Mar Biotechnol (2013) 15:63–72
DOI 10.1007/s10126-012-9455-2

ORIGINAL ARTICLE

Genomic Mining for Novel $FADH_2$-Dependent Halogenases in Marine Sponge-Associated Microbial Consortia

Kristina Bayer · Matthias Scheuermayer · Lars Fieseler · Ute Hentschel

Received: 24 November 2011 / Accepted: 15 April 2012 / Published online: 6 May 2012

Abstract Many marine sponges (Porifera) are known to contain large amounts of phylogenetically diverse microorganisms. Sponges are also known for their large arsenal of natural products, many of which are halogenated. In this study, 36 different $FADH_2$-dependent halogenase gene fragments were amplified from various Caribbean and Mediterranean sponges using newly designed degenerate PCR primers. Four unique halogenase-positive fosmid clones, all containing the highly conserved amino acid motif "GxGxxG", were identified in the microbial metagenome of *Aplysina aerophoba*. Sequence analysis of one halogenase-bearing fosmid revealed notably two open reading frames with high homologies to efflux and multidrug resistance proteins. Single cell genomic analysis allowed for a taxonomic assignment of the halogenase genes to specific symbiotic lineages. Specifically, the halogenase cluster S1 is predicted to be produced by a deltaproteobacterial symbiont and halogenase cluster S2 by a poribacterial sponge symbiont. An additional halogenase gene is possibly produced by an actinobacterial symbiont of marine sponges. The identification of three novel, phylogenetically, and possibly also functionally distinct halogenase gene clusters indicates that the microbial consortia of sponges are a valuable resource for novel enzymes involved in halogenation reactions.

Keywords $FADH_2$-dependent halogenases · Metagenomics · Single cell genomics · Sponges · Porifera · Microbial consortia · Microbial symbionts

Kristina Bayer and Matthias Scheuermayer contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s10126-012-9455-2) contains supplementary material, which is available to authorized users.

K. Bayer · U. Hentschel (✉)
Julius-von-Sachs Institute for Biological Sciences,
University of Wuerzburg,
Julius-von-Sachs Platz 3,
D-97082 Wuerzburg, Germany
e-mail: ute.hentschel@uni-wuerzburg.de

M. Scheuermayer
Center for Infectious Diseases, University of Wuerzburg,
Josef-Schneider-Straße 2,
D-97080 Wuerzburg, Germany

L. Fieseler
Institute for Food and Beverage Innovation,
Zurich University of Applied Sciences,
Campus Reidbach, P.O. Box, CH-8820, Wädenswil, Switzerland

Introduction

Sponges (phylum Porifera) are an extraordinarily rich source for bioactive metabolites. Several thousand compounds have been isolated from these marine invertebrates which contain antitumor, antimicrobial, antiviral, and generally cytotoxic activities (Scheuermayer et al. 2006; Blunt et al. 2010; Thomas et al. 2010). Many of the sponge-derived compounds, such as fatty acids, pyrroles, indoles, phenol derivatives, tyrosine derivatives, terpenes, diphenyl ethers, and dioxins are halogenated (Gribble 2003). Sponges and also other sessile animals lack physical defenses and therefore probably experience an enormous predator and epibiont pressure which, in turn, is thought to have provoked the evolution of structurally highly diverse, effective, and sophisticated chemical defenses. A number of these marine invertebrate-derived compounds are currently being developed as medical and pharmaceutical drugs; however, the supply problem is still the major obstacle in this field (Kennedy et al. 2007).

Sponges are also known to contain massive amounts of microorganisms within the mesohyl matrix which may constitute up to half of the sponge biomass. The diversity of the sponge-associated microbiota is phylogenetically complex yet highly sponge-specific and is probably permanently associated with the host sponge (for reviews, see Hentschel et al. 2006; Taylor et al. 2007; Webster and Taylor 2011). There is increasing evidence that the most important marine natural product classes, complex polyketides and non-ribosomal peptides, are truly synthesized by symbiotic bacteria rather than by the sponge itself (Salomon et al. 2004; Fisch et al. 2009; Piel 2009; Hochmuth et al. 2010). However, since the vast majority of environmental microorganisms including those from sponges are still refractory to cultivation, new experimental approaches are needed to provide information about their genomic potential for natural product biosynthesis.

One such approach is metagenomics or environmental genomics which was developed to provide access to the DNA pool of complex environmental microbial consortia in a cultivation-independent manner (Handelsman 2004; Grozdanov and Hentschel 2007; Kennedy et al. 2007). So far, polyketide synthase (Piel et al. 2004; Fieseler et al. 2007; Fisch et al. 2009) and non-ribosomal peptide synthetase gene clusters (Schirmer et al. 2005), ribosomal operons (Schleper et al. 1998; Fieseler et al. 2006), as well as functional enzymes (Schleper et al. 1997; Kennedy et al. 2011) were recovered from marine sponge-associated microbiota by ways of metagenomics that would have been inaccessible by conventional techniques.

The metagenomics approach has recently been complemented by single cell genomics (Hutchison and Venter 2006; Walker and Parkhill 2008). Here, microbial cells are sorted by fluorescence-activated cell sorting (FACS) or any other cell separation technique, and the DNA of individual microbial cells is then amplified using phi29-polymerase, resulting in a billion-fold amplification of the original DNA. Single cell genomics has provided valuable insights into the genomes of single microbial cells such as that of a poribacterial sponge symbiont (Siegl et al. 2010). Single cell genomics has also allowed the cloning of secondary metabolite gene clusters from single amplified genomes (Siegl and Hentschel 2010; Grindberg et al. 2011) and to link functional genes to their phylogenetic origin (Stepanauskas and Sieracki 2007; Siegl and Hentschel 2010) which is notoriously difficult in complex metagenomic DNA pools. The knowledge of the phylogenetic origin of any product is also important for practical purposes, such as choosing the right strain for heterologous expression studies.

$FADH_2$-dependent halogenases represent the biggest class of specific halogenating enzymes known to date. This class is responsible for the halogenation of a multitude of bacterial secondary metabolites including balhimycin, rebeccamycin, and pyrrolnitrin. Nearly all known $FADH_2$-dependent halogenases are involved in the halogenation of aromatic or heteroaromatic ring systems. Two distinct subgroups of enzymes exist, one of which uses phenol or pyrrol as substrate and the other tryptophan (Murphy 2006). Homologues of $FADH_2$-dependent halogenases have been found in various bacterial phyla, including the *Actinobacteria* (Pelzer et al. 1999), *Cyanobacteria* (Rouhiainen et al. 2000), *Planctomycetes* (Glöckner et al. 2003), and the *Proteobacteria* (Murphy 2006).

