, 251–262. doi: 10.1269/jrr.39.251
- Takaichi, S. (1993). Usefulness of field desorption mass spectrometry in determining molecular masses of carotenoids, natural carotenoid derivatives and their chemical derivatives. *Org. Mass Spectrom.* 28, 785–788. doi: 10.1002/oms.1210280711
- Takaichi, S. (2000). Characterization of carotenoid in a combination of a C18 HPLC column with isocratic elution and absorption spectra with a photodiode-array detector. *Photosynth. Res.* 65, 93–99. doi: 10.1023/A:1006445503030
- Takaichi, S., and Ishidzu, J. (1992). Carotenoid glycoside ester from *Rhodococcus rhodochros*. *Methods Enzymol.* 213, 366–374. doi: 10.1016/0076-6879(92)13138-N
- Takaichi, S., and Shimada, K. (1992). Characterization of carotenoids in photosynthetic bacteria. *Methods Enzymol.* 213, 374–385. doi: 10.1016/0076-6879(92)13139-O
- Takashina, T., Hamamoto, T., Otozai, K., Grant, W. D., and Horikoshi, K. (1990). *Haloarcula japonica* sp. nov., a new triangular halophilic archaeabacterium. *Syst. Appl. Microbiol.* 13, 177–181. doi: 10.1016/S0723-2020(11)80165-7
- Tian, B., Xu, Z., Sun, Z., Lin, J., and Hua, Y. (2007). Evaluation of the antioxidant effects of carotenoids from *Deinococcus radiodurans* through targeted mutagenesis, chemiluminescence, and DNA damage analysis. *Biochim. Biophys. Acta* 1770, 902–911. doi: 10.1016/j.bbagen.2007.01.016
- Yatsunami, R., Kawakami, T., Ohtani, H., and Nakamura, S. (1997). Primary structure of the novel bacterial rhodopsin from extremely halophilic archaeon *Haloarcula japonica* strain TR-1. *Nucleic Acids Symp. Ser.* 37, 111–112.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 26 December 2013; accepted: 25 February 2014; published online: 17 March 2014.

Citation: Yatsunami R, Ando A, Yang Y, Takaichi S, Kohno M, Matsumura Y, Ikeda H, Fukui T, Nakasone K, Fujita N, Sekine M, Takashina T and Nakamura S (2014) Identification of carotenoids from the extremely halophilic archaeon *Haloarcula japonica*. *Front. Microbiol.* 5:100. doi: 10.3389/fmicb.2014.00100

This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.

Copyright © 2014 Yatsunami, Ando, Yang, Takaichi, Kohno, Matsumura, Ikeda, Fukui, Nakasone, Fujita, Sekine, Takashina and Nakamura. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Prebiotic protein design supports a halophile origin of foldable proteins

Liam M. Longo and Michael Blaber\*

Department of Biomedical Sciences, College of Medicine, Florida State University, Tallahassee, FL, USA

\*Correspondence: michael.blaber@med.fsu.edu

**Edited by:**

R. Thane Papke, University of Connecticut, USA

**Reviewed by:**

Dominique Madern, Institut de Biologie Structurale, France

Johann P. Gogarten, University of Connecticut, USA

Greg Fournier, Massachusetts Institute of Technology, USA

**Keywords:** halophile, abiogenesis, proteogenesis, prebiotic, protein folding

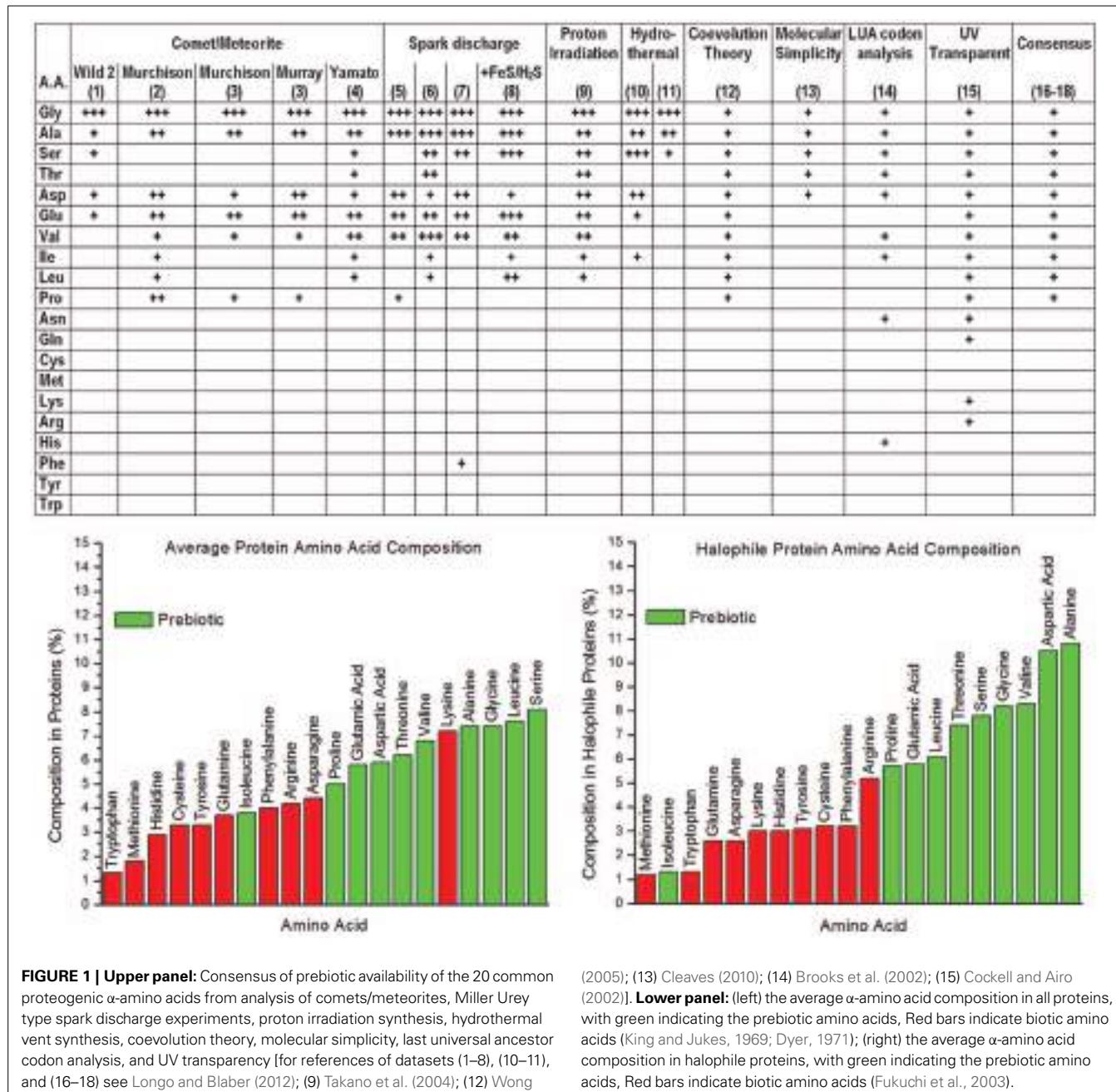
There are significant challenges in forming testable hypotheses regarding abiogenesis (i.e., the origin of life); for example, the original environment on the early Earth during the process of abiogenesis is a matter of debate [although it was significantly different from the current environment (Oparin, 1952; Hazen et al., 2008)]. Furthermore, the process of abiogenesis occurred over a time scale that is impractical to replicate as a laboratory experiment. More difficult still is the likelihood that current life forms are far removed from the earliest “living systems”—which may well have utilized entirely different initial energetic, biochemical, and “genetic” systems. Despite such difficulties, there are potentially testable hypotheses regarding the origin of *important classes* of biomolecules from abiotic processes. A key biomolecule that emerged in abiogenesis is the foldable polypeptide, which ultimately evolved to provide essentially all of the important biochemical and structural machinery in living systems. Each of the generally-acknowledged abiotic chemical processes present during abiogenesis, including atmospheric spark discharge, hydrothermal vent chemistry, as well as deep-space synthesis and delivery of organic material via comet and asteroid bombardment, can produce a subset of the 20 common  $\alpha$ -amino acids [for a summary see Longo and Blaber (2012)]. Such prebiotic amino acids would have provided the raw material for the earliest polypeptides (i.e., “proteogenesis”); thus, the properties of such amino acids and polypeptides are of special interest. As with all things related to abiogenesis,

the set of prebiotic amino acids available for proteogenesis has been a matter of debate; however, a compilation of broad and diverse analyses is arguably converging upon a consensus set of 10 prebiotic  $\alpha$ -amino acids (**Figure 1**).

The alphabet size and chemical properties of the prebiotic  $\alpha$ -amino acids are critical parameters as regards the capability to form foldable polypeptides. The rules of protein folding are not fully understood; however, some essential requirements of amino acids to promote folding are known. The tertiary structure of folded proteins is an assemblage of common secondary structure elements—including  $\alpha$ -helix,  $\beta$ -strand, and reverse-turns. Thus, a foldable set of amino acids includes representatives with a high propensity to form each of the common secondary structure elements. Additionally, soluble globular proteins typically fold so as to sequester hydrophobic side chains within the protein interior, and this forms a significant energetic contribution to the overall stability of the folded protein; thus, a foldable set of amino acids contains both hydrophobic and hydrophilic members. Finally, functional considerations require that among the amino acids is a representative that can act as a nucleophile, and thereby provide useful chemical activity to an otherwise benign structural scaffold. While folding requirements are demanding, it is clear that the extant set of 20 common  $\alpha$ -amino acids is redundant in this regard. Thus, the question of the *minimum*  $\alpha$ -amino acid alphabet necessary to enable protein foldability has also been investigated, with a proposed minimum

alphabet size of 9–10 amino acids (Romero et al., 1999; Murphy et al., 2000). Thus, with regard to set size, the prebiotic  $\alpha$ -amino acid alphabet is located on the very cusp of foldable potential—a precarious position indeed, as it would provide essentially no redundancy in the requirements for protein foldability. Viewed in these terms, the prebiotic set is *remarkable* in containing all necessary elements for foldability—including high-propensity amino acids for each of the basic types of protein secondary structure, hydrophobic and hydrophilic amino acids, as well as several nucleophilic amino acids [for a discussion of such properties see Longo and Blaber (2012)]. However, the characteristics of a protein comprised of the prebiotic set of  $\alpha$ -amino acids present a *stark deviation from the majority of extant proteins*, since aromatic residues, key contributors to the critical hydrophobic effect that drives protein folding, are absent in the prebiotic alphabet. Furthermore, there are no basic amino acids in the prebiotic set (McDonald and Storrie-Lombardi, 2010), thus restricting protein design to exclusively acidic polypeptides—limiting the presence of salt bridge interactions and resulting in acidic pI.

A number of successful studies of simplified protein design have been reported whereby foldable proteins have been constructed from a reduced  $\alpha$ -amino acid alphabet, and relevance for proteogenesis have been described. However, such studies have focused exclusively upon achieving minimization of the *alphabet size*, without regard to the *prebiotic relevance* of the included amino acid alphabet.



Thus, without exception, such minimal foldable proteins have depended upon critical aromatic amino acids within the core, as well as stabilizing salt bridges (dependent upon basic amino acids), to achieve a stable structure—no minimal protein design has utilized a plausible prebiotic alphabet. Thus, while minimal alphabets can yield foldable polypeptides, the foldable potential of the set of prebiotic amino acids has not been explored with the necessary rigor.

To determine the folding potential of the set of prebiotic  $\alpha$ -amino acids our lab evaluated the consequences of enriching for the prebiotic set in a designed  $\beta$ -trefoil protein. The  $\beta$ -trefoil is a common protein architecture that has been the subject of much study as regards its evolutionary emergence from a simple 42-mer peptide motif (Lee and Blaber, 2011; Lee et al., 2011). Two “primitive” versions of the  $\beta$ -trefoil protein were subsequently designed; in primitive version 1 (PV1) the

amino acid alphabet was reduced to 13 unique amino acids with an overall prebiotic composition of 74%; in primitive version 2 (PV2) the amino acid alphabet was reduced to 12 unique amino acids with an overall prebiotic composition of 79%. Notably, PV2 is devoid of any aromatic amino acids. The hydrophobic core of the  $\beta$ -trefoil architecture involves a substantial number of positions (21 total; or ~17% of total positions), and with PV2 this important region of the protein is

entirely prebiotic and comprised of only 3 different amino acids (Leu, Ile, and Val). In addition to reduced core hydrophobicity the PV1 and PV2 proteins exhibit a substantial increase in acidic property (due to the exclusively acidic nature of the prebiotic set of amino acids). A high negative surface charge density is a characteristic feature of halophilic proteins, enabling them to remain soluble in high salt via carboxylate binding of solvated metal cations. Additionally, halophile proteins are characterized as having reduced hydrophobicity, and denaturation under low salt conditions. This property of prebiotic amino acid composition and compatibility with a high salt environment is understood principally in terms of the biophysics of protein solubility in high salt enabled by surface acidic charge that binds hydrated salt cation and the effect of salt in stabilizing the hydrophobic effect. In low salt buffer PV2 is only fractionally folded even at its temperature of maximum stability. However, high salt stabilizes the PV2 protein, shifting its melting temperature into the region of high-mesophile/low-thermophile stability and exhibiting >99% fractional native state; thus, by the criteria of efficient foldability PV2 is an *obligate halophilic protein* (Longo et al., 2013). While high salt also stabilizes PV1, it is not essential for folding stability as the aromatic residues in the core result in efficient hydrophobic packing within the  $\beta$ -trefoil architecture. The crystal structure of the PV2 protein showed a substantial acidic surface charge characteristic of halophile proteins, and distinctly different from the initial mesophile  $\beta$ -trefoil protein from which PV2 was derived. Thus, by several defining criteria the enrichment of prebiotic amino acids in creating the PV2 protein had produced a halophile protein (Longo et al., 2013). The PV2 protein, however, is not 100% prebiotic in its amino acid composition, and further work to achieve an entirely prebiotic foldable protein is needed to support the hypotheses put forth in this opinion article.

A reasonable postulate of abiogenesis is that some residual aspect of the process may still be identifiable in extant organisms. Protein machinery in extant

organisms can be profoundly complex—as can be seen in molecular assemblies such as ATPase, ribosomes, cilia, the proteasome, pyruvate dehydrogenase complex, myosin, and others; however, such complex protein assemblages are built up from *remarkably simple  $\alpha$ -amino acids that can be synthesized by abiotic chemical processes*. The amino acid composition of proteins is enriched for the prebiotic set with 64% of amino acids being prebiotic (Figure 1); however, the composition of halophile proteins shows a substantially greater enrichment (72%) of prebiotic amino acids (Figure 1). Thus, it is compelling to speculate that this signature is a legacy of abiogenesis—in that *the properties of the halophile environment are highly compatible with foldable polypeptides derived from available prebiotic  $\alpha$ -amino acids*. The halophile environment is typically thought of as a curious niche that mesophiles adapted into; however, it has also been proposed as an appropriate environment for abiogenesis and proteogenesis (Dundas, 1998; Rode, 1999). Studies of the folding potential of the set of prebiotic  $\alpha$ -amino acids suggest that the halophile environment was a potentially ideal cradle for the proteogenic process in abiogenesis.

## REFERENCES

- Brooks, D. J., Fresco, J. R., Lesk, A. M., and Singh, M. (2002). Evolution of amino acid frequencies in proteins over deep time: inferred order of introduction of amino acids into the genetic code. *Mol. Biol. Evol.* 19, 1645–1655. doi: 10.1093/oxfordjournals.molbev.a003988
- Cleaves, H. J. I. (2010). The origin of the biologically coded amino acids. *J. Theor. Biol.* 263, 490–498. doi: 10.1016/j.jtbi.2009.12.014
- Cockell, C. S., and Airo, A. (2002). On the plausibility of a UV transparent biochemistry. *Orig. Life Evol. Biosph.* 32, 255–274. doi: 10.1023/A:1016507810083
- Dundas, I. (1998). Was the environment for primordial life hypersaline? *Extremophiles* 2, 375–377. doi: 10.1007/s007920050081
- Dyer, K. F. (1971). The quiet revolution: a new synthesis of biological knowledge. *J. Biol. Educ.* 5, 15–24. doi: 10.1080/00219266.1971.9653663
- Fukuchi, S., Yoshimune, K., Wakayama, M., Moriguchi, M., and Nishikawa, K. (2003). Unique amino acid composition of proteins in halophilic bacteria. *J. Mol. Biol.* 327, 347–357. doi: 10.1016/S0022-2836(03)00150-5
- Hazen, R. M., Papineau, D., Bleeker, W., Downs, R. T., Ferry, J. M., McCoy, T. J., et al. (2008). Mineral evolution. *Am. Mineral.* 93, 1693–1720. doi: 10.2138/am.2008.2955
- King, J. L., and Jukes, T. H. (1969). Non-Darwinian evolution. *Science* 164, 788–798. doi: 10.1126/science.164.3881.788
- Lee, J., and Blaber, M. (2011). Experimental support for the evolution of symmetric protein architecture from a simple peptide motif. *Proc. Natl. Acad. Sci. U.S.A.* 108, 126–130. doi: 10.1073/pnas.1015032108
- Lee, J., Blaber, S. I., Dubey, V. K., and Blaber, M. (2011). A polypeptide “building block” for the  $\beta$ -trefoil fold identified by “top-down symmetric deconstruction.” *J. Mol. Biol.* 407, 744–763. doi: 10.1016/j.jmb.2011.02.002
- Longo, L., Lee, J., and Blaber, M. (2013). Simplified protein design biased for pre-biotic amino acids yields a foldable, halophilic protein. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2135–2139. doi: 10.1073/pnas.1219530110
- Longo, L. M., and Blaber, M. (2012). Protein design at the interface of the pre-biotic and biotic worlds. *Arch. Biochem. Biophys.* 526, 16–21. doi: 10.1016/j.abb.2012.06.009
- McDonald, G. D., and Storrie-Lombardi, M. C. (2010). Biochemical constraints in a protobiotic Earth devoid of basic amino acids: the “BAA(-) World.” *Astrobiology* 10, 989–1000. doi: 10.1089/ast.2010.0484
- Murphy, L. R., Wallqvist, A., and Levy, R. M. (2000). Simplified amino acid alphabets for protein fold recognition and implications for folding. *Protein Eng.* 13, 149–152. doi: 10.1093/protein/13.3.149
- Oparin, A. I. (1952). *The Origin of Life*. New York, NY: Dover.
- Rode, B. M. (1999). Peptides and the origin of life. *Peptides* 20, 773–786. doi: 10.1016/S0196-9781(99)00062-5
- Romero, P., Obradovic, Z., and Dunker, A. K. (1999). Folding minimal sequences: the lower bound for sequence complexity of globular proteins. *FEBS Lett.* 462, 363–367. doi: 10.1016/S0014-5793(99)01557-4
- Takano, Y., Ohashi, A., Kaneko, T., and Kobayashi, K. (2004). Abiotic synthesis of high-molecular-weight organics from an inorganic gas mixture of carbon monoxide, ammonia, and water by 3 MeV proton irradiation. *Appl. Phys. Lett.* 84, 1410–1412. doi: 10.1063/1.1646757
- Wong, J. T.-F. (2005). Coevolution theory of the genetic code at age thirty. *Bioessays* 27, 406–425. doi: 10.1002/bies.20208

Received: 09 November 2013; accepted: 19 December 2013; published online: 06 January 2014.

Citation: Longo LM and Blaber M (2014) Prebiotic protein design supports a halophile origin of foldable proteins. *Front. Microbiol.* 4:418. doi: 10.3389/fmicb.2013.00418

This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.

Copyright © 2014 Longo and Blaber. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Glass-forming property of hydroxyectoine is the cause of its superior function as a desiccation protectant

Christoph Tanne<sup>1</sup>, Elena A. Golovina<sup>2</sup>, Folkert A. Hoekstra<sup>2</sup>, Andrea Meffert<sup>1</sup> and Erwin A. Galinski<sup>1\*</sup>

<sup>1</sup> Institute of Microbiology and Biotechnology, Rheinische Friedrich-Wilhelms-University Bonn, Bonn, Germany

<sup>2</sup> Laboratory of Plant Physiology, Wageningen University, Wageningen, Netherlands

**Edited by:**

Aharon Oren, The Hebrew University of Jerusalem, Israel

**Reviewed by:**

R. Thane Papke, University of Connecticut, USA

Aharon Oren, The Hebrew University of Jerusalem, Israel

**\*Correspondence:**

Erwin A. Galinski, Institute of Microbiology and Biotechnology, Rheinische Friedrich-Wilhelms-University Bonn, Meckenheimer Allee 168, Bonn 53115, Germany  
e-mail: galinski@uni-bonn.de

We were able to demonstrate that hydroxyectoine, in contrast to ectoine, is a good glass-forming compound. Fourier transform infrared and spin label electron spin resonance studies of dry ectoine and hydroxyectoine have shown that the superior glass-forming properties of hydroxyectoine result from stronger intermolecular H-bonds with the OH group of hydroxyectoine. Spin probe experiments have also shown that better molecular immobilization in dry hydroxyectoine provides better redox stability of the molecules embedded in this dry matrix. With a glass transition temperature of 87°C (vs. 47°C for ectoine) hydroxyectoine displays remarkable desiccation protection properties, on a par with sucrose and trehalose. This explains its accumulation in response to increased salinity and elevated temperature by halophiles such as *Halomonas elongata* and its successful application in “anhydrobiotic engineering” of both enzymes and whole cells.

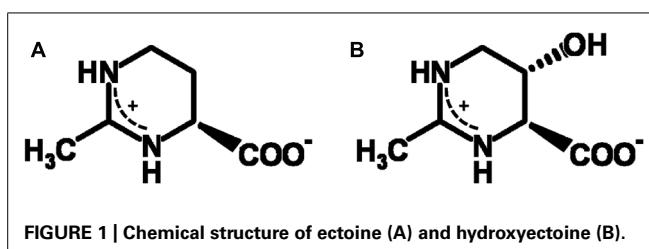
**Keywords:** hydroxyectoine, desiccation, glass transition temperature, enzyme stabilization, ESR, FTIR

## INTRODUCTION

Compatible solutes (organic osmolytes) are low-molecular mass water-binding organic solutes, which are accumulated in the cytoplasm of halophiles for osmotic equilibrium, either as a replacement for or in combination with inorganic salts. They are also known as versatile stress-protecting compounds, in particular for the stabilization of proteins, membranes and whole cells. The cyclic amino acid derivative ectoine (**Figure 1A**) is one of the most common compatible solutes among halophilic heterotrophic *Bacteria* and has found diverse applications, above all as ingredient of skin care products (Graf et al., 2008). Organisms able to synthesize ectoine are often also able to convert this compound into S,S-1,4,5,6-tetrahydro-2-methyl-5-hydroxypyrimidine-4-carboxylate, hydroxyectoine (**Figure 1B**), by means of a 2-oxoglutarate dependent non-heme-iron(II) containing dioxygenase (Bursy et al., 2007; Reuter et al., 2010; Widderich et al., 2014). In *Chromohalobacter salexigens*, a member of the *Halomonadaceae*, the relative proportion of the hydroxylated derivative increases with salinity and/or temperature (Garcia-Estepa et al., 2006; Vargas et al., 2008). It has long been known that hydroxyectoine is a superior stress protectant against desiccation for both whole cells and enzymes (Lippert and Galinski, 1992; Louis et al., 1994). This knowledge has subsequently been applied in “anhydrobiotic engineering” of *Escherichia coli* and *Pseudomonas putida* (Manzanera et al., 2002, 2004a,b). In addition, a comparative enzyme protection study with heat-stabilizing compounds from extreme thermophiles (Borges et al., 2002) has revealed superb heat-stabilizing properties. However, the biophysical basis of the difference between ectoine and hydroxyectoine, of which only the latter is a good heat and desiccation protectant, has until now not been resolved. The sugars sucrose and trehalose, on the other hand, are well known desiccation protectants in all domains of life and their remarkable function has been linked

to the ability to form glasses, which in selected cases ensures conservation of biological functions in an (almost) completely dry state. This phenomenon of “anhydrobiosis” (life without water; Clegg, 2001) is apparent in many higher forms of life (e.g., seeds, resurrection plants, tardigrada, the chironomid *Polypedilum vanderplanki*). Cytoplasmic glasses are considered one of the main mechanisms of desiccation tolerance (Crowe et al., 1998; Hoekstra et al., 2001; Potts, 2001; Buitink and Leprince, 2004; Ballesteros and Walters, 2011). The outstanding role of disaccharides (possibly in combination with intrinsically disordered proteins, IDPs) has given rise to a number of biophysical models, of which the “water entrapment” (Belton and Gil, 1994; Cottone et al., 2002) and “anchorage” hypothesis (Allison et al., 1999; Francia et al., 2008) are the most comprehensive because they encompass glass-formation of solutes and simultaneous entrapment of small water clusters, possibly anchored to critical sites at the interface with biomolecules.

The preservation of biological structures and bioactivity over a long period of time is a main challenge in life science with impacts on medicine, pharmacy and biohybrid technologies (e.g., biosensors). As the halophilic *Halomonas elongata* is unable to synthesize trehalose and/or sucrose, but appears to convert ectoine into hydroxyectoine in response to heat and water stress, we investigated whether this provides an alternative adaptation strategy for survival of cells and biomolecules in the dry state. This required a biophysical comparison of the glass-forming abilities of ectoine and hydroxyectoine and, in particular, the thermal stability of anhydrous glasses as characterized by the glass transition temperature (Tg). As intermolecular H-bonding plays an essential role in thermal properties of glasses (Omaya et al., 2008; Zhou et al., 2013), temperature-controlled infrared spectroscopy is widely used to study hydrogen bonds in solids. A compatible approach in the study of anhydrous



glasses is spin probe electron spin resonance (ESR). As the motion of spin probe molecules is influenced by their intermolecular H-bonding with surrounding molecules, the dynamics of spin probe molecules is obtained from the analysis of the shape of their ESR spectra. Both approaches were used here to characterize differences in glass properties of ectoine and hydroxyectoine.

