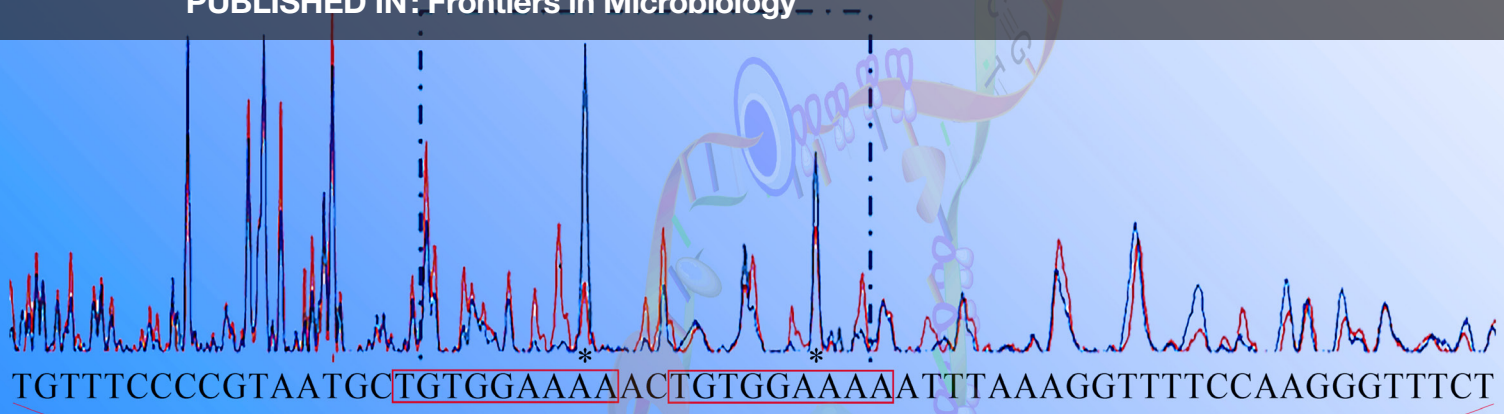
DNA REPLICATION ORIGINS IN MICROBIAL GENOMES

EDITED BY: Feng Gao
PUBLISHED IN: Frontiers in Microbiology



frontiers Research Topics

frontiers

Frontiers Copyright Statement

© Copyright 2007-2016 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 978-2-88919-779-8
DOI 10.3389/978-2-88919-779-8

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

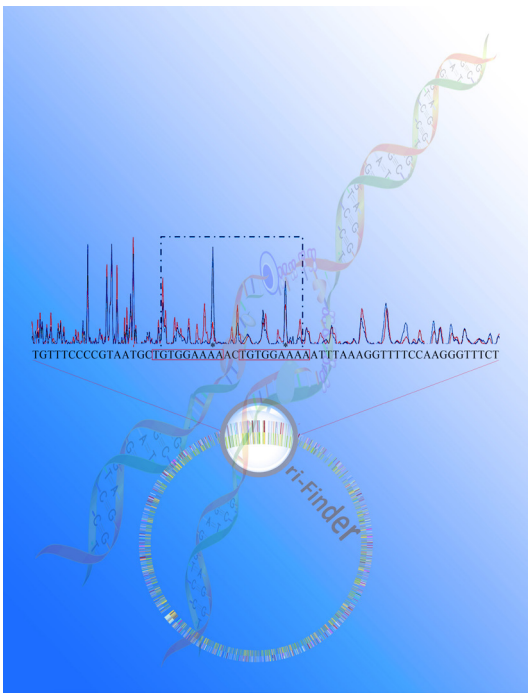
Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public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

DNA REPLICATION ORIGINS IN MICROBIAL GENOMES

Topic Editor:
Feng Gao, Tianjin University, China



In spite of significant differences among bacteria, archaea, and eukaryotes in the process of DNA replication, they all have the same core components of replication machines: DNA polymerases, circular sliding clamps, a pentameric clamp loader, helicase, primase, and single-strand binding protein (SSB). In all three domains of life, DNA replication initiates on defined genome sites, termed replication origins. Intensive studies have been carried out by in silico analyses as well as in vivo and in vitro experiments in the last two decades. As a study from in silico to in vitro, replication origins in Cyanothece ATCC 51142 have been predicted initially by Ori-Finder, a web-based system for finding oriCs in bacterial genomes, and the experimental supports including DNase I footprint assay are provided for the identified replication origins and their interactions with initiator protein DnaA.

Image based on Figure 7 from: Huang H, Song C-C, Yang Z-L, Dong Y, Hu Y-Z and Gao F (2015) Identification of the Replication Origins from Cyanothece ATCC 51142 and Their Interactions with the DnaA Protein: From In Silico to In Vitro Studies. Front. Microbiol. 6:1370.
doi: 10.3389/fmicb.2015.01370

DNA replication, a central event for cell proliferation, is the basis of biological inheritance. Complete and accurate DNA replication is integral to the maintenance of the genetic integrity of organisms. In all three domains of life, DNA replication begins at replication origins. In bacteria,

replication typically initiates from a single replication origin (oriC), which contains several DnaA boxes and the AT-rich DNA unwinding element (DUE). In eukaryotic genomes, replication initiates from significantly more replication origins, activated simultaneously at a specific time. For eukaryotic organisms, replication origins are best characterized in the unicellular eukaryote budding yeast Saccharomyces cerevisiae and the fission yeast Schizosaccharomyces pombe. The budding yeast origins contain an essential sequence element called the ARS (autonomously replicating sequence), while the fission yeast origins consist of AT-rich sequences. Within the archaeal domain, the multiple replication origins have been identified by a predict-and-verify approach in the hyperthermophilic archaeon Sulfolobus. The basic structure of replication origins is conserved among archaea, typically including an AT-rich unwinding region flanked by several short repetitive DNA sequences, known as origin recognition boxes (ORBs). It appears that archaea have a simplified version of the eukaryotic replication apparatus, which has led to considerable interest in the archaeal machinery as a model of that in eukaryotes.

The research on replication origins is important not only in providing insights into the structure and function of the replication origins but also in understanding the regulatory mechanisms of the initiation step in DNA replication. Therefore, intensive studies have been carried out in the last two decades. The pioneer work to identify bacterial oriCs in silico is the GC-skew analysis. Later, a method of cumulative GC skew without sliding windows was proposed to give better resolution. Meanwhile, an oligomer-skew method was also proposed to predict oriC regions in bacterial genomes. As a unique representation of a DNA sequence, the Z-curve method has been proved to be an accurate and effective approach to predict bacterial and archaeal replication origins. Budding yeast origins have been predicted by Oriscan using similarity to the characterized ones, while the fission yeast origins have been identified initially from AT content calculation. In comparison with the in silico analysis, the experimental methods are time-consuming and labor-intensive, but convincing and reliable. To identify microbial replication origins in vivo or in vitro, a number of experimental methods have been used including construction of replicative oriC plasmids, microarray-based or high-throughput sequencing-based marker frequency analysis, two-dimensional gel electrophoresis analysis and replication initiation point mapping (RIP mapping). The recent genome-wide approaches to identify and characterize replication origin locations have boosted the number of mapped yeast replication origins. In addition, the availability of increasing complete microbial genomes and emerging approaches has created challenges and opportunities for identification of their replication origins in silico, as well as in vivo and in vitro.

The Frontiers in Microbiology Research Topic on DNA replication origins in microbial genomes is devoted to address the issues mentioned above, and aims to provide a comprehensive overview of current research in this field.

Citation: Gao, F., ed. (2016). DNA Replication Origins in Microbial Genomes. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-779-8

Dedicated to the 120th Anniversary of
Tianjin University (formerly Peiyang University),
the first modern higher education university in China

Table of Contents

06 ***Editorial: DNA Replication Origins in Microbial Genomes***
Feng Gao

08 ***DNA replication origins in archaea***
Zhenfang Wu, Jingfang Liu, Haibo Yang and Hua Xiang

15 ***Ori-Finder 2, an integrated tool to predict replication origins in the archaeal genomes***
Hao Luo, Chun-Ting Zhang and Feng Gao

21 **oriC*-encoded instructions for the initiation of bacterial chromosome replication***
Marcin Wolański, Rafał Donczew, Anna Zawilak-Pawlik and Jolanta Zakrzewska-Czerwińska

35 ***Redefining bacterial origins of replication as centralized information processors***
Gregory T. Marczynski, Thomas Rolain and James A. Taylor

45 ***Control regions for chromosome replication are conserved with respect to sequence and location among* Escherichia coli *strains***
Jakob Frimodt-Møller, Godefroid Charbon, Karen A. Krogfelt and Anders Løbner-Olesen

60 ***Bacteria may have multiple replication origins***
Feng Gao

64 ***Choosing a suitable method for the identification of replication origins in microbial genomes***
Chengcheng Song, Shaocun Zhang and He Huang

82 ***Identification of the Replication Origins from* Cyanothece *ATCC 51142 and Their Interactions with the DnaA Protein: From* In Silico *to* In Vitro *Studies***
He Huang, Cheng-Cheng Song, Zhi-Liang Yang, Yan Dong, Yao-Zhong Hu and Feng Gao

93 ***Sequence analysis of origins of replication in the* Saccharomyces cerevisiae *genomes***
Wen-Chao Li, Zhe-Jin Zhong, Pan-Pan Zhu, En-Ze Deng, Hui Ding, Wei Chen and Hao Lin

99 ***An integrated overview of spatiotemporal organization and regulation in mitosis in terms of the proteins in the functional supercomplexes***
Yueyuan Zheng, Junjie Guo, Xu Li, Yubin Xie, Mingming Hou, Xuyang Fu, Shengkun Dai, Rucheng Diao, Yanyan Miao and Jian Ren

109 ***Recent advances in the genome-wide study of DNA replication origins in yeast***
Chong Peng, Hao Luo, Xi Zhang and Feng Gao

frontiers in Microbiology

EDITORIAL
published: 08 January 2016
doi: 10.3389/fmicb.2015.01545

Editorial: DNA Replication Origins in Microbial Genomes

Feng Gao [1, 2, 3]*

[1] Department of Physics, Tianjin University, Tianjin, China, [2] Key Laboratory of Systems Bioengineering, Ministry of Education, Tianjin University, Tianjin, China, [3] SynBio Research Platform, Collaborative Innovation Center of Chemical Science and Engineering, Tianjin, China

Keywords: archaea, bacteria, yeast, replication origin, DNA replication, replication regulation, orisome, regulatory proteins

The Editorial on the Research Topic

DNA Replication Origins in Microbial Genomes

In all three domains of life, DNA replication initiates on defined genome sites, termed replication origins. In bacteria, replication typically initiates from a single replication origin (*oriC*). In eukaryotic genomes, replication initiates from significantly more replication origins, ranging from hundreds in yeast to tens of thousands in human (Gao et al., 2012). Within the archaeal domain, multiple replication origins have been identified in *Sulfolobus* species, haloarchaea etc. (Lundgren et al., 2004; Robinson et al., 2004; Wu et al.; Yang et al., 2015). The research on replication origins is important not only in providing insights into the structure and function of the replication origins but also in understanding the regulatory mechanisms of the initiation step in DNA replication. Therefore, intensive studies, by *in silico* analyses as well as *in vivo* and *in vitro* experiments, have been carried out in the last two decades.

Based on the sequence-derived features, various *in silico* approaches have been developed to identify microbial replication origins (Frank and Lobry, 2000; Breier et al., 2004; Mackiewicz et al., 2004; Zhang and Zhang, 2005; Worning et al., 2006; Gao and Zhang, 2007, 2008a; Gao et al., 2013; Gao, 2014). For example, the locations of replication origins sites have been predicted for thousands of bacterial genomes by Ori-Finder, a web-based system for finding *oriC*s in bacterial genomes (Gao and Zhang, 2007; Gao et al., 2013). A new version of Ori-Finder for archaea, Ori-Finder 2, has been developed to predict *oriC*s in archaeal genomes automatically (Luo et al.). To confirm the predicted replication origins, it is important to choose a most suitable experimental strategy. Song et al. summarize the main existing experimental methods to determine the replication origin regions and their practical applications (Song et al.). As a study from *in silico* to *in vitro*, the experimental supports are provided for the identified replication origins in *Cyanothece* ATCC 51142 (Gao and Zhang, 2008b), and their interactions with the initiator protein DnaA (Huang et al.).

In spite of a great variety of origin sequences across species, all bacterial replication origins contain the information necessary to guide assembly of the DnaA protein complex at *oriC*, triggering the unwinding of DNA and the beginning of replication. Therefore, *oriC*-encoded instructions should be interpreted particularly in the context of replication initiation and its regulation (Wolanski et al.). Wolanski et al. show that *oriC*-encoded instructions allow not only for initiation but also for precise regulation of replication initiation and coordination of chromosomal replication with the cell cycle (also in response to environmental signals; Wolanski et al.). Frimodt-Moller et al. find control regions for chromosome replication are conserved with respect to sequence and location among *Escherichia coli* strains (Frimodt-Moller et al.). Based on the single origin usage strategy that distinguishes bacteria, Marczynski et al. redefine the bacterial origins

OPEN ACCESS

Edited and reviewed by:
John R. Battista, Louisiana State University, USA

***Correspondence:**
Feng Gao
fgao@tju.edu.cn

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *18 November 2015*
Accepted: *21 December 2015*
Published: *08 January 2016*

Citation:
Gao F (2016) Editorial: DNA Replication Origins in Microbial Genomes. Front. Microbiol. 6:1545. doi: 10.3389/fmicb.2015.01545

as centralized information processors, and describe how negative-feedback, phospho-relay, and chromosome-partitioning systems act to regulate chromosome replication (Marczynski et al.). On the other hand, the *in silico* analyses show that some bacteria, although very few, may have multiple origins of replication per chromosome (Frank et al., 2015; Gao), and the recent work also suggests that there are multiple replication origins in *Synechocystis* that fire asynchronously, as in eukaryotic nuclear chromosomal replication (Ohbayashi et al., 2015).

For eukaryotic organisms, replication origins are best characterized in the unicellular eukaryote budding yeast *Saccharomyces cerevisiae* and the fission yeast *Schizosaccharomyces pombe*. With the recent development of genome-wide approaches, the number of yeast species involved in ORIs research has increased dramatically, which has created opportunities for the sequence, protein, and comparative analysis of replication origins in yeast genomes (Li et al.; Zheng et al.; Peng et al.).

The Frontiers in Microbiology Research Topic on DNA replication origins in microbial genomes is devoted to address the issues mentioned above, and aims to provide a comprehensive overview of the current research in this field.

DEDICATION

This article is dedicated to the 120th Anniversary of Tianjin University (formerly Peiyang University), the first modern higher education university in China.

ACKNOWLEDGMENTS

The author would like to thank Prof. Chun-Ting Zhang for the invaluable assistance and inspiring discussions, the international editors and reviewers from over 10 countries for their excellent assistance and constructive comments on the manuscripts, and the Frontiers in Microbiology team for their continued support and assistance. The present work was supported in part by National Natural Science Foundation of China (Grant Nos. 31571358, 31171238, 30800642, and 10747150), Program for New Century Excellent Talents in University (No. NCET-12-0396), and the China National 863 High-Tech Program (2015AA020101).

REFERENCES

Breier, A. M., Chatterji, S., and Cozzarelli, N. R. (2004). Prediction of *Saccharomyces cerevisiae* replication origins. *Genome Biol.* 5:R22. doi: 10.1186/gb-2004-5-4-r22

Frank, A. C., and Lobry, J. R. (2000). Oriloc: prediction of replication boundaries in unannotated bacterial chromosomes. *Bioinformatics* 16, 560–561. doi: 10.1093/bioinformatics/16.6.560

Frank, O., Göker, M., Pradella, S., and Petersen, J. (2015). Ocean's Twelve: flagellar and biofilm chromids in the multipartite genome of *Marinovum algicola* DG898 exemplify functional compartmentalization. *Environ. Microbiol.* 17, 4019–4034. doi: 10.1111/1462-2920.12947

Gao, F. (2014). Recent advances in the identification of replication origins based on the Z-curve method. *Curr. Genomics* 15, 104–112. doi: 10.2174/1389202915999140328162938

Gao, F., Luo, H., and Zhang, C. T. (2012). DeOri: a database of eukaryotic DNA replication origins. *Bioinformatics* 28, 1551–1552. doi: 10.1093/bioinformatics/bts151

Gao, F., Luo, H., and Zhang, C. T. (2013). DoriC 5.0: an updated database of *oriC* regions in both bacterial and archaeal genomes. *Nucleic Acids Res.* 41, D90–D93. doi: 10.1093/nar/gks990

Gao, F., and Zhang, C. T. (2007). DoriC: a database of *oriC* regions in bacterial genomes. *Bioinformatics* 23, 1866–1867. doi: 10.1093/bioinformatics/btm255

Gao, F., and Zhang, C. T. (2008a). Ori-Finder: a web-based system for finding *oriC*s in unannotated bacterial genomes. *BMC Bioinformatics* 9:79. doi: 10.1186/1471-2105-9-79

Gao, F., and Zhang, C. T. (2008b). Origins of replication in *Cyanothece* 51142. *Proc. Natl. Acad. Sci. U.S.A.* 105, E125; author reply E126-127. doi: 10.1073/pnas.0809987106

Lundgren, M., Andersson, A., Chen, L., Nilsson, P., and Bernander, R. (2004). Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. *Proc. Natl. Acad. Sci. U.S.A.* 101, 7046–7051. doi: 10.1073/pnas.0400656101

Mackiewicz, P., Zakrzewska-Czerwinska, J., Zawilak, A., Dudek, M. R., and Cebrat, S. (2004). Where does bacterial replication start? Rules for predicting the oriC region. *Nucleic Acids Res.* 32, 3781–3791. doi: 10.1093/nar/gkh699

Ohbayashi, R., Watanabe, S., Ehira, S., Kanesaki, Y., Chibazakura, T., and Yoshikawa, H. (2015). Diversification of DnaA dependency for DNA replication in cyanobacterial evolution. *ISME J.* doi: 10.1038/ismej.2015.194. [Epub ahead of print].