In this study, we have focussed our efforts on the discovery of $FADH_2$-dependent halogenase genes from marine sponge-associated microbiota, because sponges contain structurally diverse halogenated (brominated) metabolites, thus representing a clinically important group of bioactive metabolites. Furthermore, owing to their integration in biosynthetic gene clusters, halogenases serve as anchors in a genomic mining strategy, which has been instrumental for the discovery of novel biosynthetic gene clusters (Hornung et al. 2007; Gao and Huang 2009). This study contributes to a continued effort to characterize the genomic potential of sponge-associated microorganisms for biotechnologically relevant enzymes and natural products.

Materials and Methods

Sponge Collection

The Mediterranean sponges *Aplysina aerophoba*, *Axinella polypoides*, and *Axinella damicornis* were collected by scuba diving offshore Banyuls-sur-Mer, France (42 °29′ N; 03 °08′ E) and *Aplysina cavernicola* was collected offshore Elba, Italy (42 °43′ N; 10 °08′ E). The Caribbean sponges *Aplysina archeri*, *Aplysina cauliformis*, *Aplysina lacunosa*, *Agelas conifera*, *Agelas clathrodes*, *Agelas dispar*, *Ectyoplasia ferox*, *Xestospongia muta*, *Pseudoceratina crassa*, and *Ptilocaulis* sp. were collected offshore Little San Salvador Island, Bahamas (24 °32′ N; 75 °55′ W). Individual specimens were brought to the surface in plastic bags, frozen in liquid nitrogen on board ship, and stored at −80 °C until further use. For single cell genomic analyses, *A. aerophoba* sponges were also collected offshore Rovinj, Croatia (45 °05′N, 13 °38′E). For whole genome amplification (WGA) screenings, the sponges were maintained in seawater aquaria for few days prior to preparation for cell sorting.

Degenerate Halogenase Primer Design

The deduced amino acid sequences of the following $FADH_2$-dependent halogenases were retrieved from the GenBank database: ComH (AAK81830) from *Streptomyces lavendulae*, CmlS (AAK08979) from *Streptomyces venezuelae*, Asm12

(AAM54090) from *Actinosynnema pretiosum* subsp. *auranticum*, StaK (AAM80530) and StaI (AAM80532) from *Streptomyces toyocaensis*, PltA (Q9X3R1) and PltM (Q9X3Q8) from *Pseudomonas fluorescens*, AdpC (CAC01605) from *Anabaena circinalis 90*, XAC4089 (NP_644388) from *Xanthomonas axonopodis*, BhaA (CAA76550) from *Amycolatopsis balhimycina* and an unnamed protein (AAL06656) from *Streptomyces globisporus.* The sequences were submitted to Block Maker Server (http://blocks.fhcrc.org/) (Henikoff et al. 1995) and five conserved amino acid regions (block A–E) were identified. The CODEHOP program (http://blocks.fhcrc.org/codehop.html) (Rose et al. 1998) was used to generate partially degenerate oligonucleotides. Table 1 summarizes conserved amino acid regions used for primer design and provides sequences of designed primer HALOforD and HALOrevE. Tryptophan-specific halogenases were excluded from the primer design.

PCR screening of Sponge-Associated Microbiota

Genomic DNA was isolated from liquid nitrogen frozen sponge tissues that were ground in a mortar and were further processed using the Fast DNA spin for soil Kit (MP Biomedicals). For amplification of halogenase genes, the primers HALOforD and HALOrevE were used, and PCR conditions were as follows: initial denaturation (95 °C for 5 min) followed by 30 cycles of denaturation (95 °C for 30 s), primer annealing (63 °C for 30 s), primer extension (72 °C for 30 s), and a final extension step of 2 min at 72°C. One unit of Taq DNA Polymerase (Qiagen) was used with 100 pmol of each primer and 0.2 pmol of each dNTP in a reaction volume of 50 µl. The amplification products were run on 2 % agarose gels, stained with ethidium bromide and visualized by UV light. Products of the appropriate size were purified with the QIAquick PCR Purification Kit (Qiagen), ligated into the pGEMTeasy vector (Promega) and transformed into electrocompetent *Escherichia coli* XL1-Blue cells. For each PCR product, 20 clones were selected at random and RFLP-analyses using the restriction endonuclease *Hae*III were conducted. Plasmid DNA of clones with an individual banding pattern was isolated by the alkaline lysis method and the correct insert sizes were confirmed by *Eco*RI endonuclease digestion and agarose gel electrophoresis (Sambrook and Russel 2001).

Metagenomic Library Screening and Sequencing of the Halogenase-Bearing Fosmid 3 K12

The metagenomic library, "Apae02", harboring an estimated 1.1 Gb of microbial community DNA was previously constructed from *A. aerophoba* (Fieseler et al. 2004, 2006). Altogether, 29,108 individual clones are filed in 76 384-well microtiter plates. Fosmid DNA was isolated from each 384-well plate (each plate representing one "pool") using the QIAprep Spin Miniprep Kit (Qiagen). Following identification of a positive pool by PCR using the halogenase primers HALOforD and HALOrevE, clones were transferred to 12-cm

Table 1 Degenerate primer design for $FADH_2$-dependent halogenases

Halogenase	Forward primer	Reverse primer
ComH	(217) AFESGWFWYIPLSD	(303) LVGDAACFVDPVFSSGV
CmlS	(210) TFEDGWVWMIPIKD	(292) LCGDAACFTDPIFRRSV
XAC4089	(231) CGPGYWFWLIPLSS	(315) LTGEAGVFLDPFYSPGS
Asm12	(212) SHSDGWIWAIPLSA	(291) MVGDAGCFGDPMFSGGV
StaK	(216) TFDGGWFWYVPLFG	(302) LVGDAACSVDPLFSSGV
StaI	(238) AFDSGWFWYIPLSD	(324) LVGDAACFVDPVFSSGV
AAL06656	(217) TFEHGWFWYIPLSP	(303) LVGDAACFVDPVFSSGV
BhaA	(215) AFDSGWFWYIPLSD	(301) LVGDAACFVDPVFSSGV
PltA	(232) PMTNTWVWQIPITE	(313) LIGDAARFVDPIFSSGV
PltM	(232) IFEEGWLWVIPFNN	(318) LLPQATGFIDPLFSRGL
AdpC	(297) CGEGYWVWTIPLST	(379) CVGEAGTFPDPFYSPGT
Consensus	CFEGGWFWYIPLSD	LVGDAACFVDPFFSRGV
Primer sequence	5′-G GAC GGC TGG TTC TGG NHN ATH CC-3′	5′-CAC GCC GCG GGA GWA NAN NGG RTC-3′
Primer name	HALOforD	HALOrevE
Properties of primer	24 mer, degeneracy 144-fold	24 mer, degeneracy 256-fold