High molecular immobilization in glasses provides chemical stability of the system due to restricted molecular diffusion. Chemical stability together with structural stability are the main factors which determine survival of organisms in the dry state. The spin probe approach provides an opportunity to obtain information about molecular immobilization and redox activity from the same ESR spectra. Spin probe molecules can participate in different redox-reactions resulting in non-paramagnetic species (Belkin et al., 1987). Therefore, the spin probe approach is a unique technique which allows direct observation of the relationships between structural and chemical stability of dry systems.

## MATERIALS AND METHODS

### CHEMICALS

Antibiotic broth medium No. 3 (complex medium) was purchased from Oxoid LTD (Hampshire, Great Britain). Pyruvic acid, sucrose,  $\text{KH}_2\text{PO}_4$ , KOH, and  $(\text{NH}_4)_2\text{SO}_4$  were purchased from Roth (Karlsruhe, Germany). D-glucose,  $\text{MgSO}_4 \cdot 7 \text{ H}_2\text{O}$ , and  $\text{FeSO}_4 \cdot 7 \text{ H}_2\text{O}$  were obtained from Merck (Darmstadt, Germany). Trehalose and NaCl were purchased from Fluka (Buchs, Switzerland). Ectoine ( $\geq 99\%$ ) for lactate dehydrogenase stress experiments was purchased from bitop AG (Witten, Germany). Hydroxyectoine ( $\geq 99\%$ ) for lactate dehydrogenase (LDH) stress experiments was isolated from *H. elongata* strain DSM 2581<sup>T</sup> in our laboratories. Freeze-dried LDH (rabbit muscle) and nicotinamide adenine dinucleotide were purchased from Sigma (Steinheim, Germany). Phosphate buffered saline (PBS) was purchased from AppliChem (Darmstadt, Germany). Ectoine ( $\geq 99.0\%$ ) and hydroxyectoine ( $\geq 98\%$ ) for spin lable and Fourier transform infrared (FTIR) experiments were purchased from Sigma (USA). Perdeuterated spin probe Tempone-d<sub>16</sub> (4-Oxo-2,2,6,6-tetramethylpiperidine-d<sub>16</sub>-1-oxyl; **Figure 7 inset**) was a kind gift of Prof. I. Grigoriev (Institute of Organic Chemistry of the Russian Academy of Sciences, Novosibirsk, Russia).

### STRAINS, MEDIA, AND CULTIVATION

*Halomonas elongata* strain DSM 2581<sup>T</sup> was obtained from DSMZ (Braunschweig, Germany). Complex media were used for precultures (5.0 g/L peptone, 1.5 g/L yeast extract, 1.5 g/L “Lab-Lemco”

powder, 1.0 g/L glucose, 3.68 g/L  $\text{K}_2\text{HPO}_4$ , 1.32 g/L  $\text{KH}_2\text{PO}_4$  and 150 g/L NaCl; pH was adjusted to 7.4). Minimal media MM63 were used for second pre- and for main culture [13.61 g/L  $\text{KH}_2\text{PO}_4$ , 4.21 g/L KOH, 0.25 g/L  $\text{MgSO}_4 \cdot 7 \text{ H}_2\text{O}$ , 1.98 g/L  $(\text{NH}_4)_2\text{SO}_4$ , 0.0011 g/L  $\text{FeSO}_4 \cdot 7 \text{ H}_2\text{O}$ , 5 g/L glucose, and the required amount of NaCl, pH was adjusted to 7.4; Larsen et al., 1987].

Bacterial cultures were grown in shake flasks. Cell growth was tracked by optical density measurement of incubated media at a wavelength of 600 nm. For solute content analysis *H. elongata* was grown in MM63 with 10 or 15% NaCl (w/w) at 30, 40, and 45°C, respectively. In addition, to surpass its temperature limit of growth, a thermal shock from 37 to 50°C was applied in the mid exponential growth phase. For the determination of desiccation survival, *H. elongata* was grown in two main cultures MM63 with 15% NaCl at 30°C. In the mid exponential growth phase one of the main cultures was shocked to 50°C the other remained at 30°C. After 4 h in stationary phase samples were taken for survival experiments.

### SOLUTE CONTENT ANALYSIS

Bacterial cell material was harvested in the stationary growth phase and dried in a rotational vacuum concentrator (Speed-Vac) for at least 8 h at 45°C and 10 mbar. Solute extraction was achieved following the Bligh and Dyer protocol (Bligh and Dyer, 1959), as modified by Galinski and Herzog (1990). Homogenized samples (30 mg) of dried bacterial biomass were extracted by vigorous shaking (10 min) with 500  $\mu\text{L}$  of modified Bligh and Dyer solution [methanol/chloroform/water 10:5:4 (v/v)]. Following the addition of 130  $\mu\text{L}$  chloroform and 130  $\mu\text{L}$  water, phase separation was assisted by centrifugation at approximately 9300 g (5 min) and the resulting polar upper phase was used for high performance liquid chromatography (HPLC) analysis. For the detection of neutral zwitter-ionic or polar uncharged water soluble compatible solutes an aminopropyl-modified silica column was used (Grom-Sil Amin-1PR, 3  $\mu\text{m}$ , 125  $\times$  4 mm, LiChrocart-System, Alltech Grom GmbH). The isocratic eluent was 80% acetonitrile-water (v/v) at a flow rate of 1 mL/min. A combination of UV- and RI-detector was used for peak identification and solute quantification.

### DESICCATION-SURVIVAL EXPERIMENT

Stationary phase samples were diluted with glucose-free medium to an optical density of 0.1. From this dilution, 100  $\mu\text{L}$  were taken to determine the initial cell number (threefold). The same volume (100  $\mu\text{L}$ ) of each sample was transferred into both a closed 1.5 mL microcentrifuge tube (undried control) and an open 1.5 mL microcentrifuge tube. The samples were then subjected to vacuum drying for 3 h at 45°C and 10 mbar in a SpeedVac (RVC 2-25 CD plus, Christ, Osterode am Harz). Subsequently, dried samples (from open microcentrifuge tubes) were rehydrated by adding 1 mL of glucose-free MM63. Samples were further diluted to  $10^{-6}$  or  $10^{-7}$  for viable cell number determination. This procedure was repeated three times for every sample. Appropriate dilutions of the samples were plated on a complex medium agar plate with 15% NaCl. After 48 h of incubation at 30°C colony forming units (CFU) were counted and expressed as percentage of initial cell number.

## CASTING OF SOLUTE GLASS MATRICES

To cast solute glass matrices 3  $\mu\text{L}$  of 2 M solute solution in ultrapure water were placed on a clean polystyrene plate and dried at 60°C for 2 h. Transmission light microscopy was used for optical characterization of glass matrices.

## LACTATE DEHYDROGENASE (LDH) STRESS EXPERIMENT AND ACTIVITY ASSAY

Ten microliter of 2 M solute solution or ultrapure water with 0.05 mg/mL LDH were placed carefully into separated wells of a 96 well plate. Solidification of solute solutions was accomplished by air drying at 60°C for 2, 4, and 6 h. Rehydration to the original volume was accomplished within 2 min of incubation in PBS buffer, pH 7.5 (AppliChem, Darmstadt) at room temperature. The working volume (200  $\mu\text{L}$ ) for the activity determination of undried (control) and rehydrated LDH solute matrices was obtained by diluting (20-fold) to a final enzyme concentration of 2.5  $\mu\text{g}/\text{mL}$  LDH with 160  $\mu\text{L}$  PBS buffer (pH 7.5), 20  $\mu\text{L}$  10 mM pyruvic acid (final concentration 1 mM) and by addition of 20  $\mu\text{L}$  of 7.5 mM NADH/H<sup>+</sup> (final concentration of 0.75 mM) to start the reaction. Decrease in absorbance at 340 nm was monitored.

## SPIN LABEL EXPERIMENTS

Tempone was added to aqueous solutions of ectoine or hydroxyectoine (32 mg/mL) in a final concentration of 1 mol%, so that in a dry state the proportion of (hydroxy) ectoine: Tempone = 100:1. At such proportion the concentration broadening of the ESR spectra would not be observed under conditions of uniform distribution of spin probe molecules. Samples of labeled solution (100  $\mu\text{L}$ ) were spread over chemically inert glass beads (80–110  $\mu\text{m}$  diameter) on a glass slide and allowed to dry for 5 days in a stream of dry air (3% RH) at room temperature in an air-dry box. The dried material was transferred to 2-mm capillaries in an air-dry box (3% RH), to prevent rehydration on air, and then flame-sealed.

The capillary with the sample was placed in an ESR quartz tube for spectrum recording. ESR spectra were recorded with an X-band ESR spectrometer (Elexsys model E 500; Bruker Analytik, Rheinstetten, Germany) equipped with a temperature unit using regular air within the temperature range 295–400K and liquid N<sub>2</sub> for temperatures below 295 K. The spectra were recorded at 5° increments with equilibration for 1 min at each temperature. The scan range was 100 G for all spectra. To prevent over-modulation and saturation of ESR signal, the modulation amplitude was 2.5 G for solid-like spectra and 1 G for fluid-like spectra. The microwave power was limited to 5 mW.

## FTIR EXPERIMENTS

Small volumes (5  $\mu\text{L}$ ) of aqueous solution of ectoine or hydroxyectoine (32 mg/ml) were dried on circular CaF<sub>2</sub> windows (2 × 13 mm) in a stream of dry air (3% RH) at room temperature in an air-dry box. Although most of the water was removed fast, the samples were further air-dried at 3% RH for 5 days in order to achieve equilibrium water potential. Each sample was hermetically sealed between two CaF<sub>2</sub> windows using a rubber O-ring and mounted into a temperature-controlled brass cell.

Infrared spectra of dry ectoine and hydroxyectoine were obtained with a Perkin-Elmer series 1725 FTIR spectrometer equipped with an external beam facility to which a Perkin-Elmer IR-microscope was attached. The microscope was equipped with a narrow band mercury-cadmium-telluride liquid nitrogen-cooled IR-detector. The samples between two CaF<sub>2</sub> windows were tightly mounted into a temperature-controlled brass cell that was cooled by liquid nitrogen. The temperature of the cell was regulated by a computer-controlled device that activated a liquid nitrogen pump in conjunction with a power supply for heating the cell. The temperature of the sample was recorded separately using two PT-100 elements that were located very close to the sample windows. The optical bench was purged with dry CO<sub>2</sub>-free air. Spectra were recorded starting with the lowest temperature with a scanning rate of 1.5°C/min. The acquisition parameters were: 4  $\text{cm}^{-1}$  resolution, 32 co-added interferograms, 3500–1000  $\text{cm}^{-1}$  wavenumber range.

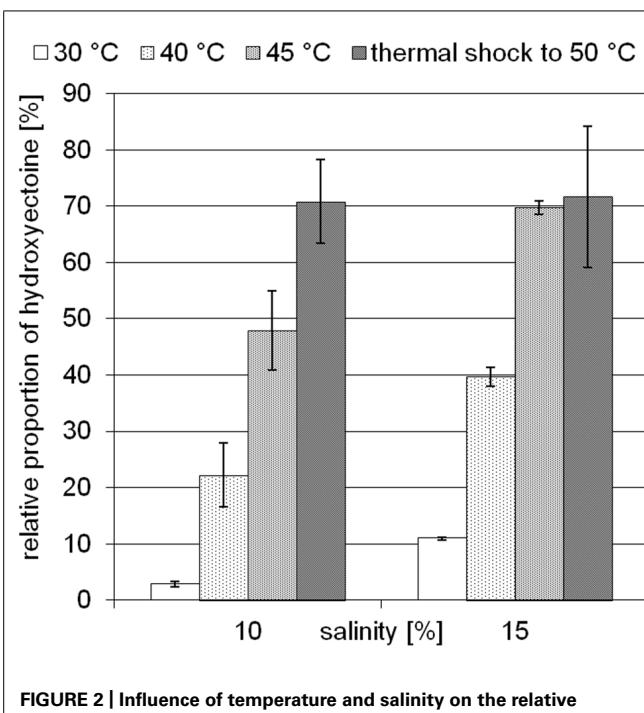
Spectral analysis and display were carried out using the Infrared Data Manager Analytical software, version 3.5 (Perkin-Elmer). The temperature-induced changes in dry ectoine and hydroxyectoine matrixes were monitored by observing the position of the bands around 1388 and 1088  $\text{cm}^{-1}$ . The band around 1388  $\text{cm}^{-1}$  is present in both ectoine and hydroxyectoine, while the band around 1088  $\text{cm}^{-1}$  is present only in hydroxyectoine (Figure 11B). The band positions were calculated as the average of spectral positions ( $n = 50$ ) at 75% of the total peak height (Wolkers et al., 1998). Breaks in the temperature dependence of this peak position were determined as a point of intersection of two regression lines below and above the temperature of the break (Figure 12A).

All ESR and FTIR experiments were conducted on the same samples prepared under the same conditions. Each model experiment was repeated at least twice, and the results of the single experiments are presented.

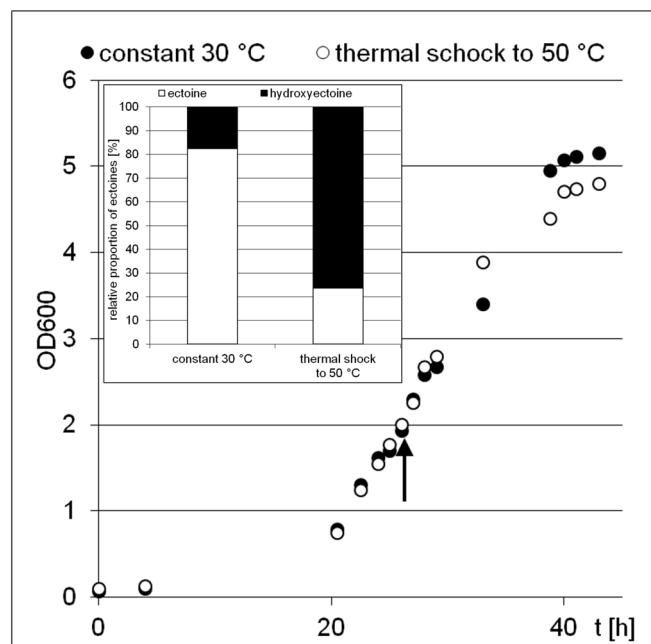
## RESULTS

### INFLUENCE OF WATER AND TEMPERATURE STRESS ON INTRACELLULAR ECTOINE LEVELS

The moderately halophilic (optimum 3–5% NaCl) but extremely halotolerant *H. elongata* employs the compatible solutes ectoine/hydroxyectoine for osmotic adaptation by increasing their cytoplasmic concentration in a near-linear fashion. At a salinity of 15% NaCl, *H. elongata* experiences severe water stress and, as a consequence, its growth rate is reduced to 0.1 (approximately 25% of maximum; Dötsch et al., 2008). In contrast to other members of the family Halomonadaceae (e.g., *C. salexigens*), *H. elongata* is not able to synthesize the well-known desiccation protectant trehalose, which makes it a good model for investigating the role of hydroxyectoine. As depicted in Figure 2, increase in both temperature and salinity leads to a higher relative proportion of hydroxyectoine. At its maximum growth temperature of 45°C and a salinity of 15% NaCl, the hydroxyectoine level reached 70% of total ectoines. Although the organism is unable to grow at 50°C from inoculum, it was possible to increase the hydroxyectoine level even further by upshock experiments (i.e., raising the temperature in mid-exponential phase to 50°C). We used the combination of high salinity (15%) and temperature upshock to simulate a dehydration event and investigated its impact on survival rates of *H. elongata*.



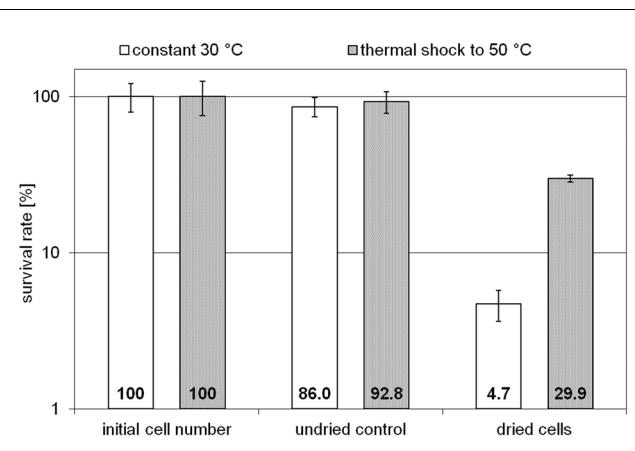
**FIGURE 2 | Influence of temperature and salinity on the relative proportion of hydroxyectoine in *H. elongata*.** Error bars represents lowest and highest value of duplicate measurements for both salinities at 30°C and for 15% NaCl at 45°C; all others represent standard deviations of three replicates (in the case of 15% NaCl, upshock) or at least four replicates.



**FIGURE 3 | Growth of *H. elongata* in minimal medium MM63 with 15% sodium chloride.** The experiment was performed with two parallel culture, one at constant 30°C (black dots), the other with a rapid temperature upshift at OD 2 (arrow) from 30 to 50°C (white dots). The inset shows the relative proportions of ectoines in bacterial cells at the early stationary growth phase (point of harvest).

#### DRY STABILIZATION OF *H. elongata* CELLS AT ELEVATED HYDROXYECTOINE LEVELS

Two *H. elongata* cultures were grown at 15% NaCl and 30°C to an optical density of approximately 2, when one of them was heat shocked to 50°C (arrow in Figure 3). Subsequently, the solute content of both cultures was analyzed at early stationary phase. It is clearly seen that the applied temperature increase had little effect on the organisms growth and yield. The relative proportion of hydroxyectoine, however, increased from approximately 17% (at 30°C) to approximately 75% (at 50°C; Figure 3, inset), indicating that the conversion of ectoine into hydroxyectoine is enhanced further by temperature upshift. To compare the survival rates of desiccated stationary phase cells, colony-forming units of initial cell numbers (control before drying process), undried controls (from closed vials) and dried samples (vacuum drying for 3 h at 45°C and 10 mbar) were determined (Figure 4). It was demonstrated that heat-stressed cells with 75% hydroxyectoine had much higher survival rates. A survival rate of nearly 30% was achieved as compared to only 4.7%, for untreated cells. Although it cannot be excluded that other adaptational processes may also be responsible for improved survival of heat-shocked cells (see Discussion), we concluded that the hydroxylation of ectoine plays an important part in this improved desiccation survival. Provided that a simple hydroxylation step is indeed able to alter the properties of a common compatible solute (ectoine) in such a way as to provide desiccation protection, then drying of the model enzyme LDH should show a similar response and the

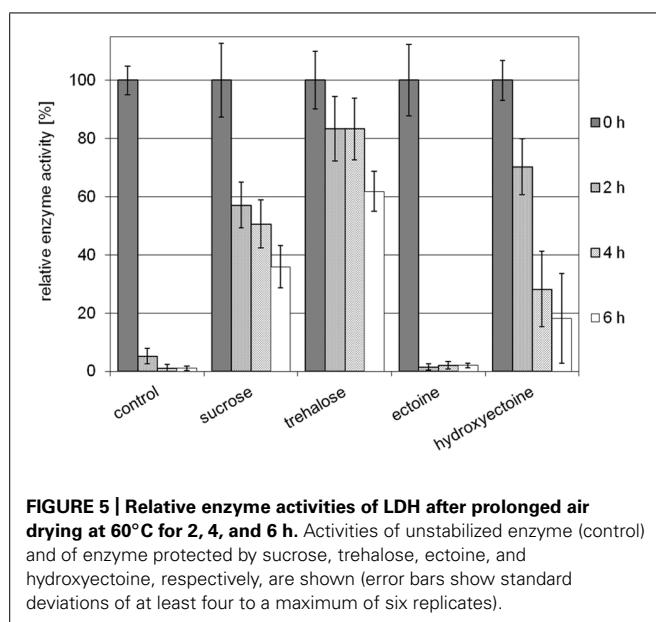


**FIGURE 4 | Survival rates of unstressed *H. elongata* cells (initial cell number), undried control cells and cells which survived harsh drying for 3 h at 45°C and 10 mbar (error bars show standard deviations of triplicates).**

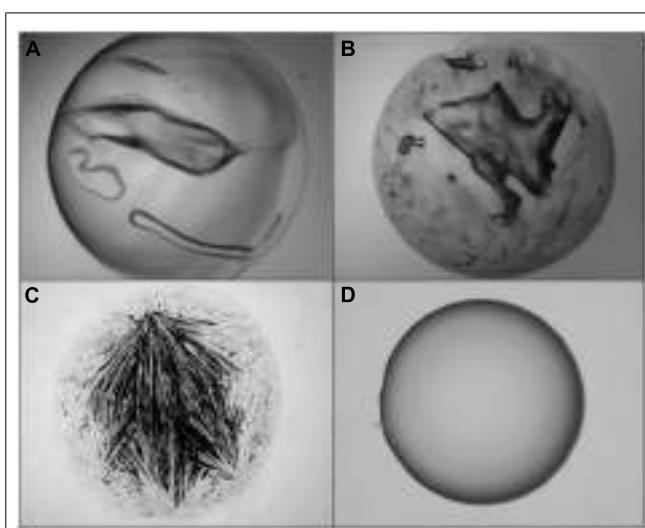
protective effect of hydroxyectoine should compare favorably with the well-known desiccation protectants sucrose and trehalose.

#### STRESS PROTECTION BY HYDROXYECTOINE DURING HEATING/DRYING OF LACTATE DEHYDROGENASE (LDH)

Stabilization of model enzyme LDH by compatible solutes against heat and freeze-drying has been investigated before, and for both



**FIGURE 5 | Relative enzyme activities of LDH after prolonged air drying at 60°C for 2, 4, and 6 h.** Activities of unstabilized enzyme (control) and of enzyme protected by sucrose, trehalose, ectoine, and hydroxyectoine, respectively, are shown (error bars show standard deviations of at least four to a maximum of six replicates).



**FIGURE 6 | Light microscopic photographs of solute matrices which were casted from 2 M solutions of trehalose (A), sucrose (B), ectoine (C), and hydroxyectoine (D) by air drying of 10 μL at 60°C for 2 h.**

stress factors the superiority of hydroxyectoine over ectoine has been demonstrated (Lippert and Galinski, 1992; Borges et al., 2002). Here a small volume (10  $\mu$ L) of LDH at a concentration of 0.05 mg/mL in 2 M solutes (ectoine, hydroxyectoine, sucrose, and trehalose, respectively) was exposed to air-drying at 60°C. The dried protein was subsequently diluted in buffer and checked for residual activity. As shown in **Figure 5**, the unprotected enzyme loses approximately 95% of its original activity after 2 h of drying. Prolonged drying destroyed activity almost completely. It is worthy of note that the presence of ectoine had no stabilizing effect under the conditions employed, whereas hydroxyectoine after 2 h drying at 60°C displayed a residual activity of approximately 70%, which lies between the values of sucrose (58%) and trehalose (83%) as benchmarks. Upon further drying, however, the stabilizing effect of hydroxyectoine declined more rapidly than with disaccharides. Nonetheless the observed differences between ectoine (no stabilization) and hydroxyectoine (on a par with disaccharides after 2 h of drying) are remarkable and put hydroxyectoine into the same category as the glass-forming disaccharides. We therefore expanded our comparison of both ectoines to include their glass-forming abilities.

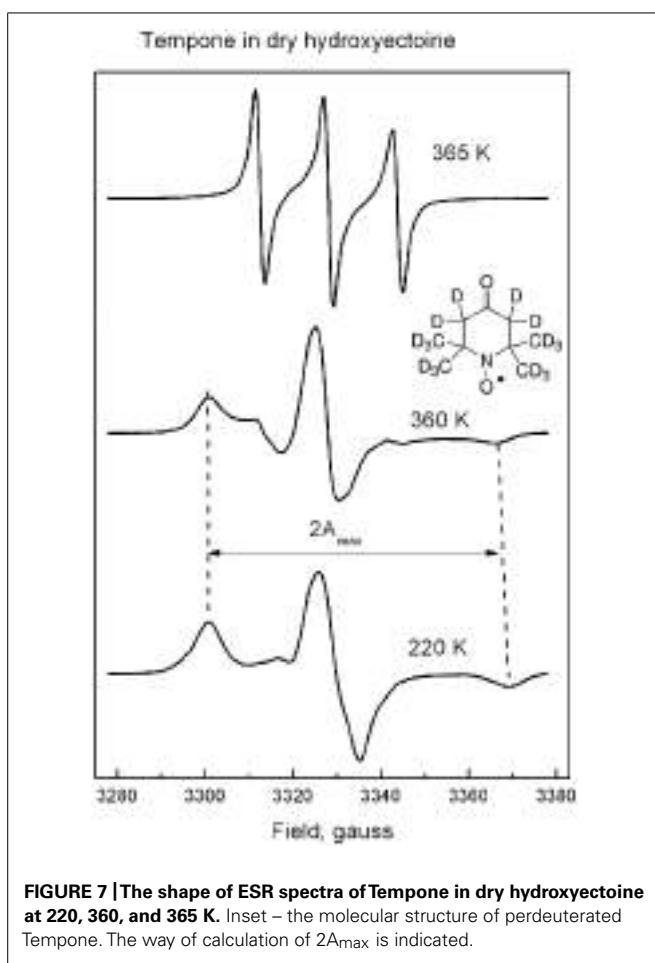
#### GLASS-FORMING ABILITIES OF ECTOINES

A simple glass-casting experiment from 2 M solution at 60°C disclosed a striking difference between both ectoines (**Figures 6C,D**). While hydroxyectoine formed a clear and transparent solid, visual examination of ectoine samples revealed an inhomogeneous structure with crystalline inclusions, indicating a mixture of glassy and crystalline states. It can also be seen that the benchmark glass-formers sucrose and trehalose formed solids with cracks under the drying conditions employed (**Figures 6A,B**). These were never observed with solid hydroxyectoine samples. Thus it can be concluded that hydroxyectoine, in contrast to ectoine, is a good glass

former and that this property is probably related to the superior desiccation protection of hydroxyectoine on biological structures.