Robinson, N. P., Dionne, I., Lundgren, M., Marsh, V. L., Bernander, R., and Bell, S. D. (2004). Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell* 116, 25–38. doi: 10.1016/S0092-8674(03)01034-1

Worning, P., Jensen, L. J., Hallin, P. F., Staerfeldt, H. H., and Ussery, D. W. (2006). Origin of replication in circular prokaryotic chromosomes. *Environ. Microbiol.* 8, 353–361. doi: 10.1111/j.1462-2920.2005.00917.x

Yang, H., Wu, Z., Liu, J., Liu, X., Wang, L., Cai, S., et al. (2015). Activation of a dormant replication origin is essential for *Haloferax mediterranei* lacking the primary origins. *Nat. Commun.* 6, 8321. doi: 10.1038/ncomms9321

Zhang, R., and Zhang, C. T. (2005). Identification of replication origins in archaeal genomes based on the Z-curve method. *Archaea* 1, 335–346. doi: 10.1155/2005/509646

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 © 2016 Gao. 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.

frontiers in
MICROBIOLOGY

REVIEW ARTICLE
published: 29 April 2014
doi: 10.3389/fmicb.2014.00179

DNA replication origins in archaea

Zhenfang Wu[1,2], Jingfang Liu[1]*, Haibo Yang[1,2] and Hua Xiang[1]*

[1] State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China
[2] University of Chinese Academy of Sciences, Beijing, China

Edited by:
Feng Gao, Tianjin University, China

Reviewed by:
Jonathan H. Badger, J. Craig Venter Institute, USA
Qunxin She, University of Copenhagen, Denmark

****Correspondence:***
Hua Xiang and Jingfang Liu, State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, No. 1 Beichen West Road, Chaoyang District, Beijing 100101, China
e-mail: xiangh@im.ac.cn; liujf@im.ac.cn

DNA replication initiation, which starts at specific chromosomal site (known as replication origins), is the key regulatory stage of chromosome replication. Archaea, the third domain of life, use a single or multiple origin(s) to initiate replication of their circular chromosomes. The basic structure of replication origins is conserved among archaea, typically including an AT-rich unwinding region flanked by several conserved repeats (origin recognition box, ORB) that are located adjacent to a replication initiator gene. Both the ORB sequence and the adjacent initiator gene are considerably diverse among different replication origins, while *in silico* and genetic analyses have indicated the specificity between the initiator genes and their cognate origins. These replicator–initiator pairings are reminiscent of the *oriC-dnaA* system in bacteria, and a model for the negative regulation of origin activity by a downstream cluster of ORB elements has been recently proposed in haloarchaea. Moreover, comparative genomic analyses have revealed that the mosaics of replicator-initiator pairings in archaeal chromosomes originated from the integration of extrachromosomal elements. This review summarizes the research progress in understanding of archaeal replication origins with particular focus on the utilization, control and evolution of multiple replication origins in haloarchaea.

Keywords: DNA replication origin, origin recognition box, archaea, control, evolution, haloarchaea

INTRODUCTION

DNA replication is a fundamental cellular process that is functionally conserved across all three domains of life (bacteria, archaea, and eukaryote). The precise regulation of DNA replication ensures the accurate duplication of genomic information, and replication initiation is the first and most important stage of this regulation. The first model of DNA replication initiation was proposed for *Escherichia coli* in 1963, postulating that a trans-acting factor binds to a cis-acting site which triggers replication initiation (Jacob et al., 1963). In the subsequent 50 years, this "replicon model" has been demonstrated to be essentially true in all organisms, and the cis-acting site is now known as the replication origin. Bacterial chromosomes are typically replicated from a single origin, whereas the replication of eukaryotic chromosomes initiates from a number of discrete origins (Leonard and Mechali, 2013). DNA replication origins have been well-defined in bacteria and unicellular eukaryotes, and relative topics are covered in a number of excellent reviews (Messer, 2002; Mott and Berger, 2007; Zakrzewska-Czerwinska et al., 2007; Mechali, 2010; Aparicio, 2013). In contrast, focus on DNA replication origins in archaea, the third domain of life, commenced only approximately a decade ago. DNA replication origins have been mapped primarily for a few representatives of archaeal species distributed in the three main phyla, Euryarchaeota, Crenarchaeota, and Thaumarchaeota (Myllykallio et al., 2000; Lundgren et al., 2004; Robinson et al., 2004; Grainge et al., 2006; Norais et al., 2007; Majernik and Chong, 2008; Coker et al., 2009; Pelve et al., 2012, 2013; Wu et al., 2012, 2014). In addition, more detailed characterization has been revealed in several model systems, such as *Pyrococcus* species (Myllykallio et al., 2000; Matsunaga et al., 2001, 2003), *Sulfolobus* species (Lundgren et al., 2004; Robinson et al., 2004; Duggin et al., 2008; Samson et al., 2013), *Haloferax volcanii* (Norais et al., 2007; Hawkins et al., 2013) and *Haloarcula hispanica* (Wu et al., 2012, 2014). It is now known that archaea use a single or multiple origin(s) to replicate their circular chromosomes (Kelman and Kelman, 2004; Robinson and Bell, 2005; Hyrien et al., 2013). The multiple origins together with their adjacent initiator genes in certain archaeal chromosomes may have arisen from the capture of extrachromosomal elements and appear to be mosaics of distinct replicator–initiator pairings (Robinson and Bell, 2007; Wu et al., 2012).

This replicator–initiator system consists of an origin region and an initiator gene (the *cdc6* gene in most cases and *whiP* in the *oriC3* of *Sulfolobus* species). The origin region usually has a high content of adenine and thymine residues (AT-rich) flanked by several conserved repeated motifs known as origin recognition boxes (ORBs). The initiator protein Cdc6 (also denoted Orc or Orc1/Cdc6 in other papers) shows homology to both Orc1 and Cdc6 of eukaryotes and therefore is considered to be involved in both specific recognition of the origin region and loading of the minichromosome maintenance helicase (MCM; Robinson and Bell, 2005). Despite the conservation of the replicator-initiator structure, archaeal replication origins exhibit considerable diversity in terms of both the ORB elements within different origins and their adjacent initiator genes. The specificity of the initiator genes and their cognate origins was recently established by means of *in silico* and genetic analyses in *Sulfolobus* species (Samson et al., 2013) and *Haloarcula hispanica* (Wu et al., 2012, 2014). The *cis* organization of the

replication origin and the initiator gene (replicator–initiator) is reminiscent of the *oriC-dnaA* system in bacteria. Recently, we revealed that bacterial-like control mechanisms may be used by different replication origins in haloarchaea, and a model has been proposed for the negative regulation of *oriC2* by a downstream cluster of ORB elements in *Haloarcula hispanica* (Wu et al., 2014).

The goal of this review is to present an overview of progress made over the past decade in our understanding of DNA replication origins of archaeal genomes, including the identification (mapping), characterization and evolution of multiple replication origins on the chromosomes. We focus on the utilization and control of multiple replication origins in haloarchaea, as well as comparisons of replication origins from different archaeal species to draw the generality and evolution of multiple replication origins in archaea.

IDENTIFICATION (MAPPING) OF REPLICATION ORIGINS

Similar to bacteria, archaea have simple circular chromosomes (and also several extrachromosomal elements in some archaea); however, many archaea characterized to date harbor multiple replication origins. The approaches developed in bacteria or eukaryotes have been employed to investigate replication origins in archaea, such as GC-skew analysis (Myllykallio et al., 2000), the Z-curve method (Zhang and Zhang, 2003), autonomously replicating sequence (ARS) assay (Berquist and DasSarma, 2003; Norais et al., 2007; Wu et al., 2012), replication initiation point mapping (RIP mapping; Matsunaga et al., 2003), two-dimensional gel analysis (Matsunaga et al., 2001; Robinson et al., 2004), and marker frequency analysis (MFA; Lundgren et al., 2004; Coker et al., 2009; Pelve et al., 2012; Hawkins et al., 2013; Wu et al., 2014). DNA replication origins have been mapped in about a dozen archaeal species.

SINGLE REPLICATION ORIGIN IN *Pyrococcus* SPECIES

The first description of DNA replication origins of archaeal genomes was reported by Myllykallio et al. (2000). These researchers identified a single replication origin (*oriC*) in *Pyrococcus abyssi* by means of cumulative skew of GGGT, and the study found that the *oriC* is flanked with the only *cdc6* gene and several eukaryotic-like replication genes (Myllykallio et al., 2000). The origin organization was observed to be highly conserved in two other *Pyrococcus* species, *Pyrococcus horikoshii* and *Pyrococcus furiosus* (Myllykallio et al., 2000). The *oriC* was then experimentally confirmed via two-dimensional (2D) gel analysis (Matsunaga et al., 2001) and RIP mapping (Matsunaga et al., 2003), and the studies demonstrated that the *oriC* has several repeated sequences (now named ORBs) and is directly upstream of the *cdc6* gene, reminiscent of the *oriC-dnaA* origin system in bacteria. Furthermore, the specific interaction of the Cdc6 protein with the *oriC* was detected via chromatin immunoprecipitation assays (Matsunaga et al., 2001). All of these data indicated that the circular chromosome of the *Pyrococcus* species is bidirectionally replicated from a bacterial mode of replication origin by eukaryotic-type machinery (**Figure 1A**).

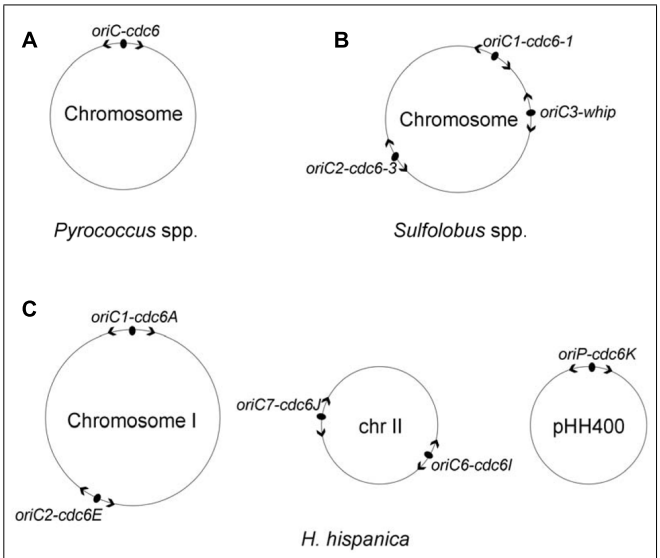


FIGURE 1 | Distribution of DNA replication origins in three well-studied archaeal model systems, *Pyrococcus* species (A), *Sulfolobus* species (B) and *Haloarcula hispanica* (C). Origins are indicated with filled ovals, and arrowheads designate bidirectional replication. Replicator-initiator indicates that each origin is specified by its proximally encoded initiator. Both *Pyrococcus* species and *Sulfolobus* species contain a single chromosome; the chromosome of *Pyrococcus* species is replicated from a single origin (Myllykallio et al., 2000), whereas the chromosome of *Sulfolobus* species is replicated from three origins in near synchrony (Duggin et al., 2008). The *Haloarcula hispanica* genome consists of a main chromosome and two extrachromosomal elements with five active replication origins: *oriC1-cdc6A* and *oriC2-cdc6E* in the main chromosome I, *oriC6-cdc6I* and *oriC7-cdc6J* in the minichromosome II, and *oriP-cdc6K* in the megaplasmid pHH400 (Wu et al., 2012).

THREE REPLICATION ORIGINS IN *Sulfolobus* SPECIES

The first example of archaeal chromosomes with multiple replication origins was the identification of three replication origins in the single chromosome of *Sulfolobus* species using 2D gel analysis (Robinson et al., 2004, 2007) and microarray-based MFA (Lundgren et al., 2004), and the results demonstrated that bidirectional replication initiated from three origins in both *Sulfolobus acidocaldarius* and *Sulfolobus solfataricus* (*oriC1*, *oriC2*, and *oriC3*; **Figure 1B**). The *oriC1* and *oriC2*, in each species, are located directly upstream of *cdc6-1* and *cdc6-3*, respectively, which have previously been identified by 2D gel electrophoresis in *S. solfataricus* (Robinson et al., 2004). The third origin, *oriC3*, is adjacent to the *whiP* (Winged-helix initiator protein) gene that is related to the eukaryotic *cdt1* gene. An origin comparison between *Aeropyrum* and *Sulfolobus* suggested that the *oriC3-whiP* might have originated from the capture of extrachromosomal elements (Robinson and Bell, 2007). Using synchronized cultures, MFA results indicated that all three origins fire once per cell cycle and are initiated in near synchrony but with a slightly later activation of *oriC2* (Lundgren et al., 2004; Duggin et al., 2008). Recently, three replication origins were also mapped in another *Sulfolobus* species, *Sulfolobus islandicus*, and a combination of genetic and MF analyses demonstrated that the

three origins are specifically dependent on their adjacent initiator genes (two *cdc6* genes and one *whiP* gene; Samson et al., 2013).

MULTIPLE REPLICATION ORIGINS IN HALOARCHAEA

Haloarchaeal genomes are generally composed of multiple genetic elements (chromosome, minichromosome, and megaplasmids) with multiple Cdc6 homologs (Capes et al., 2011), which is suggestive of the occurrence of multiple replication origins. Recently, multiple replication origins were predicted in 15 completely sequenced haloarchaeal genomes by searching for putative ORBs associated with *cdc6* genes (Wu et al., 2012), and active replication origins have been experimentally studied in three model systems, *Halobacterium* sp. NRC-1 (Berquist and DasSarma, 2003; Coker et al., 2009), *Haloferax volcanii* (Norais et al., 2007; Hawkins et al., 2013) and *Haloarcula hispanica* (Wu et al., 2012, 2014).

The first prediction of multiple DNA replication origins in haloarchaeal genomes came from Z curve method analysis of the genome of *Halobacterium* sp. NRC-1, which revealed two *cdc6*-adjacent replication origins in its chromosome (Zhang and Zhang, 2003). However, only one replication origin was verified to have ARS activity (Berquist and DasSarma, 2003). Whole-genome MFA was employed to map the activation of replication origins *in vivo* in *Halobacterium* sp. NRC-1, which demonstrated multiple discrete origin sites in the chromosome, with two being located in the vicinity of *cdc6* genes (denoted *orc7* and *orc10* in the original paper; Coker et al., 2009).

Eleven *cdc6* genes are encoded in *Haloarcula hispanica*, and eight of them have surrounding ORB-like elements. ARS activity assays demonstrated that only five predicted origins, *oriC1-cdc6A* and *oriC2-cdc6E* in the main chromosome, *oriC6-cdc6I*, and *oriC7-cdc6J* in the minichromosome and *oriP-cdc6K* in the megaplasmid (pHH400), were able to confer replication ability to a non-replicating plasmid (**Figure 1C**; Wu et al., 2012). Recently, we combined extensive gene deletion and microarray-based MFA to map the activation of replication origins *in vivo* in *Haloarcula hispanica*, demonstrating that the chromosome is bidirectionally replicated from the two initially proven origins, *oriC1-cdc6A*, and *oriC2-cdc6E* (Wu et al., 2014). Importantly, our results indicated that one active *ori-cdc6* pairing on each replicon, i.e., *oriC1-cdc6A* or *oriC2-cdc6E* on the chromosome, *oriC6-cdc6I* or *oriC7-cdc6J* on the minichromosome, and *oriP-cdc6K* on pHH400, is essential for genome replication in *Haloarcula hispanica* (Wu et al., 2014).

Five replication origins were initially identified in *Haloferax volcanii* using a combination of bioinformatics and genetic approaches: two within the chromosome and one each within the three megaplasmids pHV1, pHV3, and pHV4 (Norais et al., 2007). Recently, aside from the previously identified origins, a sixth replication origin was mapped in the chromosome via high-throughput sequencing-based MFA (Hawkins et al., 2013). All six replication origins are adjacent to *cdc6* genes. Furthermore, four chromosomal replication origins were mapped in the laboratory H26 strain with integration of pHV4 into the chromosome (Hawkins et al., 2013). Surprisingly, the four origins can be deleted simultaneously, and in the absence of these replication origins, the strain even grew 7.5% faster than the wild-type strain (Hawkins et al., 2013). Because the *radA* gene (the archaeal *recA*/*rad51* homologue) was determined to be essential in the absence of all four origins, the authors proposed that the replication of the origin-less *Haloferax volcanii* chromosome is dependent on homologous recombination (Hawkins et al., 2013). However, this mode of recombination-dependent replication of the *Haloferax volcanii* chromosome was not yet observed in other investigated archaea. In contrast, at least one active replication origin has been proven to be essential for chromosome replication in *Haloarcula hispanica* (Wu et al., 2014), and triple-deletion mutant was not available for the three initiators in the chromosome of *S. islandicus* (Samson et al., 2013). It would be interesting to investigate how the RadA-dependent replication (if any) efficiently replicates the *Haloferax volcanii* chromosome, or if there are undetected replication origins functioned in the chromosome lacking the main origins.

MAPPING OF REPLICATION ORIGINS IN OTHER ARCHAEA

DNA replication origins have been well-defined in several bacterial model systems, and have been predicted and/or identified in more than 1300 bacterial genomes (Gao and Zhang, 2007, 2008). Similarly, to understand the general nature of replication origins in archaea, it is necessary to determine and compare replication origins from a broad selection of archaeal species. Fortunately, the genomes of 100s of archaea distributed in different phyla have been sequenced and are publically available, allowing the prediction and mapping of replication origins in these genomes. To date, replication origins have been demonstrated in a dozen archaeal species. Similar to *Pyrococcus* species, *Archaeoglobus fulgidus* has been shown to contain a single replication origin (Maisnier-Patin et al., 2002). Two replication origins have been identified in *Aeropyrum pernix* by using a combination of biochemical and two-dimensional gel electrophoresis (Grainge et al., 2006; Robinson and Bell, 2007). Studies of DNA replication in methanogens have demonstrated that a single origin is responsible for replication initiation of the chromosome of *Methanothermobacter thermautotrophicus* (Capaldi and Berger, 2004; Majernik and Chong, 2008). Recently, four replication origins were mapped in the single chromosome of *Pyrobaculum calidifontis* via high-throughput sequencing-based MFA (Pelve et al., 2012). To generate a broader view of modes of origin replication in archaea, Pelve et al. (2013) further completed origin mapping in a thaumarchaeon, revealing a single replication origin in the *Nitrosopumilus maritimus* chromosome.

DISTINCT REPLICATOR-INITIATOR SYSTEMS IN ARCHAEA

The initiator protein DnaA is highly conserved in bacteria, and bacterial replication origins generally possess conserved sequence elements, DnaA boxes. In contrast, the three replication origins in *Sulfolobus* species differ from each other. Each of the three origins is specifically recognized by its proximally encoded initiator protein, two distinct Cdc6 proteins and WhiP (Dueber et al., 2011; Samson et al., 2013). In addition, the recognition mechanisms appear to be different, as classic ORB and its shorter version (miniORB) are, respectively, observed in the *oriC1* and *oriC2* regions, while neither is observed in the *oriC3* region (Robinson et al., 2004; Samson et al., 2013).