Degenerate nucleotide code as follows : N = A + C + G + T, H = A + C + T, W = A + T, R = A + G. The sequences of amino acid blocks D (left) and E (right) after BLOCK MAKER analyses are presented. The numbers in brackets give the respective starting point of the amino acid sequence. Target sites of the degenerate primers are shown in bold

square petri dishes using a pin replicator. The pool was divided in subpools and re-screened by PCR. PCR products were cloned as described above and sequenced. Of the four unique halogenase-bearing fosmids identified, fosmid 3K12 was chosen for full sequencing in order to characterize the genomic context of the halogenase gene. Shotgun cloning and sequencing of 3K12 (AGOWA GmbH, Berlin, Germany) provided an at least six-fold coverage per base pair. Open reading frames (ORFs) were identified with the computer programs Vector NTI (version 8, Invitrogen) and ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). If multiple putative ORFs were identified in the same sequence region, the one showing highest sequence similarity to known homologues as revealed by translated BLAST (blastx) search (Altschul et al. 1997) was selected.

Screening of Whole Genome Amplifications

Sponge-associated microbial consortia were extracted from *A. aerophoba* biomass following the differential centrifugation protocol developed by Fieseler et al. (2004, 2006). Cell sorting was performed by FACS with freshly prepared material following the protocol of Siegl and Hentschel (2010). WGAs were screened for halogenase genes using the primers HALOforD and HALOrevE, and the PCR products were sequenced directly. In order to determine the phylogenetic identity of the halogenase-positive WGAs, eubacterial 16S rRNA gene primers (27f and 1492r; Lane 1991), poribacterial-specific 16S rRNA gene primers (389f and 1130r; Fieseler et al. 2006), archaeal 16S rRNA gene primers (Arch21f and Arch958r; DeLong 1992), and 18S rRNA gene primers (A and B; Medlin et al. 1988) were applied. PCR products were cleaned using the NucleoSpin Extract II Kit (Macherey-Nagel) and cloned into the pGEMTeasy vector (Promega). RFLP analysis using enzymes *Bsu*RI (*Hae*III) and *Msp*I (Fermentas) of at least 32 randomly picked clones per WGA were performed. Inserts of clones showing an individual banding pattern were sequenced.

Sequencing

Sequencing was performed using the BigDye Terminator v1.1 Cycle Sequencing Kit and the ABI 377XL automated sequencer (Applied Biosystems). The sequence data were edited with the Vector NTI program (Invitrogen). 16S rRNA genes were checked for chimeras using the program pintail (Ashelford et al. 2005). The halogenase gene sequences from marine sponge material were deposited into GenBank under the accession numbers ACY25399–ACY25432. The halogenase gene sequences of the metagenomic clones 26C24, 29G1, and 42I22 were deposited into GenBank under the accession numbers ACY25433–ACY25435. The complete sequence of metagenome clone 3K12 was deposited into GenBank under the accession numbers ACY25436–ACY25467. Halogenase and 16S rRNA gene sequences from whole genome amplifications were deposited under accession numbers JN093133–JN093138 and JN002373–JN002382, respectively.

Phylogenetic Analyses

The alignment of amino acid sequences was performed over 134 positions including gaps using clustalX (Thompson et al. 1997) and checked manually using the program align (Hepperle 2002). For phylogenetic analysis of halogenase amino acid sequences, the program MrBayes (Huelsenbeck and Ronquist 2001) was used. The WAG model was applied and a gamma-shaped rate variation with a proportion of invariable sites was estimated. Temperature was set to 0.1. The analysis was run until standard deviation of split frequencies was below 0.01. After a burn-in phase every 100th sample of 5,380,000 generations was considered for amino acid tree construction. The phylogenetic tree was edited using the program FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

Results

Primer Validation

PCR analysis of genomic DNA from potentially $FADH_2$-dependent halogenase-bearing bacteria was performed in order to validate the new primers, HALOforD and HALOrevE (Table 1). Amplification products of the deduced size of around 300 bp were obtained from DNA of *Amycolatopsis orientalis* DSM 40040^T and from *Pseudomonas syringae* (potato rhizosphere isolate strain Lüsewitz) and *Pseudomonas aeruginosa* (wt), taken from the ZINF strain collection, University of Wuerzburg. Sequencing of the respective PCR products confirmed the presence of $FADH_2$-dependent halogenase genes in these bacteria. As expected, amplification products were not obtained from several *Enterobacteriaceae* strains (*E. coli*, *Shigella flexneri*, *Yersinia enterocolytica,* and *Salmonella typhimurium*) which are not known to contain halogenase genes (data not shown).

Halogenase Genes from Sponge-Associated Microbiota

Altogether, 36 different putative $FADH_2$-dependent halogenase gene fragments of about 300 bp in size were found from the Mediterranean species *A. aerophoba* (n=5) and *A. cavernicola* (n=2) as well as from the Caribbean species *A. archeri* (n=5), *A. cauliformis* (n=5), *A. lacunosa* (n=5), *A. conifera* (n=4), *A. clathrodes* (n=1), *X. muta* (n=4), and *E. ferox* (n=5). No amplification products were recovered from *A. dispar* and *P. crassa* as well as from the low microbial abundance (LMA) sponges *A. polypoides*, *A. damicornis,* and *Ptilocaulis* sp.. Seawater samples were also negative.

Table 2 Conserved amino acid motifs of $FADH_2$-dependent halogenases

Halogenase	Conserved motifs	
ClohaI	(13) GGGPGG	(218) WWWCIP
AdpC	(71) GGGLAG	(302) WVWTIP
CmlS	(10) GGGPAG	(215) WVWMIP
StaK	(12) GGGPAG	(221) WFWYVP
PltA	(12) GGGPAG	(237) WVWQIP
3 K12	(09) GGGPAG	(213) WSWAIP
26 C29	(16) GGGPAG	(221) WFWYIP
29 G1	(36) GGGPGG	(239) WSWAIP
42I22	(11) GGGPAG	(217) WLWNIP
Consensus	GxGxxG	WxWxI

Putative halogenases of the fosmid clones 3K12, 26C24, 29G1, and 42I22 are highlighted

Halogenase-Containing Metagenome Clones

PCR screening of a previously established, >29,000 clone member containing metagenomic library (Fieseler et al. 2006) revealed nine $FADH_2$-dependent halogenase-positive fosmid clones of which four showed unique gene sequences. All four halogenases contained the highly conserved amino acid motif "GxGxxG", which is located close to the N-terminus of the enzymes and which is essential for the binding of the nucleotide cofactor $FADH_2$ (Scrutton et al. 1990; Wynands and van Pée 2004). The presence of this conserved amino acid motif, together with the second characteristic halogenase motif "WxWxI" within the sponge-derived sequences supports their identity as halogenases (Table 2).