#### SPIN PROBE STUDY OF GLASS PROPERTIES OF DRY ECTOINE AND HYDROXYECTOINE

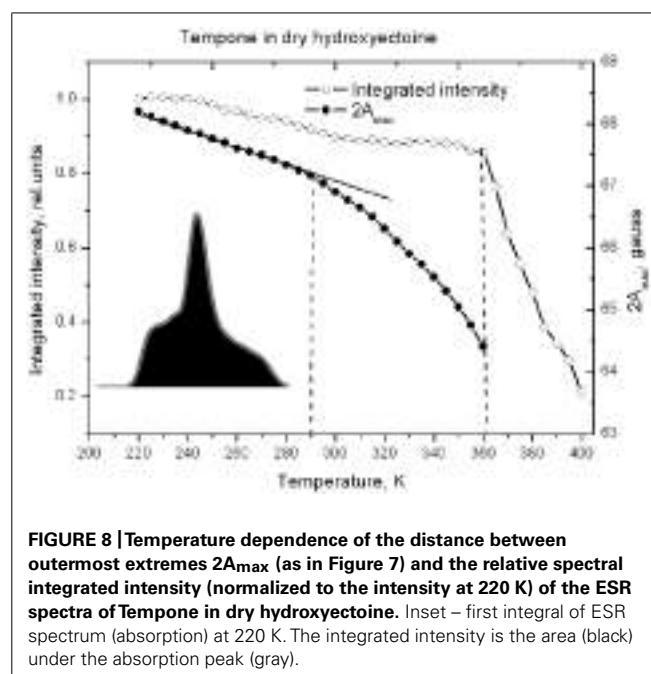
Hydrogen bonds and packing density are the key factors which determine the properties of anhydrous glasses (Omayu et al., 2008; Wang et al., 2009; Zhou et al., 2013). Spin probe Tempone as a reporter molecule was used to study the glass properties of dry ectoine and hydroxyectoine. Tempone is a small water-soluble stable free radical, which has the shape of a sphere with a radius around 3 Å. It has one ketone group  $>\text{C}=\text{O}$  (**Figure 7**, inset), which can be an acceptor for H-bonds. The uncoupled electron nitroxide group N-O is surrounded by 4 bulky methyl groups (**Figure 7**, inset) and is probably less available for H-bonding. One molecule of ectoine can provide two donors for hydrogen bonds – two N-H groups from ring N – and one acceptor  $\text{C}=\text{O}$  from  $\text{COO}^-$  group (**Figure 1**). Hydroxyectoine has one additional potential donor. This is a hydroxyl group  $-\text{OH}$ , which is attached to the heterocyclic ring (**Figure 1**). The H-bonds between  $>\text{C}=\text{O}$  of Tempone and N-H groups of both ectoine and hydroxyectoine are less strong than the H-bond between  $>\text{C}=\text{O}$  of Tempone and OH group of hydroxyectoine because oxygen is more electronegative than nitrogen. Being hydrogen-bonded with ectoine and hydroxyectoine, the spin probe molecule moves collectively with the solute molecules as long as such H-bonds exist. Tempone has an ESR spectrum consisting of three lines due to hyperfine interactions of its uncoupled electron with nitrogen spin (Knowles et al., 1976). Because of spectral anisotropy, ESR spectra of spin probes are sensitive to motion and are, therefore, suitable for a temperature-dependent study of molecular immobilization in a dry matrix caused by H-bonds.



**FIGURE 7 |**The shape of ESR spectra of Tempone in dry hydroxyectoine at 220, 360, and 365 K. Inset – the molecular structure of perdeuterated Tempone. The way of calculation of  $2A_{\max}$  is indicated.

Tempone in dry hydroxyectoine has a solid-like spectrum up to 360 K (Figure 7). This spectral shape is typical for highly immobilized spin probe molecules. The distance between outer extremes  $2A_{\max}$  (Figure 7) is used to characterize the degree of immobilization of the nitroxide moiety of the spin probe (Knowles et al., 1976). Above 360 K a sudden change in the spectral shape is observed (Figure 7). The spectrum of Tempone at 365 K in Figure 7 has three equidistant narrow lines. Such a spectral shape is typical for fast isotropic rotation of the spin probe molecule (Knowles et al., 1976).

Figure 8 shows that  $2A_{\max}$  of Tempone spectra decreases slowly up to 290 K, after which the rate of disordering increases, but the spin probe remains immobile up to 360 K. Obviously, the break in the temperature dependence of  $2A_{\max}$  is caused by some structural rearrangements in the solvent matrix (Dzuba et al., 1993, 2005). Above 360 K,  $2A_{\max}$  cannot be determined because the spectrum becomes isotropic (Figure 7). The ESR spectrum is the first derivative of the absorption spectrum. To estimate the number of paramagnetic centers in a sample, the ESR spectrum has to be double-integrated to obtain the area under the absorption peak (Figure 8, inset). The double integral of the ESR spectra is called integrated intensity. The onset of a sharp decrease in integrated intensity coincides with the point of fast isotropic rotation at 360 K. Such an increase of the molecular freedom may result

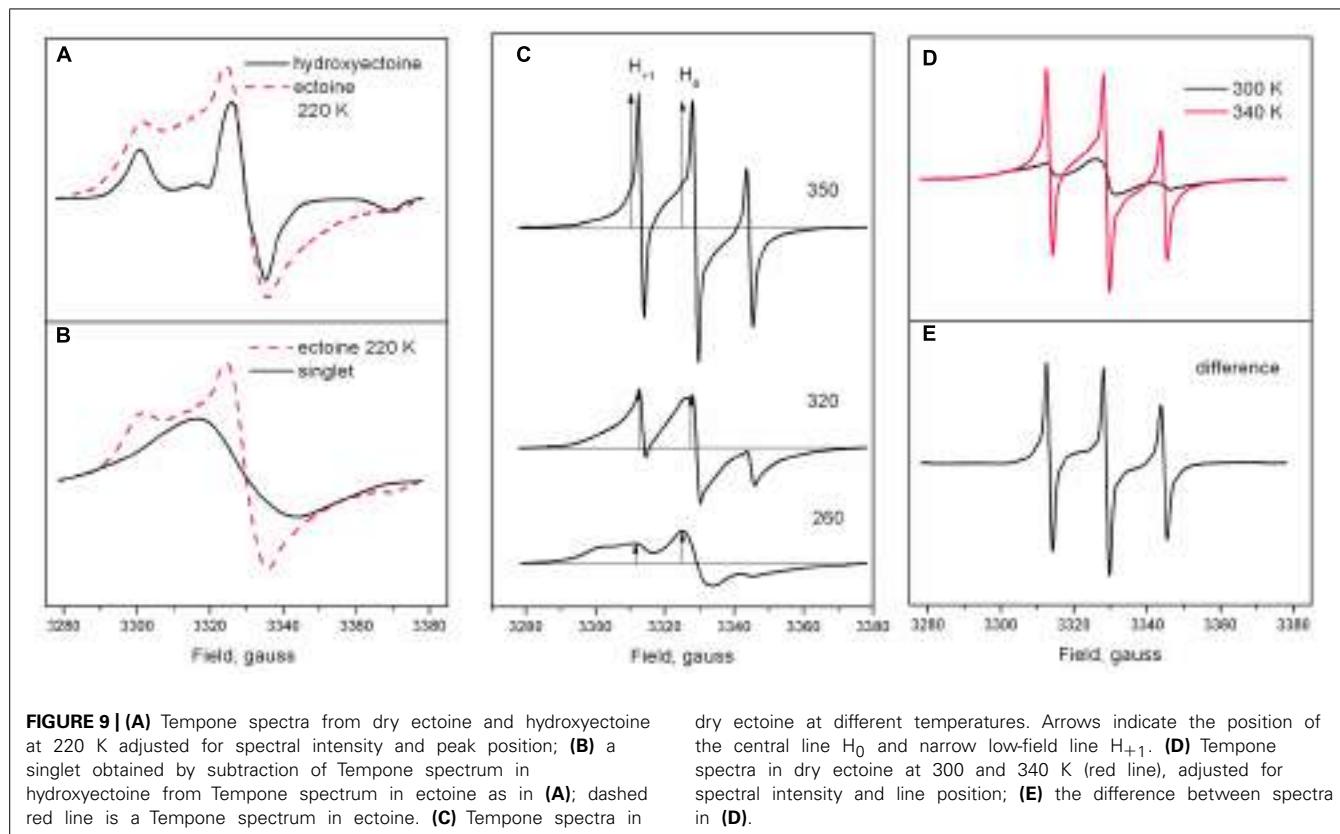


**FIGURE 8 |**Temperature dependence of the distance between outermost extremes  $2A_{\max}$  (as in Figure 7) and the relative spectral integrated intensity (normalized to the intensity at 220 K) of the ESR spectra of Tempone in dry hydroxyectoine. Inset – first integral of ESR spectrum (absorption) at 220 K. The integrated intensity is the area (black) under the absorption peak (gray).

from the breaking of hydrogen bonds between the ketone group of Tempone and the hydroxyl groups of hydroxyectoine.

Figure 9A shows the shape of Tempone spectra in dry ectoine and hydroxyectoine at 220 K. The shapes of the spectra are different. The spectrum from hydroxyectoine was subtracted from that in dry ectoine after spectral titration (adjustment of spectral position and amplitudes). The difference spectrum is a singlet (Figure 9B). The spectrum subtraction shows that the ESR spectrum of Tempone in ectoine is the superposition of a solid-like triplet (as in hydroxyectoine) and a singlet. The solid-like triplet spectrum (as in Figures 7 and 9A) is caused by Tempone molecules, which are spatially separated (i.e., in a glassy state of ectoine). Singlet spectra (Figure 9B) are caused by spin exchange between highly concentrated Tempone molecules (Knowles et al., 1976). This results from Tempone molecules, which are excluded from crystalline ectoine and therefore locally concentrated. It is therefore possible to conclude that under the conditions employed (5 days of air-drying at room temperature), solid ectoine samples are a mixture of glassy and crystalline states.

The superposition of singlet and triplet in Tempone spectra in dry ectoine does not allow the correct determination of  $2A_{\max}$ . However, it is still possible to characterize the temperature-induced dynamic changes in Tempone spectra in dry ectoine at a semi-quantitative level by plotting the ratio of the heights of the positive peaks of the low-field narrow line  $H_{+1}$  and the central line  $H_0$  (arrows in Figure 9C) against temperature ( $H_{+1}/H_0$  vs. temperature). The central line  $H_0$  is a superposition of central lines from both solid-like and fluid-like spectra, while the low-field narrow line  $H_{+1}$  represents only the fluid-like spectrum (Figure 9C).  $H_{+1}/H_0$  is an approximate estimation of the proportion of the fluid-like spectral component. Subtraction of the Tempone spectrum of ectoine at 300 K from that at 340 K (Figure 9D) shows the



narrow-line spectrum, which has a shape typical for fast isotropic motion of the spin probe molecule (**Figure 9E**). **Figures 9C–E** clearly demonstrate the gradual increase of the fluid-like component of the spectra and thus also the increase in the proportion of the freely rotating Tempone molecules in the sample.

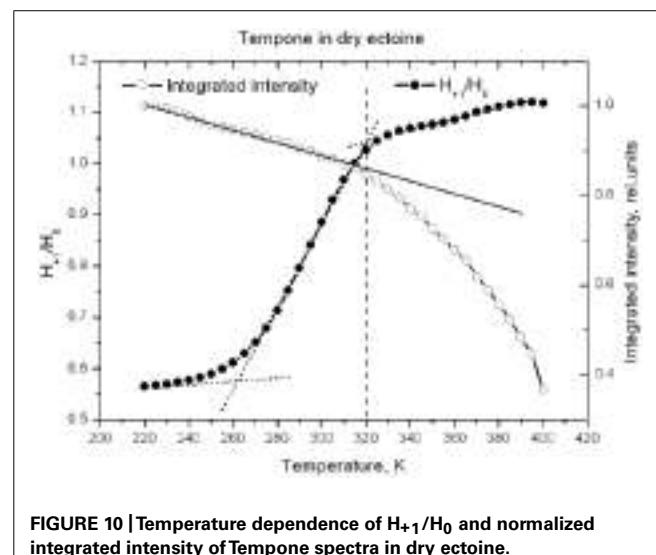
**Figure 10** shows the temperature-induced changes in  $H_{+1}/H_0$  of Tempone spectra in dry ectoine. In contrast to hydroxyectoine (**Figure 8**), the fluid-like component of Tempone spectra from ectoine appears already around 240–250 K as the increase of  $H_{+1}/H_0$ , but the main changes occur between 270 and 320 K. At  $T > 320$  K the spectral shape is presented mainly by a narrow line isotropic spectrum as in **Figure 9C top**. This point of change (320 K) again coincides with the onset of a decrease in integrated intensity of ESR spectra.

Comparison of data in **Figures 8** and **10** suggests that H-bonds between Tempone ( $>\text{C}=\text{O}$ ) and hydroxyectoine exist up to 360 K ( $87^\circ\text{C}$ ) and are completely broken above this temperature allowing free rotation of spin probe molecules. In ectoine the break in hydrogen bonds between Tempone and solute molecules begins below room temperature and is completed at 320 K ( $47^\circ\text{C}$ ). These data show that hydroxyectoine glasses are more thermostable than ectoine glasses.

#### FTIR STUDY OF SOLID MATRIX OF ECTOINE AND HYDROXYECTOINE

Spin probe ESR shows only interactions between guest molecule and solute molecules. Temperature-controlled infrared spectroscopy provides information about H-bonding intra- and intermolecular interactions between solute molecules. The FTIR study

dry ectoine at different temperatures. Arrows indicate the position of the central line  $H_0$  and narrow low-field line  $H_{+1}$ . **(D)** Tempone spectra in dry ectoine at 300 and 340 K (red line), adjusted for spectral intensity and line position; **(E)** the difference between spectra in **(D)**.

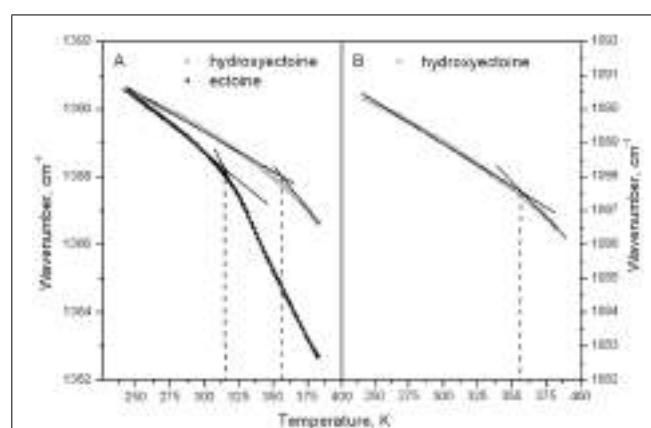


**FIGURE 10 |** Temperature dependence of  $H_{+1}/H_0$  and normalized integrated intensity of Tempone spectra in dry ectoine.

of dry ectoine and hydroxyectoine was performed parallel to the spin probe study. Visually, partial macro-crystallization of dry ectoine was observed. For IR spectra recording an amorphous area,  $500 \mu\text{m} \times 500 \mu\text{m}$  in size, was selected under attached Perkin-Elmer IR- microscope. This partial crystallization of ectoine had already been concluded from the shape of ESR spectra of Tempone in dry ectoine (**Figure 9**). In the case of hydroxyectoine, no visual crystallization of the dry solid was observed under the microscope.

**Figure 11A** shows the IR spectra of dry amorphous ectoine and hydroxyectoine. Both spectra contain broad overlapping bands in the hydrogen stretching region N-H, C-H, and N-H ( $>2500\text{ cm}^{-1}$ ) and a set of narrow lines in the finger print region (**Figure 11B** in details). The OH stretching vibration band around  $3300\text{ cm}^{-1}$  as a function of temperature was previously used to determine Tg in dry sugars and their mixtures (Wolkers et al., 1998, 2004; Kets et al., 2004; Imamura et al., 2006). The overlapping of different broad peaks in the hydrogen stretching region  $2500\text{--}3500\text{ cm}^{-1}$  in ectoine and hydroxyectoine spectra (**Figure 11A**) does not allow the use of OH stretching band at around  $3300\text{ cm}^{-1}$  for characterization of glasses formed by these compounds. The fingerprint region of the FTIR spectra contains several peaks, which are shown in detail in **Figure 11B**. The identification of peaks in the fingerprint region is difficult and needs additional information (Coates, 2000). The fingerprint area of FTIR spectra from dry hydroxyectoine has some additional peaks which are not present in ectoine, and some similar peaks (**Figure 11B**). The differences are caused by the presence of additional OH group in hydroxyectoine, which is attached to the heterocyclic ring.

The information obtained from ESR experiments (**Figures 8** and **10**) can be used for assignment of some bands in the fingerprint region of the spectra in **Figure 11B**. Our spin probe study showed that the H-bonds between  $>\text{C=O}$  of Tempone and OH-group of dry hydroxyectoine are broken at approximately  $360\text{ K}$ . The inspection of the wavenumber-temperature dependences of all major peaks in the fingerprint area of hydroxyectoine showed that only the bands around  $1088$  and  $1388\text{ cm}^{-1}$  (**Figure 11B**) have a break at approximately  $360\text{ K}$  (**Figures 12A,B**). The peak around  $1088\text{ cm}^{-1}$  does not exist in the ectoine spectrum, and can thus be attributed to OH group attached to heterocyclic ring in hydroxyectoine. On the other hand, the FTIR spectrum of dry ectoine has the same band around  $1388\text{ cm}^{-1}$  as hydroxyectoine (**Figure 11B**) and can be assigned to NH groups in both ectoine and hydroxyectoine. However, the break in the wavenumber-temperature dependence



**FIGURE 12 |** Temperature dependence of the frequency vibration at  $1382\text{--}1392\text{ cm}^{-1}$  in dry ectoine and hydroxyectoine (A) and at  $1082\text{--}1092\text{ cm}^{-1}$  in dry hydroxyectoine (B).

for this band in ectoine occurs not at  $360\text{ K}$ , as in hydroxyectoine, but at approximately  $320\text{ K}$  (**Figure 12A**). This coincides with the temperature at which all Tempone molecules start free rotation (**Figures 9** and **10**). The fact that IR inflections of hydroxyectoine are less distinct, is explained by experimental limitations (i.e., upper temperature limit of  $100^\circ\text{C}$  und, therefore, fewer data points beyond Tg) and higher strength of OH hydrogen bonds resulting in a less pronounced slope of the wavenumber vs. temperature plot.

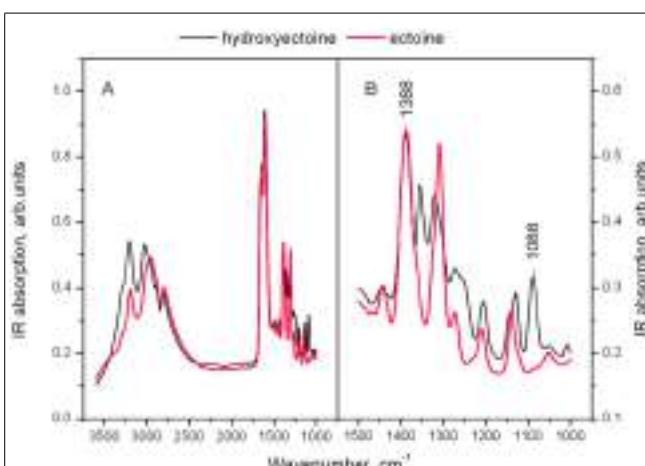
## DISCUSSION

### PROTEIN STABILIZATION EFFECTS OF COMPATIBLE SOLUTES IN SOLUTION

Compatible solutes (organic osmolytes) of halophilic bacteria have long been known as versatile stress-protecting compounds, in particular for the stabilization of proteins and whole cells. The molecular nature of their stabilizing function has been explained by preferential interaction with water and subsequent exclusion from a protein hydration shell (Arakawa and Timasheff, 1985). This original concept has been expanded by others, who revealed that the unfavorable interaction of the peptide backbone is the main driving force for this stabilization phenomenon, named the “osmophobic effect” (Bolen and Baskakov, 2001). In order to elucidate the molecular features which make a solute a stabilizing compound, a quantitative solvation model was proposed (Street et al., 2006), in which backbone/solute interaction energy is a function of the interactants’ polarity and surface area. On the basis of this model, we are now able to quantify the chemical features, which make a solute “compatible” and rank them according to their protein stabilization in solution. The superiority of hydroxyectoine over ectoine in desiccation protection of cells and biomolecules has, however, so far not been explained.

### ESR STUDIES PROVE SUPERIOR GLASS PROPERTIES OF HYDROXYECTOINE

The presented ESR data show that dry hydroxyectoine has considerable advantages as a matrix for guest molecules over ectoine. Hydroxyectoine forms stable glasses, while ectoine is more prone



**FIGURE 11 |** (A) FTIR spectra of dry amorphous ectoine (red line) and hydroxyectoine at room temperature. The intensity of the spectra are normalized to the height of the peak at  $1600\text{ cm}^{-1}$ ; (B) enlarged finger print region of FTIR spectrum.

to crystallization. The difference is caused by one additional OH group in hydroxyectoine. This group reduces the degree of molecule symmetry and thus probably prevents the molecules from tight packing (Wang et al., 2009). This creates the space for optimized orientation of molecules for H-bonding, the formation of which needs not only the presence of donor and acceptor, but also specific geometry (Steiner, 2002). The more bulky shape of hydroxyectoine in dry matrix provides better conditions for H-bonding with other molecules than the more symmetric ectoine, which has a tendency to aggregate and crystallize. As a result, the additional hydroxyl provides more possibilities for H-bonds by increasing the number of donors, the strength of intermolecular interactions and the energy barrier for the reorganization of the molecules. It has already been shown that the presence of hydroxyl groups increases thermal stability of glasses (Omayu et al., 2008; Zhou et al., 2013).

The large temperature difference between molecular fluidizing in ectoine (320 K) and hydroxyectoine (360 K) of 40 degrees, as shown by ESR and FTIR studies (**Figures 8, 10, and 12**), results from different types of intermolecular H-bonds that cause molecular immobilization. In ESR experiments, the motional behavior of Tempone in hydroxyectoine is determined by the strongest H-bonds between CO of Tempone and OH of the solute. In ectoine, motional behavior of Tempone is determined by weaker H-bonds with NH groups of the solute.

Stronger H-bonds provide not only better structural stability but also chemical stability of Tempone molecules in dry hydroxyectoine. Chemical stability of dry material is associated with molecular immobilization, which considerably reduces the probability of diffusion-controlled reactions. The probability of chemical reactions increases above glass transition (Craig et al., 2001). The spin probe approach provides an opportunity to obtain information about molecular immobilization and redox activity from the same ESR spectra. Spin probe molecules can participate in different redox-reactions resulting in non-paramagnetic species. The most probable are reduction to hydroxylamine and irreversible reaction with other radicals to stable non-radical products (Belkin et al., 1987). In the case of redox conversion of spin probe molecules, the number of paramagnetic species decreases resulting in the decrease of the integral intensity of the spectra (**Figures 8 and 10**).

The integrated intensity of the ESR spectra of Tempone in dry hydroxyectoine does not significantly change up to the point of breaking H-bonds between 360 and 365 K. Above this temperature, the integrated intensity of the spectra sharply decreases (**Figure 8**). This coincides with the transformation of the spectral shape of Tempone to fluid-like type (**Figure 7**). The break in the temperature dependence of the integrated intensity of Tempone spectra in dry ectoine was observed at 320 K (**Figure 10**). As in the case of hydroxyectoine, the redox conversion of Tempone in dry ectoine sharply increases when spin probe molecules obtain a motional freedom similar to that in a fluid phase (**Figure 9E**). Such motional freedom was obviously caused by the break of H-bonds  $>\text{C}=\text{O}\dots\text{HN}$  ( $\text{CO}$  belongs to Tempone,  $\text{NH}$  belongs to ectoine). Clearly, redox conversion of spin probe molecules becomes possible when they obtain motional freedom. The fluid-type spectrum of Tempone indicates the possibility of translational diffusion and

increased probability of chemical reactions. Therefore, the presence of hydroxyectoine in the dry cytoplasm of anhydrobiotic organisms would improve the structural and chemical stability of glasses as a result of increased number and strength of H-bonds between molecules.