Haloarchaeal genomes generally contain multiple *cdc6* genes and replication origins. Recently, we conducted a comparison of the origin-associated Cdc6 homologs and the corresponding predicted ORB elements. Our results suggested that the replication origins from haloarchaeon are notably diverse in terms of ORB elements and their adjacent *cdc6* genes, which could be sorted into distinct families. Based on this phylogenetic analysis, linkage-specificity of Cdc6 homologs and the corresponding ORB elements was proposed, suggestive of their specific interaction (Wu et al., 2012). Very recently, we employed comprehensive genetic studies to investigate the specificity of multiple replication origins and *cdc6* genes in *Haloarcula hispanica*, and our results indicated that each Cdc6 protein specifically recognizes its proximal origin (Wu et al., 2014). Thus, multiple replication origins along with their adjacent *cdc6* genes appear to be distinct *ori-cdc6* systems. These distinct *ori-cdc6* systems in haloarchaeon may have many evolutionary advantages: first, it ensures the compatibility of multiple replication origins, which accounts for the observations that multiple Cdc6 proteins from a haloarchaeal genome are distributed into different families (Wu et al., 2012) and that the *oriC2*-containing plasmid is incompatible with *Haloarcula hispanica* (Wu et al., 2014); second, distinct *ori-cdc6* pairings help minimize competition among multiple origins for initiators and maintain independent control of replication initiation at different origins. Importantly, as haloarchaeal genomes generally contain multiple replicons, distinct *ori-cdc6* origins may be favorable for replicon-specific replication control, similar to the different modes of replication origin adopted by the two chromosomes of *Vibrio cholerae* (Egan and Waldor, 2003).

To understand the molecular mechanisms involved in the specific recognition of origins by initiators, the structures of two origin-bound Cdc6 proteins from *Aeropyrum pernix* (Gaudier et al., 2007) and *S. solfataricus* (Dueber et al., 2007) were crystallized. Both of the two Cdc6 proteins contain an N-terminal AAA^+ domain and a C-terminal WH domain. Intriguingly, both of the studies demonstrated that, in addition to the canonical DNA binding WH domain, the AAA^+ domains of these two initiators are responsible for recognizing origins (Dueber et al., 2007; Gaudier et al., 2007). Subsequently, biochemical data also demonstrated that both the WH domain and AAA^+ domain contribute to the origin-binding specificity of the Cdc6 protein (Dueber et al., 2011).

CONTROL OF REPLICATION INITIATION AT MULTIPLE ORIGINS IN ARCHAEA

Multiple mechanisms that regulate replication initiation have been well-characterized in both bacteria and unicellular eukaryotes, and are summarized in a number of excellent reviews (Mott and Berger, 2007; Mechali, 2010; Rajewska et al., 2012; Aparicio, 2013). In contrast, the mechanisms by which archaea regulate replication initiation at multiple origins, either on the same chromosome or from different genetic elements, are far less understood. All of the archaeal replication origins characterized to date are dependent on their adjacent initiator gene (the *cdc6* gene in most cases; Samson et al., 2013; Wu et al., 2014), and these distinct *ori-cdc6* pairings may contribute to their independent control. In addition, the *cis* location of the *cdc6* gene and the origin is proved to not be required for ARS activity in both *Haloferax volcanii* and *Haloarcula hispanica* (Norais et al., 2007; Wu et al., 2014). Therefore, we have proposed that direct linkage of the initiator gene to the origin may facilitate its transcription after replication initiation to sequentially control its cognate origin.

Using the *Haloarcula hispanica* model system, we suggested that some bacterial-like mechanisms may be employed at different replication origins in haloarchaea (Wu et al., 2014). A G-rich inverted-repeat directly inside each ORB element of *Haloarcula hispanica oriC1* was shown to be a replication enhancer that stimulated origin activation at *oriC1*. Because of the repeat's close location to ORB elements, we proposed that the G-rich inverted-repeat enhances the binding of initiator or regulatory factors at *oriC1*, similar to many repeated sequences in bacteria that are binding sites for initiation proteins or regulatory factors, playing a crucial role in the control of replication initiation (Rajewska et al., 2012). In addition, a model has been proposed, and partly tested, for the negative regulation of *oriC2* by a downstream cluster of Cdc6 binding elements (ORBs), likely *via* Cdc6E titration, similar to the negative control of replication initiation via a *datA* locus exhibiting DnaA-titrating activity in *E. coli* (Kitagawa et al., 1998). More interestingly, many additional predicted replication origins have the *oriC2*-like structure, suggesting that this strategy of negative replication origin control is used generally by haloarchaea.

Despite the bacterial-like structure of archaeal replication origins, archaea use eukaryotic-type replication machinery (Robinson and Bell, 2005), indicating that archaea may adopt eukaryotic-like mechanisms to control replication proteins and thus replication initiation. Interestingly, genome-wide transcription mapping indicated that serine–threonine protein kinases show cyclic induction in *Sulfolobus* species, indicating that regulatory factors similar to eukaryotic cyclin-dependent kinase (CDK) complexes may be present in archaea (Lundgren and Bernander, 2007). Recently, an ATP-ADP binary switch model for Cdc6-mediated replication control was proposed in *S. islandicus*, postulating that binding of ATP remodels Cdc6 conformation for efficient MCM recruitment, and subsequent ATP hydrolysis renders Cdc6 incapable of further recruiting MCM (Samson et al., 2013). In addition, as almost all replication origins are dependent on Cdc6 proteins, conformational changes of Cdc6 proteins may play important roles in coordinating replication initiation at different origins within a cell.

EVOLUTION OF MULTIPLE REPLICATION ORIGINS IN ARCHAEA

Although considerable diversity of replication origins has been observed in haloarchaea, comparison analysis revealed a conserved replication origin, *oriC1*, which is positioned in the main chromosome of all analyzed haloarchaeal genomes (Coker et al., 2009; Wu et al., 2012). Both the ORBs within *oriC1* and the *oriC1*-associated Cdc6 homologs are highly conserved. In addition, gene order analysis found that genes around *oriC1* are highly syntenic among haloarchaea (**Figure 2**; Capes et al., 2011). Notably, other studies (Robinson et al., 2004; Coker et al., 2009) and our results indicated that the *oriC1* replication origin is broadly conserved in archaea, in terms of both function and structure, which

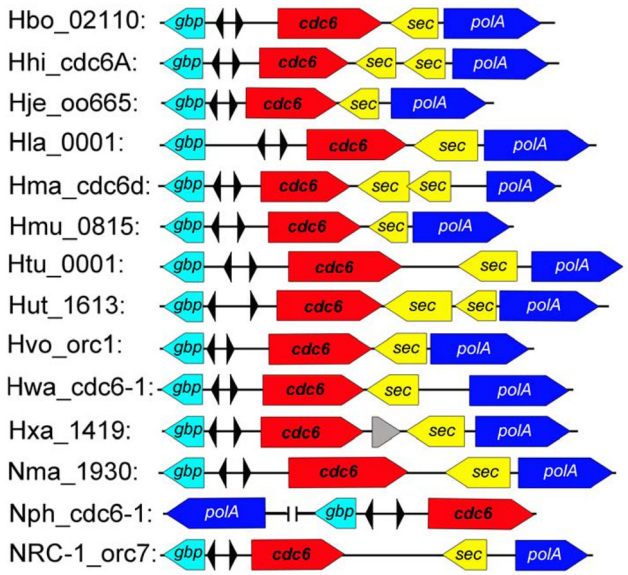


FIGURE 2 | The conserved *oriC1* origin of replication in sequenced haloarchaeal genomes. The *oriC1* context region was mapped as shown in the sequenced haloarchaea. The colored boxed arrows represent different genes as follows: GTP-binding protein (*gbp*, teal), initiator protein (*cdc6*, red), signal sequence peptidase (*sec*, yellow) and DNA-directed DNA polymerase (*polA*, blue). The inverted ORB elements are indicated by small triangles.

strongly suggested that the ancestral chromosome was dependent on *oriC1*. Variations were observed in *oriC1* homologs from different archaeal phyla, which may contribute to the adaptability of archaea to different extreme environments. For example, an extended halophile-specific "G-string" element has been identified at the end of each ORB in haloarchaea, and these "G-string" elements have been proven to be essential for autonomous replication based on the *oriC1* in *Haloarcula hispanica* (Wu et al., 2014).

Multiple replication origins along with their adjacent *cdc6* genes appear to be mosaics of distinct replicator–initiator systems. A comparison between *Aeropyrum* and *Sulfolobus* origins suggested that the capture of extrachromosomal elements accounts for replicon evolution (Robinson and Bell, 2007). In particular, it has been proposed that the three replication origins of the *Sulfolobus* species arose by the integration of extrachromosomal elements into a single-origin ancestral chromosome (*oriC1-cdc6-1*), and the acquisition of *oriC3-whiP* occurred prior to the integration of *oriC2-cdc6-3* (Samson et al., 2013). Similarly, genomic context analyses of *ori-cdc6* systems in haloarchaea revealed that 40% of predicted replication origins were observed with transposases or integrases nearby, indicative of the translocation of a subset of replication origins among haloarchaea. In addition, comparative analyses of the selected replication origins suggested that different evolutionary mechanisms, including ancestral conservation and coupled acquisition and deletion events, may account for the current mosaics of multiple replication origins in the haloarchaeal genomes. Importantly, a comparative genomic analysis of two *Haloarcula* species, *Haloarcula hispanica* and *Haloarcula marismortui*, revealed that the species-specific origins are located in extremely variable regions, suggesting that these novel origins were recently acquired, via either integration into the chromosome or rearrangement of extrachromosomal elements (Wu et al., 2012). Further work may focus on comparisons of replication origins from closely related species to reveal the dynamics of origin evolution and whether origin evolution alters the mode of genomic replication.

PERSPECTIVES

To date, the number of archaea with mapped replication origins is still limited, which to some extent has affected us to get a panoramic view of the generality and evolution of replication origins in archaea. In addition to the mapping of replication origins, the development of prediction algorithms for replication origins in archaeal genomes and the construction of databases with these predicted origins (Gao et al., 2013) will be useful for comparing replication origins from a broader range of archaeal species. Fortunately, the rapid increase in the number of complete archaeal genomic sequences that are publically available will promote our studies of archaeal replication origins.

In addition, the control and coordination of replication initiation at multiple origins in archaea is far less understood. The multireplicon structure of haloarchaeal genomes allows for precise control and coordination of replication initiation at multiple origins. As the chromosome and extrachromosomal elements within a haloarchaeon are generally different sizes and have different copy numbers (Breuert et al., 2006; Liu et al., 2013), it will be interesting to reveal whether they initiate synchronously and how they maintain different copy numbers, as well as what roles multiple replication origins play in governing polyploidy in haloarchaea. In addition, the coordination of multiple origins may play important roles in maintaining the multireplicon structure of haloarchaeal genomes. As most replication origins are dependent on Cdc6 proteins in haloarchaea (excluding the origins of small plasmids), we propose that the coordination of replication initiation at different origins may be obtained by conformational changes of Cdc6 proteins via an ATP-ADP binary switch, which has recently been proposed for chromosome replication in *S. islandicus* (Samson et al., 2013). Thus, more exhaustive work should be taken into account to uncover the control and coordination of the replication initiation from multiple origins, either on the same chromosome or from different genetic elements, in haloarchaeal multireplicon genomes.

ACKNOWLEDGMENTS

This work was partially supported by grants from the National Natural Science Foundation of China (30925001, 31100893, 31271334).

REFERENCES

Aparicio, O. M. (2013). Location, location, location: it's all in the timing for replication origins. *Genes Dev.* 27, 117–128. doi: 10.1101/gad.209999.112

Berquist, B. R., and DasSarma, S. (2003). An archaeal chromosomal autonomously replicating sequence element from an extreme halophile, *Halobacterium* sp.

strain NRC-1. *J. Bacteriol.* 185, 5959–5966. doi: 10.1128/JB.185.20.5959-5966.2003

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

Capaldi, S. A., and Berger, J. M. (2004). Biochemical characterization of Cdc6/Orc1 binding to the replication origin of the euryarchaeon *Methanothermobacter thermoautotrophicus*. *Nucleic Acids Res.* 32, 4821–4832. doi: 10.1093/nar/gkh819

Capes, M. D., Coker, J. A., Gessler, R., Grinblat-Huse, V., Dassarma, S. L., Jacob, C. G., et al. (2011). The information transfer system of halophilic archaea. *Plasmid* 65, 77–101. doi: 10.1016/j.plasmid.2010.11.005

Coker, J. A., Dassarma, P., Capes, M., Wallace, T., Mcgarrity, K., Gessler, R., et al. (2009). Multiple replication origins of *Halobacterium* sp. strain NRC-1: properties of the conserved *orc7*-dependent *oriC1*. *J. Bacteriol.* 191, 5253–5261. doi: 10.1128/JB.00210-09

Dueber, E. C., Costa, A., Corn, J. E., Bell, S. D., and Berger, J. M. (2011). Molecular determinants of origin discrimination by Orc1 initiators in archaea. *Nucleic Acids Res.* 39, 3621–3631. doi: 10.1093/nar/gkq1308

Dueber, E. L., Corn, J. E., Bell, S. D., and Berger, J. M. (2007). Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. *Science* 317, 1210–1213. doi: 10.1126/science.1143690

Duggin, I. G., Mccallum, S. A., and Bell, S. D. (2008). Chromosome replication dynamics in the archaeon *Sulfolobus acidocaldarius*. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16737–16742. doi: 10.1073/pnas.0806414105

Egan, E. S., and Waldor, M. K. (2003). Distinct replication requirements for the two *Vibrio cholerae* chromosomes. *Cell* 114, 521–530. doi: 10.1016/S0092-8674(03)00611-1

Gao, F., Luo, H., and Zhang, C. T. (2013). DoriC 5.0: an updated database of *oriC* regions in both bacterial and archaeal genomes. *Nucleic Acids Res.* 41, D90–D93. doi: 10.1093/nar/gks990

Gao, F., and Zhang, C. T. (2007). DoriC: a database of *oriC* regions in bacterial genomes. *Bioinformatics* 23, 1866–1867. doi: 10.1093/bioinformatics/btm255

Gao, F., and Zhang, C. T. (2008). Ori-Finder: a web-based system for finding *oriC*s in unannotated bacterial genomes. *BMC Bioinformatics* 9:79. doi: 10.1186/1471-2105-9-79

Gaudier, M., Schuwirth, B. S., Westcott, S. L., and Wigley, D. B. (2007). Structural basis of DNA replication origin recognition by an ORC protein. *Science* 317, 1213–1216. doi: 10.1126/science.1143664

Grainge, I., Gaudier, M., Schuwirth, B. S., Westcott, S. L., Sandall, J., Atanassova, N., et al. (2006). Biochemical analysis of a DNA replication origin in the archaeon *Aeropyrum pernix*. *J. Mol. Biol.* 363, 355–369. doi: 10.1016/j.jmb.2006.07.076

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

Hyrien, O., Rappailles, A., Guilbaud, G., Baker, A., Chen, C. L., Goldar, A., et al. (2013). From simple bacterial and archaeal replicons to replication N/U-domains. *J. Mol. Biol.* 425, 4673–4689. doi: 10.1016/j.jmb.2013.09.021

Jacob, F., Cuzin, F., and Brenner, S. (1963). On Regulation of DNA Replication in Bacteria. *Cold. Spring Harb. Symp. Quant. Biol.* 28, 329–348. doi: 10.1101/SQB.1963.028.01.048

Kelman, L. M., and Kelman, Z. (2004). Multiple origins of replication in archaea. *Trends Microbiol.* 12, 399–401. doi: 10.1016/j.tim.2004.07.001

Kitagawa, R., Ozaki, T., Moriya, S., and Ogawa, T. (1998). Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for *Escherichia coli* DnaA protein. *Genes Dev.* 12, 3032–3043. doi: 10.1101/gad.12.19.3032

Leonard, A. C., and Mechali, M. (2013). DNA replication origins. *Cold Spring Harb. Perspect. Biol.* 5:a010116. doi: 10.1101/cshperspect.a010116

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

Lundgren, M., Andersson, A., Chen, L., Nilsson, P., and Bernander, R. (2004). Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. *Proc. Natl. Acad. Sci. U.S.A.* 101, 7046–7051. doi: 10.1073/pnas.0400656101

Lundgren, M., and Bernander, R. (2007). Genome-wide transcription map of an archaeal cell cycle. *Proc. Natl. Acad. Sci. U.S.A.* 104, 2939–2944. doi: 10.1073/pnas.0611333104

Maisnier-Patin, S., Malandrin, L., Birkeland, N. K., and Bernander, R. (2002). Chromosome replication patterns in the hyperthermophilic euryarchaea *Archaeoglobus fulgidus* and *Methanocaldococcus* (*Methanococcus*) *jannaschii*. *Mol. Microbiol.* 45, 1443–1450. doi: 10.1046/j.1365-2958.2002.03111.x

Majernik, A. I., and Chong, J. P. (2008). A conserved mechanism for replication origin recognition and binding in archaea. *Biochem. J.* 409, 511–518. doi: 10.1042/BJ20070213

Matsunaga, F., Forterre, P., Ishino, Y., and Myllykallio, H. (2001). *In vivo* interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11152–11157. doi: 10.1073/pnas.191387498

Matsunaga, F., Norais, C., Forterre, P., and Myllykallio, H. (2003). Identification of short "eukaryotic" Okazaki fragments synthesized from a prokaryotic replication origin. *EMBO Rep.* 4, 154–158. doi: 10.1038/sj.embor.embor732

Mechali, M. (2010). Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat. Rev. Mol. Cell Biol.* 11, 728–738. doi: 10.1038/nrm2976

Messer, W. (2002). The bacterial replication initiator DnaA. DnaA and *oriC*, the bacterial mode to initiate DNA replication. *FEMS Microbiol. Rev.* 26, 355–374.