To gain insights into the genomic context of the halogenase genes, fosmid 3K12 from the metagenomic library of *A. aerophoba* was completely sequenced (Fig. 1). The fosmid contained an insert of 40.365 bp in size. The G + C content was of 64.2 mol % and the coding density was 90.4 %. Altogether, 31 potential protein encoding genes and one truncated ORF (ORF32) were identified. A phylogenetic marker such as a ribosomal RNA gene or a housekeeping gene was absent on the sequenced insert. Approximately 30 % of the identified genes were affiliated to genes with unknown function.

Genes encoding proteins possibly involved in secondary metabolism are the following (Supplementary Table 1): ORF1 encodes a putative alpha/beta-hydrolase, homologs of which are involved in carbazole catabolism and in naphthalene degradation. ORF2 encodes a putative enzyme which shows homologies to a monooxygenase of *Pirellula staleyi* (YP_003372867) involved in dibenzothiophene transformation. The deduced halogenase protein of ORF3 showed high similarity to ClohaI (AAN65327) from the clorobiocin gene cluster of *Streptomyces roseochromogenes* subsp. *oscitans*. An InterPro scan (Apweiler et al. 2001) of the deduced amino acid sequence of ORF27 (ACY25462) revealed the protein domain "IPR012338" which is typically present in penicillin-binding proteins. The putative protein exhibited homologies to beta-lactamases and to ORF2 of the epothilone biosynthetic gene cluster. The ORF27-encoded protein showed also sequence similarity to the esterase (MitL) of the mitomycin C gene cluster of *S. lavendulae* NRRL 2564.

ORF31 and ORF32 showed high similarity to each other (76 % identical bases over 778 bp) (Supplementary Table 1). Both had significant homologies to integral membrane efflux proteins of the major facilitator superfamily PF07690 which transports small solutes across chemiosmotic gradients. The putative protein encoded by ORF31 exhibited additionally high similarities to an antibiotic efflux protein of the leinamycin biosynthetic gene cluster (AAN85487, 53 %), to a multidrug resistance protein of *Streptomyces pristinaespiralis* (CAA58879, 54 %), and to a putative resistance protein of the monensin biosynthetic gene cluster (AAO65793, 53 %). The protein encoded by the truncated ORF32 showed high similarity to the virginiamycin resistance protein VarS (BAA78678, 60 %) of *Streptomyces virginiae*, to the multidrug resistance protein from *S. pristinaespiralis* (CAA58879, 61 %), and to RifP (AAC01725, 58 %), a putative efflux transport protein of the rifampicin biosynthetic gene cluster.

Halogenase-Containing WGAs

From >300 screened WGAs derived from *A. aerophoba* microbiota, 12 were positive for halogenase genes. For six of the halogenase-positive WGAs, a corresponding bacterial 16S rRNA gene was identified (Table 3). All six WGAs tested negative for 18S rRNA genes and for archaeal 16S rRNA genes. WGA 1-4E contained only a single phylotype

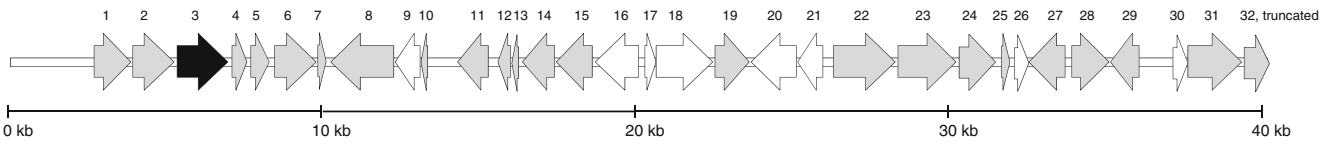


Fig. 1 Physical map of the insert of fosmid 3 K12 ORFs with unknown function are presented in *white*, functionally annotated ORFs are in *gray*, and the halogenase-bearing ORF is presented in *black* color. *ORF numbers and* orientation correspond to Supplementary Table 1

Table 3 Phylogenetic affiliation of halogenase-positive WGAs

WGAs	16S rRNA Gene - Acc #	Top Blast hits, Acc#, (length of overlap/identity in %) host sponge	Phylum
WGA 1-4E	JN002375	Uncultured bacterium clone PK025, EF076115 (97/97) *Plakortis* sp.	*Deltaproteobacteria*
WGA 2.3_2B	JN002379	Uncultured bacterium clone XA3H07F, HQ270280 (100/98) *Xestospongia testudinaria*	*Deltaproteobacteria*
	JN002380	Uncultured Chloroflexi bacterium 1i19, FJ560485 (100/99) *Aplysina aerophoba*	*Chloroflexi*
WGA 7.12._5G	JN002377	Uncultured bacterium clone XA1B01F, FJ229909 (100/77) *Xestospongia testudinaria*	*Actinobacteria*
WGA 1-3 G	JN002373	Uncultured Poribacteria bacterium clone, P30 EU071663 (100/99) *Plakortis* sp.	*Poribacteria*
WGA 19.1._1E	JN002376	Uncultured Poribacteria bacterium clone S2, EU071665 (100/99) *Ircinia* sp.	*Poribacteria*
	JN002378	Uncultured Chloroflexi bacterium 1i19, FJ560485 (100/99) *Aplysina aerophoba*	*Chloroflexi*
WGA 7.12._5E	JN002374	Uncultured Poribacteria bacterium clone S6, EU071666 (100/99) *Ircinia* sp.	*Poribacteria*
	JN002382	Uncultured bacterium clone XA2B06F, HQ270226 (100/98) *Xestospongia testudinaria*	*Deltaproteobacteria*