#### FTIR STUDIES CONFIRM GLASS TRANSITION TEMPERATURES OF ECTOINE AND HYDROXYECTOINE

The broad region of OH stretching vibrations around  $3380\text{ cm}^{-1}$  in FTIR spectra is commonly used to determine  $T_g$  in dry sugars and their mixtures (Wolkers et al., 1998, 2004; Kets et al., 2004; Imamura et al., 2006). Ectoine does not contain OH groups and therefore does not contain a distinct band in this region (**Figure 11A**). Hydroxyectoine contains one OH group attached to the heterocyclic ring, and the FTIR spectrum is expected to display the band around  $3380\text{ cm}^{-1}$  from OH stretching vibrations as for dry sugars. However, only a weak shoulder is present in this region in the FTIR spectrum of hydroxyectoine (**Figure 11A**). Obviously, the broad bands from NH stretching and CH stretching vibrations ( $2800\text{--}3200\text{ cm}^{-1}$ ) mask the OH band at  $3380\text{ cm}^{-1}$ . Close inspection of bands within the fingerprint region helped to find other bands, which can be used to characterize glass transition in dry (hydroxy)ectoine. The decrease of the wavenumber with temperature for both bands at  $1388$  and  $1088\text{ cm}^{-1}$  (**Figures 12A,B**) is characteristic for bending deformations of the functional groups, which are involved in H-bonding (Vanderkooi et al., 2005). All data together allow the assignment of the band around  $1388\text{ cm}^{-1}$  to bending deformations of NH groups of the ring, and the band around  $1088\text{ cm}^{-1}$  to bending deformations of OH attached to the heterocyclic ring. The same position of the break of the temperature dependence for OH and NH bending vibrations in hydroxyectoine (360 K, **Figures 12A,B**) indicates that in this dry matrix NH groups form intermolecular H-bonds with OH groups. In ectoine, OH groups are absent and such H-bonds are not possible. Weaker intermolecular H-bonds between NH and CO of  $\text{COO}^-$  group and, even weaker still, H-bonds between NH groups determine the lower position of the break in the temperature dependence of the wavenumber of NH bending vibration in dry ectoine (**Figure 12A**).

In conclusion, it can be said that FTIR data corroborated the ESR observations, which demonstrate a remarkable difference in  $T_g$  of approximately  $40^\circ\text{C}$  between ectoine and hydroxyectoine. The high  $T_g$  of the latter ( $87^\circ\text{C}$ ) places hydroxyectoine somewhere between sucrose ( $65^\circ\text{C}$ ) and trehalose ( $117^\circ\text{C}$ ).

#### RESPONSE OF *H. elongata* TO SALT AND TEMPERATURE STRESS

Earlier studies have shown increasing hydroxyectoine levels in response to salinity and temperature (Wohlfarth et al., 1990; Severin et al., 1992) in the *Halomonadaceae*. This upregulation of hydroxyectoine synthesis has subsequently been investigated in greater depth with the related *C. salexigens*. Under comparable growth conditions ( $1.5\text{ M NaCl} = 8.7\%$  and  $2.5\text{ M NaCl} = 14.5\%$ , at a temperature of  $45^\circ\text{C}$ ), relative proportions of hydroxyectoine of approximately 60 and 70% respectively have been reported (Garcia-Estepa et al., 2006; Rodriguez-Moya et al., 2013). By genetic manipulation (transcriptional fusion, gene duplication,

and overexpression on vector), this proportion was increased even further to 77% (8.7% NaCl, 37°C), albeit at the expense of a reduced growth rate (Rodríguez-Moya et al., 2013). As demonstrated by **Figure 2**, *H. elongata* grown at 15% NaCl similarly increases its relative proportion of hydroxyectoine in response to both salinity and temperature, with a maximum of approximately 70% at the highest tolerated growth temperature of 45°C. However, we report here for the first time that a sudden temperature upshift to 50°C (beyond the organism's temperature maximum) also leads to a rapid increase in the proportion of hydroxyectoine (up to 75%). Surprisingly, such a temperature shift, when applied in the exponential phase, did not markedly impair growth and yield (**Figure 3**).

### DESICCATION PROTECTION OF WHOLE CELLS

Given the superior glass properties of hydroxyectoine, it is not surprising that (water-stressed) halophilic ectoine-type organisms have adopted a rather economical strategy to protect themselves from desiccation damage. Instead of synthesizing distinct glass-forming compounds such as sucrose and/or trehalose, they simply convert, by a single enzymatic step, one compatible solute into another, hydroxyectoine, with a much higher T<sub>g</sub>. The consequences of such a conversion from 17 to 75% hydroxyectoine (experimentally enforced by salt and temperature shock) is clearly demonstrated by a six-times increase in desiccation survivors under drastic drying conditions (3 h at 45°C and 10 mbar; **Figure 4**). This remarkable improvement was probably accomplished by the high cytoplasmic concentrations of intracellular glass-forming hydroxyectoine. Desiccation tolerance of the related *C. salexigens* has been investigated with cells grown under similar conditions (14.5% NaCl and 45°C) and applying the drying protocol by Manzanera et al. (2002), which entails vacuum drying at 30°C and 313 mbar for 30 h (Reina-Bueno et al., 2012). These conditions are less drastic than those applied here, except for length of time. The authors observed a surprisingly low desiccation tolerance of approximately 5% survivors, which was reduced even further by at least one order of magnitude in knock-out mutants unable to synthesize either hydroxyectoine or trehalose. From this they concluded that both compounds probably play a role in desiccation tolerance, although other factors must be considered. As the relative hydroxyectoine content in *C. salexigens* is similar, the much better performance of *H. elongata* is either explained as a consequence of more favorable drying conditions or of other factors triggered by temperature upshock (see below).

As we did not want to exceed the maximum growth temperature during the drying process, the temperature applied experimentally happened to be close to the T<sub>g</sub> of ectoine (47°C), but much below that of hydroxyectoine (87°C). One would therefore assume that a slightly increased drying temperature (e.g., 50°C) would enhance the difference between ectoine and hydroxyectoine as desiccation protectants.

The experimental set-up presented here enforced the cytoplasmic conversion of ectoine into hydroxyectoine, and purposely avoided the use of external excipient as an additional protective element, although one would expect a positive influence from additional external protection. Others have improved the desiccation tolerance of whole cells (both *E. coli* and *P. putida*) by

external application of 1 M trehalose or hydroxyectoine, both in combination with 1.5% polyvinyl pyrrolidone (PVP) as a thickening agent (Manzanera et al., 2002, 2004a). For *E. coli*, they obtained survival rates of approximately 60% after vacuum drying with very little difference between trehalose and hydroxyectoine. In the case of *P. putida*, hydroxyectoine (approximately 40% survival) performed even better than trehalose (<20% survival). Hence the positive effect of hydroxyectoine as an external drying excipient (comparable to that of trehalose) appears to be met by a similarly striking effect when accumulated intracellularly. It remains to be shown whether the external addition of glass-formers (i.e., a combination of both strategies) would increase the survival rates even more (de Castro et al., 2000).

### OTHER POTENTIAL FACTORS INVOLVED IN DESICCATION TOLERANCE

Inorganic ions, and in particular polyphosphates, are suspected of also playing a role in stress adaptation, either alone or in combination with other solutes (Seufferheld et al., 2008). All attempts using HPLC and <sup>31</sup>P-NMR techniques to detect any major changes in inorganic cations (Mg, Ca, Na, K) and anions (Cl, NO<sub>3</sub>, SO<sub>4</sub>, PO<sub>4</sub>), including polyphosphates, were however, unsuccessful. It therefore appears that the conversion of ectoine into hydroxyectoine is the major observable change within low-molecular mass compounds when *H. elongata* cells prepare for a pending heat stress/desiccation event. The notable exception is potassium glutamate, the only other organic compound present in large amounts (ratio of approximately 1:4 glutamate/hydroxyectoine). Glutamate accumulation is one of the first physiological responses to salinity-induced dehydration in many microorganisms. Whether its H-donor and H-acceptor groups also contribute to the formation of a stabilizing H-bond network in the glassy state still needs to be clarified. These findings indicate that in particular hydroxyectoine is the crucial low-molecular mass compound for desiccation survival of *H. elongata*. It seems that the presence of at least one such glass-forming compound is essential to enable anhydrobiotic stabilization of biological systems (Tunnacliffe et al., 2001).

We cannot of course exclude that, besides hydroxylation of ectoine, other factors triggered by the combination of high salt and temperature upshock, such as heat-stress proteins, may play a role in increased desiccation protection. We were, however, unable to detect any temperature-induced proteins at the expected molecular mass. Similarly, very small IDPs (2–30 kDa) may also play a role. Among these, late embryogenesis abundant (LEA) proteins, essential for glass formation in dormant seeds, are of particular interest (Tunnacliffe and Wise, 2007). These proteins are highly charged (due to the abundance of glutamate and lysine residues), highly hydrated and form random coils in solution. During desiccation, the formation of secondary structures releases water. Therefore, IDPs probably serve as water-retaining molecules and as scavengers for inorganic ions (ion sequestration). As such they slow down removal of water and prevent adverse effects of increased ionic strength (Tompa et al., 2006). Garay-Arroyo et al. (2000) coined the term "hydrophilins," an expanded definition of IDPs which comprises LEA proteins and dehydrins. Such proteins are characterized by a hydrophilicity index >1 and a glycine content

>6% as a distinguishing feature. In *E. coli*, five proteins conform to this definition (four of which respond to osmotic stress; Garay-Arroyo et al., 2000). One of the *E. coli* hydrophilins (YCIG) has been checked for its stabilization properties with LDH during vacuum drying, and performed well (comparable to 100 mM trehalose) at a molar ratio of 1:1 (hydrophilin:enzyme) up to 98% water loss (Reyes et al., 2005). A bioinformatic study on the *H. elongata* genome revealed only four uncharacterized small proteins which conform to the above criteria of hydrophilins (unpublished); however, a possible involvement in desiccation response has not yet been shown. We therefore have at present no proof that small water-binding/ ion sequestering proteins, as discussed for desiccation tolerant plants, invertebrates, cyanobacteria, fungi, and other anhydrobiotic organisms (Hand et al., 2011), may also participate in the desiccation protection of the moderately halophilic, extremely halotolerant *H. elongata*.

#### AIR-DRYING OF MODEL ENZYME LACTATE DEHYDROGENASE AT HIGH TEMPERATURE

The parameters of the drying process applied to model enzyme LDH were chosen so as to display the consequences of the vastly different Tg of the two ectoines. The drying temperature of 60°C lies above the Tg of ectoine (47°C) but below that of hydroxyectoine (87°C). The concentration of the enzyme (50 µg/mL) was chosen to minimize self-stabilizing effects (Lippert and Galinski, 1992) and excludes the interference of stabilizers from the commercial preparations. Under these conditions, hydroxyectoine performed very well (70% residual activity, as compared to approximately 60% for sucrose and approximately 80% for trehalose) after 2 h drying (Figure 5). These values compare with those previously obtained after freeze-drying in 1 M solution, all approximately 70% (Lippert and Galinski, 1992), and those of a comparative experiment with hydrophilins and trehalose when vacuum-dried down to 2% residual water, approximately 80%, as described by Reyes et al. (2005). The most striking difference here, however, lies in the fact that ectoine had no stabilizing effect whatsoever. This is very likely due to the high drying temperature, well above the Tg of ectoine. The fact that prolonged drying (4 and 6 h) reduced residual activity, and in the case of hydroxyectoine, more severely than in the presence of sucrose and trehalose, may be explained by differences in their water-retaining capacity. As has been shown in vacuum-drying experiments of LDH, the final water content has a great influence on residual activity, which rapidly decreases beyond 97% water loss (Reyes et al., 2005). The water-retaining capacity of hydroxyectoine has so far not been investigated, whereas the remarkable anhydrobiotic properties of trehalose have been studied in depth and accredited to its unusual polymorphism (several transitions between polymorphic crystalline and amorphous states). In particular, the dihydrate crystallites allow diffusion of water in and out of channels (Jones et al., 2006). The crystalline-glassy nanocomposite matrix of trehalose traps residual water molecules and makes it such a good material for the preservation of protein conformation (Jain and Roy, 2009; Sussich et al., 2010).

Without doubt the kinetics of water removal have a strong influence on the formation of the glassy state, and in particular the amount of retained water (Willart et al., 2002; Cesaro,

2006). This becomes also apparent when the results of the air-drying of a droplet of solute are closely investigated. Not only does this depend on the hydrophobicity of the surface, which alters microflow patterns and influences crack formation upon continued drying (Adams et al., 2008), but also on droplet size and the speed of water removal at the air-liquid interface. The evaporation of solvent creates steep solute concentrations and temperature gradients, which in turn yield aggregation of solutes, skin formation, crystallization, and most importantly, a spatial heterogeneity in the desiccated droplet (Ragoonanan and Aksan, 2008). In addition, shearing forces (resulting from crystallization effects) and mechanical stress (from cracking) may also damage biological structures in the dry state (Aksan and Toner, 2004; Jones et al., 2006). All these physical, rheological and chemical factors, when combined, tend to create an almost unpredictable outcome. As the glass-forming properties of hydroxyectoine, although suspected for a long time, have only now been proven, we still need to resolve to what extent speed of drying and residual water content influence its stabilizing effect on biomolecules and whole cells. Nevertheless, we are now able to present a novel glass-forming desiccation protectant, hydroxyectoine, which may even be able to challenge the best-known sugar-type stabilizers.

#### ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the Graduiertenkolleg GRK 1572 “Bionics – Interactions across Boundaries to the Environment”.

#### REFERENCES

- Adams, D. R., Toner, M., and Langer, R. (2008). Microflow and crack formation patterns in drying sessile droplets of liposomes suspended in trehalose solutions. *Langmuir* 24, 7688–7697. doi: 10.1021/la703835w
- Aksan, A., and Toner, M. (2004). Isothermal desiccation and vitrification kinetics of trehalose-dextran solutions. *Langmuir* 20, 5521–5529. doi: 10.1021/la0355186
- Allison, S. D., Chang, B., Randolph, T. W., and Carpenter, J. F. (1999). Hydrogen bonding between sugar and protein is responsible for inhibition of dehydration-induced protein unfolding. *Arch. Biochem. Biophys.* 365, 289–298. doi: 10.1006/abbi.1999.1175
- Arakawa, T., and Timasheff, S. N. (1985). The stabilization of proteins by osmolytes. *Biophys. J.* 47, 411–414. doi: 10.1016/S0006-3495(85)83932-1
- Ballesteros, D., and Walters, C. (2011). Detailed characterization of mechanical properties and molecular mobility within dry seed glasses: relevance to the physiology of dry biological systems. *Plant J.* 68, 607–619. doi: 10.1111/j.1365-313X.2011.04711.x
- Belkin, S., Mehlhorn, R. J., Hideg, K., Hankovsky, O., and Packer, L. (1987). Reduction and destruction rates of nitroxide spin probes. *Arch. Biochem. Biophys.* 256, 232–243. doi: 10.1016/0003-9861(87)90441-3
- Belton, P. S., and Gil, A. M. (1994). IR and Raman spectroscopic studies of the interaction of trehalose with hen egg white lysozyme. *Biopolymers* 34, 957–961. doi: 10.1002/bip.360340713
- Bligh, E. G., and Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* 37, 911–917. doi: 10.1139/o59-099
- Bolen, D. W., and Baskakov, I. V. (2001). The osmophobic effect: natural selection of a thermodynamic force in protein folding. *J. Mol. Biol.* 310, 955–963. doi: 10.1006/jmbi.2001.4819
- Borges, N., Ramos, A., Raven, N. D. H., Sharp, R. J., and Santos, H. (2002). Comparative study of the thermostabilizing properties of mannosylglycerate and other compatible solutes on model enzymes. *Extremophiles* 6, 209–216. doi: 10.1007/s007920100236
- Buitink, J., and Leprince, O. (2004). Glass formation in plant anhydrobiotes: survival in the dry state. *Cryobiology* 48, 215–228. doi: 10.1016/j.cryobiol.2004.02.011

- Bursy, J., Pierik, A. J., Pica, N., and Bremer, E. (2007). Osmotically induced synthesis of the compatible solute hydroxyectoine is mediated by an evolutionarily conserved ectoine hydroxylase. *J. Biol. Chem.* 282, 31147–31155. doi: 10.1074/jbc.M704023200
- Cesáro, A. (2006). Carbohydrates: all dried up. *Nat. Mater.* 5, 593–594. doi: 10.1038/nmat1701
- Clegg, J. S. (2001). Cryptobiosis – a peculiar state of biological organization. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 128, 613–624. doi: 10.1016/S1096-4959(01)00300-1
- Coates, J. (2000). “Interpretation of infrared spectra, a practical approach,” in *Encyclopedia of Analytical Chemistry*, ed. R. A. Meyers (Chichester: John Wiley & Sons Ltd), 10815–10837.
- Cottone, G., Ciccotti, G., and Cordone, L. (2002). Protein–trehalose–water structures in trehalose coated carboxy-myoglobin. *J. Chem. Phys.* 117, 9862. doi: 10.1063/1.1518960
- Craig, I. D., Parker, R., Rigby, N. M., Cairns, P., and Ring, S. G. (2001). Mail-lard reaction kinetics in model preservation systems in the vicinity of the glass transition: experiment and theory. *J. Agric. Food Chem.* 49, 4706–4712. doi: 10.1021/jf0100752
- Crowe, J. H., Carpenter, J. F., and Crowe, L. M. (1998). The role of vitrification in anhydrobiosis. *Annu. Rev. Physiol.* 60, 73–103. doi: 10.1146/annurev.physiol.60.1.73
- de Castro, A. G., Bredholt, H., Strøm, A. R., and Tunnacliffe, A. (2000). Anhydrobiotic engineering of gram-negative bacteria. *Appl. Environ. Microbiol.* 66, 4142–4144. doi: 10.1128/AEM.66.9.4142-4144.2000
- Dötsch, A., Severin, J., Alt, W., Galinski, E. A., and Kreft, J.-U. (2008). A mathematical model for growth and osmoregulation in halophilic bacteria. *Microbiology* 154, 2956–2969. doi: 10.1099/mic.0.2007/012237-0
- Dzuba, S., Golovina, Y., and Tsvetkov, Y. (1993). Echo-induced EPR spectra of spin probes as a method for identification of glassy states in biological objects. *J. Magn. Reson. B* 101, 134–138. doi: 10.1006/jmrb.1993.1022
- Dzuba, S., Kirilina, E. P., Salnikov, E. S., and Kulik, L. V. (2005). Restricted orientational motion of nitroxides in molecular glasses: direct estimation of the motional time scale basing on the comparative study of primary and stimulated electron spin echo decays. *J. Chem. Phys.* 122, 94702. doi: 10.1063/1.1856926
- Francia, F., Dezi, M., Mallardi, A., Palazzo, G., Cordone, L., and Venturoli, G. (2008). Protein–matrix coupling/uncoupling in “dry” systems of photosynthetic reaction center embedded in trehalose/sucrose: the origin of trehalose peculiarity. *J. Am. Chem. Soc.* 130, 10240–10246. doi: 10.1021/ja801801p
- Galinski, E. A., and Herzog, R. M. (1990). The role of trehalose as a substitute for nitrogen-containing compatible solutes (*Ectothiorhodospira halochloris*). *Arch. Microbiol.* 153, 607–613. doi: 10.1007/BF00245273
- Garay-Arroyo, A., Colmenero-Flores, J. M., Garcíarrubio, A., and Covarrubias, A. A. (2000). Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *J. Biol. Chem.* 275, 5668–5674. doi: 10.1074/jbc.275.8.5668
- García-Estepa, R., Argandoña, M., Reina-Bueno, M., Capote, N., Iglesias-Guerra, F., Nieto, J. J., et al. (2006). The ectD gene, which is involved in the synthesis of the compatible solute hydroxyectoine, is essential for thermoprotection of the halophilic bacterium *Chromohalobacter salexigens*. *J. Bacteriol.* 188, 3774–3784. doi: 10.1128/JB.00136-06
- Graf, R., Anzali, S., Buenger, J., Pfluecker, F., and Driller, H. (2008). The multifunctional role of ectoine as a natural cell protectant. *Clin. Dermatol.* 26, 326–333. doi: 10.1016/j.jcidermatol.2008.01.002
- Hand, S. C., Menze, M. A., Toner, M., Boswell, L., and Moore, D. (2011). LEA proteins during water stress: not just for plants anymore. *Annu. Rev. Physiol.* 73, 115–134. doi: 10.1146/annurev-physiol-012110-142203
- Hoekstra, F. A., Golovina, E. A., and Buitink, J. (2001). Mechanisms of plant desiccation tolerance. *Trends Plant Sci.* 6, 431–438. doi: 10.1016/S1360-1385(01)02052-0
- Imamura, K., Sakaura, K., Ohyama, K.-I., Fukushima, A., Imanaka, H., Sakiyama, T., et al. (2006). Temperature scanning FTIR analysis of hydrogen bonding states of various saccharides in amorphous matrixes below and above their glass transition temperatures. *J. Phys. Chem. B* 110, 15094–15099. doi: 10.1021/jp057270
- Jain, N. K., and Roy, I. (2009). Effect of trehalose on protein structure. *Protein Sci.* 18, 24–36. doi: 10.1002/pro.3
- Jones, M. D., Hooton, J. C., Dawson, M. L., Ferrie, A. R., and Price, R. (2006). Dehydration of trehalose dihydrate at low relative humidity and ambient temperature. *Int. J. Pharm.* 313, 87–98. doi: 10.1016/j.ijpharm.2006.01.026
- Kets, E. P. W., IJpelhaar, P. J., Hoekstra, F. A., and Vromans, H. (2004). Citrate increases glass transition temperature of vitrified sucrose preparations. *Cryobiology* 48, 46–54. doi: 10.1016/j.cryobiol.2003.12.002
- Knowles, P. F., Marsh, D., and Rattle, H. W. E. (1976). *Magnetic Resonance of Biomolecules: An Introduction to the Theory and Practice of NMR and ESR in Biological Systems*. London: John Wiley & Sons Ltd.
- Larsen, P. I., Sydnes, L. K., Landfald, B., and Strøm, A. R. (1987). Osmoregulation in *Escherichia coli* by accumulation of organic osmolytes: betaines, glutamic acid, and trehalose. *Arch. Microbiol.* 147, 1–7. doi: 10.1007/BF00492896
- Lippert, K., and Galinski, E. A. (1992). Enzyme stabilization by ectoine-type compatible solutes: protection against heating, freezing and drying. *Appl. Microbiol. Biotechnol.* 37, 61–65. doi: 10.1007/BF00174204
- Louis, P., Trüper, H. G., and Galinski, E. A. (1994). Survival of *Escherichia coli* during drying and storage in the presence of compatible solutes. *Appl. Microbiol. Biotechnol.* 41, 684–688. doi: 10.1007/BF00167285
- Manzanera, M., de Castro, A. G., Tøndervik, A., Rayner-Brandes, M., Strøm, A. R., and Tunnacliffe, A. (2002). Hydroxyectoine is superior to trehalose for anhydrobiotic engineering of *Pseudomonas putida* KT2440. *Appl. Environ. Microbiol.* 68, 4328–4333. doi: 10.1128/AEM.68.9.4328-4333.2002
- Manzanera, M., Vilchez, S., and Tunnacliffe, A. (2004a). High survival and stability rates of *Escherichia coli* dried in hydroxyectoine. *FEMS Microbiol. Lett.* 233, 347–352. doi: 10.1016/j.femsle.2004.03.005
- Manzanera, M., Vilchez, S., and Tunnacliffe, A. (2004b). Plastic encapsulation of stabilized *Escherichia coli* and *Pseudomonas putida*. *Appl. Environ. Microbiol.* 70, 3143–3145. doi: 10.1128/AEM.70.5.3143-3145.2004
- Omayu, A., Ueno, T., and Matsumoto, A. (2008). The role of intermolecular hydrogen bonding on thermal properties of maleimide–isobutene alternating copolymers with polar groups. *Macromol. Chem. Phys.* 209, 1503–1514. doi: 10.1002/macp.200800156
- Potts, M. (2001). Desiccation tolerance: a simple process? *Trends Microbiol.* 9, 553–559. doi: 10.1016/S0966-842X(01)02231-4
- Ragoonianan, V., and Aksan, A. (2008). Heterogeneity in desiccated solutions: implications for biostabilization. *Biophys. J.* 94, 2212–2227. doi: 10.1529/biophysj.107.110684
- Reina-Bueno, M., Argandoña, M., Salvador, M., Rodríguez-Moya, J., Iglesias-Guerra, F., Csonka, L. N., et al. (2012). Role of trehalose in salinity and temperature tolerance in the model halophilic bacterium *Chromohalobacter salexigens*. *PLoS ONE* 7:e33587. doi: 10.1371/journal.pone.0033587
- Reuter, K., Pittelkow, M., Bursy, J., Heine, A., Craan, T., Bremer, E., et al. (2010). Synthesis of 5-hydroxyectoine from ectoine: crystal structure of the non-heme iron(II) and 2-oxoglutarate-dependent dioxygenase EctD. *PLoS ONE* 5:e10647. doi: 10.1371/journal.pone.0010647
- Reyes, J. L., Rodrigo, M.-J., Colmenero-Flores, J. M., Gil, J.-V., Garay-Arroyo, A., Campos, E., et al. (2005). Hydrophilins from distant organisms can protect enzymatic activities from water limitation effects in vitro. *Plant Cell Environ.* 28, 709–718. doi: 10.1111/j.1365-3040.2005.01317.x
- Rodríguez-Moya, J., Argandoña, M., Iglesias-Guerra, F., Nieto, J. J., and Vargas, C. (2013). Temperature- and salinity-decoupled overproduction of hydroxyectoine by *Chromohalobacter salexigens*. *Appl. Environ. Microbiol.* 79, 1018–1023. doi: 10.1128/AEM.02774-12
- Seufferheld, M. J., Alvarez, H. M., and Farias, M. E. (2008). Role of polyphosphates in microbial adaptation to extreme environments. *Appl. Environ. Microbiol.* 74, 5867–5874. doi: 10.1128/AEM.00501-08
- Severin, J., Wohlfarth, A., and Galinski, E. A. (1992). The predominant role of recently discovered tetrahydropyrimidines for the osmoadaptation of halophilic eubacteria. *J. Gen. Microbiol.* 138, 1629–1638. doi: 10.1099/00221287-138-8-1629
- Steiner, T. (2002). The hydrogen bond in the solid state. *Angew. Chem. Int. Ed. Engl.* 41, 49–76. doi: 10.1002/1521-3773(20020104)41:1<48::AID-ANIE48>3.0.CO;2-U
- Street, T. O., Bolen, D. W., and Rose, G. D. (2006). A molecular mechanism for osmolyte-induced protein stability. *Proc. Natl. Acad. Sci. U.S.A.* 103, 13997–14002. doi: 10.1073/pnas.0606236103
- Sussich, F., Skopec, C. E., Brady, J. W., and Cesáro, A. (2010). Water mobility in the dehydration of crystalline trehalose. *Food Chem.* 122, 388–393. doi: 10.1016/j.foodchem.2009.08.014