Mott, M. L., and Berger, J. M. (2007). DNA replication initiation: mechanisms and regulation in bacteria. *Nat. Rev. Microbiol.* 5, 343–354. doi: 10.1038/nrmicro1640

Myllykallio, H., Lopez, P., Lopez-Garcia, P., Heilig, R., Saurin, W., Zivanovic, Y., et al. (2000). Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* 288, 2212–2215. doi: 10.1126/science.288.5474.2212

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 *Haloferax volcanii*. *PLoS Genet.* 3:e77. doi: 10.1371/journal.pgen.0030077

Pelve, E. A., Lindas, A. C., Knoppel, A., Mira, A., and Bernander, R. (2012). Four chromosome replication origins in the archaeon *Pyrobaculum calidifontis*. *Mol. Microbiol.* 85, 986–995. doi: 10.1111/j.1365-2958.2012.08155.x

Pelve, E. A., Martens-Habbena, W., Stahl, D. A., and Bernander, R. (2013). Mapping of active replication origins *in vivo* in thaum- and euryarchaeal replicons. *Mol. Microbiol.* 90, 538–550. doi: 10.1111/mmi.12382

Rajewska, M., Wegrzyn, K., and Konieczny, I. (2012). AT-rich region and repeated sequences – the essential elements of replication origins of bacterial replicons. *FEMS Microbiol. Rev.* 36, 408–434. doi: 10.1111/j.1574-6976.2011.00300.x

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

Robinson, N. P., and Bell, S. D. (2007). Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5806–5811. doi: 10.1073/pnas.0700206104

Robinson, N. P., Blood, K. A., McCallum, S. A., Edwards, P. A., and Bell, S. D. (2007). Sister chromatid junctions in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *EMBO J.* 26, 816–824. doi: 10.1038/sj.emboj.7601529

Robinson, N. P., Dionne, I., Lundgren, M., Marsh, V. L., Bernander, R., and Bell, S. D. (2004). Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell* 116, 25–38. doi: 10.1016/S0092-8674(03)01034-1

Samson, R. Y., Xu, Y., Gadelha, C., Stone, T. A., Faqiri, J. N., Li, D., et al. (2013). Specificity and function of archaeal DNA replication initiator proteins. *Cell Rep.* 3, 485–496. doi: 10.1016/j.celrep.2013.01.002

Wu, Z., Liu, H., Liu, J., Liu, X., and Xiang, H. (2012). Diversity and evolution of multiple *orc/cdc6*-adjacent replication origins in haloarchaea. *BMC Genomics* 13:478. doi: 10.1186/1471-2164-13-478

Wu, Z., Liu, J., Yang, H., Liu, H., and Xiang, H. (2014). Multiple replication origins with diverse control mechanisms in *Haloarcula hispanica*. *Nucleic Acids Res.* 42, 2282–2294. doi: 10.1093/nar/gkt1214

Zakrzewska-Czerwinska, J., Jakimowicz, D., Zawilak-Pawlik, A., and Messer, W. (2007). Regulation of the initiation of chromosomal replication in bacteria. *FEMS Microbiol. Rev.* 31, 378–387. doi: 10.1111/j.1574-6976.2007.00070.x

Zhang, R., and Zhang, C. T. (2003). Multiple replication origins of the archaeon *Halobacterium* species NRC-1. *Biochem. Biophys. Res. Commun.* 302, 728–734. doi: 10.1016/S0006-291X(03)00252-3

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: 01 April 2014; published online: 29 April 2014.
Citation: Wu Z, Liu J, Yang H and Xiang H (2014) DNA replication origins in archaea. Front. Microbiol. ***5****:179. doi: 10.3389/fmicb.2014.00179*
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology.
Copyright © 2014 Wu, Liu, Yang and Xiang. 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.

frontiers in
MICROBIOLOGY

METHODS ARTICLE
published: 15 September 2014
doi: 10.3389/fmicb.2014.00482

Ori-Finder 2, an integrated tool to predict replication origins in the archaeal genomes

Hao Luo[1], Chun-Ting Zhang[1]* and Feng Gao[1,2,3]*

[1] *Department of Physics, Tianjin University, Tianjin, China*
[2] *Key Laboratory of Systems Bioengineering (Ministry of Education), Tianjin University, Tianjin, China*
[3] *SynBio Research Platform, Collaborative Innovation Center of Chemical Science and Engineering, Tianjin, China*

Edited by:
Eric Altermann, AgResearch Ltd., New Zealand

Reviewed by:
Dirk Linke, Max Planck Society, Germany
Andrew F. Gardner, New England Biolabs, USA

****Correspondence:***
Feng Gao and Chun-Ting Zhang, Department of Physics, Tianjin University, Tianjin 300072, China
e-mail: fgao@tju.edu.cn; ctzhang@tju.edu.cn

DNA replication is one of the most basic processes in all three domains of cellular life. With the advent of the post-genomic era, the increasing number of complete archaeal genomes has created an opportunity for exploration of the molecular mechanisms for initiating cellular DNA replication by *in vivo* experiments as well as *in silico* analysis. However, the location of replication origins (*oriC*s) in many sequenced archaeal genomes remains unknown. We present a web-based tool Ori-Finder 2 to predict *oriC*s in the archaeal genomes automatically, based on the integrated method comprising the analysis of base composition asymmetry using the Z-curve method, the distribution of origin recognition boxes identified by FIMO tool, and the occurrence of genes frequently close to *oriC*s. The web server is also able to analyze the unannotated genome sequences by integrating with gene prediction pipelines and BLAST software for gene identification and function annotation. The result of the predicted *oriC*s is displayed as an HTML table, which offers an intuitive way to browse the result in graphical and tabular form. The software presented here is accurate for the genomes with single *oriC*, but it does not necessarily find all the origins of replication for the genomes with multiple *oriC*s. Ori-Finder 2 aims to become a useful platform for the identification and analysis of *oriC*s in the archaeal genomes, which would provide insight into the replication mechanisms in archaea. The web server is freely available at http://tubic.tju.edu.cn/Ori-Finder2/.

Keywords: archaea, replication origins, Z-curve, origin recognition box, DNA replication

INTRODUCTION

DNA replication is one of the essential and conserved features among all three domains of life. In bacteria, DNA replication initiates from a single replication origin (*oriC*), which is often adjacent to the replication-related genes and distributed with the DnaA box motifs, whereas eukaryotic organisms exploit significantly more replication origins, ranging from hundreds in yeast to tens of thousands in human (Gao et al., 2012). Archaea are classified as a separate domain in the three-domain system, and share some similar features with both bacteria and eukaryotes (Woese and Fox, 1977). Similar to the bacteria, the *oriC*s in archaea are located in the intergenic regions around the replication-related proteins and distributed with the origin recognition boxes (ORBs). The ORB motifs are the conserved sequences and recognition sites for the Orc1/Cdc6 initiation proteins (Barry and Bell, 2006). In some organisms, G-stretches are also observed at the end of ORBs. On the other hand, the origin binding proteins in archaea are homologous to the related eukaryotic Orc1/Cdc6 proteins, and some archaea could also adopt more than one *oriC* to initiate DNA replication. With the increasing availability of complete archaeal genomes, identification of their *oriC*s would provide further insight into the mechanism of DNA replication in archaea and reveal the evolutionary history between bacteria and eukaryotes (Barry and Bell, 2006; Wu et al., 2014b).

The first putative *oriC* of archaea was identified in *Halobacterium* sp. strain NRC-1 by GC-skew method and demonstrated by cloning into a non-replicating plasmid (Myllykallio et al., 2000). The Z-curve method is an alternative technique that detects the asymmetrical nucleotide distribution around replication origins. The three components of the Z-curve, x_n, y_n, and z_n display the distributions of purine versus pyrimidine (R vs. Y), amino versus keto (M vs. K) and strong H-bond versus weak H-bond (S vs. W) bases along the sequence, respectively. The x_n and y_n components are termed the RY and MK disparity curves, respectively. The AT and GC disparity curves are defined by $(x_n + y_n)/2$ and $(x_n - y_n)/2$, which shows the excess of A over T and G over C, respectively, along the sequence (Zhang and Zhang, 2005; Gao, 2014). Based on the Z-curve analysis, we have identified single *oriC* in *Methanocaldococcus jannaschii* and *Methanosarcina mazei*, double *oriC*s in *Halobacterium* sp. strain NRC-1, and three *oriC*s in *Sulfolobus solfataricus* P2, which are consistent with the subsequent experiments (Soppa, 2006). Recently, multiple *orc1/cdc6*-associated *oriC*s in all the available haloarchaeal genomes have been predicted by identification of putative ORBs (Wu et al., 2012). Based on these discoveries, several basic features of the *oriC*s could be summarized in archaea. Firstly, most *oriC*s are located in proximity to the genes encoding archaeal replication-related proteins, such as archaeal Orc/Cdc6 protein, Whip (Winged-Helix Initiator

Protein) and DNA primase. Secondly, *oriC*s are often located around the extremes of disparity curves. Finally, most of the *oriC*s contains the AT-rich unwinding elements and conserved ORBs (Zhang and Zhang, 2005; Barry and Bell, 2006; Wu et al., 2014a).

Our group has developed a web-based system Ori-Finder 1 to find *oriC*s in the bacterial genomes based on the Z-curve method with high accuracy and reliability (Gao and Zhang, 2008). Now with the knowledge of *oriC*s in the archaeal genomes, we present an online tool, Ori-Finder 2, to identify the *oriC*s in the archaeal genomes, based on the integrated method comprising the analysis of base composition asymmetry using the Z-curve method, the distribution of ORB elements identified by FIMO tool, and the occurrence of genes frequently close to replication origins, which is available at http://tubic.tju.edu.cn/Ori-Finder2/.

METHODS AND IMPLEMENTATION

Ori-Finder 2 utilizes an integrated approach to predict *oriC*s in the user-supplied archaeal genomes automatically. **Figure 1** presents the workflow of Ori-Finder 2. Users submit an annotated or unannotated genome sequence to the web server. For the annotated genome, we recommend that users submit the sequence file in GenBank format or upload the sequence file in FASTA format as well as its corresponding protein table (PTT) file. The web server is also able to analyze the unannotated genomes by integrating two gene prediction pipelines, ZCURVE1.02 and Glimmer3 (Guo et al., 2003; Delcher et al., 2007), for gene identification and BLAST program for functional annotations of genes. Then all the intergenic sequences are scanned by Find Individual Motif Occurrences (FIMO), a software tool for scanning DNA or protein sequences with motifs described as position-specific scoring matrices (Grant et al., 2011), to obtain the ORB sequences, and also by REPuter program, a classic pipeline to compute exact repeats and palindromes in complete genomes (Kurtz et al., 2001), to identify the repeats. Finally, all the intergenic sequences adjacent to the replication-related genes with the ORB sequences are predicted as *oriC*s. Since the approach relies on the prior knowledge of *oriC*s in archaea, it may fail to identify the *oriC*s adjacent to the unknown genes which might be involved in DNA replication. In order to overcome the drawback, the intergenic sequences, which contain more than two conserved motifs, will be also predicted as *oriC*s. BLAST searches are performed against DoriC, a database of bacterial and archaeal replication origins, to search the homologs (Gao and Zhang, 2007; Gao et al., 2013). Here, the conserved motifs of ORB sequences used in FIMO were obtained from DoriC. All the records in DoriC were organized into several taxonomic clusters, including *Methanobacteriaceae*, *Methanomicrobia*, *Methanococcaceae*, *Sulfolobaceae* and *Thermococcaceae*. And the conserved ORB motifs were calculated from the corresponding clusters by Multiple EM for Motif Elicitation (MEME) program, a tool used to discover motifs in a group of related DNA or protein sequences (Bailey et al., 2009). **Table 1** displays the regular expressions of ORB motifs. Note that the common motif is calculated from all the records in DoriC. The motif logos are shown in the submission form, and the position specific probability matrix (PSPM) is available in the document webpage. Each job of Ori-Finder 2 is assigned a unique ID, and the whole process will take several minutes to complete. Users could retrieve their results with the job ID or be notified by email if specified in the submission page.

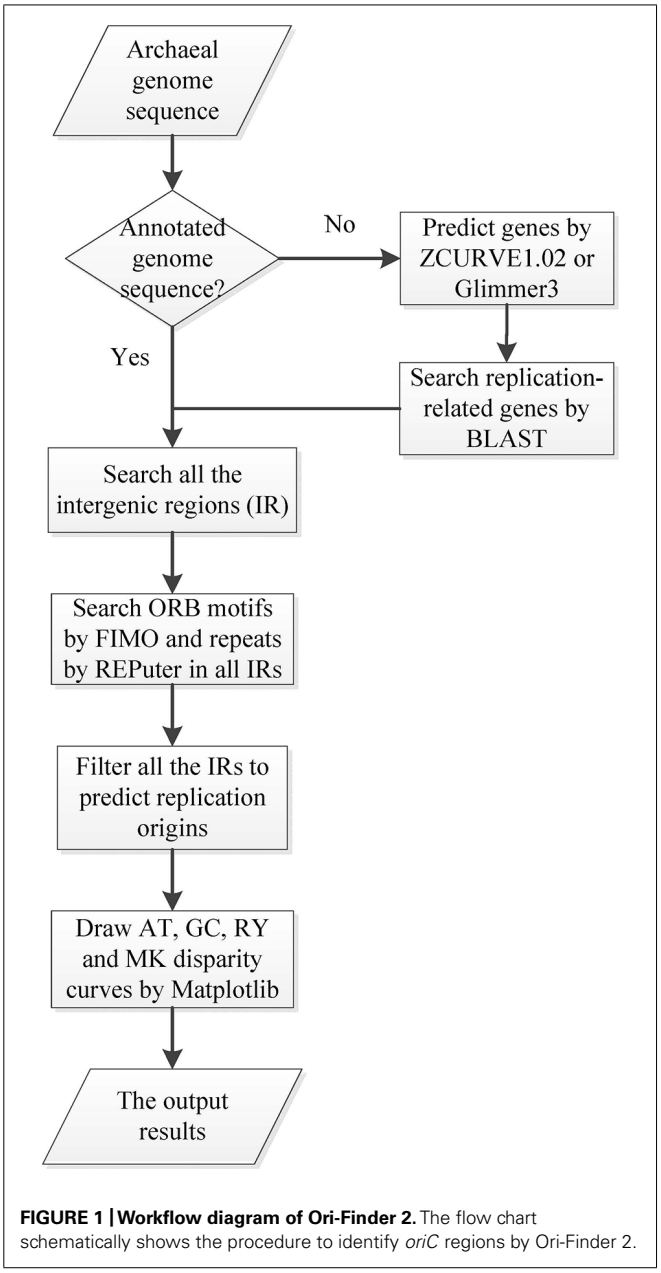


FIGURE 1 | Workflow diagram of Ori-Finder 2. The flow chart schematically shows the procedure to identify *oriC* regions by Ori-Finder 2.

In the result, the information including genome size, GC content, the locations of replication-related genes and the predicted *oriC*s, as well as the Z-curve (AT, GC, RY, and MK disparity curves) for the input genome is displayed as an HTML table. In addition, the detailed information about the repeats identified by REPuter program, ORBs recognized by FIMO and the homologs in DoriC are also presented in the corresponding subtable. The ORB motifs in all the intergenic regions are also available for download from the provided URL. Users could also click to enlarge the embedded figure to obtain the high

Table 1 | The regular expressions of the ORB motifs identified by MEME.

Taxonomy	*E*-value	Regular expressions
Halobacteriaceae	4.5E-180	TT[TC]CACC[GCT]GAAAC[GA][AC][GA]G[GT]G[GT]
Methanobacteriaceae	4.20E-68	TT[TA]CACTTGAAAT[GTA]T[CT][CG]TC
Methanomicrobia	1.50E-202	TCCA[GC]T[GT]GAAA[CT][AG]A[AT]GGGGT
Methanococcaceae	4.80E-90	TT[TA][GT]ATTCA[TC][GA]AT[AT]T[AT]T[AT]
Sulfolobacea	7.50E-296	[GC]GGCCGG[AG]A[GT][CT][GT]T[CG]A[CA]CC[TC]GG
	2.40E-286	TCCA[AG][AT][TG]GAA[CA][CT][GA]AAGGGGT
	8.20E-120	GAGTGC[GT]CGGTT[CGT]GCA[ATC]CC[AG]
Thermococcaceae	9.10E-223	[TC]TCCAGTGGAAA[TC][GA]AA[AG]CTC
	6.90E-56	[CAG]TTTCCA[CT][TA]GGA[AT][CGA][CT]
	2.00E-52	AATG[ACT]ACA[AT]A[AGT]ATG[TA][TG]CATT
Common[a]	1.20E-225	TCCA[CG]T[TG]GAAA[TC][GA]AAGGGGT

[a] *Note that the Common motif is calculated from all the records in DoriC by MEME. In Halobacteriaceae, Methanobacteriaceae, Methanomicrobia, Sulfolobacea, and Thermococcaceae, they share the consensus sequences "TCCA—GAAAC" similar to the common motif. In Methanomicrobia and Sulfolobacea, "G-string" (GGGGT) is observed obviously at the end of ORB motifs.*

Table 2 | The prediction results of 13 archaeal chromosomes[a].

Organism	Refseq	*OriC*s in DoriC	*OriC*s predicted by Ori-Finder 2	True positive
Aeropyrum pernix K1	NC_000854	2	2	1
Pyrococcus abyssi GE5	NC_000868	1	1	1
Methanothermobacter thermautotrophicus str. Delta H chromosome	NC_000916	1	2	1
Archaeoglobus fulgidus DSM 4304	NC_000917	1	1	0
Pyrococcus horikoshii OT3	NC_000961	1	1	1
Halobacterium sp. NRC-1	NC_002607	2	4	2
Pyrococcus furiosus DSM 3638	NC_003413	1	1	1
Hyperthermus butylicus DSM 5456	NC_008818	2	1	1
Pyrobaculum calidifontis JCM 11548	NC_009073	1	1	0
Haloferax volcanii DS2	NC_013967	5	6	5
Haloarcula hispanica ATCC 33960 chromosome II	NC_015943	4	7	3
Haloarcula hispanica ATCC 33960 chromosome I	NC_015948	5	1	1
Nitrosopumilus maritimus SCM1	NC_010085	1	1	1
Total	–	27	29	18

[a] *Note that the detailed information is available at http://tubic.tju.edu.cn/Ori-Finder2/doc.php#9.*

resolution one which displays the RY, MK, GC, AT disparity curves, replication-related proteins, and the predicted *oriC*s. The result webpage and figures will be stored in 7 days on the web server.

Ori-Finder 2 is developed using Python and PHP on a Unix platform with an Apache web-server. The web interface is implemented using Common Gateway Interface (CGI) python scripts, and the webpage is designed with HTML, CSS, and JavaScript. The pipeline of Ori-Finder 2 uses the Biopython library, and the output graphs are generated by the Python module Matplotlib (Hunter, 2007; Cock et al., 2009).