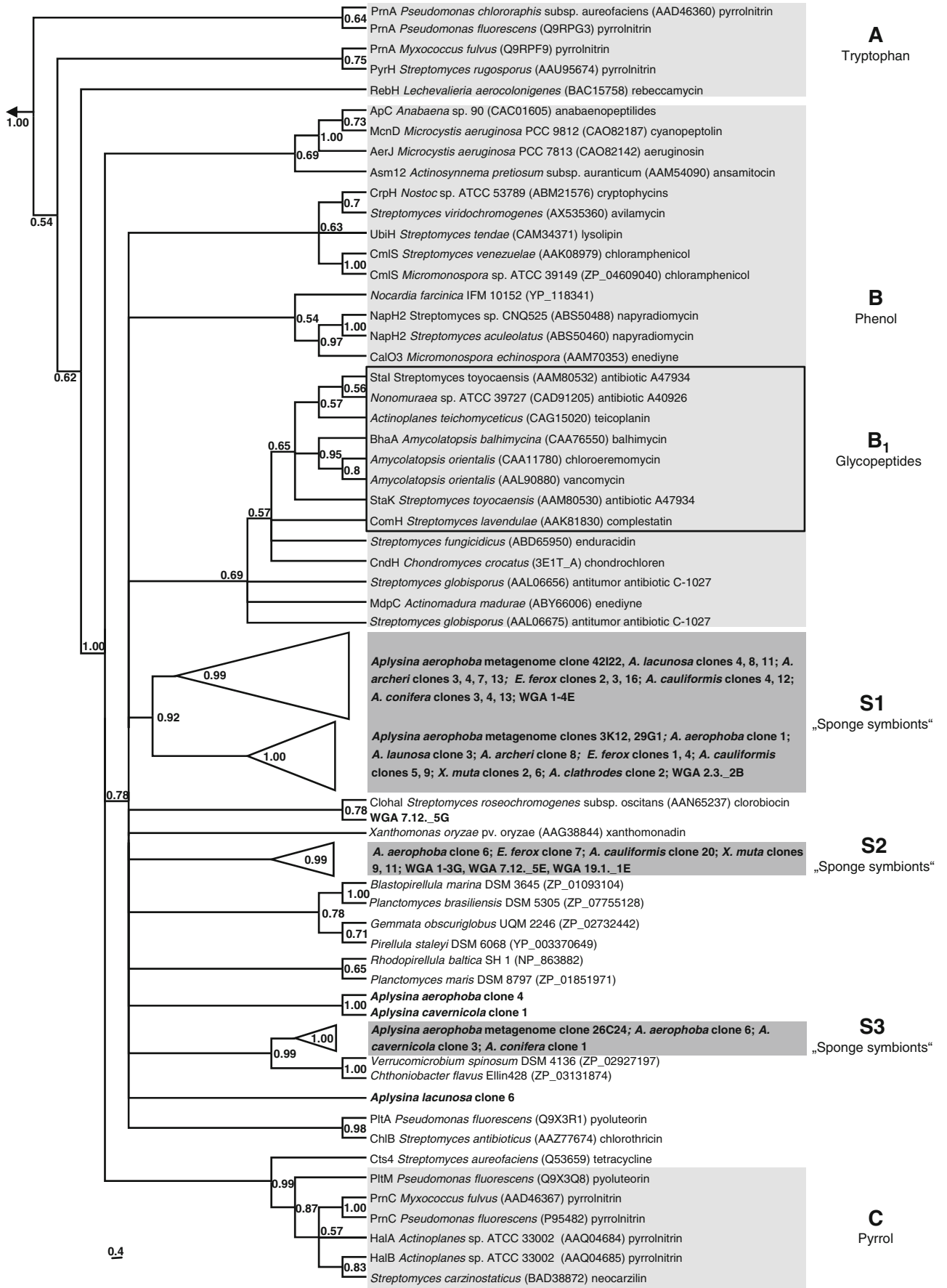
which was most closely related to a deltaproteobacterial 16S rRNA gene. WGA 2.3_2B contained two phylotypes; one of which was most closely related to a deltaproteobacterial 16S rRNA gene sequence and the other one to a *Chloroflexi* 16S rRNA gene previously identified by single cell genomic analysis (Siegl and Hentschel 2010). WGA 7.12._5G contained a single actinobacterial phylotype. WGA 1-3 G contained a single phylotype of poribacterial origin. WGA 19.1._1E contained two phylotypes; one of which related to candidate phylum Poribacteria and the other one to *Chloroflexi*. WGA 7.12._5E also contains a mix of two cells affiliated with the Poribacteria and *Deltaproteobacteria*.

Phylogenetic Analysis of Halogenase Genes

The overall layout of the phylogenetic tree revealed largely a substrate specific clustering which is consistent with previous reports (Fig. 2) (Hornung et al. 2007; Gao and Huang 2009). Cluster A contains exclusively enzymes responsible for the halogenation of tryptophan- or indole-derived residues such as pyrrolnitrin and rebeccamycin (van Pée and Patallo 2006). Cluster B is formed by enzymes that halogenate phenol-residues such as ansamitocin, substance C-1027, and enduracidin. The halogenase CmlS, involved in the biosynthesis of chloramphenicol, is an exception in this clade as it halogenates aliphatics (Piraee et al. 2004). A distinct subcluster, "B_1", is present in clade "B" that is formed by enzymes involved in the biosynthesis of vancomycin-like glycopeptides. These include for example, complestatin, substances A47934 and A40926, balhimycin, chloroeremomycin, and teicoplanin. A third cluster (C) of known halogenases comprises enzymes responsible for the halogenation of pyrrol-rings, with the exception of the halogenase from the neocarzilin gene cluster which halogenates aliphatic substances (Otsuka et al. 2004).

The majority of sponge-derived halogenases form one dominant cluster (S1) with two subgroups that is phylogenetically distinct from all previously known halogenases (Fig. 2). This cluster contains 25 halogenase gene fragments from six different sponge species, all of which belong to the "high microbial abundance" group of sponges (Hentschel et al. 2003). The cluster furthermore contains three metagenomic halogenase gene sequences, of which one full metagenomic clone sequence is presented in this study (Supplementary Table 1). Two halogenase-positive WGAs (WGA 1-4E and WGA 2.3._2B) are also affiliated with the S1 cluster. A second cluster (S2) of halogenase genes contains five halogenase gene fragments from different sponges. Moreover, three halogenase genes from WGA screenings (WGA 1-3 G, WGA 7.12._5E, and WGA 19.1._1E) fall into this clade. A third, sponge-derived cluster of halogenases (S3) contains one metagenome clone (26 C24) and three PCR fragments from *Aplysina* and *Agelas* sponges. One WGA reaction (WGA 7.12._5G) yielded both, a halogenase gene and a 16S rRNA gene, both of which were most closely affiliated with members of the phylum *Actinobacteria*. Finally, two clustered halogenase gene sequences from *A. aerophoba* and *A. cavernicola* and one independent sequence from *A. lacunosa* are present in