- Tompa, P., Bánki, P., Bokor, M., Kamasa, P., Kovács, D., Lasanda, G., et al. (2006). Protein-water and protein-buffer interactions in the aqueous solution of an intrinsically unstructured plant dehydrin: NMR intensity and DSC aspects. *Biophys. J.* 91, 2243–2249. doi: 10.1529/biophysj.106.084723
- Tunncliffe, A., de Castro, A. G., and Manzanera, M. (2001). Anhydrobiotic engineering of bacterial and mammalian cells: is intracellular trehalose sufficient? *Cryobiology* 43, 124–132. doi: 10.1006/cryo.2001.2356
- Tunncliffe, A., and Wise, M. J. (2007). The continuing conundrum of the LEA proteins. *Naturwissenschaften* 94, 791–812. doi: 10.1007/s00114-007-0254-y
- Vanderkooi, J. M., Dashnau, J. L., and Zelent, B. (2005). Temperature excursion infrared (TEIR) spectroscopy used to study hydrogen bonding between water and biomolecules. *Biochim. Biophys. Acta* 1749, 214–233. doi: 10.1016/j.bbapap.2005.03.008
- Vargas, C., Argandoña, M., Reina-Bueno, M., Rodríguez-Moya, J., Fernández-Aunión, C., and Nieto, J. J. (2008). Unravelling the adaptation responses to osmotic and temperature stress in *Chromohalobacter salexigens*, a bacterium with broad salinity tolerance. *Saline Syst.* 4, 14. doi: 10.1186/1746-1448-4-14
- Wang, R., Pellerin, C., and Lebel, O. (2009). Role of hydrogen bonding in the formation of glasses by small molecules: a triazine case study. *J. Mater. Chem.* 19, 2747–2753. doi: 10.1039/b820294j
- Widderich, N., Pittelkow, M., Höppner, A., Mulnaes, D., Buckel, W., Gohlke, H., et al. (2014). Molecular dynamics simulations and structure-guided mutagenesis provide insight into the architecture of the catalytic core of the ectoine hydroxylase. *J. Mol. Biol.* 426, 586–600. doi: 10.1016/j.jmb.2013.10.028
- Willart, J. F., De Gusseme, A., Hemon, S., Descamps, M., Leveiller, F., and Rameau, A. (2002). Vitrification and polymorphism of trehalose induced by dehydration of trehalose dihydrate. *J. Phys. Chem. B.* 106, 3365–3370. doi: 10.1021/jp012836+
- Wohlfarth, A., Severin, J., and Galinski, E. A. (1990). The spectrum of compatible solutes in heterotrophic halophilic eubacteria of the family Halomonadaceae. *J. Gen. Microbiol.* 136, 705–712. doi: 10.1099/00221287-136-4-705
- Wolkers, W. F., Oldenhof, H., Alberda, M., and Hoekstra, F. A. (1998). A Fourier transform infrared microspectroscopy study of sugar glasses: application to anhydrobiotic higher plant cells. *Biochim. Biophys. Acta* 1379, 83–96. doi: 10.1016/S0304-4165(97)00085-8
- Wolkers, W. F., Oliver, A. E., Tablin, F., and Crowe, J. H. (2004). A Fourier-transform infrared spectroscopy study of sugar glasses. *Carbohydr. Res.* 339, 1077–1085. doi: 10.1016/j.carres.2004.01.016
- Zhou, Q.-H., Li, M., Yang, P., and Gu, Y. (2013). Effect of hydrogen bonds on structures and glass transition temperatures of maleimide-isobutene alternating copolymers: molecular dynamics simulation study. *Macromol. Theor. Simul.* 22, 107–114. doi: 10.1002/mats.201200057

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 23 January 2014; paper pending published: 07 February 2014; accepted: 21 March 2014; published online: 04 April 2014.

Citation: Tanne C, Golovina EA, Hoekstra FA, Meffert A and Galinski EA (2014) Glass-forming property of hydroxyectoine is the cause of its superior function as a desiccation protectant. *Front. Microbiol.* 5:150. doi: 10.3389/fmicb.2014.00150

This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.

Copyright © 2014 Tanne, Golovina, Hoekstra, Meffert and Galinski. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Glutamine synthetase 2 is not essential for biosynthesis of compatible solutes in *Halobacillus halophilus*

Anna Shyan<sup>1†</sup>, Melanie Thompson<sup>1†</sup>, Saskia Köcher<sup>1</sup>, Michaela Tausendschön<sup>1</sup>, Helena Santos<sup>2</sup>, Inga Hänelt<sup>1</sup> and Volker Müller<sup>1\*</sup>

<sup>1</sup> Molecular Microbiology and Bioenergetics, Institute of Molecular Biosciences, Johann Wolfgang Goethe-University of Frankfurt am Main, Frankfurt am Main, Germany

<sup>2</sup> Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Oeiras, Portugal

**Edited by:**

R. Thane Papke, University of Connecticut, USA

**Reviewed by:**

Aharon Oren, The Hebrew University of Jerusalem, Israel

James A. Coker, University of Maryland–University College, USA

**\*Correspondence:**

Volker Müller, Molecular Microbiology and Bioenergetics, Institute of Molecular Biosciences, Johann Wolfgang Goethe-University of Frankfurt am Main, Max-von-Laue-Straße 9, 60438 Frankfurt am Main, Germany  
e-mail: vmueller@bio.uni-frankfurt.de

<sup>†</sup>Anna Shyan and Melanie Thompson have contributed equally to this work.

*Halobacillus halophilus*, a moderately halophilic bacterium isolated from salt marshes, produces various compatible solutes to cope with osmotic stress. Glutamate and glutamine are dominant compatible solutes at mild salinities. Glutamine synthetase activity in cell suspensions of *Halobacillus halophilus* wild type was shown to be salt dependent and chloride modulated. A possible candidate to catalyze glutamine synthesis is glutamine synthetase A2, whose transcription is stimulated by chloride. To address the role of GlnA2 in the biosynthesis of the osmolytes glutamate and glutamine, a deletion mutant ( $\Delta glnA2$ ) was generated and characterized in detail. We compared the pool of compatible solutes and performed transcriptional analyses of the principal genes controlling the solute production in the wild type strain and the deletion mutant. These measurements did not confirm the hypothesized role of GlnA2 in the osmolyte production. Most likely the presence of another, yet to be identified enzyme has the main contribution in the measured activity in crude extracts and probably determines the total chloride-modulated profile. The role of GlnA2 remains to be elucidated.

**Keywords:** *Halobacillus halophilus*, glutamine synthetase, compatible solutes, osmoregulation, halophile

## INTRODUCTION

Moderately halophilic bacteria are truly fascinating microorganisms that can grow over a wide range of salinities (from 0.5 to 3.0 M NaCl) with identical growth rates demonstrating a high flexibility in coping with salt stress. The molecular basis for this extraordinary capability of bacterial cells to adapt to these huge changes in salinity has been studied in several moderate halophiles such as the Gram negative *Halomonas elongata* (Vreeland et al., 1980; Cánovas et al., 1996, 1998; Ono et al., 1999; Grammann et al., 2002; Kurz et al., 2006) or the Gram positive *Halobacillus halophilus* (for a recent review, see Hänelt and Müller, 2013). The most prominent challenge a moderate halophile faces in its habitat is the loss of water from the cytoplasm at high salinities (Ventosa et al., 1998; Oren, 2008). This is combated by the accumulation of compatible solutes, small molecules that do not interfere with the primary metabolism (Galinski, 1995; Kempf and Bremer, 1998; Roeßler and Müller, 2001b; Roberts, 2005). *Halobacillus halophilus* accumulates glycine betaine, glutamate, glutamine, proline, and ectoine (Roeßler and Müller, 2001a; Müller and Saum, 2005; Saum et al., 2006; Saum and Müller, 2007, 2008a; Burkhardt et al., 2009). Most interestingly, *Halobacillus halophilus* switches its osmolyte strategy according to the salt concentration in the growth medium. Cells grown at moderate salinities (around 1 M NaCl) mainly synthesize glutamate and glutamine while at higher salinities (2.0–3.0 M NaCl) proline is the dominant compatible solute (Saum and Müller, 2007). In addition, *Halobacillus halophilus* not only accumulates compatible solutes but also chloride up to molar concentrations in the cytoplasm

(Roeßler and Müller, 2002). Growth of *Halobacillus halophilus* is strictly dependent on the anion Cl<sup>-</sup> (Roeßler and Müller, 1998). In line with this, various studies (Dohrmann and Müller, 1999; Roeßler and Müller, 2001a, 2002; Sewald et al., 2007; Köcher et al., 2009) unraveled a chloride modulon that mediates sensing of the external salt concentration and transmission of information to enzymes whose activities are modulated by chloride or to genes whose transcription is regulated by the anion (Saum and Müller, 2008b).

The routes for the biosynthesis of compatible solutes and their regulation in *Halobacillus halophilus* were examined in recent years. Biosynthesis of glutamate and glutamine occurs via glutamate dehydrogenase or the GOGAT cycle. The genome of *Halobacillus halophilus* contains two genes potentially encoding glutamate dehydrogenases and one encoding a glutamate synthase (Saum et al., 2012). Though their enzymes are probably involved in osmolyte production, transcriptional analysis did not reveal any effect of salt on their expression (Saum et al., 2006). Two genes potentially encoding glutamine synthetases, *glnA1* and *glnA2*, were identified in the genome of *Halobacillus halophilus*. Expression of *glnA2* but not *glnA1* increased up to 4-fold in cells adapted to high salt and was stimulated by chloride. Furthermore, glutamine synthetase activity increased with increasing salinities in the growth media in a chloride-dependent manner. These observations raised the hypothesis that GlnA2 is involved in the synthesis of the solutes glutamate and glutamine while the not upregulated GlnA1 most likely is part of the nitrogen metabolism (Saum et al., 2006). We decided to follow up the hints

given by these observations and to address the role of GlnA2 in solute biosynthesis in *Halobacillus halophilus* using the recently established genetic system (Köcher et al., 2011). Consequently, the *glnA2* gene was deleted and the phenotype of the resulting mutant was characterized.

## MATERIALS AND METHODS

### ORGANISMS AND CULTIVATION

All strains used in this study are listed in **Table 1**. *Escherichia coli* DH5 $\alpha$  was used as a general cloning strain (Hanahan, 1983) and grown under standard conditions (Ausubel et al., 1992). *Halobacillus halophilus* (DSMZ 2266) was routinely grown in glucose minimal medium (G10 medium) containing 50 mM glucose, 37 mM NH<sub>4</sub>Cl, 36  $\mu$ M FeSO<sub>4</sub>  $\times$  7 H<sub>2</sub>O, 100 mM Tris base, 3 mM K<sub>2</sub>HPO<sub>4</sub>, yeast extract (0.1 g·l<sup>-1</sup>), DSM 141 vitamin solution (1 ml·l<sup>-1</sup>), and DSM 79 artificial seawater (250 ml·l<sup>-1</sup>). The final concentration of NaCl varied depending on the assay conditions (values are given in the text). The pH was adjusted to 7.8 with H<sub>2</sub>SO<sub>4</sub>. *Halobacillus halophilus* was cultivated aerobically on a rotary shaker with 125 rpm at 30°C. Growth was monitored by measuring the optical density of the cultures at 578 nm (OD<sub>578</sub>). For protoplast transformation *Halobacillus halophilus* was grown in MB medium as specified by the manufacturer (Roth, Karlsruhe, Germany). For regeneration of protoplasts and selection

of clones MB3 agar plates containing MB medium (Roth, Karlsruhe, Germany) supplemented with 0.5 M Na-succinate, 0.01% bovine serum albumin (BSA), 0.05% casamino acids (CAA), 0.5% glucose, and 0.8% agar were used. For selection of *Halobacillus halophilus* carrying the chloramphenicol acetyltransferase gene (*cat*) chloramphenicol was added from a sterile stock to a final concentration of 5  $\mu$ g·ml<sup>-1</sup>.

### CONSTRUCTION OF pHH $\Delta$ *glnA2*

Standard methods were used for construction of all plasmids. The primers and plasmids are listed in **Tables 2** and **3**, respectively. To construct a  $\Delta$ *glnA2* DNA fragment, upstream and downstream regions of the gene were amplified using specific primers (*glnA2\_up\_for*, *glnA2\_up\_rev* and *glnA2\_do\_for*, *glnA2\_do\_rev*; **Table 2**). The downstream region of 1028 bp and the upstream region of 1044 bp were then fused together in a Fusion PCR (Dillon and Rosen, 1990). For this purpose, the 3' end of the upstream fragment pro\_up included a 26 bp overhang with homology to the 5' end of the downstream fragment pro\_do. A Fusion PCR with primers *glnA2\_up\_for* and *glnA2\_do\_rev* was performed as followed: one cycle at 94°C for 2 min, 25 cycles at 94°C for 30 s, 58°C for 1 min and 72°C for 2.5 min, followed by one cycle at 72°C for 10 min. The expected fragment  $\Delta$ *glnA2* was purified from the gel using the High Pure PCR Product Purification

**Table 1 | Strains used in this study.**

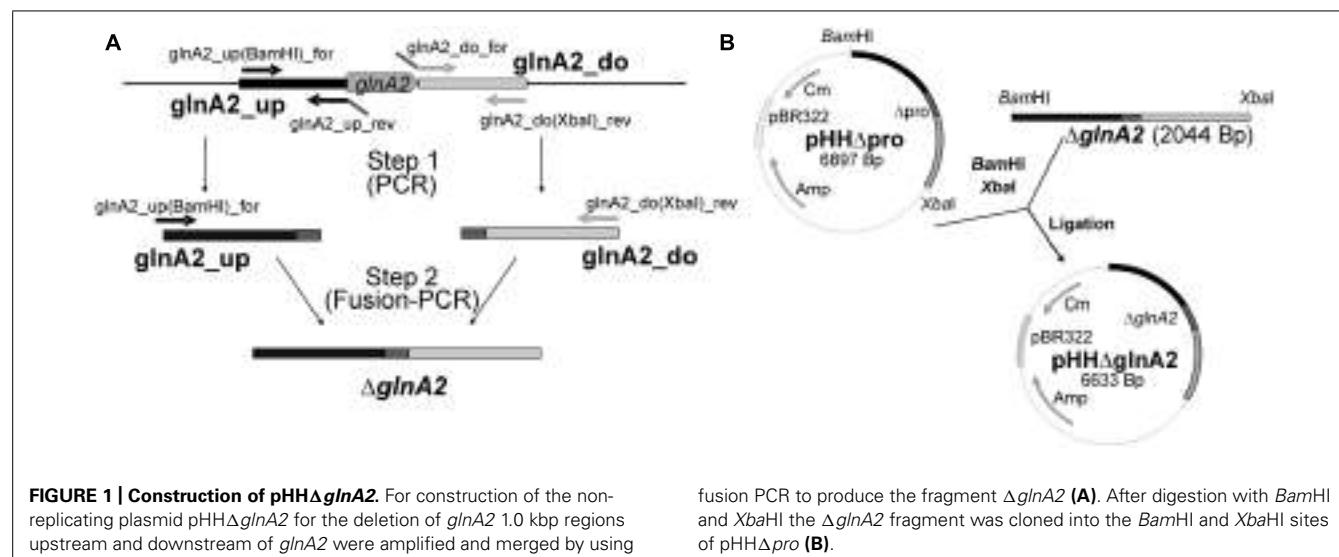
Strain	Genotype	Reference
<i>Halobacillus halophilus</i> (DSMZ 2266)	Wild type	Claus et al. (1983)
<i>Halobacillus halophilus</i> $\Delta$ <i>glnA2</i>	$\Delta$ <i>glnA2</i>	This study
<i>Halobacillus halophilus</i> $\Delta$ <i>pro</i>	$\Delta$ <i>pro</i>	Köcher et al. (2011)
<i>Escherichia coli</i> DH5 $\alpha$ (DSMZ 6893)	F $\Phi$ 80, <i>lacZ</i> $\Delta$ M15, <i>endA1</i> , <i>recA1</i> , <i>hsdR17</i> (r <sub>k</sub> <sup>-</sup> m <sub>k</sub> <sup>+</sup> ), <i>supE44</i> , <i>thi-1</i> , $\lambda^-$ , <i>gyrA96</i> , <i>relA1</i> , $\Delta$ ( <i>lacZYA</i> <sup>-</sup> <i>argF</i> )U196	Hanahan (1983)

**Table 2 | Oligonucleotides used in this study.**

Name	Application	Sequence 5' $\rightarrow$ 3'
<i>glnA2_up(BamHI)_for</i>	Amplification of the region upstream <i>glnA2</i> gene	CGGGATCCCGCATGTGCGTAGTTATGTACGTC
<i>glnA2_up_rev</i>	Amplification of the region upstream <i>glnA2</i> gene	CGATACAATTCTCGTAGGCCTCGGGTATCTAAGTCTGAAACAAGG
<i>glnA2_do_for</i>	Amplification of the region downstream <i>glnA2</i> gene	CGAGGCCCTACGAGGAATTGATCGCACACAGCGCAGGCTTATATTG
<i>glnA2_do(XbaI)_rev</i>	Amplification of the region downstream <i>glnA2</i> gene	GATGACCCAGTCTCTACGGGCTCTAGAGC
<i>glnA2_downB</i>	Verification of pHH $\Delta$ <i>glnA2</i> segregants	ATTGATACATGTTCAACGACGATAAGAG
<i>glnA2_upA</i>	Verification of pHH $\Delta$ <i>glnA2</i> segregants	ATTATGCCGGCTATACAGTGGAGGACCCATCAA
<i>glnA2_rev</i>	Amplification of the <i>glnA2</i> probe (Southern blot)	TGGTTCTGCAGTTTATTGTTCATATATTGTCTG
<i>glnA2_for</i>	Amplification of the <i>glnA2</i> probe (Southern blot)	CGTAGTGCATATGACTTACACGAAAGAACGATTAAGC

**Table 3 | Plasmids used in this study.**

Plasmid	Description	Reference
pHH $\Delta$ <i>pro</i>	pTOPO derivative for generation of deletion mutants	Köcher et al. (2011)
pHH $\Delta$ <i>glnA2</i>	pTOPO derivative for generation of deletion mutants	This study



Kit (Roche, Mannheim, Germany) and cloned into pHH $\Delta$ pro using the restriction enzymes BamHI and XbaI (Figure 1). The construct with the size of 6633 bp was confirmed by sequencing. DNA sequences were retrieved from the genome sequence of *Halobacillus halophilus* (Saum et al., 2012).

#### PROTOPLAST TRANSFORMATION OF *Halobacillus halophilus*

The transformation procedure was performed as described before (Köcher et al., 2011). *Halobacillus halophilus* was grown in 200 ml MB medium in one liter Erlenmeyer flasks to the exponential growth phase (OD<sub>578</sub> 0.6–0.8). After harvesting, the cells were washed in 2 ml SM3B buffer, which consists of 0.5 volume of twofold SMM buffer (1.0 M sucrose, 0.04 M maleate buffer, pH 6.5, and 0.04 M MgCl<sub>2</sub>) and 0.5 volume of twofold concentrated MB medium. The cell pellet was resuspended in 8 ml SM3B buffer containing 0.4 mg·ml<sup>-1</sup> lysozyme. The suspension was incubated at 37°C with mild shaking for 1 h. Protoplast formation was monitored by phase-contrast microscopy. Protoplasts were harvested by centrifugation at 1,000 × g for 30 min. The precipitated protoplasts were resuspended in 1 ml SM3B and centrifuged again. The washed protoplasts were then resuspended in 500 µl SM3B and transferred to a 15-ml tube. 10–30 µg of pHH $\Delta$ glnA2 were gently added to the protoplasts and mixed. 1.5 ml polyethylene glycol (PEG) 4000 (25% w/v, solved in 1× SMM) were added to the protoplast-DNA mixture, the suspensions were mixed and incubated at room temperature for 10 min. Subsequently, 5 ml SM3B were added and the protoplasts sedimented by centrifugation at 1,000 × g for 30 min. The precipitated protoplasts were resuspended in 2 ml SM3B buffer supplemented with 0.01% BSA and 0.05 µg·ml<sup>-1</sup> chloramphenicol. For regeneration, the protoplasts were incubated aerobically at 30°C for 2 h before being plated on regeneration medium (MB3) agar plates containing 5 µg·ml<sup>-1</sup> chloramphenicol. Regeneration plates were incubated at 30°C for 5–6 days. Due to the plasmid's inability to replicate in *Halobacillus halophilus*, selection of the pHH $\Delta$ glnA2 transformants for Cm<sup>R</sup> resulted in glnA2/ $\Delta$ glnA2 merodiploid strains with the plasmid integrated

into the chromosome via single homologous recombination. Resistant clones were isolated and the genotypes were verified by Southern blot analyses. For segregation, the strains were grown under non-selective conditions (MB medium, without chloramphenicol added) over 90 generations to allow resolution of the merodiploid state by recombination between the homologous sequences upstream and downstream of the glnA2 gene. Different dilutions were plated on MB medium and single colonies were isolated and tested for Cm<sup>S</sup>. The genotypes of Cm<sup>S</sup> clones obtained by screening via replica plating were verified by Southern blot analysis.

#### SOUTHERN BLOT ANALYSIS

Sequence-specific probes were generated using “PCR DIG Labeling Mix” (Roche, Mannheim, Germany). Labeling reaction was performed according to the protocol supplied by the manufacturer. Two different probes were used for Southern blot hybridization, one against the flanking region of the mutated locus and one against the locus glnA2, which should be deleted. PCR primers that were used to generate labeled DNA fragments are listed in Table 2. Southern blot hybridization was performed as described by Ausubel et al. (1992) and signals were detected using the luminescent detection substrate CSPD as recommended by the manufacturer (Roche, Mannheim, Germany).

#### DETERMINATION OF COMPATIBLE SOLUTES

Cells were grown in G10 medium to an OD<sub>578</sub> of 0.3–0.6, harvested and freeze-dried. Secondary and tertiary amines were isolated and analyzed by HPLC as described previously (Kunte et al., 1993; Saum et al., 2006). Ectoine was quantified directly (Saum and Müller, 2008a). NMR analyses of compatible solutes with the same cells were carried out as previously described (Saum et al., 2009).

#### SDS-PAGE AND IMMUNOBLOTS

Cells were resuspended in lysis buffer [50 mM Tris-HCl (pH 7.8), 20 mM NaCl, 0.1 mg·ml<sup>-1</sup> lysozyme], incubated at 37°C for

15 min, and disrupted by sonication (four pulses; duty cycle: 50%; output control: 5) using a Branson Sonifier 250 (G. Heinemann Ultraschall- and Labortechnik, Schwäbisch Gmünd, Germany). Cell debris was separated by centrifugation for 10 min at 20,000 g. The supernatant was recovered and the protein content was determined using the assay described by Bradford (1976) with BSA as standard. 20 µg of protein was separated on a denaturing SDS gel (Laemmli, 1970) and blotted on a nitrocellulose membrane as described previously (Kyhse-Andersen, 1984). For detection, blot membranes were incubated in a mixture of 4 ml of solution A (200 ml containing 0.1 M Tris/HCl, pH 6.8, 50 mg luminol), 400 ml of solution B (10 ml dimethylsulfoxide containing 11 mg p-hydroxycoumaric acid) and 1.2 ml H<sub>2</sub>O<sub>2</sub> for 2 min before exposure to WICORex film (Typon Imaging AG, Burgdorf, Switzerland).

#### REAL-TIME PCR ANALYSIS

For real-time PCR analysis, *Halobacillus halophilus* cells were harvested in the early exponential growth phase (OD<sub>578</sub> 0.15–0.3). RNA isolation and qPCR were done according to the standard protocol as described before (Saum et al., 2006). The primers used to amplify *glnA1*, *ectA*, and *proH* were published previously (Saum and Müller, 2007). Data analysis was accomplished applying the 2<sup>-ΔΔCT</sup> method (Livak and Schmittgen, 2001). Real-time PCR analysis was done with three independent physiological parallels to ensure statistical relevance. The open reading frame encoding malate dehydrogenase served as an internal normalizer. The expression of this gene did not change with changing salinities of the medium (Saum et al., 2006).