RESULTS AND DISCUSSION

Based on this online system, we predicted the *oriC*s for all the available complete archaeal genomes in GenBank. For example, *Pyrococcus abyssi* is a classical model of DNA replication in the archaeal organisms. Similar to bacteria, there is only one *oriC* in its circular chromosome, which has been identified by

A

Ori-Finder 2 Result - jobID: uE3ZImnzUd (motif: Common; motif count: 3; *P*-value: 1E-04)	
Your Query:	Pyrococcus abyssi GE5 chromosome, complete genome.
Chromosome size	1765118 nt
Chromosome GC Level	44.71%
The location of *oriC*s with ORB sequence	122701..123499;
The location of DNA replication genes	121402..122700;

B

OriC: 122701..123499							
The size of *oriC*	799 nt						
The GC Level of *oriC*	34.42%						
DoriC	ORI10010003 ORI10010007 ORI10010101 ORI10010186 ORI10010016 ORI10010121						
The information of ORB	Pattern name	Start	End	Strand	Score	P value	Matched Sequence
	Common	426	445	+	-9.82037	2.33e-05	CCCCCAGGGTTTCATTTCCA
	Common	431	450	+	-10.2665	2.57e-05	AGGGTTTCATTTCCACTGGA
	Common	709	728	-	-11.605	3.4e-05	AGAGTTTCATTTCCAGTGGA
	Common	714	733	-	-16.2897	8.64e-05	CCCCCAGAGTTTCATTTCCA

The sequence of *oriC*:

```
GACCTAAATTTCCAGTGGAAACATAACACTTTTCGCTAACATATTTTCATTGTGCATGACATAATGTGAC
ATTTAACACGACTATTTAAGCCAAACCGTAGAGAGTCTCCATTGGAAGCATTGGAACCAGAGGAAAGCAA
CGTAATAATCTATGGCCCACTTTTTATCCTTTGCCTTGACAAGAACATGTTAATGAAGATAATTTCATCT
GAAATTTATTCTCGAGAATTTCCACAGGAAACCTCCACAGGACATGAACAAGGATGTGCGTTGAATCTTT
ACATTTTCCAGTGGCCATAACCAATTTCCGATGGAAGGAAAATGAACATATATGTTCATTCAGATAAGAG
CATTTTTACACAAATAAGTGCATTAGGTTAACTTAATTTGTTTAATATTTTGCCAAAGTTAAACAGAAGT
GAAGTCCCCCAGGGTTTCATTTCCACTGGAACCGGGTTGCGACACCGAGAATATTCTATTCTTTTTTCTG
ACAAGACACTCTCCCATATTAGTTGGAACTCCTATTTAAATATTTATCTATGAACAGTTTTATAAGTAAA
TTGAGAAGCAGAGAAGCAAATGTTCACATATTTATGCACAAAATAGTACATCAACTGAAAAAATGTTCAT
TTGCAAAAGTAAAAATTTTTGGCTTTTAATTTTACGTTTGTAAACCAACGAAGAATTGCATGCCTTTCTC
AACCCAGTTCCACTGGAAATGAAACTCTGGGGGGTTCTTTAGATGGCTATCTGAATGCTAGCATATAAAT
CGAAAATATACAAGAAAATGTCAAAAAAA
```

C

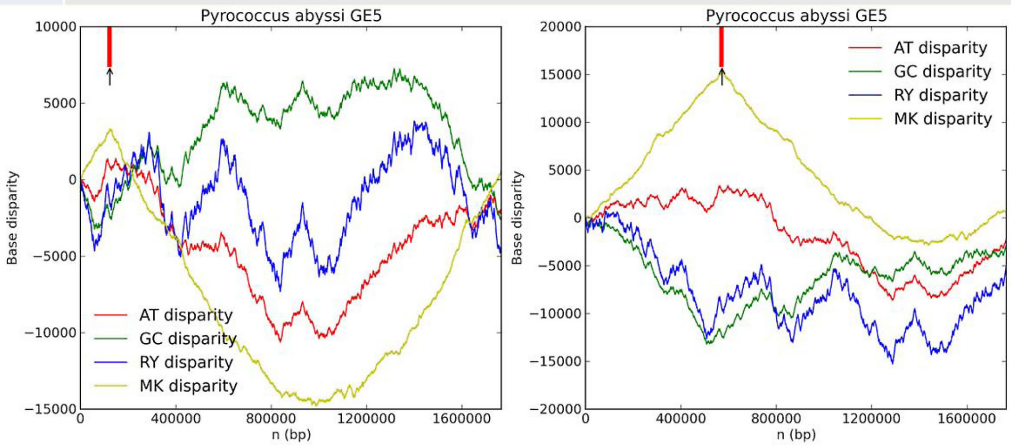


FIGURE 2 | Example of Ori-Finder 2 result for *Pyrococcus abyssi* GE5. **(A)** The information of genome size, GC content, the locations of replication-related genes and the predicted *oriC*s. **(B)** The detailed information of the predicted *oriC* region including size, GC content, homologs in DoriC and sequence, as well as the information of the identified ORBs including the ORB motif (also referred to as "Pattern name"), location, strand, the associated log-likelihood ratio score, *P* value and the matched sequences. Note that the log-likelihood ratio score and *P* value are computed by FIMO to measure the similarity between the ORB motif and the matched sequence, and the *P* value cutoff for FIMO motif searching is 10^{-4}. The ORB motif used here is the common motif. **(C)** The left figure shows the Z-curves (AT, GC, RY, and MK disparity curves) for the original sequence, and the right figure shows the Z-curves (AT, GC, RY, and MK disparity curves) for the rotated sequence beginning and ending in the maximum of the GC disparity curve. The short vertical red line indicates the location of replication-related protein. The black arrow is the predicted *oriC* region.

cumulative oligomer skew and confirmed by *in vivo* method. With the annotated genome file, the *oriC* predicted by Ori-Finder 2 is in accordance with the experimental result and located at the peak of the MK disparity curve. Several ORB sequences are recognized in the *oriC*. **Figure 2** is a screenshot of the result by Ori-Finder 2. In addition, some archaea adopt more than one *oriC* during the DNA replication. For this situation, Ori-Finder 2 also predicted multiple *oriC*s in their genomes. *Haloferax volcanii* DS2 has a chromosome with multiple *oriC*s. Five *oriC*s were identified *in silico*, and three of them have been confirmed *in vitro* (Norais et al., 2007; Wu et al., 2012; Hawkins et al., 2013). With the annotated genome file, all the five *oriC*s mentioned above have been predicted by Ori-Finder 2 successfully, and another *oriC* with three ORB motifs is also found, which is adjacent to the genes *purO* and *cgi*. Besides that, the *oriC*s identified in the unannotated genomes are consistent with the previous results. In order to estimate the performance of Ori-Finder 2, we used 13 annotated archaeal chromosomes, whose *oriC*s have been confirmed by experimental method or identified *in silico* by other groups (**Table 2**). Compared with the records in DoriC, the sensitivity and precision are 66.7% and 62.1%, respectively. The reason of the lower precision and sensitivity compared with the programs to detect bacterial origins, such as Ori-Finder 1, is that bacteria have only one *oriC* in their chromosomes, but archaea tend to have more than one. Furthermore, *oriC*s in archaea show more diversity than those in bacteria, such as more complex ORBs in comparison with the DnaA boxes, and more unknown species-specific replication-related genes. It is difficult to predict the *oriC*s in archaea with high precision and sensitivity due to the limited amount of experimental data. For example, not all the *oriC*s in the genomes with multiple *oriC*s are found, and the ORBs with unique features need to be further explored by experimental methods. For the convenience of users' query, the *oriC*s confirmed by *in vivo* or *in silico* methods have been collected into DoriC, which is freely available at http://tubic.tju.edu.cn/doric/.

CONCLUSION

Here, we presented a user-friendly interactive web-based platform Ori-Finder 2 to predict the *oriC*s in the archaeal genomes. The tool integrated several genomic pipelines, including FIMO, BLAST, ZCURVE, Glimmer, and REPuter, to comprehensively annotate and analyze the *oriC*s. Moreover, the ORB motifs are also calculated by MEME and organized by taxonomy. The software presented here does not necessarily find all the origins of replication in cases where there are multiple ones in a genome. However, we will continually strive to improve our approach to make it more accurate and sensitive with the increase of the *oriC*s confirmed experimentally in archaea. As the only currently available auto-annotation system for the archaeal replication origins at the sequence level, we believe that Ori-Finder 2 will be helpful to predict the archaeal replication origins and provide insight into DNA replication in archaea.

AUTHOR CONTRIBUTIONS

Hao Luo designed the computer program and drafted the manuscript. Chun-Ting Zhang and Feng Gao supervised the study and revised the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors thank Dr. Kurtz for providing the REPuter binaries. They also would like to thank Dr. Ren Zhang for invaluable assistance. The present work was supported in part by National Natural Science Foundation of China (Grant Nos. 31171238 and 30800642), and Program for New Century Excellent Talents in University (No. NCET-12-0396).

REFERENCES

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335

Barry, E. R., and Bell, S. D. (2006). DNA replication in the archaea. *Microbiol. Mol. Biol. Rev.* 70, 876–887. doi: 10.1128/MMBR.00029-06

Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423. doi: 10.1093/bioinformatics/btp163

Delcher, A. L., Bratke, K. A., Powers, E. C., and Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23, 673–679. doi: 10.1093/bioinformatics/btm009

Gao, F. (2014). Recent advances in the identification of replication origins based on the Z-curve method. *Curr. Genomics* 15, 104–112. doi: 10.2174/1389202915999140328162938

Gao, F., Luo, H., and Zhang, C. T. (2012). DeOri: a database of eukaryotic DNA replication origins. *Bioinformatics* 28, 1551–1552. doi: 10.1093/bioinformatics/bts151.

Gao, F., Luo, H., and Zhang, C. T. (2013). DoriC 5.0: an updated database of oriC regions in both bacterial and archaeal genomes. *Nucleic Acids Res.* 41, D90–D93. doi: 10.1093/nar/gks990

Gao, F., and Zhang, C. T. (2007). DoriC: a database of oriC regions in bacterial genomes. *Bioinformatics* 23, 1866–1867. doi: 10.1093/bioinformatics/btm255

Gao, F., and Zhang, C. T. (2008). Ori-Finder: a web-based system for finding oriCs in unannotated bacterial genomes. *BMC Bioinformatics* 9:79. doi: 10.1186/1471-2105-9-79

Grant, C. E., Bailey, T. L., and Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018. doi: 10.1093/bioinformatics/btr064

Guo, F. B., Ou, H. Y., and Zhang, C. T. (2003). ZCURVE: a new system for recognizing protein-coding genes in bacterial and archaeal genomes. *Nucleic Acids Res.* 31, 1780–1789. doi: 10.1093/nar/gkg254

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

Hunter, J. D. (2007). Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. doi: 10.1109/MCSE.2007.55

Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29, 4633–4642. doi: 10.1093/nar/29.22.4633

Myllykallio, H., Lopez, P., Lopez-Garcia, P., Heilig, R., Saurin, W., Zivanovic, Y., et al. (2000). Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* 288, 2212–2215. doi: 10.1126/science.288.5474.2212

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 *Haloferax volcanii*. *PLoS Genet.* 3:e77. doi: 10.1371/journal.pgen.0030077

Soppa, J. (2006). From genomes to function: haloarchaea as model organisms. *Microbiology* 152(Pt 3), 585–590. doi: 10.1099/mic.0.28504-0

Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5088–5090. doi: 10.1073/pnas.74.11.5088

Wu, Z., Liu, H., Liu, J., Liu, X., and Xiang, H. (2012). Diversity and evolution of multiple orc/cdc6-adjacent replication origins in haloarchaea. *BMC Genomics* 13:478. doi: 10.1186/1471-2164-13-478

Wu, Z., Liu, J., Yang, H., Liu, H., and Xiang, H. (2014a). Multiple replication origins with diverse control mechanisms in *Haloarcula hispanica*. *Nucleic Acids Res.* 42, 2282–2294. doi: 10.1093/nar/gkt1214

Wu, Z., Liu, J., Yang, H., and Xiang, H. (2014b). DNA replication origins in archaea. *Front. Microbiol.* 5:179. doi: 10.3389/fmicb.2014.00179

Zhang, R., and Zhang, C. T. (2005). Identification of replication origins in archaeal genomes based on the Z-curve method. *Archaea* 1, 335–346. doi: 10.1155/2005/509646

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: 05 August 2014; accepted: 27 August 2014; published online: 15 September 2014.

*Citation: Luo H, Zhang C-T and Gao F (2014) Ori-Finder 2, an integrated tool to predict replication origins in the archaeal genomes. Front. Microbiol. **5**:482. doi: 10.3389/fmicb.2014.00482*

This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology.

Copyright © 2014 Luo, Zhang and Gao. 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.

frontiers in MICROBIOLOGY

REVIEW ARTICLE
published: 06 January 2015
doi: 10.3389/fmicb.2014.00735

oriC-encoded instructions for the initiation of bacterial chromosome replication

Marcin Wolański[1†], Rafał Donczew[2†], Anna Zawilak-Pawlik[2] and Jolanta Zakrzewska-Czerwińska[1,2]*

[1] Department of Molecular Microbiology, Faculty of Biotechnology, University of Wrocław, Wrocław, Poland
[2] Department of Microbiology, Ludwik Hirszfeld Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, Wrocław, Poland

Edited by:
Feng Gao, Tianjin University, China

Reviewed by:
Gregory Marczynski, McGill University, Canada
Dhruba Chattoraj, National Institutes of Health, USA

***Correspondence:**
Jolanta Zakrzewska-Czerwińska, Department of Molecular Microbiology, Faculty of Biotechnology, University of Wrocław, ul. Joliot-Curie 14A, 50-383 Wroclaw, Poland
e-mail: jolanta.zakrzewska@uni.wroc.pl

† These authors have contributed equally to this work.

Replication of the bacterial chromosome initiates at a single origin of replication that is called *oriC*. This occurs via the concerted action of numerous proteins, including DnaA, which acts as an initiator. The origin sequences vary across species, but all bacterial *oriCs* contain the information necessary to guide assembly of the DnaA protein complex at *oriC*, triggering the unwinding of DNA and the beginning of replication. The requisite information is encoded in the unique arrangement of specific sequences called DnaA boxes, which form a framework for DnaA binding and assembly. Other crucial sequences of bacterial origin include DNA unwinding element (DUE, which designates the site at which *oriC* melts under the influence of DnaA) and binding sites for additional proteins that positively or negatively regulate the initiation process. In this review, we summarize our current knowledge and understanding of the information encoded in bacterial origins of chromosomal replication, particularly in the context of replication initiation and its regulation. We show that *oriC* encoded instructions allow not only for initiation but also for precise regulation of replication initiation and coordination of chromosomal replication with the cell cycle (also in response to environmental signals). We focus on *Escherichia coli*, and then expand our discussion to include several other microorganisms in which additional regulatory proteins have been recently shown to be involved in coordinating replication initiation to other cellular processes (e.g., *Bacillus*, *Caulobacter, Helicobacter, Mycobacterium,* and *Streptomyces*). We discuss diversity of bacterial *oriC* regions with the main focus on roles of individual DNA recognition sequences at *oriC* in binding the initiator and regulatory proteins as well as the overall impact of these proteins on the formation of initiation complex.

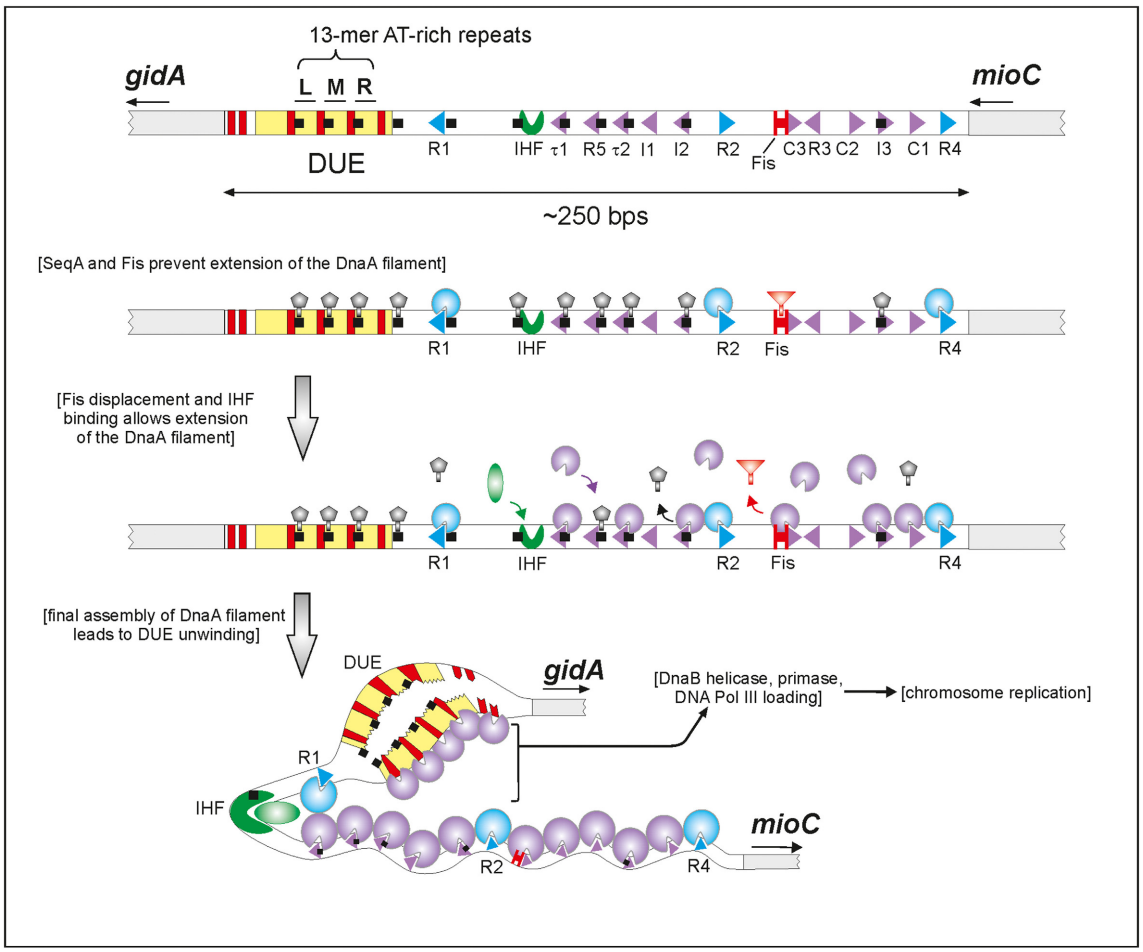
Keywords: *oriC*, DnaA, initiation of chromosome replication, orisome, replication regulation, regulatory proteins, bacteria

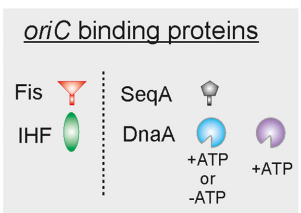
INTRODUCTION

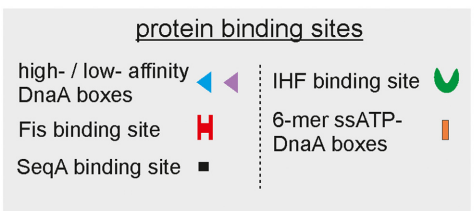
In contrast to the situation in Eukaryotes, chromosomal replication in bacteria begins at a single site on the chromosome: the origin of replication (*oriC*) (Leonard and Méchali, 2013). Using various *in silico* approaches (Mackiewicz et al., 2004; Gao et al., 2013), researchers have predicted the locations of the *oriCs* for more than 1500 bacterial chromosomes. However, *in vivo* replication activity has been confirmed for only a dozen such origins. Over the last 30 years, researchers have made considerable progress in understanding the mechanisms of replication initiation, particularly the organization and function of the *oriC* region in *Escherichia coli* (**Figure 1**), which is a model microorganism for the study of chromosomal replication (for reviews, see references, Fuller et al., 1984; Hwang and Kornberg, 1992; Messer, 2002; Kaguni, 2006, 2011; Leonard and Méchali, 2013). These studies have shown that replication is initiated through the cooperative binding of the initiator protein, DnaA, to multiple DnaA-recognition sites (boxes) within the *oriC* region. This triggers separation of the DNA strands at the AT-rich DNA unwinding element (DUE), providing an entry site for helicase and later on the other enzymes (e.g., primase and DNA Pol III) that are responsible for DNA synthesis.