Fig. 2 Phylogeny of halogenase genes reconstructed from 134 amino acid positions using MrBayes. Sequences derived from this study are shown in *bold*. MrBayes values are indicated. Tree was rooted with FAD dependent oxidoreductase (YP_284125) from *Dechloromonas aromatica* RCB and geranylgeranyl reductase (ZP_03127665) from *Chthoniobacter flavus* Ellin428, *arrow to outgroup*. *Scale bar* indicates 40 % sequence divergence. The *lighter gray boxes* indicate a substrate specific clustering. Exceptions are CmlS of *S. venezuelae* and *Micromonospora* sp. as well as a halogenase of *S. carzinostaticus* which halogenate aliphatic substances. A distinct subcluster "B_1" is marked by a *box*, that is formed by enzymes involved in the biosynthesis of vancomycin-like glycopeptides. The *dark gray boxes* indicate "sponge cluster" ▶


PrnA Pseudomonas chlororaphis subsp. aureofaciens (AAD46360) pyrrolnitrin
PrnA Pseudomonas fluorescens (Q9RPG3) pyrrolnitrin
PrnA Myxococcus fulvus (Q9RPF9) pyrrolnitrin
PyrH Streptomyces rugosporus (AAU95674) pyrrolnitrin
RebH Lechevalieria aerocolonigenes (BAC15758) rebeccamycin
ApC Anabaena sp. 90 (CAC01605) anabaenopeptilides
McnD Microcystis aeruginosa PCC 9812 (CAO82187) cyanopeptolin
AerJ Microcystis aeruginosa PCC 7813 (CAO82142) aeruginosin
Asm12 Actinosynnema pretiosum subsp. auranticum (AAM54090) ansamitocin
CrpH Nostoc sp. ATCC 53789 (ABM21576) cryptophycins
Streptomyces viridochromogenes (AX535360) avilamycin
UbiH Streptomyces tendae (CAM34371) lysolipin
CmlS Streptomyces venezuelae (AAK08979) chloramphenicol
CmlS Micromonospora sp. ATCC 39149 (ZP_04609040) chloramphenicol
Nocardia farcinica IFM 10152 (YP_118341)
NapH2 Streptomyces sp. CNQ525 (ABS50488) napyradiomycin
NapH2 Streptomyces aculeolatus (ABS50460) napyradiomycin
CalO3 Micromonospora echinospora (AAM70353) enediyne
StaI Streptomyces toyocaensis (AAM80532) antibiotic A47934
Nonomuraea sp. ATCC 39727 (CAD91205) antibiotic A40926
Actinoplanes teichomyceticus (CAG15020) teicoplanin
BhaA Amycolatopsis balhimycina (CAA76550) balhimycin
Amycolatopsis orientalis (CAA11780) chloroeremomycin
Amycolatopsis orientalis (AAL90880) vancomycin
StaK Streptomyces toyocaensis (AAM80530) antibiotic A47934
ComH Streptomyces lavendulae (AAK81830) complestatin
Streptomyces fungicidicus (ABD65950) enduracidin
CndH Chondromyces crocatus (3E1T_A) chondrochloren
Streptomyces globisporus (AAL06656) antitumor antibiotic C-1027
MdpC Actinomadura madurae (ABY66006) enediyne
Streptomyces globisporus (AAL06675) antitumor antibiotic C-1027
Aplysina aerophoba metagenome clone 42l22, A. lacunosa clones 4, 8, 11; A. archeri clones 3, 4, 7, 13; E. ferox clones 2, 3, 16; A. cauliformis clones 4, 12; A. conifera clones 3, 4, 13; WGA 1-4E
Aplysina aerophoba metagenome clones 3K12, 29G1; A. aerophoba clone 1; A. launosa clone 3; A. archeri clone 8; E. ferox clones 1, 4; A. cauliformis clones 5, 9; X. muta clones 2, 6; A. clathrodes clone 2; WGA 2.3._2B
ClohaI Streptomyces roseochromogenes subsp. oscitans (AAN65237) clorobiocin
WGA 7.12._5G
Xanthomonas oryzae pv. oryzae (AAG38844) xanthomonadin
A. aerophoba clone 6; E. ferox clone 7; A. cauliformis clone 20; X. muta clones 9, 11; WGA 1-3G, WGA 7.12._5E, WGA 19.1._1E
Blastopirellula marina DSM 3645 (ZP_01093104)
Planctomyces brasiliensis DSM 5305 (ZP_07755128)
Gemmata obscuriglobus UQM 2246 (ZP_02732442)
Pirellula staleyi DSM 6068 (YP_003370649)
Rhodopirellula baltica SH 1 (NP_863882)
Planctomyces maris DSM 8797 (ZP_01851971)
Aplysina aerophoba clone 4
Aplysina cavernicola clone 1
Aplysina aerophoba metagenome clone 26C24; A. aerophoba clone 6; A. cavernicola clone 3; A. conifera clone 1
Verrucomicrobium spinosum DSM 4136 (ZP_02927197)
Chthoniobacter flavus Ellin428 (ZP_03131874)
Aplysina lacunosa clone 6
PltA Pseudomonas fluorescens (Q9X3R1) pyoluteorin
ChlB Streptomyces antibioticus (AAZ77674) chlorothricin
Cts4 Streptomyces aureofaciens (Q53659) tetracycline
PltM Pseudomonas fluorescens (Q9X3Q8) pyoluteorin
PrnC Myxococcus fulvus (AAD46367) pyrrolnitrin
PrnC Pseudomonas fluorescens (P95482) pyrrolnitrin
HalA Actinoplanes sp. ATCC 33002 (AAQ04684) pyrrolnitrin
HalB Actinoplanes sp. ATCC 33002 (AAQ04685) pyrrolnitrin
Streptomyces carzinostaticus (BAD38872) neocarzilin
A
Tryptophan
B
Phenol
B1
Glycopeptides
S1
„Sponge symbionts“
S2
„Sponge symbionts“
S3
„Sponge symbionts“
C
Pyrrol
0.4

the tree; however, their phylogenetic position remains unresolved.

Discussion

The primer set developed in this study represents a suitable tool for the detection of $FADH_2$-halogenase genes in environmental samples as well as in cultivated bacteria. It builds upon a previous publication, where primers were designed to amplify and clone a $FADH_2$-dependent halogenase gene fragment from *Streptomyces venezuelae* (Piraee and Vining 2002). Eleven different halogenase amino acid sequences were included in the primer design in the present study. Using the primers HALOforD and HALOrevE, $FADH_2$-halogenases could also be amplified from *A. orientalis* DSM 40040^T as well as from *P. syringae* and *P. aeruginosa*.