#### ENZYMATIC ASSAY

Glutamine synthetase activity at whole cells was measured as described previously (Saum et al., 2006). For the preparation of cell suspensions, growth of cell cultures was stopped in the mid-exponential growth phase by adding 0.1% (w/v) cetyltrimethylammonium bromide (CTAB) and incubating the cells for 10 min on a shaker at 30°C. Cells were then harvested and washed in 0.2 volumes 1 M KCl corresponding to the NaCl concentration of 1 M in the growth medium. Finally, the pellet was resuspended in 1.5 M KCl to an OD<sub>578</sub> of 70 and the cell suspension was stored on ice. The standard reaction mixture (4 ml) contained 126 mM imidazole hydrochloride, 17 mM hydroxylamine hydrochloride, 0.26 mM MnCl<sub>2</sub>, 24 mM potassium arsenite, 84 g·ml<sup>-1</sup> CTAB, 0.37 mM Na-ADP, and varying final KCl concentration as indicated in the text. The pH was adjusted to 7.0. The cell suspension (0.5 ml) was preincubated with this mixture for 2 min at 37°C on a shaker, and the reaction was started by the addition of L-glutamine to a final concentration of 25 mM. Samples (0.5 ml) were withdrawn for 20 min in 5-min intervals, the reaction was stopped by the addition of 1 ml stop mix (0.2 M FeCl<sub>3</sub>, 0.15 M trichloroacetic acid, 0.25 M HCl), and the samples were incubated on ice for 30 min. Cells were removed by centrifugation (2 min, 15,000 × g). The formation of γ-glutamyl hydroxamate, which forms a brownish complex together with FeCl<sub>3</sub>, was measured at 540 nm. The purified enzyme was measured in forward reaction. The reaction mixture (800 µl) contained 94 mM Tris, 47 mM hydroxylamine hydrochloride, 56 mM MgCl<sub>2</sub>, 168 mM L-glutamate, and varying final KCl concentrations

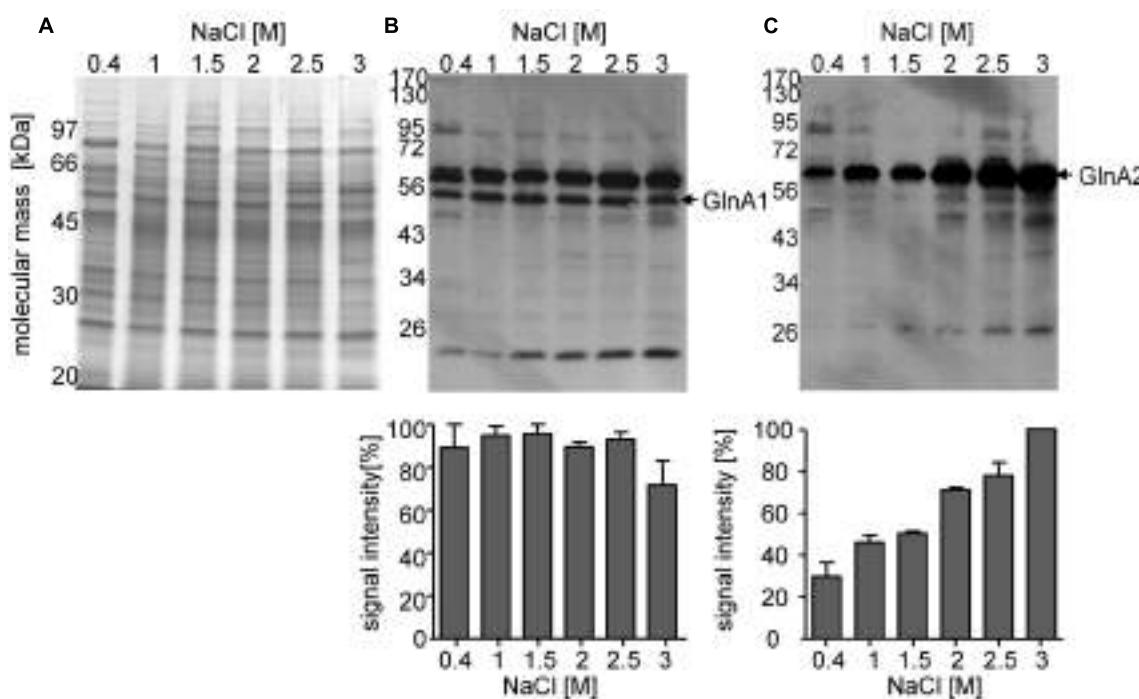
as indicated. The pH was adjusted to 7.0. The purified enzyme was preincubated for 5 min at 37°C with this mixture in a waterbath. The reaction was started by the addition of 10 mM ATP (final concentration). Samples (80 µl) were withdrawn for 10 min in 2.5 min intervals, and the reaction was stopped by the addition of 200 µl stop mix (0.2 M FeCl<sub>3</sub>, 0.15 M trichloroacetic acid, 0.25 M HCl). The samples were incubated on ice for 30 min and precipitated protein removed by centrifugation (2 min, 15,000 × g). Formation of the γ-glutamyl-hydroxamate/FeCl<sub>3</sub>-complex was measured at 540 nm.

#### PROTEIN CONTENT

80 µl perchloric acid (3 M) was added to 200 µl cell suspension. This mixture was incubated for 10 min at 100°C and then cooled on ice. After the addition of 1120 µl H<sub>2</sub>O and 466 µl trichloroacetic acid [25% (w/v)], precipitated protein was separated by centrifugation (15,000 × g for 15 min). The pellet was then resuspended in 400 µl Na<sub>2</sub>HPO<sub>4</sub> buffer (20 mM) and 200 µl NaOH (0.1 M). The protein concentrations of these solutions as well as of cell extracts were determined by the method of Bradford (Bradford, 1976) using BSA as the standard. The protein content of CTAB permeabilized cells was determined before the addition of CTAB.

#### PURIFICATION OF GlnA1 AND GlnA2 FROM *Halobacillus halophilus*

*Halobacillus halophilus* was grown in G10 medium with 1.5 M NaCl to the late-exponential growth phase. Cells were harvested via centrifugation and resuspended in buffer containing 25 mM Tris, 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 10% glycerol, pH 8.0. After incubation with a final concentration of 0.1 mg·ml<sup>-1</sup> lysozyme, cells were disrupted via french press in three passages at 1000 PSIG. Cellular debris was removed via centrifugation at 13,000 × g for 20 min. Cytoplasm and membranes were separated via ultracentrifugation at 48,000 × g for 1.5 h. The cytoplasm was removed carefully and submitted to a PEG-fractionation. First, 7% PEG<sub>8000</sub> (final concentration) was added and the mixture stirred for 20 min. Precipitated protein was removed by centrifugation at 13,000 × g for 20 min. The glutamine synthetases of *Halobacillus halophilus* remained in the supernatant at this step. Protein pellet and supernatant were separated and the GlnA1- and GlnA2-containing supernatant was adjusted to 17% PEG<sub>8000</sub> (final concentration). The mixture was again stirred for 20 min and the now precipitating glutamine synthetases pelleted via centrifugation (13,000 × g for 20 min). The GlnA1- and GlnA2-containing protein pellet was resuspended in fresh buffer (25 mM Tris, 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 10% glycerol, pH 8.0) and the proteins were separated on an anion exchanger ResourceQ by a salt-gradient ranging from 100 to 550 mM NaCl. During this step, it was possible to separate GlnA1 and GlnA2, with GlnA1 eluting at low NaCl concentrations of about 220 mM, while GlnA2 eluted at higher NaCl concentrations of about 520 mM. Fractions that were shown to contain the corresponding glutamine synthetases by Western blot analysis and activity measurements were pooled and separated via Blue Sepharose column. Finally, each glutamine synthetase was separated from contaminating protein via gel filtration using a Superose 6 column.



**FIGURE 2 | Cellular concentration of GlnA1 and GlnA2 in dependence of different salinities in the growth media.** Cells of *Halobacillus halophilus* were cultivated in G10 minimal medium in the presence of varying NaCl concentrations (0.4–3 M). Samples were used to prepare cell-free extracts for SDS-Pages and Western blotting followed by densitometric analysis. All quantifications were carried out in duplicate using two independent cell

cultures. A Coomassie blue-stained SDS-Page containing 20 µg protein per lane (A). Western blot analysis using specific antibodies against GlnA1 (B) and GlnA2 (C) showing the cellular concentration of GlnA1 (B) and GlnA2 (C), respectively, in dependence of salinities in the growth media. Averaged values of two independent Western blot analyses are given in the lower panel. The highest signal intensities were set to 100%.

## RESULTS

### CELLULAR CONCENTRATION OF GlnA2 IS DEPENDENT ON THE SALINITY

In a previous study we found that the relative abundance of *glnA2* transcript increased at elevated salinities (Saum et al., 2006). Here, we analyzed the cellular abundance of GlnA1 and GlnA2 in relation to the external salinity. Cells were grown in G10 minimal media with NaCl concentrations ranging from 0.4 to 3 M, harvested, and disrupted by cell lysis and sonication. Consequently, 20 µg of protein from the cellular extract was separated by SDS-PAGE and the presence of GlnA1 and 2 were detected via Western Blot. With similar quantities of protein loaded per lane (Figure 2A), the amount of GlnA1 appeared to be consistent across all salinities tested (Figure 2B). GlnA2 appeared to increase with the salinity (Figure 2C). The threefold increase of the GlnA2 observed by increasing the salinity from 0.4 to 3 M NaCl is in good agreement with the fourfold increase in expression of *glnA2* gene determined previously (Saum et al., 2006) and supports the hypothesis that GlnA2 is the essential glutamine synthetase in solute production.

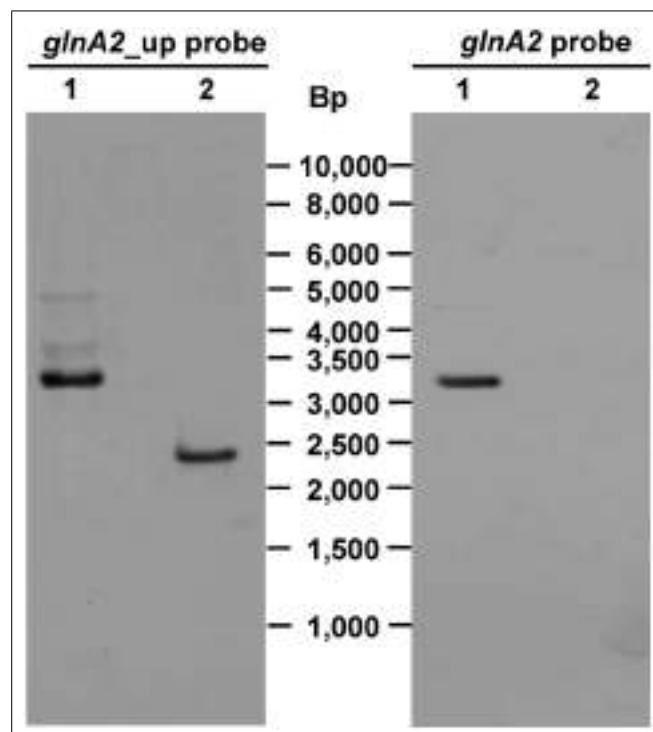
### CONSTRUCTION OF *Halobacillus halophilus* Δ*glnA2*

To delete the chromosomal copy of the *glnA2* gene, fragments of 1028 and 1044 bp upstream and downstream of *glnA2*, respectively, were amplified from the *Halobacillus halophilus* genomic DNA and fused together via fusion PCR to create the DNA fragment Δ*glnA2*. This fragment was then digested and cloned into

pHHΔ*pro* resulting in the generation of pHHΔ*glnA2* (Figure 1). Aliquots of pHHΔ*glnA2* were used to transform *Halobacillus halophilus* by protoplast fusion. The transformation and subsequent segregation procedure was performed as described recently (Köcher et al., 2011). The genotype of the transformants was verified by Southern blot analysis (Figure 3). In the wild type the probe against the *glnA2* upstream region hybridized to a 3343 bp fragment, whereas the size of the complementary fragment was only 2444 bp in the mutant. The size of both fragments met the expectations from the calculated sizes (Figure 3, left panels). Using the probe against *glnA2*, no signal was detected in the Δ*glnA2* mutant, while a signal of the expected size (3343 bp) was detected in the wild type (Figure 3, right panels). These data demonstrate that pHHΔ*glnA2* had segregated from the chromosome and that the *glnA2* gene had been deleted. Similarly, we attempted to generate a Δ*glnA1* as well as a Δ*glnA1*Δ*glnA2* mutant. However, several attempts with varying conditions failed, which hints that the *glnA1* gene is essential.

### DELETION OF *glnA2* HAS NO INFLUENCE ON THE GROWTH OF *Halobacillus halophilus*

To test the effect of deleting *glnA2* on the cells' ability to adapt to various salt concentrations, growth of the mutant was compared with that of the wild type in the presence of 1.0, 2.0, and 3.0 M NaCl dissolved in G10 medium. At all salinities tested, growth of *Halobacillus halophilus* Δ*glnA2* was similar to that of the wild-type

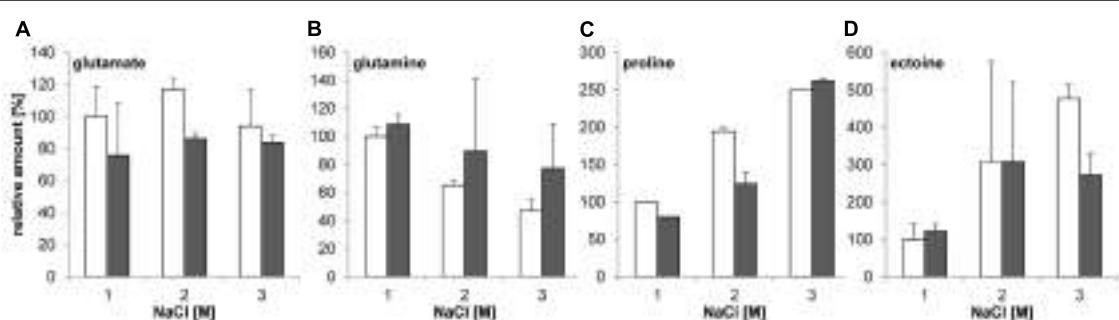


**FIGURE 3 | Genotype of *Halobacillus halophilus*  $\Delta$ glnA2.** *Pst*I/*Bgl*I digested genomic DNA from *Halobacillus halophilus* wild type (lane 1) or *Halobacillus halophilus*  $\Delta$ glnA2 (lane 2) was separated by gel electrophoresis, transferred to a nylon membrane and probed with specific DIG-labeled DNA fragments against one flanking region of the mutated *glnA2* or the *glnA2* gene. Numbers in the middle indicate the migration positions of standard DNA fragments.

strain in that the final optical densities and the growth rates were the same (data not shown). The fact that the loss of *glnA2* did not affect growth at different salt concentrations raised the question whether glutamate and glutamine are still accumulated to the same extent as in the wild type or whether other solutes compensate the deficiency of these osmolytes, especially at intermediate salinities (below 2 M NaCl).

#### THE $\Delta$ glnA2 DELETION HAS NO EFFECT ON THE COMPATIBLE SOLUTE POOL OF *Halobacillus halophilus* AT DIFFERENT SALINITIES

Since no growth phenotype was observed in the  $\Delta$ glnA2 mutant, the next goal was to test if there is any difference in the spectrum and the pool of compatible solutes in *Halobacillus halophilus*  $\Delta$ glnA2 compared to the wild type. Both strains were cultivated in G10 medium with 1.0, 2.0, and 3.0 M NaCl, respectively, to the exponential growth phase, harvested and lyophilized. The amount of glutamate, glutamine, proline, and ectoine was quantified by HPLC analyses as described previously (Saum et al., 2006). The results shown in Figures 4A,B clearly indicate that *Halobacillus halophilus*  $\Delta$ glnA2 is still able to produce glutamine and glutamate comparable to the wild type. Interestingly, production of glutamate and glutamine was not impaired indicating the presence of another enzyme responsible for their synthesis. With increasing salt concentrations proline became the main solute in the wild type, whereas the levels of glutamate and glutamine stayed constant (Saum and Müller, 2007). The same behavior was observed in *Halobacillus halophilus*  $\Delta$ glnA2. Glutamate and glutamine concentrations in the wild type and in the  $\Delta$ glnA2 mutant stayed relatively constant over the whole range of salinities tested (Figures 4A,B). The proline content increased by a factor of 2.5 by elevating the salinity from 1 to 3 M NaCl in the wild type and by a factor of 2.63 in the  $\Delta$ glnA2 mutant at the same conditions (Figure 4C). As can be seen in Figure 4D, the intracellular ectoine concentration in the  $\Delta$ glnA2 mutant was similar to that of the wild type under identical conditions. The cellular level of ectoine increased by a factor of 4.75 in the wild type and by a factor of 2.73 in the  $\Delta$ glnA2 strain upon a salinity shift from 1 to 3 M NaCl. Thus, ectoine as well as proline accumulation were also not affected by the *glnA2* deletion. In addition, we addressed the question if the production of any alternative osmolytes is altered in *Halobacillus halophilus*  $\Delta$ glnA2 compared to the wild type. The wild type can synthesize minor amounts of alanine, N-acetyl lysine, and N-acetyl ornithine (Saum et al., 2009). To answer this question, NMR analysis of the same lyophilized cells used for HPLC analysis was conducted, but as for the most common solutes, no altered production of alternative solutes compared to the wild type was detected (data not shown).



**FIGURE 4 | Relative amounts of solutes in *Halobacillus halophilus* wild type and  $\Delta$ glnA2 mutant.** Cells were cultivated in G10 medium in the presence of 1.0, 2.0, or 3.0 M NaCl and harvested in the exponential growth phase. Compatible solutes were extracted and concentrations of glutamate (A), glutamine (B), proline (C), and ectoine (D) were determined for wild type

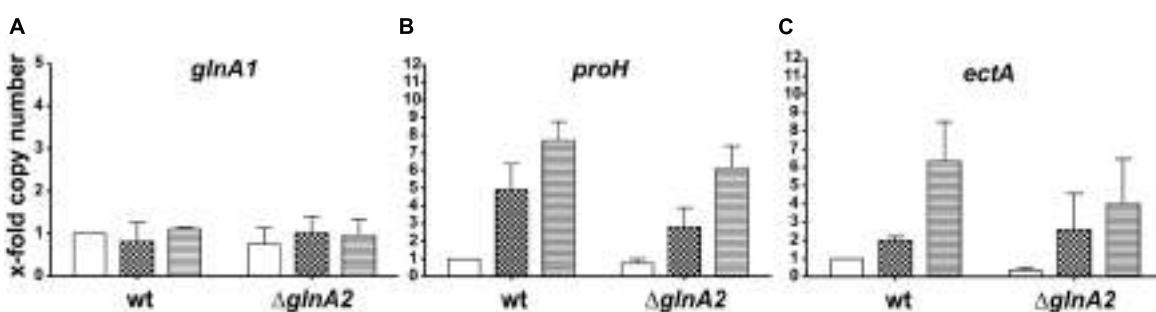
(white) and  $\Delta$ glnA2 mutant (gray) by HPLC. The presented relative quantification of solutes was conducted using the value of "wild type, 1.0 M NaCl" sample as a reference. The values represent the means and the standard deviations of the mean (S.D.s) of at least two physiologically independent parallels.

### ***glnA2* DELETION HAS NO EFFECT ON THE EXPRESSION LEVEL OF *glnA1*, *proH*, AND *ectA* IN *Halobacillus halophilus***

Surprisingly, the glutamine and glutamate concentrations in *Halobacillus halophilus* were not affected by the deletion of *glnA2*. Therefore, we intended to identify the enzymes which keep the biosynthesis of these osmolytes at the same level as in the wild type. In order to determine if there is any impact of the loss of the *glnA2* gene on the expression of other genes responsible for the production of compatible solutes, relative transcription levels of the following genes were measured: *glnA1*, *proH* (the first gene of the *pro* operon), and *ectA* (the first gene of the *ect* operon). To quantify transcript levels of these genes in the  $\Delta glnA2$  mutant, cells were grown in G10 medium in the presence of 1.0, 2.0, and 3.0 M NaCl, respectively, and harvested in the early exponential growth phase. Consequently, RNA was isolated, transcribed into cDNA and subjected to qRT-PCR. The levels of transcription in the wild type at 1 M NaCl were used as standard quantifiers for each gene and were set to 1. Since the most likely candidate to take over the function of glutamate and glutamine production in the  $\Delta glnA2$  mutant is another glutamine synthetase, we expected an increase of the *glnA1* transcription level, especially at moderate salinities. In contrast, like in the wild type, constant *glnA1* expression levels were observed in this experiment (**Figure 5A**) regardless of the salinity in the growth medium. *glnA1* is not upregulated to complement the loss of GlnA2, but probably serves as the only glutamine synthetase crucial for glutamate and glutamine biosynthesis. Also for *ectA*, the first gene of the *ect* operon responsible for ectoine biosynthesis, the relative expression levels at all salinities tested were the same as in the wild type. Salinity-dependent dynamics did not change as a result of the deletion of *glnA2*. Due to an increase of the NaCl concentration from 1 to 3 M, the relative expression of *ectA* increased up to sevenfold in the wild type as well as in the  $\Delta glnA2$  mutant (**Figure 5B**). The relative transcription level of *proH*, the first gene of the *pro* operon, was stimulated in the  $\Delta glnA2$  strain up to eightfold by an increase of the salt concentration as it was observed in the wild type (**Figure 5C**). In conclusion, also a comparison of the relative transcriptional levels of *glnA1*, *proH*, and *ectA* in the  $\Delta glnA2$  mutant and the wild type did not reveal any differences.

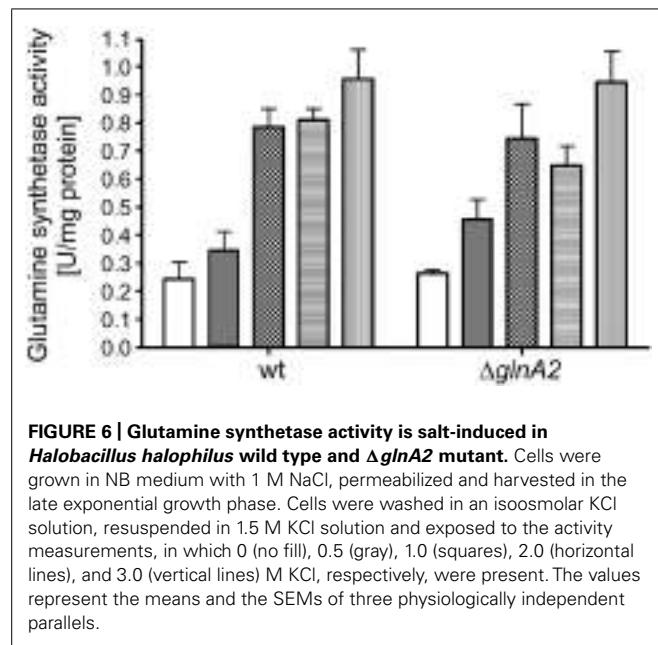
### **GLUTAMINE SYNTHETASE ACTIVITY IN *Halobacillus halophilus* WILD TYPE AND IN THE $\Delta glnA2$ MUTANT IS SALT DEPENDENT BUT NEITHER GlnA2 NOR GlnA1 ARE CHLORIDE-MODULATED**

In previous studies it was suggested that GlnA2 is responsible for the observed salt-dependent biosynthesis of glutamine and glutamate in cells since it was upregulated on a transcriptional level (Saum et al., 2006). However, as pointed out above, the accumulation of glutamate and glutamine was not affected in the  $\Delta glnA2$  mutant at all salt concentrations tested. Consequently, one could hypothesize that possibly GlnA1 and not GlnA2 is the chloride-dependent glutamine synthetase in *Halobacillus halophilus* being involved in both, solute production and the nitrogen metabolism. To test this hypothesis we initially analyzed the glutamine synthetase activity at whole cells of both *Halobacillus halophilus* wild type and the  $\Delta glnA2$  mutant. Cells were cultivated to the late-exponential growth phase in NB medium with 1 M NaCl. Activity measurements at whole cells were performed without salt or in the presence of 0.5, 1.0, 1.5, and 2.0 M KCl in the assay, respectively. In **Figure 6** the glutamine synthetase activities of the  $\Delta glnA2$  mutant and of the wild type are demonstrated. The glutamine synthetase activity increased upon an increase of the KCl concentration in both strains and the activity values in the  $\Delta glnA2$  mutant and in the wild type were very similar at the same conditions. Glutamine synthetase activity of the wild type which was  $0.25 \pm 0.1$  U mg<sup>-1</sup> protein without salt behaved in a salt-dependent manner and was upregulated to  $0.95 \pm 0.1$  U mg<sup>-1</sup> at 2 M KCl. The  $\Delta glnA2$  mutant showed an increase of the glutamine synthetase activity from  $0.25 \pm 0.02$  to  $0.98 \pm 0.15$  U mg<sup>-1</sup> upon the respective salinity increase. Apparently, the salt-stimulated enzyme was still present in the  $\Delta glnA2$  mutant and thus the enzyme of choice was likely to be GlnA1. To further address that question on a molecular level, we decided to purify both GlnA1 and GlnA2 from crude cell extract to finally test the glutamine synthetase activity of the purified enzymes. For this purpose, *Halobacillus halophilus* cells were grown in G10 medium containing 1.5 M NaCl, harvested in the late-exponential growth phase and broken by three French press cycles. Subsequently, the broken cells were fractionated by a series of low speed and high speed centrifugation. The resulting cytoplasm contained both glutamine synthetases (data not shown). By two consecutive PEG precipitation steps,



**FIGURE 5 |** Cellular transcript levels of *glnA1* (A), *proH* (B), and *ectA* (C) in *Halobacillus halophilus* wild type and  $\Delta glnA2$  mutant. Cellular transcript levels were detected in cells grown in G10 medium containing 1.0 (no fill), 2.0 (squares), or 3.0 (horizontal lines) M NaCl and harvested in the

early exponential growth phase. The presented relative quantification of transcript levels was conducted using the value of "wild type, 1.0 M NaCl" sample as a reference. The experiment was repeated in two or three independent parallels to ensure statistical relevance.