Comparative sequence analysis has demonstrated that the origin regions with confirmed *in vivo* functions differ in their sequences, organizations, and sizes, with only closely related organisms exhibiting fairly high overall similarities in their *oriC* sequences (Jakimowicz et al., 1998; Zawilak-Pawlik et al., 2005). In addition to a diverse repertoire of DnaA boxes, *oriC* regions also include various binding sites for accessory and regulatory proteins.

Chromosomal replication is mainly controlled at the initiation step (Mott and Berger, 2007; Zakrzewska-Czerwińska et al., 2007; Katayama et al., 2010; Leonard and Grimwade, 2010, 2011; Skarstad and Katayama, 2013). Therefore, the activities of the *oriC* region must be tightly regulated to guarantee that chromosomal DNA is entirely replicated only once per cell cycle. This is achieved by regulating the accessibility of *oriC* to DnaA, which occurs mainly via the binding of other proteins. Additionally, replication initiation is regulated by the modulation of DnaA protein activity.







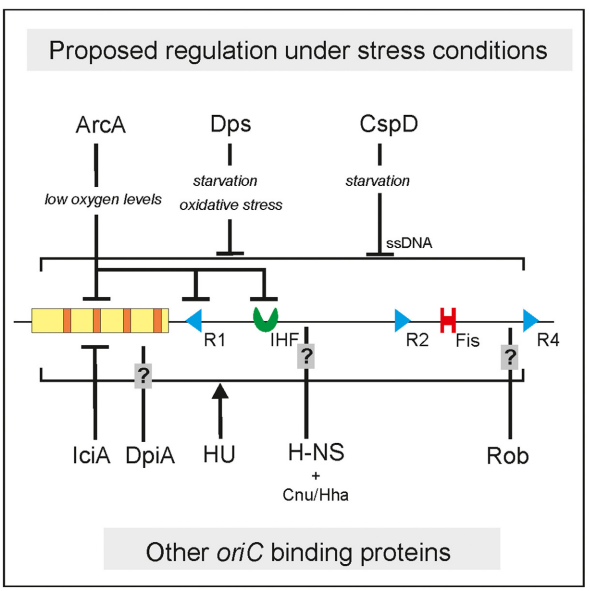


FIGURE 1 | A model of initiation replication and its regulation in *E. coli* by origin binding proteins (oriBPs). Large panel presents assumed sequence of events during the replication initiation and roles of particular oriBPs. The unwound DUE is accessible to the replication proteins complex (e.g., helicase DnaB, primase, and DNA Pol III). Small panel shows additional

(Continued)

FIGURE 1 | Continued
oriBPs divided in two subgroups, those involved in alternative scenarios that may occur under environmental stress conditions (upper part of the panel) and others, including those of unknown function (bottom part of the panel). Triangles' directions represent orientations of DnaA binding sites. Nucleotide bound status of DnaA is represented by blue and violet incomplete circles. Small arrows below gene names indicate gene orientations. In the small panel, different types of vertical lines represent type of action, activation (arrow), inhibition (bar-headed line) or unknown (question mark line). Horizontal lines indicate unspecific binding to *oriC*.

The main goal of this review is to highlight the diversity of *oriC* regions in the context of replication regulation and the bacterial cell cycle. We focus on how *oriC* regions have adjusted to coordinate the regulation of chromosome replication and the progression of the cell cycle. We postulate that *oriC* regions encode species- and genus-specific instructions for the orderly binding of DnaA and other proteins responsible for forming the functional initiation complex (orisome) and/or regulating the assembly of this complex.

oriC CHROMOSOMAL LOCALISATION AND NUCLEOTIDE SEQUENCE ARE NOT STRICTLY CONSERVED

The *oriC* regions are usually flanked by the *dnaA* gene and sometimes also the *dnaN* gene (**Figure 2**). These genes encode two pivotal proteins for initiating and continuing replication in the bacterial chromosome: DnaA, which is described above, and DnaN, which encodes a beta sliding clamp responsible for the processivity of DNA polymerase III. In linear chromosomes, such as those of *Streptomyces coelicolor* and (presumably) *Borrelia burgdorferi*, *oriC* is located in the center of the chromosome (Zakrzewska-Czerwińska and Schrempf, 1992; Mackiewicz et al., 2004). The region of gene synteny around *oriC*, which includes the highly preserved gene cluster of *rnpA-rpmH-dnaA-dnaN-recF-gyrB-gyrA*, is conserved in some (even distantly related) bacterial species. For a long time, the presence of these genes was assumed to mark the chromosomal localization of *oriC* (Ogasawara et al., 1991; Briggs et al., 2012). However, in many bacteria, including the model bacterium, *E. coli*, the *oriC* region is located in another gene context, indicating that this conserved gene cluster is not important for *oriC* function (Briggs et al., 2012). The localization of *oriC* is not random either as recent studies have indicated that the *oriC*-proximal gene context is conserved in certain group(s) of bacteria (e.g., genus or species). This is thought to enable a robust response to unfavorable conditions by allowing bacteria to increase the gene dosage in response to stress-induced initiation events (Moriya et al., 2009; Slager et al., 2014).

Interestingly, a few obligate endosymbiotic bacteria, such as *Wigglesworthia glossinidia*, *Blochmannia floridanus*, and *Candidatus Endolissoclinum faulkneri*, lack the *dnaA* gene (Akman et al., 2002; Gil et al., 2003; Mackiewicz et al., 2004; Kwan and Schmidt, 2013). It is not known how these bacteria initiate chromosome replication and whether their *oriC*s resemble the bacterial origins characterized to date. Indeed, it has been postulated that the typical DnaA box cluster might not exist/be functional in these bacteria.

The nucleotide sequences of the *oriC* regions are highly diverse across unrelated species. Thus, they are not active in or interchangeable between unrelated bacteria (O'Neill and Bender, 1988), and sequence homology alignment of *oriC* is not used to identify unknown origins in bacterial genomes. In closely related bacteria, however, the sequence homology (and thus the organization) of the entire *oriC* region might be high enough to enable the *oriC* region from one species to autonomously replicate in another species (Harding et al., 1982; Takeda et al., 1982; Zyskind et al., 1983; Roggenkamp, 2007). It might also be possible to substitute the origin of one species with that of a related species, as was shown in the successful substitution of the *E. coli oriC* for the *Vibrio cholerae* origin of replication in chromosome I (Demarre and Chattoraj, 2010; Koch et al., 2010). It is important to note that when bacteria possess two or more chromosomes, only one undergoes replication initiated by the DnaA protein at the origin typical for bacterial chromosomes (e.g., the *V. cholerae oriC*I). The replication origin on the other chromosome is plasmid like, and it is activated by initiators that lack homology to DnaA (e.g., *V. cholerae oriC*II is initiated by a RctB protein) (Egan and Waldor, 2003; Duigou et al., 2006). Such diversification avoids the need for the chromosomes to compete for initiator proteins, and yields better control of their separate but coordinated replications (Duigou et al., 2006; Jha et al., 2012; Baek and Chattoraj, 2014).

THE *oriC* REGION CAN BE CONTINUOUS OR BIPARTITE AND CONSISTS OF FUNCTIONAL MODULES

A replication origin may be continuous or bipartite. Most of the bacteria studied to date contain a continuous *oriC* region that includes all of the functional modules within a single intergenic region (**Figure 2**). The divided origins, in contrast, are composed of two subregions, each of which contains a cluster of DnaA boxes, and one of which harbors the DUE region (**Figure 2**). They also differ in length: the continuous origins range from ∼250 (*E. coli*) to ∼950 bps (*Streptomyces*) (Zakrzewska-Czerwińska and Schrempf, 1992; Jakimowicz et al., 1998), while the bipartite origins are longer, up to ∼2000 bps, because they contain a spacer gene (usually *dnaA*) between the *oriC* subregions (**Figure 2**; see below for mollicute origins). We do not yet know why some origins are split. Experimental data have shown that the spacer is important *per se*, although it may be altered to some extent without the loss of *oriC* activity. For example, a study showed that the spacer linking the *oriC*1 and *oriC*2 regions in *Bacillus subtilis* can be shortened (Moriya et al., 1992). Up until recently, the bipartite origin was assumed to be characteristic of a few Gram-positive bacteria (*B. subtilis* and *Streptococcus pyogenes*) and *Mollicutes* (*Mycoplasma* sp., *Spiroplasma* sp.) (Krause et al., 1997; Moriya et al., 1999; Suvorov and Ferretti, 2000; Lee et al., 2008; Briggs et al., 2012). However, bipartite origins were also recently identified in Gram-negative bacteria (e.g., *Helicobacter pylori*) (Donczew et al., 2012, 2014b), suggesting that bipartite origins might be more common than previously thought in diverse bacterial species. The origins of mollicutes were reported to have unusual properties, with interchangeability observed between species having divergent organizations of their *oriC*s

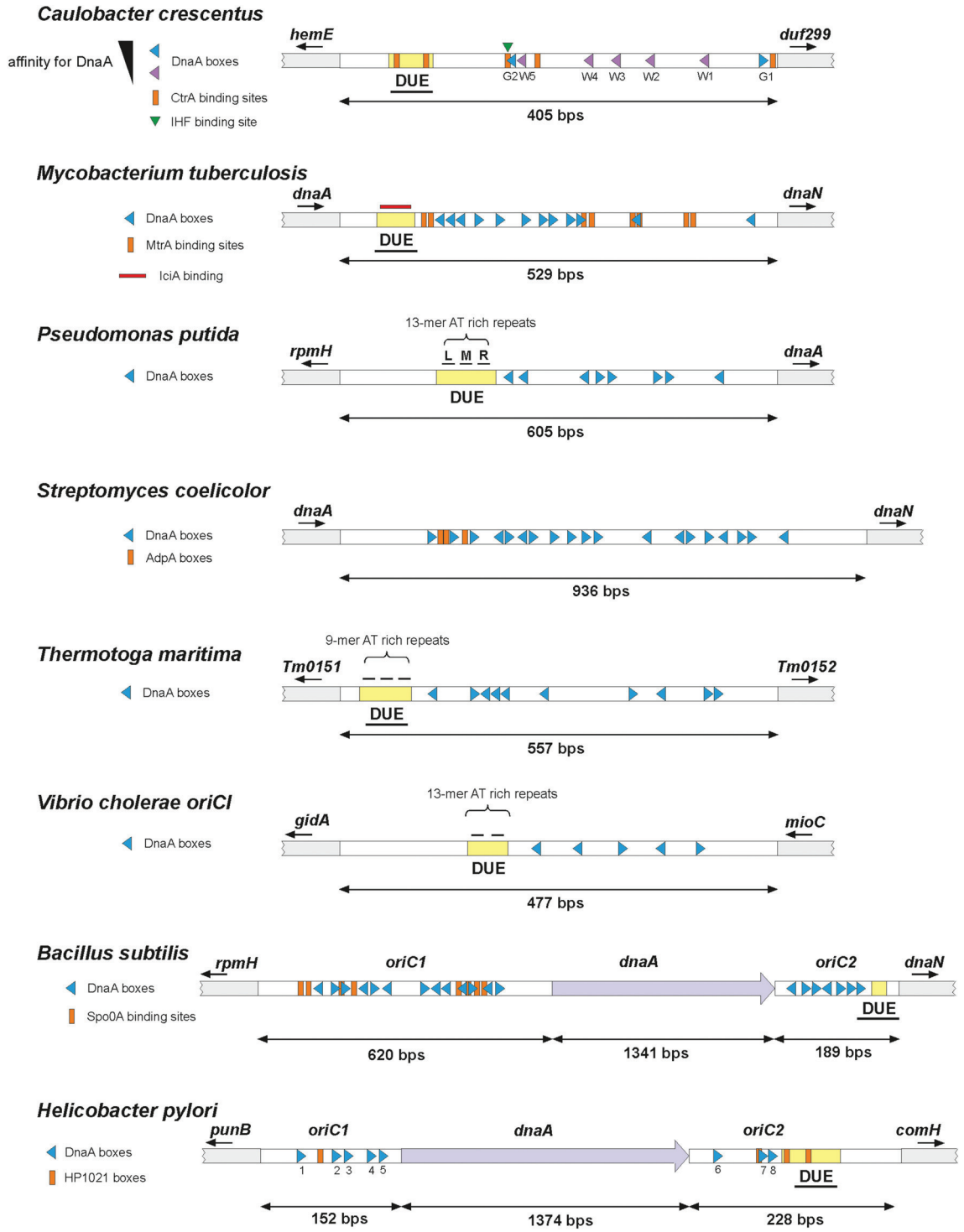


FIGURE 2 | The structures of selected bacterial origins. Continuous origin – upper part of the panel; bipartite origin – lower part of the panel. OriBP regulators' binding sites are presented if described in the literature. DUE – DNA unwinding element, underlined DUE indicates experimentally confirmed unwinding. Triangles' directions represent orientations of particular DnaA boxes. Small arrows below gene names indicate gene orientations.

(e.g., differences in the number, orientation and sequence of the DnaA boxes and/or the localization of the AT-rich regions) (Lartigue et al., 2003; Lee et al., 2008). The origins in mollicutes should be analyzed with caution, however, because they were identified solely by the minichromosome approach and no detailed characterization has yet been performed.

The study of the bacterial replication origin has progressed beyond the characterization of bacterial DnaA proteins, their

assembly onto *oriC*, and the modes of initiation complex (orisome) formation in various bacterial species. However, it remains difficult to interpret the differences in bacterial *oriC*s in the context of their activities. As discussed above, the *oriC* regions characterized to date are quite diverse in terms of their chromosomal loci, genetic contexts, nucleotide sequences, lengths and continuities. However, they are all composed of three basic functional modules: a cluster (or multiple clusters) of DnaA boxes, the DUE region, and other sequences that are recognized by regulatory proteins. These modules constitute the central management system for orisome formation, but their numbers and relative localizations vary across organisms (**Figure 2**).

In each species, this organized information provides a perfect molecular scaffold for DnaA oligomerization, controls DNA opening, and regulates the initiation of chromosomal replication. These processes are detailed in the following sections.

THE ARRANGEMENTS OF THE DnaA BOXES AND "DUE" ARE CRUCIAL FOR *oriC* ACTIVITY

The role of particular *oriC* modules has been widely studied in *E. coli*, providing a comprehensive example of a replication initiation mechanism and its interplay with cellular regulatory circuits. The DnaA boxes constitute a framework for the binding of DnaA monomers, which interact with *oriC* to form a structure that is able to disturb the DNA double-helix. The unique layout of low- and high-affinity DnaA boxes in *E. coli oriC* regulates the formation of a specific DnaA oligomer, which (according to the current model) adopts the structure of a right-handed helical filament to directly stimulate DNA unwinding (Erzberger et al., 2006; Zorman et al., 2012). The particular DnaA molecules involved in the filament introduce a bend in the DNA helix, which is gradually wrapped around the filament's outer surface (Fujikawa et al., 2003; Erzberger et al., 2006). The final complex introduces a superhelical tension in the DNA helix; this is likely to be focused in the DUE region, and triggers the initial unwinding (Erzberger et al., 2006). In the subsequent steps, the ATP-DnaA oligomer is believed to bind the newly formed single-stranded DNA segments, stabilizing and stretching them to promote further extension of the initiation bubble (Duderstadt et al., 2011; Ozaki and Katayama, 2012). The formation of a similar helical DnaA oligomer was recently shown for *B. subtilis* in the presence of both single-stranded and double-stranded DNA (Scholefield et al., 2012). Scholefield et al. suggested that separate oligomers may be involved in the unwinding and subsequent stabilization of the single-stranded DUE in this case, but we do not yet understand the mechanism underlying oligomer formation in *B. subtilis* in terms of the bipartite structure of its origin or the subsequent steps of the initiation process. It also remains to be seen whether other bacterial initiation complexes involve the formation of similar higher-order structures. Nevertheless, it is plausible that the formation of a DnaA-containing oligomer is essential for the unwinding of DNA at the DUE region, and is thus a common feature of all bacterial origins.