We report here on the discovery of at least three phylogenetically and possibly also functionally distinct clades of novel $FADH_2$-type halogenases in the microbial metagenomes of marine sponges (Fig. 2). The halogenase genes were identified in five *Aplysina* and in two *Agelas* species as well as in *X. muta* and *E. ferox*, all of which belong to the "high microbial abundance" sponge category (Hentschel et al. 2003). $FADH_2$-type halogenase genes were not identified in the LMA sponge species *A. polypoides*, *A. damicornis*, and *Ptilocaulis* sp., suggesting that the microbial producers are members of the sponge-specific symbiotic microbial consortium (Taylor et al. 2007).

The genomic context surrounding the putative halogenase on metagenomic fosmid 3 K12 revealed several ORFs that might be involved in secondary metabolism (Supplementary Table 1). Of particular interest are two putative efflux proteins (ORFs 31 and 32) that might be involved in drug resistance. However, since $FADH_2$-dependent halogenases are known to be integrated within secondary metabolite gene clusters, at least from what has been reported in actinomycetes (Hornung et al. 2007; Gao and Huang 2009), cyanobacteria (Cadel-Six et al. 2008), gram-negative bacteria (Costa et al. 2009), and also in *Rhodopirellula baltica* (Glöckner et al. 2003), the lack of integration of the described halogenase in a typical secondary metabolite gene cluster is somewhat surprising. It may be postulated that the halogenases from the microbial metagenomes of sponges are located *in trans* to the biosynthetic operons whose products they halogenate.

In order to obtain information on the original producers of the different halogenase clusters, a single cell genomics-based approach was undertaken. More than 300 WGAs were screened by PCR and six WGAs were shown to contain both halogenase and 16S rRNA genes (Table 3). The presented data allow for the following predictions: halogenase cluster S1 may be produced by a deltaproteobacterial symbiont, and the halogenase cluster S2 may be produced by a poribacterial symbiont of sponges. Furthermore, the data strongly suggest that the halogenase gene obtained from WGA 7.12._5G is produced by an actinobacterial sponge symbiont. The halogenase-positive WGAs, which contained only one phylotype each, are particularly informative in this analysis. It will however only become possible to validate these predictions once more sponge symbiont genomes become available. Taken together, genomic mining such as undertaken here by the combination of metagenomics and single cell genomics is a meaningful strategy to explore complex microbial consortia for biodiscovery.

Acknowledgments We gratefully acknowledge the marine operations personnel at the Laboratoire Arago (Banyuls-sur-Mer, France) and at the Ruder Boskovic Institute (Rovinj/ Croatia) for the help during sponge collection and Prof. K.-H. van Pée (Dresden, Germany) for helpful advice. Prof. W. Wohlleben, S. Pelzer, and Claudia Kittel (Tübingen, Germany) are acknowledged for interesting discussions and activity testing of selected clones in an *S. albus* background. We thank Christine Gernert (Wuerzburg, Germany) for excellent technical assistance. Financial support was provided by the SFB630 (grant TPA5) to U.H.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3404

Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MD, Durbin R, Falquet L, Fleischmann W, Gouzy J, Hermjakob H, Hulo N, Jonassen I, Kahn D, Kanapin A, Karavidopoulou Y, Lopez R, Marx B, Mulder NJ, Oinn TM, Pagni M, Servant F, Sigrist CJ, Zdobnov EM (2001) The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res 29:37–40

Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005) At least 1 in 20 16 S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl Environ Microbiol 71:7724–7736

Blunt JW, Copp BR, Munro MH, Northcote PT, Prinsep MR (2010) Marine natural products. Nat Prod Rep 27:165–237

Cadel-Six S, Dauga C, Castets AM, Rippka R, Tandeau de Marsac N, Welker M (2008) Halogenase genes in two non-ribosomal peptide synthetase gene clusters of *Microcystis* (*Cyanobacteria*): sporadic distribution and evolution. Mol Biol Evol 25:2031–2041

Costa R, van Aarle IM, Mendes R, van Elsas JD (2009) Genomics of pyrrolnitrin biosynthetic loci: evidence for conservation and whole-operon mobility within gram-negative bacteria. Environ Microbiol 11:159–175

DeLong EF (1992) Archaea in coastal marine environments. Proc Natl Acad Sci USA 89:5685–5689

Fieseler L, Horn M, Wagner M, Hentschel U (2004) Discovery of the novel candidate phylum "Poribacteria" in marine sponges. Appl Environ Microbiol 70:3724–3732

Fieseler L, Quaiser A, Schleper C, Hentschel U (2006) Analysis of the first genome fragment from the marine sponge associated, novel candidate phylum "Poribacteria" by environmental genomics. Environ Microbiol 8:612–624

Fieseler L, Hentschel U, Grozdanov L, Schirmer A, Wen G, Platzer M, Hrvatin S, Butzke D, Zimmermann K, Piel P (2007) Widespread occurrence and genomic context of unusually small polyketide synthase genes in microbial consortia associated with marine sponges. Appl Environ Microbiol 73:2144–2155

Fisch KM, Gurgui C, Heycke N, van der Sar SA, Anderson SA, Webb VL, Taudien S, Platzer M, Rubio BK, Robinson SJ, Crews P, Piel J (2009) Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. J Nat Chem Biol 5:494–501

Gao P, Huang Y (2009) Detection, distribution, and organohalogen compound discovery implications of the reduced flavin adenine dinucleotide-dependent halogenase gene in major filamentous actinomycete taxonomic groups. Appl Environ Microbiol 75:4813–4820

Glöckner FO, Kube M, Bauer M, Teeling H, Lombardot T, Ludwig W, Gade D, Beck A, Borzym K, Heitmann K, Rabus R, Schlesner H, Amann R, Reinhardt R (2003) Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. Proc Natl Acad Sci USA 100:8298–8303

Gribble GW (2003) The natural production of organobromine compounds. Environ Sci Pollut Res Intern 7:37–49

Grindberg RV, Ishoey T, Brinza D, Esquenazi E, Coates RC, Liu WT, Gerwick L, Dorrestein PC, Pevzner P, Lasken R, Gerwick WH (2011) Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage. PLoS One 6(4):e18565

Grozdanov L, Hentschel U (2007) An environmental genomics perspective on the diversity and function of marine sponge-associated microbiota. Curr Opin Microbiol 10:215–220

Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol Rev 68:669–685

Henikoff S, Henikoff JG, Alford WJ, Pietrokovski S (1995) Automated construction and graphical presentation of protein blocks from unaligned sequences. Gene 163:17–26