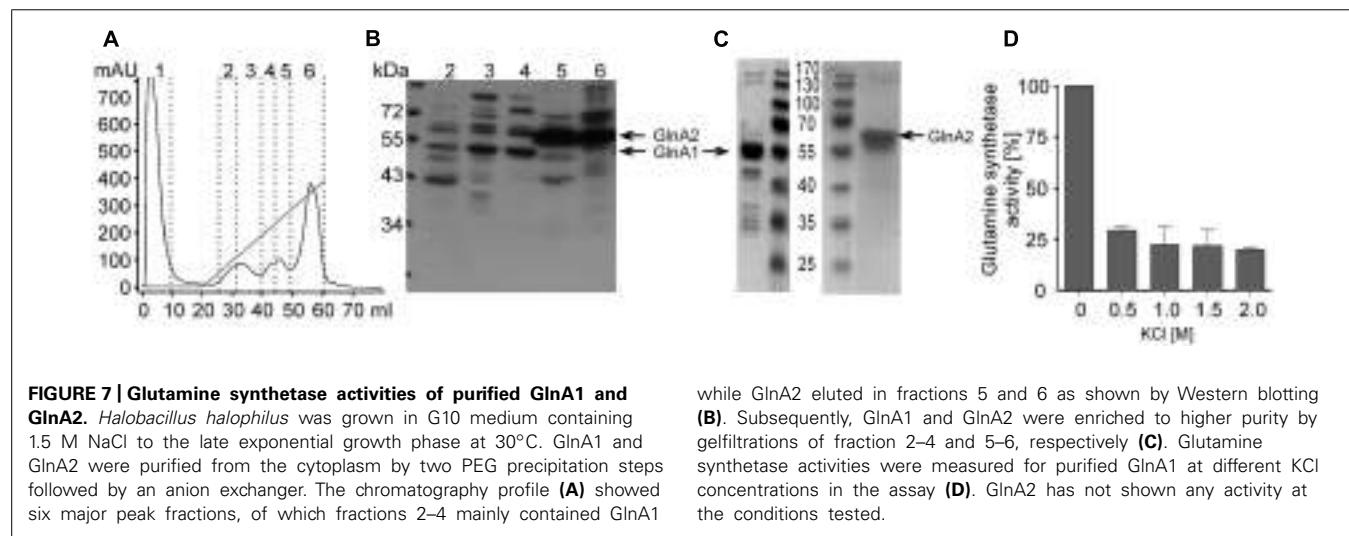


in which at the lower PEG concentration the glutamine synthetases remained in the supernatant while they precipitated at the increased PEG concentration, the enzymes were enriched. Finally, the proteins of the resuspended precipitate were separated *via* an anion exchanger by which also GlnA1 and GlnA2 were separated from each other due to their slightly different isoelectric points of 4.98 and 4.69, respectively (Figures 7A,B). The chromatography profile displayed six major peak fractions (Figure 7A), in which by Western blot analysis we identified GlnA1 mainly in fractions 2 to 4, while GlnA2 eluted later in fractions 5 and 6 (Figure 7B). To further purify GlnA1 and GlnA2, fractions 2–4 and 5–6, respectively, were pooled, concentrated and separated by gel filtration chromatography. By densitometry readings of SDS-PAGEs (Figure 7C), GlnA1 and GlnA2 were shown to be 62 and 75% pure, respectively. The enriched enzymes were analyzed

for glutamine synthetase activity at varying KCl concentrations in the assay ranging from 0 to 2 M. In this assay GlnA2 showed no enzymatic activity at various conditions tested. In contrast, GlnA1 was active with a specific activity of  $3.1 \text{ U mg}^{-1}$ . The highest activity was gained in absence of KCl while a reduction of 70% occurred in presence of 500 mM KCl already. A further increase of the salinity had no effect on the enzyme activity (Figure 7D). In summary, the deletion of *glnA2* had no effect on the glutamine synthetase activity in a whole-cell assay and at the same time, purified GlnA2 showed no activity at the conditions tested. This is in line with the observation that the deletion of *glnA2* has no effect on the glutamine and glutamate pools as well as with the finding that the loss of *glnA2* did not lead to the upregulation of the expression of the second gene encoding a glutamine synthetase, *glnA1*. GlnA1 instead showed the expected activity but as purified enzyme was not found to be salt or chloride dependent. Thus, GlnA1 most likely is the key enzyme for the synthesis of glutamine and glutamate in *Halobacillus halophilus*, but another, yet to be identified enzyme probably is responsible for the chloride-dependent activity as observed in whole-cell measurements.

## DISCUSSION

*Halobacillus halophilus* has two glutamine synthetases of the GSI type (Woods and Reid, 1993): GlnA1 and GlnA2. GlnA1 is encoded by *glnA1*, which is organized in an operon together with *glnR*. *glnR* most likely encodes a regulatory protein. This genetic organization is typical for *Bacillus sp.* and the derived amino acids sequence of *glnA1* is 81% identical to that of the corresponding protein from *Bacillus subtilis* (Saum et al., 2006), which is a classical enzyme involved in nitrogen metabolism. The second enzyme encoded by *glnA2* has a relatively unique amino acid sequence (52% identity to the glutamine synthetase from *B. subtilis*) and the respective gene lies solitary in the genome. Glutamate plays a crucial role in *Halobacillus halophilus* being involved in numerous metabolic and signaling pathways. Apart from serving as a structural unit in proteins, it can be used as single carbon and nitrogen source. It acts as a signaling molecule which triggers the production of compatible solutes proline and ectoine and can



substitute chloride during cell growth (Saum and Müller, 2007). Furthermore, together with glutamine, it is the dominant compatible solute at moderate salt concentrations (below 2 M NaCl). Therefore, two homologs of GS could possibly possess different functions; catalyze formation of glutamate and glutamine for different needs under specific conditions. Glutamine synthetase homologs in other organisms have diverse functions depending on their habitats and particular requirements. GlnA1 from *Mycobacterium tuberculosis* catalyzes the synthesis of L-glutamine, whereas GlnA2 is responsible for D-glutamine production and determines the virulence (Harth et al., 2005). Another glutamine synthetase homolog PA5508 produced in *Pseudomonas aeruginosa* is able to metabolize bulky amines rather than ammonia (Ladner et al., 2012). The nodulon/glutamine synthetase-like protein (NodGS) proteomically identified in *Arabidopsis thaliana* is a fusion protein containing a C-terminal domain of prokaryotic GS. It does not possess glutamine synthetase activity, but plays a role in root morphogenesis and microbial elicitation (Doskocilova et al., 2011).

In this study, the role of GlnA1 and GlnA2 of *Halobacillus halophilus* was studied in detail by using a *glnA2* deletion strain and by analyzing the enzymatic activity of the purified enzymes. The growth phenotype of the  $\Delta glnA2$  mutant, its accumulation of osmolytes, the expression of the genes encoding the key enzymes in osmolyte biosynthesis as well as its GS activity in comparison to the wild type were analyzed. Intriguingly, growth of the mutant at high salinities was not inhibited indicating that GlnA2 is not essential for optimal growth of *Halobacillus halophilus* at salt concentrations up to 3 M. Glutamate and glutamine pools were not altered as a result of the *glnA2* deletion. Since the proline biosynthesis occurs via glutamate, also its accumulation was investigated but no changes were detected. And finally, the possibility of secondary regulatory effects of the *glnA2* deletion on cellular concentrations of ectoine and some minor osmolytes such as N-acetyl- $\beta$ -lysine, N-acetyl-ornithine, and alanine was verified. Ectoine was shown to be a minor solute at high salinities in *Halobacillus halophilus* wild type at the exponential growth phase (Saum and Müller, 2008a). But when the *proHJA* operon was deleted, the intracellular ectoine content increased up to 300% compared to the wild type (Köcher et al., 2011). However, no changes in these compatible solute pools in the  $\Delta glnA2$  mutant compared to the wild type were observed. In accordance with pool sizes of compatible solutes, there was no influence of the *glnA2* deletion on the transcription of *proH* and *ectA*, whose gene products are responsible for the production of proline and ectoine, respectively, at different salinities. Upregulation of *glnA1* was not detected in the  $\Delta glnA2$  mutant compared to the wild type. Taken together, there is actually no direct evidence for the role of GlnA2 in glutamate and glutamine production in *Halobacillus halophilus* in principle. The data obtained in this study clearly demonstrates that GlnA2 is not essential in *Halobacillus halophilus* and the function of the enzyme remains unknown. GlnA1 instead was shown to be active as purified enzyme and it is likely to be the only key enzyme in glutamate and glutamine production. If ever achieved, analysis of the phenotypes of a  $\Delta glnA1$  and a double  $\Delta glnA1\Delta glnA2$  mutant might well clarify this question and is still considered as a next step in the investigation of glutamate and glutamine metabolism

in *Halobacillus halophilus*. Deletion of *glnA* encoding a single glutamine synthetase in *E. coli* and *glnA* together with *gudB* in *B. subtilis* led to glutamine auxotrophy (MacNeil et al., 1982; Florez et al., 2011). A  $\Delta glnA$  mutant of *Rhodopseudomonas capsulata* and a  $\Delta glnA1$  mutant of *Rhodobacter sphaeroides* also showed the Gln<sup>-</sup> phenotype, inability to assimilate ammonium and derepression of nitrogenase in presence of NH<sub>4</sub><sup>+</sup> (Scolnik et al., 1983; Li et al., 2010). Surprisingly, the *glnA* gene from *Rhodopseudomonas capsulata* could complement the Gln<sup>-</sup> phenotype of an *E. coli* *glnA* deletion strain (Scolnik et al., 1983). Taken together, these data corroborate the conservative nature of *glnA* genes reported above and highlight the structural and functional similarities of glutamine synthetases within different classes of bacteria.

The  $\Delta glnA2$  mutant and the wild type as well as the purified GlnA1 and GlnA2 proteins were additionally tested for salinity-dependent glutamine synthetase activity. Stimulation of GS activity in response to a salinity increase was clearly observed in both strains, the  $\Delta glnA2$  mutant and the wild type. Besides, relative activity values in both strains were very similar under all conditions tested leading to the hypothesis that not GlnA2 but GlnA1 is the salt-regulated enzyme. Strikingly, neither purified GlnA1 nor GlnA2 showed a salt-stimulated activity. Therefore, we raised the question whether other chloride-induced enzymes are potentially co-measured by the applied methods. There are two common ways described in literature to measure potential glutamine synthetases: one is determining the formation of  $\gamma$ -glutamylhydroxamate from glutamate and the other the transfer from glutamine (Herzfeld, 1972). Though both methods can be used to analyze glutamine synthetases, they basically test different functions. While the first reports  $\gamma$ -glutamylhydroxamate synthetase activities (GS), the second determines L-glutamine-hydroxylamine glutamyltransferase activities (GT), which are not exclusively catalyzed by glutamine synthetases. The second method, which was used for whole-cell measurements in the present work as well as in previous studies in *Halobacillus halophilus*, thus primarily measures GT activity which is attributed mainly to GTs (Herzfeld, 1972; Herzfeld and Estes, 1973). Also proline-5-kinases, like ProJ involved in proline production in *Halobacillus halophilus*, catalyze GT activity. However, the deletion of *proJ* (Köcher et al., 2011) had no significant effect on GT activity in whole cells (data not shown). Therefore, the proline-5-kinase like both glutamine synthetases does not appear to significantly contribute the most to the chloride-dependent activity profile. The stimulation of the measured activity upon a salinity increase observed in the  $\Delta glnA2$  mutant, in the  $\Delta pro$  mutant as well as in the wild type is likely to be associated to GTs in *Halobacillus halophilus*, which still need to be identified.

## ACKNOWLEDGMENT

Financial support of the project by the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

## REFERENCES

- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., et al. (1992). *Current Protocols in Molecular Biology*. New York: Green Publishing and Wiley-Interscience. doi: 10.1002/0471142727
- Bradford, M. M. (1976). A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye-binding. *Anal. Biochem.* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3

- Burkhardt, J., Sewald, X., Bauer, B., Saum, S. H., and Müller, V. (2009). Synthesis of glycine betaine from choline in the moderate halophilic *Halobacillus halophilus*: co-regulation of two divergent, polycistronic operons. *Environ. Microbiol. Rep.* 1, 38–43. doi: 10.1111/j.1758-2229.2008.00001.x
- Cánovas, D., Vargas, C., Calderon, M. I., Ventosa, A., and Nieto, J. J. (1998). Characterization of the genes for the biosynthesis of the compatible solute ectoine in the moderately halophilic bacterium *Halomonas elongata* DSM 3043. *Syst. Appl. Microbiol.* 21, 487–497. doi: 10.1016/S0723-2020(98)80060-X
- Cánovas, D., Vargas, C., Csonka, L. N., Ventosa, A., and Nieto, J. J. (1996). Osmoprotectants in *Halomonas elongata*: high-affinity betaine transport system and choline-betaine pathway. *J. Bacteriol.* 178, 7221–7226.
- Claus, D., Fahmy, F., Rolf, H. J., and Tosunoglu, N. (1983). *Sporosarcina halophila* sp. nov., an obligate, slightly halophilic bacterium from salt marsh soils. *Syst. Appl. Microbiol.* 4, 496–506. doi: 10.1016/S0723-2020(83)80007-1
- Dillon, P. J., and Rosen, C. A. (1990). A rapid method for the construction of synthetic genes using the polymerase chain reaction. *Biotechniques* 9, 298, 300.
- Dohrmann, A. B., and Müller, V. (1999). Chloride dependence of endospore germination in *Halobacillus halophilus*. *Arch. Microbiol.* 172, 264–267. doi: 10.1007/s002030050769
- Doskocilova, A., Plihal, O., Volc, J., Chumova, J., Kourova, H., Halada, P., et al. (2011). A nodulin/glutamine synthetase-like fusion protein is implicated in the regulation of root morphogenesis and in signalling triggered by flagellin. *Planta* 234, 459–476. doi: 10.1007/s00425-011-1419-7
- Florez, L. A., Gunka, K., Polania, R., Tholen, S., and Stulke, J. (2011). SPABBATS: a pathway-discovery method based on Boolean satisfiability that facilitates the characterization of suppressor mutants. *BMC Syst. Biol.* 5:5. doi: 10.1186/1752-0509-5-5
- Galinski, E. A. (1995). Osmoadaptation in bacteria. *Adv. Microb. Physiol.* 37, 272–328. doi: 10.1016/S0065-2911(08)60148-4
- Grammann, K., Volke, A., and Kunte, H. J. (2002). New type of osmoregulated solute transporter identified in halophilic members of the bacteria domain: TRAP transporter TeaABC mediates uptake of ectoine and hydroxyectoine in *Halomonas elongata* DSM 2581T. *J. Bacteriol.* 184, 3078–3085. doi: 10.1128/JB.184.11.3078-3085.2002
- Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166, 557–580. doi: 10.1016/S0022-2836(83)80284-8
- Hänelt, I., and Müller, V. (2013). Molecular mechanisms of adaptation of the moderately halophilic bacterium *Halobacillus halophilus* to its environment. *Life* 3, 234–243. doi: 10.3390/life3010234
- Harth, G., Maslesa-Galic, S., Tullius, M. V., and Horwitz, M. A. (2005). All four *Mycobacterium tuberculosis* glnA genes encode glutamine synthetase activities but only GlnA1 is abundantly expressed and essential for bacterial homeostasis. *Mol. Microbiol.* 58, 1157–1172. doi: 10.1111/j.1365-2958.2005.04899.x
- Herzfeld, A. (1972). The distribution of glutamate dehydrogenase in rat tissues. *Enzyme* 13, 246–251.
- Herzfeld, A., and Estes, N. A. III. (1973). The distinction between gamma-glutamylhydroxamate synthetase and L-glutamine-hydroxylamine glutamyl-transferase activities in rat tissues. *Biochem. J.* 133, 59–66.
- Kempf, B., and Bremer, E. (1998). Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. *Arch. Microbiol.* 170, 319–330. doi: 10.1007/s002030050649
- Köcher, S., Averhoff, B., and Müller, V. (2011). Development of a genetic system for the moderately halophilic bacterium *Halobacillus halophilus*: generation and characterization of mutants defect in the production of the compatible solute proline. *Environ. Microbiol.* 13, 2122–2131. doi: 10.1111/j.1462-2920.2011.02437.x
- Köcher, S., Breitenbach, J., Müller, V., and Sandmann, G. (2009). Structure, function and biosynthesis of carotenoids in the moderately halophilic bacterium *Halobacillus halophilus*. *Arch. Microbiol.* 191, 95–104. doi: 10.1007/s00203-008-0431-1
- Kunte, H. J., Galinski, E. A., and Trüper, H. G. (1993). A modified FMOC-method for the detection of amino acid-type osmolytes and tetrahydropyrimidines (ectoines). *J. Microbiol. Methods* 17, 129–136. doi: 10.1016/0167-7012(93)90006-4
- Kurz, M., Brunig, A. N., and Galinski, E. A. (2006). NhaD type sodium/proton-antiporter of *Halomonas elongata*: a salt stress response mechanism in marine habitats? *Saline Syst.* 2, 10. doi: 10.1186/1746-1448-2-10
- Kyhse-Andersen, J. (1984). Electrophoresis of multiple gels: a simple apparatus without buffer tank for rapid transfer of proteins from polyacrylamide to nitrocellulose. *J. Biochem. Biophys. Methods* 10, 203–209. doi: 10.1016/0165-022X(84)90040-X
- Ladner, J. E., Atanasova, V., Dolezelova, Z., and Parsons, J. F. (2012). Structure and activity of PA5508, a hexameric glutamine synthetase homologue. *Biochemistry* 51, 10121–10123. doi: 10.1021/bi3014856
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685. doi: 10.1038/227680a0
- Li, X., Liu, T., Wu, Y., Zhao, G., and Zhou, Z. (2010). Derepressive effect of NH4+ on hydrogen production by deleting the glnA1 gene in *Rhodobacter sphaeroides*. *Biotechnol. Bioeng.* 106, 564–572. doi: 10.1002/bit.22722
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-DDCt method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- MacNeil, T., Macneil, D., and Tyler, B. (1982). Fine-structure deletion map and complementation analysis of the glnA-glnL-glnG region in *Escherichia coli*. *J. Bacteriol.* 150, 1302–1313.
- Müller, V., and Saum, S. H. (2005). “The chloride regulon of *Halobacillus halophilus*: a novel regulatory network for salt perception and signal transduction in bacteria,” in *Adaption to Life at High Salt Concentrations in Archaea, Bacteria, and Eukarya*, eds N. Gunde-Cimerman, A. Oren, and A. Plemenitas (Dordrecht: Springer), 303–310.
- Ono, H., Sawada, K., Khunajakr, N., Tao, T., Yamamoto, M., Hiramoto, M., et al. (1999). Characterization of biosynthetic enzymes for ectoine as a compatible solute in a moderately halophilic eubacterium, *Halomonas elongata*. *J. Bacteriol.* 181, 91–99.
- Oren, A. (2008). Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Syst.* 4:2. doi: 10.1186/1746-1448-4-2
- Roberts, M. F. (2005). Organic compatible solutes of halotolerant and halophilic microorganisms. *Saline Syst.* 1:5. doi: 10.1186/1746-1448-1-5
- Roeßler, M., and Müller, V. (1998). Quantitative and physiological analyses of chloride dependence of growth of *Halobacillus halophilus*. *Appl. Environ. Microbiol.* 64, 3813–3817.
- Roeßler, M., and Müller, V. (2001a). Chloride dependence of glycine betaine transport in *Halobacillus halophilus*. *FEBS Lett.* 489, 125–128. doi: 10.1016/S0014-5793(01)02099-3
- Roeßler, M., and Müller, V. (2001b). Osmoadaptation in bacteria and archaea: common principles and differences. *Environ. Microbiol.* 3, 743–754. doi: 10.1046/j.1462-2920.2001.00252.x
- Roeßler, M., and Müller, V. (2002). Chloride, a new environmental signal molecule involved in gene regulation in a moderately halophilic bacterium, *Halobacillus halophilus*. *J. Bacteriol.* 184, 6207–6215. doi: 10.1128/JB.184.22.6207-6215.2002
- Saum, R., Mingote, A., Santos, H., and Müller, V. (2009). A novel limb in the osmoregulatory network of *Methanosarcina mazei* Gö1: Ne-acetyl-β-lysine can be substituted by glutamate and alanine. *Environ. Microbiol.* 11, 1056–1065. doi: 10.1111/j.1462-2920.2008.01826.x
- Saum, S. H., and Müller, V. (2007). Salinity-dependent switching of osmolyte strategies in a moderately halophilic bacterium: glutamate induces proline biosynthesis in *Halobacillus halophilus*. *J. Bacteriol.* 189, 6968–6975. doi: 10.1128/JB.00775-07
- Saum, S. H., and Müller, V. (2008a). Growth phase-dependent switch in osmolyte strategy in a moderate halophile: ectoine is a minor osmolyte but major stationary phase solute in *Halobacillus halophilus*. *Environ. Microbiol.* 10, 716–726. doi: 10.1111/j.1462-2920.2007.01494.x
- Saum, S. H., and Müller, V. (2008b). Regulation of osmoadaptation in the moderate halophile *Halobacillus halophilus*: chloride, glutamate and switching osmolyte strategies. *Saline Syst.* 4, 4. doi: 10.1186/1746-1448-4-4
- Saum, S. H., Sydow, J. F., Palm, P., Pfeiffer, F., Oesterhelt, D., and Müller, V. (2006). Biochemical and molecular characterization of the biosynthesis of glutamine and glutamate, two major compatible solutes in the moderately halophilic bacterium *Halobacillus halophilus*. *J. Bacteriol.* 188, 6808–6815. doi: 10.1128/JB.00781-06
- Saum, S., Pfeiffer, F., Palm, P., Rampp, M., Schuster, S., Müller, V., et al. (2012). Chloride and organic osmolytes: a hybrid strategy to cope with elevated salinities by the moderately halophilic, chloride-dependent bacterium *Halobacillus halophilus*. *Environ. Microbiol.* 15, 1619–1633. doi: 10.1111/j.1462-2920.2012.02770.x

- Scolnik, P. A., Virosco, J., and Haselkorn, R. (1983). The wild-type gene for glutamine synthetase restores ammonia control of nitrogen fixation to Gln- (*glnA*) mutants of *Rhodopseudomonas capsulata*. *J. Bacteriol.* 155, 180–185.
- Sewald, X., Saum, S. H., Palm, P., Pfeiffer, F., Oesterhelt, D., and Müller, V. (2007). Autoinducer-2-producing protein LuxS, a novel salt- and chloride-induced protein in the moderately halophilic bacterium *Halobacillus halophilus*. *Appl. Environ. Microbiol.* 73, 371–379. doi: 10.1128/AEM.01625-06
- Ventosa, A., Nieto, J. J., and Oren, A. (1998). Biology of moderately halophilic aerobic bacteria. *Microbiol. Mol. Biol. Rev.* 62, 504–544.
- Vreeland, D. L., Litchfield, C. D., Martin, E. L., and Elliot, E. (1980). *Halomonas elongata*, a new genus and species of extremely salt-tolerant bacteria. *Int. J. Syst. Bacteriol.* 30, 485–495. doi: 10.1099/00207713-30-2-485
- Woods, D. R., and Reid, S. J. (1993). Recent developments on the regulation and structure of glutamine synthetase enzymes from selected bacterial groups. *FEMS Microbiol. Rev.* 11, 273–284. doi: 10.1111/j.1574-6976.1993.tb00001.x

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 18 February 2014; accepted: 27 March 2014; published online: 14 April 2014.  
Citation: Shiyan A, Thompson M, Köcher S, Tausendschön M, Santos H, Hänel I and Müller V (2014) Glutamine synthetase 2 is not essential for biosynthesis of compatible solutes in *Halobacillus halophilus*. *Front. Microbiol.* 5:168. doi: 10.3389/fmicb.2014.00168

This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.

Copyright © 2014 Shiyan, Thompson, Köcher, Tausendschön, Santos, Hänel and Müller. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# N-glycosylation in *Haloferax volcanii*: adjusting the sweetness

Jerry Eichler\*, Adi Arbib, Chen Cohen-Rosenzweig, Lina Kaminski, Lina Kandiba and Zvia Konrad

Department of Life Sciences, Ben Gurion University of the Negev, Beersheva, Israel

**Edited by:**

Aharon Oren, The Hebrew University of Jerusalem, Israel

**Reviewed by:**

Aharon Oren, The Hebrew University of Jerusalem, Israel

Sonja-Verena Albers, Max Planck Institute for Terrestrial Microbiology, Germany

**\*Correspondence:**

Jerry Eichler, Department of Life Sciences, Ben Gurion University of the Negev, PO Box 653, Beersheva 84105, Israel

e-mail: jeichler@bgu.ac.il

Long believed to be restricted to Eukarya, it is now known that cells of all three domains of life perform N-glycosylation, the covalent attachment of glycans to select target protein asparagine residues. Still, it is only in the last decade that pathways of N-glycosylation in Archaea have been delineated. In the haloarchaeon *Haloferax volcanii*, a series of Agl (archaeal glycosylation) proteins is responsible for the addition of an N-linked pentasaccharide to modified proteins, including the surface (S)-layer glycoprotein, the sole component of the surface layer surrounding the cell. The S-layer glycoprotein N-linked glycosylation profile changes, however, as a function of surrounding salinity. Upon growth at different salt concentrations, the S-layer glycoprotein is either decorated by the N-linked pentasaccharide introduced above or by both this pentasaccharide as well as a tetrasaccharide of distinct composition. Recent efforts have identified Agl5–Agl15 as components of a second *Hfx. volcanii* N-glycosylation pathway responsible for generating the tetrasaccharide attached to S-layer glycoprotein when growth occurs in 1.75 M but not 3.4 M NaCl-containing medium.