The number of DnaA boxes in the studied origins ranges from five in *Pseudomonas aeruginosa* or *V. cholerae* to 19 in *S. coelicolor* (**Figure 2**). In most cases, these DnaA boxes are asymmetrical nine-nucleotide-long specific motifs (with the exception of the 12-nucleotide boxes found in *Thermotoga maritima*), whose exact sequences, numbers and layouts reflect the diversity of the various organisms. DnaA boxes from different bacteria do, however, share a common core sequence (**Table 1**). No analyzed origin contains boxes, which deviate by more than two mismatches from the so-called "perfect" box sequence (i.e., that which binds with the highest affinity) of *E. coli* (TTATCCACA), with the exception of that of *T. maritima*. In closely related organisms, the high-affinity box sequence is conserved, as seen in *E. coli*, *V. cholerae*, *Pseudomonas putida* and *P. aeruginosa*, which all belong to a branch of the γ-proteobacteria (Yee and Smith, 1990; Weigel et al., 1997; Egan and Waldor, 2003) (**Table 1**). Interestingly, even *B. subtilis*, which is evolutionarily distant from *E. coli*, shares the same conserved "perfect" box sequence (Fukuoka et al., 1990). As noted above, the "perfect" DnaA box in other species almost always differs from this *E. coli* DnaA box by only one or two nucleotides. In Actinomycetes (*M. tuberculosis* and *S. coelicolor*), which are considered high-GC organisms, the "perfect" DnaA box contains G or C at the third position: TT(G/C)TCCACA (Jakimowicz et al., 2000; Zawilak et al., 2004; Tsodikov and Biswas, 2011). Similarly, *Caulobacter crescentus* G-boxes (see **Figure 2** and **Table 1**) differ from the *E. coli* consensus sequence by a single nucleotide (in this case in the second position; TGATCCACA) (Shaheen et al., 2009). Among the studied origins, the nine-nucleotide DnaA boxes most distant from that of *E. coli* are found in *H. pylori*, which contain two mismatches (at the second and fifth positions; TCATTCACA) with respect to the "perfect" *E. coli* sequence (Zawilak et al., 2001; Donczew et al., 2014b). An interesting exception from this general rule is the *T. maritima* origin, where ten 12-nucleotide DnaA boxes were identified (consensus DnaA box: AAACCTACCACC) (Ozaki et al., 2006). As *T. maritima* is one of the most ancient bacteria, it has been proposed that its DnaA boxes may resemble a sequence recognized by the initiator protein in a last common ancestor of the unicellular organisms (Ozaki et al., 2006). Indeed, the *T. maritima* DnaA box sequence shares some similarity to the ORC-binding sites in *Saccharomyces cerevisiae* (TAAACATAAAA) and the Orc1/Cdc6-binding sequences in Archaea (e.g., in *Methanothermobacter thermoautotrophicus* – TTACAGTTGAAA) (Ozaki et al., 2006). Thus, it is probable that the *E. coli*-like nine-mer DnaA box sequence has evolved from this original 12-nucleotide sequence, becoming shortened to nine nucleotides at some point. The last six nucleotides of the *T. maritima* consensus sequence (ACCACC)

Table 1 | Sequences of high-affinity DnaA boxes from various bacteria.

Organism	High-affinity DnaA box sequence
Escherichia coli	TTATCCACA
Vibrio cholerae	TTATCCACA
Pseudomonas putida	TTATCCACA
Bacillus subtilis	TTATCCACA
Caulobacter crescentus	TGATCCACA
Helicobacter pylori	TCATTCACA
Mycobacterium tuberculosis	TTGTCCACA
Streptomyces coelicolor	TTGTCCACA
Thermotoga maritima	AAACCTACCACC

are similar to the corresponding part of the nine-nucleotide DnaA boxes. The importance of this six-nucleotide motif as an integral part of a bacterial DnaA box is further supported by the recent study of the *C. crescentus oriC*, where five six-nucleotide boxes (termed W-boxes) were identified in addition to the two known nine-nucleotide high-affinity G-boxes (Taylor et al., 2011). The W-box consensus sequence is TCCCCA, which deviates from the last six nucleotides of the *E. coli*-like box at the fourth position, and shows a very weak but detectable binding by the DnaA protein. Researchers have also identified atypical six-nucleotide-long DnaA boxes in *E. coli*; located directly in the DUE region and bound only in a single-stranded form, this consensus sequence (AGATCT) represents an alternative type of DnaA-recognized sequence (albeit so far exclusive for *E. coli*) (Speck and Messer, 2001).

The origins of closely related bacteria may be interchangeable to some degree between species, as shown *in vivo* (Koch et al., 2010) and *in vitro* (Jiang et al., 2006) for members of the γ-proteobacteria branch. Among evolutionarily distant bacteria, the situation is more complicated. Considering the similarity of DnaA boxes from different organisms, it is not surprising that the DnaA protein is often able to recognize DnaA boxes and/or whole *oriC* regions in heterologous systems *in vitro*, albeit often with a lower affinity and/or specificity. For example, *S. coelicolor oriC* (*ScoriC*) is bound efficiently by the *M. tuberculosis* and *E. coli* DnaA proteins, but neither protein was able to bend the *ScoriC* structure in the manner of the native DnaA protein (Jakimowicz et al., 2000; Zawilak-Pawlik et al., 2005). Furthermore, the origins of *S. coelicolor*, *M. tuberculosis* and *H. pylori* were not found to be active in *E. coli* cells, even though they are efficiently bound by the *E. coli* DnaA *in vitro*. Interestingly, the *H. pylori* DnaA interacts very poorly with the *E. coli oriC*, indicating that even *in vitro* DnaA/oriC systems from different species may be interchangeable in one setting (i.e., *E. coli* DnaA/*H. pylori oriC*) but not the other (i.e., *H. pylori* DnaA/*E. coli oriC*) (Zawilak-Pawlik et al., 2005). Furthermore, the DnaA proteins of *E. coli* and *B. subtilis* exhibit high affinities toward the same DnaA box sequence and were found to interact in heterologous systems *in vitro*, creating similar oligomeric structures as was visualized by EM (Krause et al., 1997). However, neither was found to trigger open-complex formation on the heterologous origin. This provides further evidence that even when there are apparent similarities, the DnaA-*oriC* systems of individual species are not easily interchangeable.

Such observations may reflect that the mode through which DnaA interacts with particular boxes can differ among bacterial organisms. For example, the *E. coli* DnaA protein interacts efficiently with single, double or multiple DnaA boxes (Weigel et al., 1997; Speck and Messer, 2001). In contrast, the DnaA proteins of some other organisms have been found to strongly prefer two or more boxes over a single box. For example, the *M. tuberculosis* DnaA does not interact with a single box, while the closely related *S. coelicolor* DnaA interacts only weakly with a single box (Zawilak-Pawlik et al., 2005). Similarly, the *H. pylori* DnaA protein has a higher affinity for two boxes vs. a single box (Zawilak et al., 2003). Such observations suggest that the joint action of multiple DnaA monomers may be required for efficient binding in many cases. This is especially true for longer origins, in which a greater number of boxes appears to correlate with an increased importance of cooperative interactions among multiple DnaA monomers, such as suggested for the origins of the Actinomycetes – *M. tuberculosis* and *S. coelicolor* (Zawilak-Pawlik et al., 2005). The affinity of individual DnaA boxes can also vary within and across bacterial origins, with low-, medium-, and high-affinity DnaA boxes present in the origins. Interestingly, the number of low-affinity boxes often exceeds the number of high-affinity sites, such as seen in the *oriC* regions of *E. coli* and *C. crescentus* (**Figures 1**, **2**; Rozgaja et al., 2011). Studies in *E. coli* showed that the number and distribution of low- and high-affinity sites is crucial for the activity of the *oriC* region, including its ability to control the frequency of initiation (Grimwade et al., 2007; Leonard and Grimwade, 2011). Low-affinity DnaA boxes provide a scaffold for DnaA oligomerization, whereas the high-affinity boxes (R1, R2, and R4; **Figure 2**) are believed to provide nucleation sites for the DnaA molecules. Low affinity-sites in *E. coli oriC* are organized into two oppositely oriented arrays separated by box R2 and flanked by boxes R1 and R4, which act as nucleation sites for DnaA oligomers (Rozgaja et al., 2011). The two DnaA oligomers were proposed to be extended by sequential interactions of DnaA monomers with arrayed low-affinity sites to finally form a contiguous DnaA filament (Rozgaja et al., 2011). It was suggested that such mode of a DnaA oligomer formation is directly implicated in origin unwinding since the two arrays of low-affinity sites are not helically phased and connection of the two halves of the oligomer would require specific twisting of the DNA strand, which would create a torsional stress (Rozgaja et al., 2011). It is worth noting that at least some of the origins of other bacteria also exhibit particular orientations of clusters of boxes (e.g., all *H. pylori* boxes share the same orientation) (**Figure 2**), which might indicate sequential binding of DnaA molecules and organized formation of a DnaA oligomer, as in *E. coli*.

The *E. coli* high-affinity boxes (R1, R2, and R4; **Figure 1**) appear to be occupied for the majority of the cell cycle, regardless of the nucleotide state of DnaA. The low-affinity boxes, on the other hand, are preferentially bound by ATP-DnaA (Miller et al., 2009; Rozgaja et al., 2011). The nucleotide state of DnaA is subjected to complex regulation system by RIDA inactivation, DARS-reactivation and rejuvenation as well as *de novo* protein synthesis (for details see Katayama et al., 2010; Leonard and Grimwade, 2011; Kasho et al., 2014) The binding of DnaA to low-affinity sites is additionally facilitated by the DiaA protein, which has been shown to stimulate the assembly of specific ATP-DnaA-*oriC* complexes (Keyamura et al., 2007), as well as specific *oriC*-binding proteins like SeqA, IHF and Fis (see "*oriC* activity is regulated by specific origin-binding proteins").

The DUE region is a typically AT-rich stretch of nucleotides (comprehensively reviewed by Rajewska et al., 2012) that often includes characteristic repeated AT-rich sequences (e.g., that of *E. coli* comprises three 13-mer repeats) separated by short, non-AT-rich insertions. DUE regions are thermodynamically unstable compared to their neighboring sequences, rendering them susceptible to superhelical stress arising from the formation of the DnaA oligomer (Erzberger et al., 2006). The initially unwound region ranges from 20 to 60 bps in size, depending on the

organism, which seems to provide sufficient space to accommodate a replicative helicase, DnaB (**Figure 1**) (Sutton et al., 1998; Abe et al., 2007; Mott et al., 2008; Keyamura et al., 2009). After the initial unwinding in *E. coli*, DnaA binds to single-stranded six-mer ATP-DnaA boxes (6-mer ssATP-DnaA boxes; **Figure 1**) located in the DUE (showing a strong preference for one of the strands), thereby stabilizing the initiation bubble prior to helicase loading (Speck and Messer, 2001). The bacterial DUE regions are always located upstream or downstream one or more DnaA box cluster(s), never in the midst of a cluster. It is important to note that the distance between the DUE and its proximal DnaA-box cluster is critical, as even slight changes were found to inhibit *oriC* unwinding (Hsu et al., 1994).

In sum, the existing evidence clearly shows that the cognate DnaA protein and DnaA boxes coevolved to achieve an optimal level of interaction. The orientation and spacing of DnaA boxes are both important for proper activity of the origin. For example, a change in the length of one helical turn between selected boxes does not affect initiation, but changes corresponding to part of a helical turn are highly detrimental (Woelker and Messer, 1993). At the level of an entire *oriC* region, the arrangement of individual boxes that differ in their affinities generates a specific order and assembly rate for the DnaA oligomer, which unwinds DNA in a precisely selected region called the DUE. From there, initiation events are further controlled by regulatory proteins that bind *oriC* at specific sites, as discussed below.

oriC ACTIVITY IS REGULATED BY SPECIFIC ORIGIN-BINDING PROTEINS

Transmission of genetic material to nascent cells requires precise regulation of chromosome replication and its coordination with the cell cycle. Since chromosomal replication is mainly regulated at the initiation stage, the principal activity of the *oriC* region (i.e., unwinding DNA) is tightly controlled. The relevant protein regulators are primarily involved in controlling the initial assembly of the DnaA oligomer along the origin of replication. The formation of an active orisome depends on the presence of proteins that: (i) regulate DnaA protein activity (e.g., Hda, which regulates the nucleotide-bound state of DnaA); (ii) facilitate the interactions between DnaA monomers (e.g., DiaA, which facilitates the assembly of the DnaA oligomer); or (iii) bind *oriC* and modulate the interaction of the DnaA protein with the origin of replication (Katayama et al., 1998; Kato and Katayama, 2001; Keyamura et al., 2007). In this section, we focus on various sequences that are targeted by the origin binding proteins (oriBPs) (other than DnaA) (**Table 2**) regulating the cell-cycle timing of replication from the *oriC* region (called "oriBP regulators").

The oriBP regulators can be divided into three classes depending on their target sequences: (i) those that interact with DnaA boxes or in their close vicinity; (ii) those that interact with AT-rich sequences within the DUE; and (iii) those that interact with other sequences within *oriC* (**Figure 1**). The oriBPs can also be classified by their sequence specificity and/or function: they may specifically or non-specifically interact with *oriC* to positively or negatively influence the unwinding of the origin. They confer their direct effects by binding to DnaA (or other oriBPs) binding sites, and exert their indirect effects by changing the DNA structure of the origin to modulate the binding of additional oriBPs.

The proteins that regulate replication initiation have been best described for *E. coli*, in which ∼11 *oriC* binding proteins have been identified (**Table 2**). However, we do not yet fully understand the roles played by all of these oriBPs in regulating replication. Here, we use the *E. coli* model to discuss the roles of particular oriBP regulators in the sequential events that are believed to occur following the initiation of replication. When possible, we also describe the roles of counterpart proteins in other bacteria and discuss alternative initiation regulators that are not found in *E. coli*.

In *E. coli*, shortly after chromosomal replication the SeqA protein binds to several sites within the *oriC* region to strictly prevent the initiation of new rounds of replication via a mechanism called "sequestration." SeqA specifically binds the short palindromic sequence, GATC, which is overrepresented within *oriC* compared to the rest of the bacterial chromosome. Newly replicated origins are hemimethylated for about 1/3 of the *E. coli* cell cycle, and SeqA preferentially binds hemimethylated GATC sequences over the fully methylated sequences. Thus, SeqA sequesters the *oriC* region until the GATC sites are fully methylated by the Dam methylase (Campbell and Kleckner, 1990; Lu et al., 1994; Brendler et al., 1995; Slater et al., 1995). SeqA predominantly inhibits replication initiation by blocking DnaA from binding to the R5, I2, I3, τ1, and τ2 sites, which overlap with the GATC sequences (Taghbalout et al., 2000; Nievera et al., 2006). This prevents the DnaA filament from being elongated from the high-affinity DnaA boxes, R1, R2, and R4, although it does not alter their occupation by DnaA (Samitt et al., 1989; Cassler et al., 1995; Nievera et al., 2006). This sequestration mechanism appears to be exclusive to a few DamMT-specifying proteobacteria, as homologs of the *seqA* gene have been identified only in this subset of Gram-negative bacteria (Brézellec et al., 2006).

Another negative regulator of initiation in *E. coli*, the Fis protein, associates with *oriC* throughout most of the cell cycle; similar to SeqA, Fis negatively influences replication initiation by regulating the occupation of DnaA on low-affinity sites (Cassler et al., 1995; Ryan et al., 2004). Fis specifically binds to a single site that is located between R2 and R3, and overlaps with the C3 DnaA binding site (**Figure 1**) (Gille et al., 1991; Filutowicz et al., 1992). Fis binding is thought to competitively inhibit the interaction of DnaA with this region (Ryan et al., 2004), and Fis exhibits a DNA-bending activity that plays a yet-unknown role (Finkel and Johnson, 1992; Ryan et al., 2004).

In addition to competing with DnaA for binding to *oriC*, both Fis and SeqA also negatively regulate the interaction of another oriBP, IHF, with the origin. In contrast to the former two proteins, IHF positively regulates replication initiation (Hwang and Kornberg, 1992; Grimwade et al., 2000; Ryan et al., 2002). As the time of initiation draws near, increasing levels of DnaA trigger the displacement of Fis and the full methylation of DNA weakens SeqA binding, ending the repressive activities of these proteins (Slater et al., 1995; Ryan et al., 2004). The release of SeqA reveals the IHF binding site; displacement of Fis promotes IHF binding; and IHF binding leads to bending of the DNA (Polaczek, 1990; Cassler et al., 1995; Rice et al., 1996; Weisberg et al., 1996;

Table 2 | OriBP (origin binding protein) regulators.

OriBP	Binding sequence and features	Reference
ESCHERICHIA COLI		
SeqA *(sequestration factor A)*	5′-GATC-3′ Binds to *oriC* and inhibits DnaA binding at low affinity sites.	Slater et al., 1995; Brendler et al., 1995; Taghbalout et al., 2000; Nievera et al., 2006
Fis *(factor for inversion stimulation)*	5′-GAACAACAGTTGTTC-3′ Binds to single site in *oriC* and inhibits DnaA binding.	Gille et al., 1991; Filutowicz et al., 1992; Ryan et al., 2004
IHF *(integration host factor)*	5′-GATCAACAACCTG-3′ Binds to single site in *oriC*, stimulates DnaA binding at low affinity sites.	Filutowicz and Roll, 1990; Polaczek, 1990; Grimwade et al., 2000
HU *(histone like U-factor)*	Binds non-specifically to *oriC* and influences: DnaA oligomer stability at *oriC*, IHF binding, and stability of ds DNA helix. Interacts with N-terminus of DnaA.	Bonnefoy and Rouvière-Yaniv, 1992; Hwang and Kornberg, 1992; Ryan et al., 2002
Dps *(DNA-binding protein from starved cells)*	Binds non-specifically to *oriC* and interacts with N-terminus of DnaA.	Chodavarapu et al., 2008b
ArcA *(aerobic respiration control)*	Binds to 13-mer AT-rich repeats, and to DnaA, IHF, IciA binding sites in *oriC*. Influences DnaA interaction with AT-rich region.	Lee et al., 2001
IciA *(inhibitor of chromosomal initiation)*	Binds to 13-mer AT-rich repeats in *oriC* and inhibits unwinding.	Hwang and Kornberg, 1990; Thöny et al., 1991; Hwang et al., 1992
CspD *(stationary phase-induced, stress response protein in the CspA family)*	No apparent target sequence, binds exclusively to ssDNA. Inhibits replication initiation.	Yamanaka et al., 2001
Rob *(right oriC-binding)*	5′-ATCGCACGATCTGTATACTT-3′ Binds to single site in *oriC*. No clear function in initiation regulation.	Skarstad et al., 1993; Martin et al., 1999
H-NS *(heat-stable nucleoid structural protein)*	5′-ATGATCGGTGATCCTG-3′ Binds to single site in *oriC*. Probably requires both Cnu and/or Hha proteins to bind DNA.	Kim et al., 2005; Yun et al., 2012a,b
DpiA *(destabilizer of plasmid inheritance)*	Binds to 13-mer AT-rich repeats in *oriC*. No function in initiation regulation.	Miller et al., 2003
BACILLUS SUBTILIS		
Spo0A *(sporulation specific sigma factor)*	5′-TG[TA]CGAA-3′ Binds to *oriC* and inhibits DnaA binding.	Strauch et al., 1990; Castilla-Llorente et al., 2006
CAULOBACTER CRESCENTUS		
CtrA *(cell transcriptional regulator)*	5′-TTAA[Nx7]TTAA-3′ Binds to *Cori* and inhibits DnaA binding.	Siam and Marczynski, 2000; Taylor et al., 2011
IHF *(integration host factor)*	5′-TAACGCTCTGTT-3′ Binds to single site in *Cori*, displaces CtrA, which facilitates bending of DNA and promotes chromosome replication.	Siam et al., 2003

(Continued)

Table 2 | Continued

OriBP	Binding sequence and features	Reference
HELICOBACTER PYLORI		
HP1021 atypical response regulator	5′-TGTT[TA]C[TA]-3 Binds to *oriC1* and *oriC2* and blocks DNA unwinding at the DUE (within the *oriC2*).	Donczew et al., 2014a
STREPTOMYCES COELICOLOR		
AdpA *(A-factor dependent protein)*	5′-TGGCSNGWWY-3′ Binds to *oriC* and inhibits DnaA binding.	Wolański et al., 2012
MYCOBACTERIUM TUBERCULOSIS		
IciA (Rv1985c) *(inhibitor of chromosomal initiation)*	Binds to AT-rich region within *oriC* and blocks DNA unwinding.	Kumar et al., 2009
MtrA *(Mycobacterium tuberculosis response regulator A)*	5′-GTCACAGCG-3′ Mechanism not known.	Rajagopalan et al., 2010

Swinger and Rice, 2004). IHF then stimulates the binding of DnaA-ATP to low-affinity sites (thus redistributing the DnaA protein) and induces the unwinding of *oriC* (Grimwade et al., 2000). Notably, the transcription of the *dnaA* gene is also subject to regulation by the SeqA protein (Campbell and Kleckner, 1990; Theisen et al., 1993; Bogan and Helmstetter, 1997). Thus, the increased DnaA concentrations that trigger the displacement of Fis displacement presumably reflect the earlier release of the *dnaA* promoter from inhibition by SeqA. In *C. crescentus*, the protein that corresponds to IHF also binds to a single site within the *oriC* of this species (*Cori*). Here, the recognition sequence for IHF overlaps the C-binding site for CtrA, which negatively regulates chromosomal replication in *C. crescentus* (for more on CtrA, see below). In this system, IHF binding leads to the displacement of CtrA from *Cori*, allowing the DNA to bend and promoting replication (Siam et al., 2003).