Hentschel U, Fieseler L, Wehrl M, Gernert C, Steinert M, Hacker J, Horn M (2003) Microbial diversity of marine sponges. In: Müller WEG (Ed) Molecular Marine Biology of Sponges, p 60–88, Springer Heidelberg

Hentschel U, Usher KM, Taylor MW (2006) Marine sponges as microbial fermenters. FEMS Microb Ecol 55:167–177

Hepperle D (2002) Align: a multicolor sequence alignment editor. http://sciencedomixde/softwarephd

Hochmuth T, Niederkrüger H, Gernert C, Siegl A, Taudien S, Platzer M, Crews P, Hentschel U, Piel J (2010) Linking chemical and microbial diversity in marine sponges: possible role for *Poribacteria* as producers of methyl-branched fatty acids. ChemBioChem 11(18):2572–2578

Hornung A, Bertazzo M, Dziarnowski A, Schneider K, Welzel K, Wohlert SE et al (2007) A genomic screening approach to the structure-guided identification of drug candidates from natural sources. Chembiochem 8:757–766

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754–755

Hutchison CA, Venter JC (2006) Single-cell genomics. Nat Biotechnol 24:657–658

Kennedy J, Marchesi JR, Dobson AD (2007) Metagenomic approaches to exploit the biotechnological potential of the microbial consortia of marine sponges. Appl Microbiol Biotechnol 75:11–20

Kennedy J, O'Leary ND, Kiran GS, Morrissey JP, O'Gara F, Selvin J, Dobson AD (2011) Functional metagenomic strategies for the discovery of novel enzymes and biosurfactants with biotechnological applications from marine ecosystems. J Appl Microbiol 111 (4):7877–7899

Lane DJ (1991) 16 S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds) Nucleic acid techniques in bacterial systematic. Wiley, London, pp 115–175

Medlin L, Elwood HJ, Stickel S, Sogin ML (1988) The characterization of enzymatically amplified eukaryotic 16 S-like rRNA-coding regions. Gene 71:491–499

Murphy CD (2006) Recent developments in enzymatic chlorination. Nat Prod Rep 23:147–152

Otsuka M, Ichinosem K, Fujii I, Ebizuka Y (2004) Cloning, sequencing, and functional analysis of an iterative type I polyketide synthase gene cluster for biosynthesis of the antitumor chlorinated polyenone neocarzilin in *Streptomyces carzinostaticus*. Antimicrob Agents Chemother 48:3468–3476

Pelzer S, Süßmuth R, Heckmann D, Recktenwald J, Huber P, Jung G, Wohlleben W (1999) Identification and analysis of the balhimycin biosynthetic gene cluster and its use for manipulating glycopeptide biosynthesis in *Amycolatopsis mediterranei* DSM5908. Antimicrob Agents Chemother 43:1565–1573

Piel J, Hui D, Wen G, Butzke D, Platzer M, Fusetani N, Matsunaga S (2004) Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. Proc Natl Acad Sci USA 101:16222–16227

Piel J (2009) Metabolites from symbiotic bacteria. Nat Prod Rep 26:338–62

Piraee M, Vining LC (2002) Use of degenerate primers and touchdown PCR to amplify a halogenase gene fragment from *Streptomyces venezuelae* ISP5230. J Ind Microbiol Biotechnol 29:1–5

Piraee M, White RL, Vining LC (2004) Biosynthesis of the dichloroacetyl component of chloramphenicol in *Streptomyces venezuelae* ISP5230: genes required for halogenations. Microbiol 150:85–94

Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. Nucleic Acids Res 26:1628–1635

Rouhiainen L, Paulin L, Suomalainen S, Hyytiäinen H, Buikema W, Haselkorn R, Sivonen K (2000) Genes encoding synthetases of cyclic depsipeptides, anabaenopeptilides, in *Anabaena* strain 90. Mol Microbiol 37:156–167

Salomon CE, Magarvey NA, Sherman DH (2004) Merging the potential of microbial genetics with biological and chemical diversity: an even brighter future for marine natural product drug discovery. Nat Prod Rep 21:105–121

Sambrook J, Russel DW (2001) In: Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Scheuermayer M, Pimentel-Elardo S, Fieseler L, Grozdanov L, Hentschel U (2006) Microorganisms of sponges: phylogenetic diversity and biotechnological potential. In: Proksch P, Mueller WEG (eds) Frontiers in marine biotechnology. Horizon Scientific Press, London, pp 289–312

Schirmer A, Gadkari R, Reeves CD, Ibrahim F, DeLong EF, Hutchinson CR (2005) Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*. Appl Environ Microbiol 71:4840–4849

Schleper C, DeLong EF, Preston CM, Feldman RA, Wu KY, Swanson RV (1998) Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum*. J Bacteriol 180:5003–5009

Schleper C, Swanson RV, Mathur EJ, DeLong EF (1997) Characterization of a DNA polymerase from the uncultivated psychrophilic archaeon *Cenarchaeum symbiosum*. J Bacteriol 179:7803–7811

Scrutton NS, Berry A, Perham RN (1990) Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. Nature 343:43–48

Siegl A, Kamke J, Hochmuth T, Piel J, Richter M, Liang C, Dandekar T, Hentschel U (2010) Single cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. ISME J 5:61–70

Siegl A, Hentschel U (2010) PKS and NRPS gene clusters from microbial symbiont cells of marine sponges by whole genome amplification. Environ Microbiol Rep 2:507–513

Stepanauskas R, Sieracki ME (2007) Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. Proc Natl Acad Sci USA 104:9052–9057

Taylor MW, Radax R, Steger D, Wagner M (2007) Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. Microbiol Mol Biol Rev 71:295–347

Thomas TRA, Kavlekar DP, LokaBharathi PA (2010) Marine drugs from sponge-microbe association—a review. Mar Drugs 8:1417–1468

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X window interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Van Pée K-H, Patallo EP (2006) Flavin-dependent halogenases involved in secondary metabolism in bacteria. Appl Microbiol Biotechnol 70:631–641

Walker A, Parkhill J (2008) Single-cell genomics. Nat Rev Microbiol 6:176–177

Webster NS, Taylor MW (2012) Marine sponges and their microbial symbionts: love and other relationships. Environ Microbiol 14 (2):517–524

Wynands I, van Pée K-H (2004) A novel halogenase gene from the pentachloropseudilin producer *Actinoplanes* sp. ATCC 33002 and detection of in vitro halogenase activity. FEMS Microbiol Lett 237:363–367