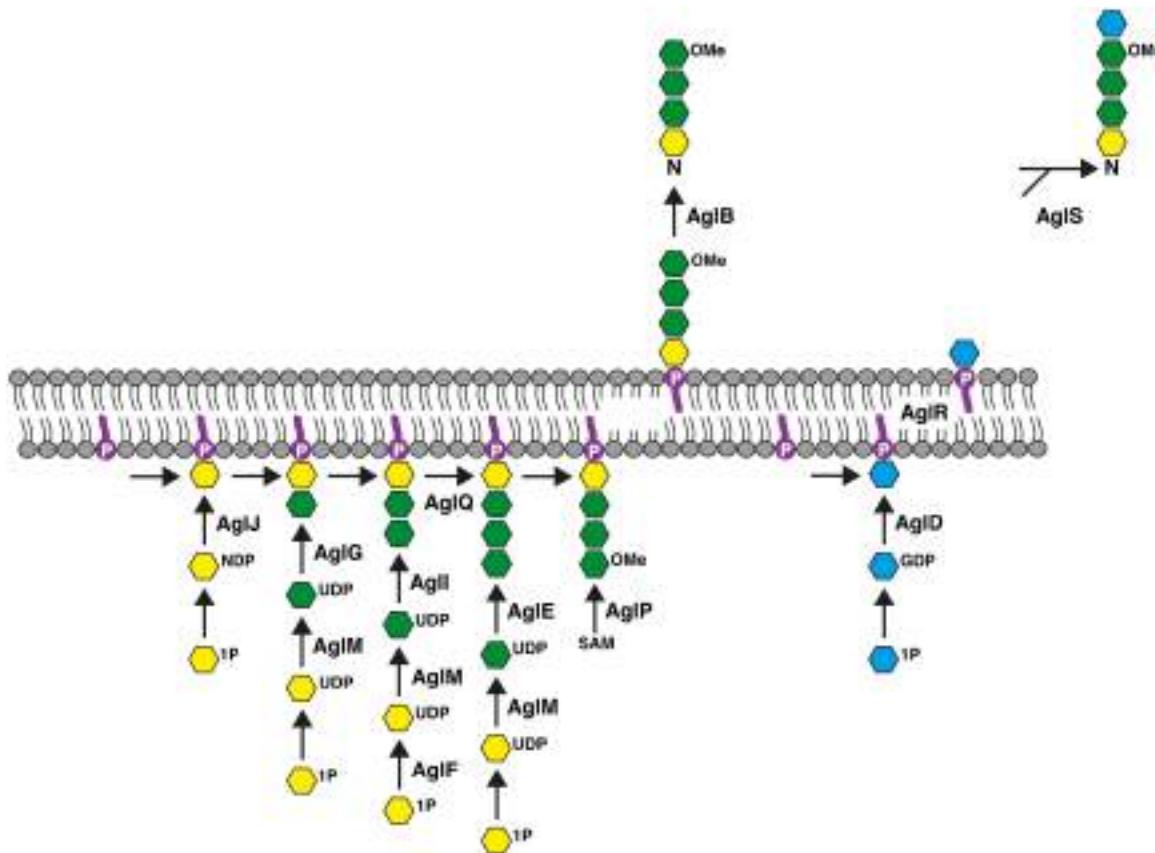
**Keywords:** Archaea, *Haloferax volcanii*, N-glycosylation, post-translational modification, protein glycosylation, S-layer glycoprotein

To cope with the challenges associated with life in a hypersaline environment, halophilic Archaea like *Haloferax volcanii* rely on a variety of strategies manifested at the molecular level. For instance, haloarchaeal proteins present more acidic residues and fewer basic residues than do their non-halophilic homologs (Lanyi, 1974; Fukuchi et al., 2003). While this approach allows haloarchaeal proteins to fold and function properly in the presence of molar concentrations of salt, modified amino acid composition does not allow such proteins to adapt to fluctuations in their surroundings. Instead, post-translational modifications offer proteins a route through which to respond to changing conditions in a transient manner. In the case of the *Hfx. volcanii* S-layer glycoprotein, the sole component of the protein shell surrounding the cell (Sumper et al., 1990), changes in environmental salinity are reflected in a modified N-glycosylation profile (Guan et al., 2012).

The *Hfx. volcanii* S-layer glycoprotein contains seven putative sites of N-glycosylation (Sumper et al., 1990), at least two of which are modified by a pentasaccharide comprising a hexose, two hexuronic acids, a methyl ester of hexuronic acid, and a mannose (Abu-Qarn et al., 2007; Guan et al., 2010; Magidovich et al., 2010). Genetic and biochemical approaches have served to identify a series of Agl (archaeal glycosylation) proteins responsible for the assembly and attachment of this N-linked glycan. AglJ, AglG, AglII, and AglE are glycosyltransferases that sequentially add the first four sugars of the N-linked pentasaccharide to a common dolichol phosphate carrier (Abu-Qarn et al., 2008; Yurist-Doutsch et al., 2008; Guan et al., 2010; Kaminski et al., 2010). Once the lipid-linked tetrasaccharide (and its precursors) has been “flipped” across the plasma membrane, the glycan is delivered to the S-layer glycoprotein Asn-13 and Asn-83

positions by AglB, an oligosaccharyltransferase (Abu-Qarn et al., 2007). The final pentasaccharide sugar, mannose, is added to a distinct dolichol phosphate carrier on the cytoplasmic face of the membrane by the glycosyltransferase AglD, delivered across the membrane to face the cell exterior in a process involving AglR, and then transferred to the Asn-linked tetrasaccharide by AglS (Plavner and Eichler, 2008; Guan et al., 2010; Calo et al., 2011; Cohen-Rosenzweig et al., 2012; Kaminski et al., 2012). In addition, other Agl proteins serve various sugar-processing or other roles that contribute to pentasaccharide assembly, such as AglF, a glucose-1-phosphate uridylyltransferase, AglM, a UDP-glucose dehydrogenase, AglP, a methyltransferase, and AglQ, an isomerase (Yurist-Doutsch et al., 2008; Magidovich et al., 2010; Yurist-Doutsch et al., 2010; Arbib et al., 2013). The most recent version of the Agl pathway is presented in Figure 1.

When first described, *Hfx. volcanii* was reported to grow at NaCl concentrations ranging from 1 M to over 4 M (Mullakhanbhai and Larsen, 1975). In deciphering the *Hfx. volcanii* pathway responsible for the assembly and attachment of the N-linked pentasaccharide decorating S-layer glycoprotein Asn-13 and Asn-83 delineated above, cells were grown in medium containing 3.4 M NaCl. However, when the S-layer glycoprotein was considered in cells grown in medium containing only 1.75 M NaCl, a different N-glycosylation profile was observed. When grown at the lower salinity, S-layer glycoprotein Asn-13 and Asn-83 were still modified by the pentasaccharide described above, although to a lesser extent than when the same cells were grown in 3.4 M NaCl-containing medium. What was more striking was that Asn-498, a position not modified when growth occurs at the higher salinity, was decorated by a novel “low salt” tetrasaccharide comprising a sulfated

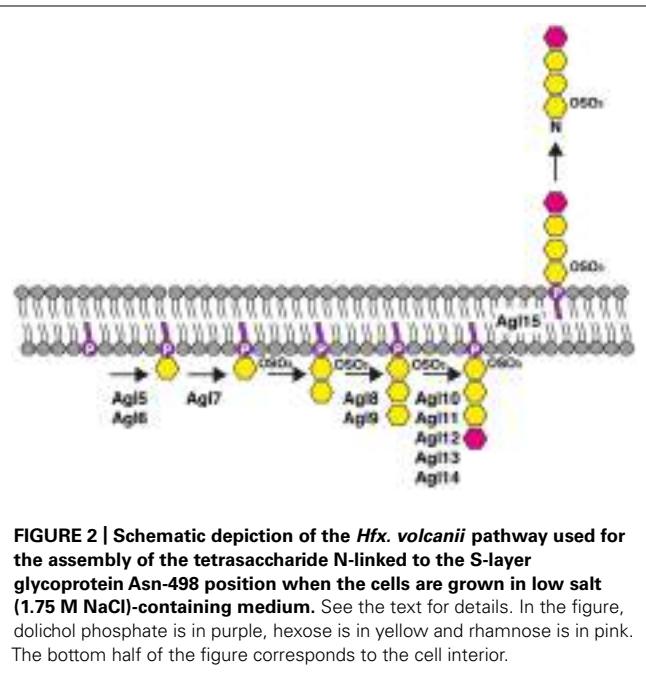


**FIGURE 1 | Schematic depiction of the *Hfx. volcanii* Agl pathway used for the assembly and attachment of the pentasaccharide N-linked to the S-layer glycoprotein Asn-13 and Asn-83 positions.** See the text for details.

In the figure, dolichol phosphate is in purple, hexose is in yellow, hexuronic acid is in green, mannose is in blue and OMe corresponds to a methyl ester group. The bottom half of the figure corresponds to the cell interior.

hexose, two hexoses and a rhamnose when cells were raised at the lower salinity (Guan et al., 2012). Moreover, the same tetrasaccharide was detected on dolichol phosphate in cells raised in 1.75 M NaCl-containing medium. Indeed, dolichol phosphate bearing the low salt tetrasaccharide had been previously reported when *Hfx. volcanii* cells were grown in medium containing only 1.25 M NaCl (Kuntz et al., 1997). Thus, both dolichol phosphate and the S-layer glycoprotein present bound glycans that differ as a function of growth medium salinity. Furthermore, medium salinity also dictated whether N-glycosylation sites in the S-layer glycoprotein were processed and to what extent. The finding that the *Hfx. volcanii* S-layer glycoprotein can be simultaneously modified by two very different N-linked glycans had also been reported to be true in a second haloarchaeon, namely *Halobacterium salinarum*. In work conducted some 30 years ago, it was reported that the S-layer glycoprotein in this organism is also modified by two distinct N-linked glycans (for a review, see Lechner and Wieland, 1989). However, unlike the situation in *Hbt. salinarum*, where relatively little is known of the pathway(s) recruited for N-glycosylation, work in the last decade has provided considerable insight into this post-translational modification in *Hfx. volcanii*, including the recently solved pathway of low salt tetrasaccharide assembly.

By combining gene deletions with mass spectrometric analysis of glycan-charged dolichol phosphate and S-layer glycoprotein-derived peptides, it was demonstrated that the Agl proteins responsible for assembly of the N-linked pentasaccharide are not involved in the biosynthesis of the low salt tetrasaccharide (Kaminski et al., 2013a). As such, efforts were directed at identifying genes encoding proteins comprising a second N-glycosylation pathway. Delineating components of the pathway responsible for generating the low salt tetrasaccharide initially relied on previous work showing that all of the *Hfx. volcanii* genes involved in the assembly of the N-linked pentasaccharide decorating S-layer glycoprotein Asn-13 and Asn-83, with the exception of *aglD*, are found in a single cluster spanning *HVO\_1517* (*aglJ*) to *HVO\_1531* (*aglM*; Yurist-Doutsch and Eichler, 2009; Yurist-Doutsch et al., 2010). As such, the *Hfx. volcanii* genome sequence (Hartman et al., 2010) was scanned for clustered open reading frames (ORFs) annotated as serving some glycosylation-related roles. Those ORFs spanning the region from *HVO\_2046* to *HVO\_2061* represent one such cluster. The involvement of the products of *HVO\_2046* to *HVO\_2061* in the biogenesis of the low salt tetrasaccharide was subsequently confirmed in a series of experiments involving gene deletions combined with mass spectrometry-based examination of dolichol phosphate and the S-layer glycoprotein. Given



their roles in N-glycosylation, these proteins were re-annotated as Agl5–Agl15 (Kaminski et al., 2013a).

Based on the effects of *agl5-agl15* deletion on dolichol phosphate and S-layer glycoprotein Asn-498 glycosylation, together with the results of a bioinformatics-based examination of the encoded proteins, a model of the pathway responsible for low salt tetrasaccharide biogenesis has been proposed (Kaminski et al., 2013a; Figure 2). In this working model, Agl5 and Agl6 are implicated in adding the linking hexose to dolichol phosphate, while Agl7 contributes to the sulfation of this lipid-linked sugar. That population of dolichol phosphate-hexose seen in cells lacking either Agl5 or Agl6 likely corresponds to the lipid carrier charged with the first sugar of the pentasaccharide transferred to Asn-13 and Asn-83, a process that also occurs in low salt conditions (Guan et al., 2012). Furthermore, because cells lacking Agl7 contain dolichol phosphate charged with a non-sulfated version of the low salt tetrasaccharide, whereas no Asn-498-fused low salt tetrasaccharide (or its di- or tri-saccharide precursors) were detected in such cells, sulfation of the dolichol phosphate-bound hexose may be required for translocation of dolichol phosphate charged with a more elaborate low salt tetrasaccharide precursor or the complete glycan itself across the plasma membrane. Clearly, additional studies are needed to precisely define the actions of Agl5, Agl6, and Agl7, as well as their order of action. While the enzyme responsible for adding the second sugar of the low salt tetrasaccharide, a hexose, to sulfated hexose-charged dolichol phosphate remains to be identified, it appears that Agl8 and Agl9 contribute to the addition of the next sugar, a hexose, to disaccharide-charged dolichol phosphate. Agl10–14 are involved in the subsequent appearance of a rhamnose to the dolichol phosphate-bound trisaccharide, yielding the complete low salt tetrasaccharide on the lipid carrier. In cells lacking Agl15, the intact low salt tetrasaccharide is assembled on dolichol

phosphate but no such glycan is detected on S-layer glycoprotein Asn-498. This observation is consistent with Agl15 serving as a flippase, mediating the translocation of low salt tetrasaccharide-charged dolichol phosphate (and likely dolichol phosphate bearing tetrasaccharide precursors) across the membrane. Indeed, Agl15 shares substantial identity (28%) and similarity (51%) with AglR, recently proposed to serve as or to assist the DolP-mannose flippase recruited in the pathway used for pentasaccharide-based glycosylation of S-layer glycoprotein Asn-13 and Asn-83 (Kaminski et al., 2012). Finally, the absence of Agl5–Agl15 did not compromise Asn-13 and Asn-83 glycosylation, arguing that these proteins are dedicated to the assembly of the low salt tetrasaccharide (Kaminski et al., 2013a).

Although *Hfx. volcanii* seemingly relies on two different pathways for the assembly of the two N-linked glycans decorating the S-layer glycoprotein, only one oligosaccharyltransferase, namely the enzyme responsible for the transfer of the lipid-linked glycan to a target protein, has been identified in this organism. In *Hfx. volcanii*, AglB is the only homolog of the eukaryal oligosaccharyltransferase catalytic subunit, Stt3, or its bacterial counterpart, PglB (Magidovich and Eichler, 2009; Kaminski et al., 2013b). As such, the absence of AglB prevented the glycosylation of S-layer glycoprotein Asn-13 and Asn-83 by the pentasaccharide normally attached at these positions (Abu-Qarn et al., 2007). On the other hand, *aglB* deletion had no effect on the appearance of the low salt tetrasaccharide added to the Asn-498 position (Kaminski et al., 2013a). Thus, a currently unidentified and novel oligosaccharyltransferase is seemingly involved in the delivery of the low salt tetrasaccharide (and its precursors) from dolichol phosphate to S-layer glycoprotein Asn-498. The same may be the case in *Hbt. salinarum*, where one of the two N-linked glycans decorating the S-layer glycoprotein in this species is transferred from a dolichol phosphate carrier while the second glycan is delivered from a dolichol pyrophosphate carrier (for review, see Lechner and Wieland, 1989; Cohen-Rosenzweig et al., 2013).

Presently, the reason why the *Hfx. volcanii* S-layer glycoprotein (and the *Hbt. salinarum* S-layer glycoprotein, for that matter) can be modified by two distinct N-linked glycans as a function of environmental salinity can only be supposed. Likewise, the reason why Asn-498 is only modified when cells are grown at a given salt concentration is not clear. One could envisage a salt concentration-related conformational change in the S-layer glycoprotein leading to the exposure of Asn-498 to the low salt tetrasaccharide N-glycosylation machinery only at the lower salinity. Alternatively, modification of Asn-498 could be a question of the availability of the low salt tetrasaccharide since only minute levels of this glycan are bound to dolichol phosphate in cells grown in high salt conditions. Another consideration that requires further study concerns how cells lacking different components of the N-linked pentasaccharide biosynthetic pathway are able to decorate Asn-498 with the low salt tetrasaccharide at elevated salinity (Kaminski et al., 2013a). Finally, as studies on species other than *Hfx. volcanii* begin to provide novel insight into archaeal N-glycosylation, it will be important to determine whether environmental concerns apart from salinity also modulate such protein modification.

## AUTHOR CONTRIBUTIONS

All authors made substantial contributions to the acquisition, analysis, and interpretation of data described in this report. All authors critically reviewed the report and approved the final version. All authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## ACKNOWLEDGMENTS

Research in the Eichler laboratory is supported by the Israel Science Foundation (grant 8/11) and the US Army Research Office (W911NF-11-1-520).

## REFERENCES

- Abu-Qarn, M., Giordano, A., Battaglia, F., Trauner, A., Hitchen, P. G., Morris, H. R., et al. (2008). Identification of AglE, a second glycosyltransferase involved in N-glycosylation of the *Haloferax volcanii* S-layer glycoprotein. *J. Bacteriol.* 190, 3140–3146. doi: 10.1128/JB.00056-08
- Abu-Qarn, M., Yurist-Doutsch, S., Giordano, A., Trauner, A., Morris, H. R., Hitchen, P., et al. (2007). *Haloferax volcanii* AglB and AglD are involved in N-glycosylation of the S-layer glycoprotein and proper assembly of the surface layer. *J. Mol. Biol.* 374, 1224–1236. doi: 10.1016/j.jmb.2007.10.042
- Arbib, A., Yurist-Doutsch, S., Guan, Z., and Eichler, J. (2013). AglQ is a novel component of the *Haloferax volcanii* N-glycosylation pathway. *PLoS ONE* 8:e81782. doi: 10.1371/journal.pone.0081782
- Calo, D., Guan, Z., Naparstek, S., and Eichler, J. (2011). Different routes to the same ending: Comparing the N-glycosylation processes of *Haloferax volcanii* and *Halorubra marismortui*, two halophilic archaea from the Dead Sea. *Mol. Microbiol.* 81, 1166–1177. doi: 10.1111/j.1365-2958.2011.07781.x
- Cohen-Rosenzweig, C., Yurist-Doutsch, S., and Eichler, J. (2012). AglS, a novel component of the *Haloferax volcanii* N-glycosylation pathway, is a dolichol phosphate-mannose mannosyltransferase. *J. Bacteriol.* 194, 6909–6916. doi: 10.1128/JB.01716-12
- Cohen-Rosenzweig, C., Guan, Z., Shaanan, B., and Eichler, J. (2013). Substrate promiscuity: AglB, the archaeal oligosaccharyltransferase, can process a variety of lipid-linked glycans. *Appl. Environ. Microbiol.* doi: 10.1128/AEM.03191-3113 [Epub ahead of print].
- Fukuchi, S., Yoshimune, K., Wakayama, M., Moriguchi, M., and Nishikawa, K. (2003). Unique amino acid composition of proteins in halophilic bacteria. *J. Mol. Biol.* 327, 347–357. doi: 10.1016/S0022-2836(03)00150-5
- Guan, Z., Naparstek, S., Calo, D., and Eichler, J. (2012). Protein glycosylation as an adaptive response in Archaea: growth at different salt concentrations leads to alterations in *Haloferax volcanii* S-layer glycoprotein N-glycosylation. *Environ. Microbiol.* 14, 743–753. doi: 10.1111/j.1462-2920.2011.02625.x
- Guan, Z., Naparstek, S., Kaminski, L., Konrad, Z., and Eichler, J. (2010). Distinct glycan-charged phosphodolichol carriers are required for the assembly of the pentasaccharide N-linked to the *Haloferax volcanii* S-layer glycoprotein. *Mol. Microbiol.* 78, 1294–1303. doi: 10.1111/j.1365-2958.2010.07405.x
- Hartman, A. L., Norais, C., Badger, J. H., Delmas, S., Haldenby, S., Madupu, R., et al. (2010). The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS ONE* 5:e9605. doi: 10.1371/journal.pone.0081782
- Kaminski, L., Abu-Qarn, M., Guan, Z., Naparstek, S., Ventura, V. V., Raetz, C. R., et al. (2010). AglJ adds the first sugar of the N-linked pentasaccharide decorating the *Haloferax volcanii* S-layer glycoprotein. *J. Bacteriol.* 192, 5572–5579. doi: 10.1128/JB.00705-10
- Kaminski, L., Guan, Z., Abu-Qarn, M., Konrad, Z., and Eichler, J. (2012). AglR is required for addition of the final mannose residue of the N-linked glycan decorating the *Haloferax volcanii* S-layer glycoprotein. *Biochim. Biophys. Acta* 1820, 1664–1670. doi: 10.1016/j.bbagen.2012.06.014
- Kaminski, L., Guan, Z., Yurist-Doutsch, S., and Eichler, J. (2013a). Two distinct N-glycosylation pathways process the *Haloferax volcanii* S-layer glycoprotein upon changes in environmental salinity. *MBio* 4, e00716–13. doi: 10.1128/mBio.00716-13
- Kaminski, L., Lurie-Weinberger, M. N., Allers, T., Gophna, U., and Eichler, J. (2013b). Phylogenetic- and genome-derived insight into the evolutionary history of N-glycosylation in Archaea. *Mol. Phylogenet. Evol.* 68, 327–339. doi: 10.1016/j.ympev.2013.03.024
- Kuntz, C., Sonnenbichler, J., Sonnenbichler, I., Sumper, M., and Zeitzer, R. (1997). Isolation and characterization of dolichol-linked oligosaccharides from *Haloferax volcanii*. *Glycobiology* 7, 897–904. doi: 10.1093/glycob/7.7.897
- Lanyi, J. K. (1974). Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriol. Rev.* 38, 272–290.
- Lechner, J., and Wieland, F. (1989). Structure and biosynthesis of prokaryotic glycoproteins. *Annu. Rev. Biochem.* 58, 173–194. doi: 10.1146/annurev.bi.58.070189.001133
- Magidovich, H., and Eichler, J. (2009). Glycosyltransferases and oligosaccharyltransferases in Archaea: putative components of the N-glycosylation pathway in the third domain of life. *FEMS Microbiol. Lett.* 300, 122–130. doi: 10.1111/j.1574-6968.2009.01775.x
- Magidovich, H., Yurist-Doutsch, S., Konrad, Z., Ventura, V. V., Dell, A., Hitchen, P. G. (2010). AglP is a S-adenosyl-L-methionine-dependent methyltransferase that participates in the N-glycosylation pathway of *Haloferax volcanii*. *Mol. Microbiol.* 76, 190–199. doi: 10.1111/j.1365-2958.2010.07090.x
- Mullakhanbhui, M. F., and Larsen, H. (1975). *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Arch. Microbiol.* 104, 207–214. doi: 10.1007/BF00447326
- Plavner, N., and Eichler, J. (2008). Defining the topology of the N-glycosylation pathway in the halophilic archaea *Haloferax volcanii*. *J. Bacteriol.* 190, 8045–8052. doi: 10.1128/JB.01200-08
- Sumper, M., Berg, E., Mengele, R., and Strobel, I. (1990). Primary structure and glycosylation of the S-layer protein of *Haloferax volcanii*. *J. Bacteriol.* 172, 7111–7118.
- Yurist-Doutsch, S., Abu-Qarn, M., Battaglia, F., Morris, H. R., Hitchen, P. G., Dell, A., et al. (2008). AglF, aglG and aglI, novel members of a gene island involved in the N-glycosylation of the *Haloferax volcanii* S-layer glycoprotein. *Mol. Microbiol.* 69, 1234–1245. doi: 10.1111/j.1365-2958.2008.06352.x
- Yurist-Doutsch, S., and Eichler, J. (2009). Manual annotation, transcriptional analysis, and protein expression studies reveal novel genes in the agl cluster responsible for N glycosylation in the halophilic archaeon *Haloferax volcanii*. *J. Bacteriol.* 191, 3068–3075. doi: 10.1128/JB.01838-08
- Yurist-Doutsch, S., Magidovich, H., Ventura, V. V., Hitchen, P. G., Dell, A., and Eichler, J. (2010). N-glycosylation in Archaea: on the coordinated actions of *Haloferax volcanii* AglF and AglM. *Mol. Microbiol.* 75, 1047–1058. doi: 10.1111/j.1365-2958.2009.07045.x

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Received: 23 November 2013; paper pending published: 06 December 2013; accepted: 06 December 2013; published online: 24 December 2013.*

*Citation: Eichler J, Arbib A, Cohen-Rosenzweig C, Kaminski L, Kandiba L and Konrad Z (2013) N-glycosylation in *Haloferax volcanii*: adjusting the sweetness. Front. Microbiol. 4:403. doi: 10.3389/fmicb.2013.00403*

*This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology.*

*Copyright © 2013 Eichler, Arbib, Cohen-Rosenzweig, Kaminski, Kandiba and Konrad. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

## ADVANTAGES OF PUBLISHING IN FRONTIERS



### FAST PUBLICATION

Average 90 days  
from submission  
to publication



### COLLABORATIVE PEER-REVIEW

Designed to be rigorous –  
yet also collaborative, fair and  
constructive



### RESEARCH NETWORK

Our network  
increases readership  
for your article



### OPEN ACCESS

Articles are free to read,  
for greatest visibility



### TRANSPARENT

Editors and reviewers  
acknowledged by name  
on published articles



### GLOBAL SPREAD

Six million monthly  
page views worldwide



### COPYRIGHT TO AUTHORS

No limit to  
article distribution  
and re-use



### IMPACT METRICS

Advanced metrics  
track your  
article's impact



### SUPPORT

By our Swiss-based  
editorial team