In *E. coli*, HU is a second positive regulator of initiation. Although this histone-like protein was believed to non-specifically bind DNA, some evidence has suggested that it may interact with *oriC* in a specific manner (Bonnefoy and Rouvière-Yaniv, 1992; Ryan et al., 2002). HU enhances the DnaA-dependent unwinding of *oriC*. This presumably occurs through its ability to bend and destabilize DNA (Hwang and Kornberg, 1992; Ryan et al., 2002). However, HU was further shown to interact with the N-terminal part of DnaA to stabilize the DnaA oligomer assembled at *oriC* (Chodavarapu et al., 2008a), suggesting that *oriC* unwinding may also be stimulated through this additional mechanism. Interestingly, HU was also shown to reduce the binding of DnaA at the DnaA-I3 site (Ryan et al., 2002), and modulate the binding of IHF to *oriC* in a manner dependent on the relative concentrations of IHF and HU (Bonnefoy and Rouvière-Yaniv, 1992).

The oriBPs, Dps, and ArcA, negatively regulate replication initiation in response to oxidative stress and oxygen depletion, respectively (Almirón et al., 1992; Lee et al., 2001; Chodavarapu et al., 2008b). Dps non-specifically binds DNA and interacts with the N-terminus of the DnaA protein to inhibit DNA unwinding (Almirón et al., 1992; Chodavarapu et al., 2008b). It has been suggested that Dps may act as a checkpoint during oxidative stress, delaying initiation until the oxidative DNA damage has been repaired (Chodavarapu et al., 2008b). Under anaerobic conditions, in contrast, ArcA is phosphorylated by a cognate kinase of the two-component system. ArcA-P transcriptionally regulates the genes required to maintain anaerobic growth (Lee et al., 2001), and it is also thought to regulate the activity of *oriC*. *In vitro*, ArcA-P binds a region that contains AT-rich 13-mers and the binding sites for IHF and DnaA (R1 box). It prevents the formation of the open complex without displacing IHF or DnaA from the DNA (Lee et al., 2001), suggesting that ArcA-P may disrupt the interaction between the DnaA protein and the AT-rich region.

Interestingly, ArcA-P is capable of displacing another oriBP, IciA, which specifically binds to the 13-mer AT-rich region and inhibits the unwinding of *oriC* (Hwang and Kornberg, 1990, 1992; Thöny et al., 1991). Interestingly, IciA is also capable of transcriptionally regulating genes known to be involved in DNA replication (e.g., *dnaA*) and amino acid metabolism (Lee et al., 1996; Nandineni and Gowrishankar, 2004; Bouvier et al., 2008). A study of the IciA counterpart in *Mycobacterium tuberculosis* showed that this protein also binds to the AT-rich region of the *oriC* and *in vitro* blocks DnaA-dependent helix opening, and may play a role in maintaining mycobacterial latency (during which DNA replication is arrested) (Kumar et al., 2009).

Regarding additional oriBPs in *E. coli*, the CspD protein reportedly inhibits both the initiation and elongation of chromosomal replication *in vitro* (Yamanaka et al., 2001). Finally, additional proteins capable of specifically binding *oriC* have been identified and described (e.g., Rob, H-NS, and DpiA), but their specific roles and contributions to the replication initiation process are not yet known (Skarstad et al., 1993; Martin et al., 1999; Miller et al., 2003; Kim et al., 2005; Yun et al., 2012a,b).

In bacteria that undergo a complex life cycle (e.g., *Bacillus, Caulobacter,* and *Streptomyces*), the regulation of replication initiation must also be adjusted to the developmental stage to ensure that each nascent cell receives a single copy of the chromosome (Wolański et al., 2014). Recently, master transcription factors known to regulate the expression levels of hundreds of genes involved in cell cycle progression and cell differentiation were

demonstrated to be also involved in controlling frequency of chromosomal replication initiation events. Examples of these are Spo0A, CtrA, and AdpA proteins, which temporally and spatially coordinate chromosome replication with developmental program in *B. subtilis, C. crescentus*, and *S. coelicolor*, respectively (Laub et al., 2000, 2002; Molle et al., 2003; Fujita and Losick, 2005; Fujita et al., 2005; Ohnishi et al., 2005; Wolański et al., 2011). They bind specifically to relevant recognition sequences within the origin of replication and inhibit the binding of DnaA, thereby disrupting assembly of the DnaA oligomer and inhibiting replication initiation (Siam and Marczynski, 2000; Castilla-Llorente et al., 2006; Taylor et al., 2011; Wolański et al., 2012; Boonstra et al., 2013; reviewed in Wolański et al., 2014). In all three cases, the binding sites for these regulators overlap with one or more DnaA binding sites, setting up a competition between the regulator and initiator for binding to *oriC* (**Figure 2**). Interestingly, the activities of Spo0A and CtrA are regulated by phosphorylation, which enhances their binding to DNA. Increasing the levels of these active proteins inhibits chromosomal replication and stimulates the expression levels of various genes responsible for differentiation.

In the pathogenic *Mycobacterium*, *M. tuberculosis*, in addition to IciA, the MtrA protein has been shown to bind the *oriC* region and regulate chromosomal replication (Rajagopalan et al., 2010). MtrA binds specifically to four MtrA boxes that are dispersed throughout the *oriC*, between the DnaA boxes (**Figure 2**). Each MtrA box consists of two direct repeats of GTCACAgcg-like sequences. Mutations in the MtrA binding sequences were found to compromise the replication of the minichromosome (an *oriC* containing plasmid), whereas increased levels of MtrA appear to be associated with deficient autonomous replication of the minichromosome (Rajagopalan et al., 2010). Thus, MtrA may play both positive and negative roles in the initiation of replication. The exact action mechanism of MtrA at *oriC* is not yet known, but it has been suggested that this protein may facilitate or hinder the ability of DnaA to oligomerize at *oriC*, rather than interfering with the direct binding of the initiator protein. MtrA has been identified as a response regulator component of the signal transduction system, MtrAB, which suggests that its role in replication initiation might depend on its phosphorylation status (Via et al., 1996; Fol et al., 2006; Rajagopalan et al., 2010). Interestingly, it has been recently shown that in other pathogenic bacterium, *H. pylori*, the orisome assembly is controlled by HP1021 protein – the orphan response regulator, which was previously shown to affect expression of nearly 80 genes (Pflock et al., 2007). HP1021 competes with DnaA for the binding sites at *oriC* and inhibits DNA unwinding at the DUE site (Donczew et al., 2014a). It suggests that HP1021 controls initiation of *H. pylori* chromosome replication in response to yet unknown stimuli. It is very likely that in numerous bacteria chromosome replication is regulated by signal transduction systems in response to cellular or external stimuli affecting bacterial growth.

CONCLUSION AND OUTLOOK

In sum, the bacterial origins differ across organisms in the organization of their DNA modules, but all origins encode comprehensive instructions for the assembly and disassembly of the orisome-forming proteins, enabling the timely regulation of this first and crucial step in chromosomal replication. The instructions direct the sequential binding of DnaA molecules to the available array of high- and low-affinity DnaA boxes to form a nucleoprotein complex that triggers the unwinding of DNA within the AT-rich region of the *oriC*. The *oriC*-encoded instructions also guide a number of other *oriC*-binding proteins that directly or indirectly respond to environmental signals and induce or repress formation of the DnaA-*oriC* complex, thereby modulating replication initiation. Tight regulation of the initiation process is achieved in all bacteria, albeit via different strategies involving various *oriC* binding proteins, many of which play additional roles in cell-cycle regulation. In pathogens, the functions of some initiation regulators may also depend on interactions with the host cell cycle; however, such interactions have not yet been thoroughly elucidated. It is important to remember that origins do not contain universal instructions. Only origins from very closely related organisms exhibit similar organizations, and the repertoire of regulatory proteins is unique for each species or group of related organisms. That enables a bacterium to perfectly adjust its replication to the cell cycle and coordinate its growth with external stimuli. As reviewed herein, we know a great deal about origins and their structures. To continue progressing in this field, we need detailed analyses of orisome formation, as has already been done for *E. coli* and (to a lesser extent) a limited number of other organisms (e.g., *B. subtilis* or *M. tuberculosis*). Future studies should examine how differences in origin structure are translated to the species-specific characteristics of DnaA oligomerization and its control by regulatory proteins. In addition, many important aspects of the replication initiation process remain to be discovered, particularly in pathogens, including the answers to questions, such as:

How is replication initiation coordinated with the cell cycles of different pathogens upon host infection?
How do host signals modulate or influence replication initiation in pathogenic bacteria?

ACKNOWLEDGMENTS

We are grateful to Dagmara Jakimowicz for providing helpful comments on the manuscript. This work was supported by the National Science Centre, Poland (Maestro, Grant 2012/04/A/NZ1/00057) and by Wroclaw Research Centre EIT+ under the project "Biotechnologies and advanced medical technologies" (BioMed; POIG.01.01.02-02-003/08), which is financed through the European Regional Development Fund (Operational Programme Innovative Economy, 1.1.2). The cost of publication was financed by the Wroclaw Centre of Biotechnology, programme the Leading National Research Centre (KNOW) for years 2014–2018.

REFERENCES

Abe, Y., Jo, T., Matsuda, Y., Matsunaga, C., Katayama, T., and Ueda, T. (2007). Structure and function of DnaA N-terminal domains: specific sites and mechanisms in inter-DnaA interaction and in DnaB helicase loading on oriC. *J. Biol. Chem.* 282, 17816–17827. doi: 10.1074/jbc.M701841200

Akman, L., Yamashita, A., Watanabe, H., Oshima, K., Shiba, T., Hattori, M., et al. (2002). Genome sequence of the endocellular obligate symbiont of

tsetse flies, *Wigglesworthia glossinidia*. *Nat. Genet.* 32, 402–407. doi: 10.1038/ng986

Almirón, M., Link, A. J., Furlong, D., and Kolter, R. (1992). A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*. *Genes Dev.* 6, 2646–2654. doi: 10.1101/gad.6.12b.2646

Baek, J. H., and Chattoraj, D. K. (2014). Chromosome I controls chromosome II replication in *Vibrio cholerae*. *PLoS Genet.* 10:e1004184. doi: 10.1371/journal.pgen.1004184

Bogan, J. A., and Helmstetter, C. E. (1997). DNA sequestration and transcription in the oriC region of *Escherichia coli*. *Mol. Microbiol.* 26, 889–896. doi: 10.1046/j.1365-2958.1997.6221989.x

Bonnefoy, E., and Rouvière-Yaniv, J. (1992). HU, the major histone-like protein of *E. coli*, modulates the binding of IHF to oriC. *EMBO J.* 11, 4489–4496.

Boonstra, M., de Jong, I. G., Scholefield, G., Murray, H., Kuipers, O. P., and Veening, J.-W. (2013). Spo0A regulates chromosome copy number during sporulation by directly binding to the origin of replication in *Bacillus subtilis*. *Mol. Microbiol.* 87, 925–938. doi: 10.1111/mmi.12141

Bouvier, J., Stragier, P., Morales, V., Rémy, E., and Gutierrez, C. (2008). Lysine represses transcription of the *Escherichia coli* dapB gene by preventing its activation by the ArgP activator. *J. Bacteriol.* 190, 5224–5229. doi: 10.1128/JB.01782-07

Brendler, T., Abeles, A., and Austin, S. (1995). A protein that binds to the P1 origin core and the oriC 13mer region in a methylation-specific fashion is the product of the host seqA gene. *EMBO J.* 14, 4083–4089.

Brézellec, P., Hoebeke, M., Hiet, M.-S., Pasek, S., and Ferat, J.-L. (2006). DomainSieve: a protein domain-based screen that led to the identification of dam-associated genes with potential link to DNA maintenance. *Bioinformatics* 22, 1935–1941. doi: 10.1093/bioinformatics/btl336

Briggs, G. S., Smits, W. K., and Soultanas, P. (2012). Chromosomal replication initiation machinery of low-G+C-content Firmicutes. *J. Bacteriol.* 194, 5162–5170. doi: 10.1128/JB.00865-12

Campbell, J. L., and Kleckner, N. (1990). *E. coli* oriC and the dnaA gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. *Cell* 62, 967–979. doi: 10.1016/0092-8674(90)90271-F

Cassler, M. R., Grimwade, J. E., and Leonard, A. C. (1995). Cell cycle-specific changes in nucleoprotein complexes at a chromosomal replication origin. *EMBO J.* 14, 5833–5841.

Castilla-Llorente, V., Munoz-Espin, D., Villar, L., Salas, M., and Meijer, W. J. J. (2006). Spo0A, the key transcriptional regulator for entrance into sporulation, is an inhibitor of DNA replication. *EMBO J.* 25, 3890–3899. doi: 10.1038/sj.emboj.7601266

Chodavarapu, S., Felczak, M. M., Yaniv, J. R., and Kaguni, J. M. (2008a). *Escherichia coli* DnaA interacts with HU in initiation at the *E. coli* replication origin. *Mol. Microbiol.* 67, 781–792. doi: 10.1111/j.1365-2958.2007.06094.x

Chodavarapu, S., Gomez, R., Vicente, M., and Kaguni, J. M. (2008b). *Escherichia coli* Dps interacts with DnaA protein to impede initiation: a model of adaptive mutation. *Mol. Microbiol.* 67, 1331–1346. doi: 10.1111/j.1365-2958.2008.06127.x

Demarre, G., and Chattoraj, D. K. (2010). DNA Adenine methylation is required to replicate both *Vibrio cholerae* chromosomes once per cell cycle. *PLoS Genet.* 6:e1000939. doi: 10.1371/journal.pgen.1000939

Donczew, R., Lurz, R., Zakrzewska-Czerwinska, J., and Zawilak-Pawlik, A. (2012). *Helicobacter pylori* oriC–the first bipartite origin of chromosome replication in Gram-negative bacteria. *Nucleic Acids Res.* 40, 9647–9660. doi: 10.1093/nar/gks742

Donczew, R., Makowski, L., Jaworski, P., Bezulska, M., Nowaczyk, M., Zakrzewska-Czerwińska, J., et al. (2014a). The atypical response regulator HP1021 controls formation of the *Helicobacter pylori* replication initiation complex. *Mol. Microbiol.* doi: 10.1111/mmi.12866

Donczew, R., Mielke, T., Jaworski, P., Zakrzewska-Czerwińska, J., and Zawilak-Pawlik, A. (2014b). Assembly of *Helicobacter pylori* initiation complex is determined by sequence-specific and topology-sensitive DnaA-oriC interactions. *J. Mol. Biol.* doi: 10.1016/j.jmb.2014.05.018

Duderstadt, K. E., Chuang, K., and Berger, J. M. (2011). DNA stretching by bacterial initiators promotes replication origin opening. *Nature* 478, 209–213. doi: 10.1038/nature10455

Duigou, S., Knudsen, K. G., Skovgaard, O., Egan, E. S., Løbner-Olesen, A., and Waldor, M. K. (2006). Independent control of replication initiation of the two *Vibrio cholerae* chromosomes by DnaA and RctB. *J. Bacteriol.* 188, 6419–6424. doi: 10.1128/JB.00565-06

Egan, E. S., and Waldor, M. K. (2003). Distinct replication requirements for the two *Vibrio cholerae* chromosomes. *Cell* 114, 521–530. doi: 10.1016/S0092-8674(03)00611-1

Erzberger, J. P., Mott, M. L., and Berger, J. M. (2006). Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nat. Struct. Mol. Biol.* 13, 676–683. doi: 10.1038/nsmb1115

Filutowicz, M., and Roll, J. (1990). The requirement of IHF protein for extrachromosomal replication of the *Escherichia coli* oriC in a mutant deficient in DNA polymerase I activity. *New Biol.* 2, 818–827.

Filutowicz, M., Ross, W., Wild, J., and Gourse, R. L. (1992). Involvement of Fis protein in replication of the *Escherichia coli* chromosome. *J. Bacteriol.* 174, 398–407.

Finkel, S. E., and Johnson, R. C. (1992). The Fis protein: it's not just for DNA inversion anymore. *Mol. Microbiol.* 6, 3257–3265. doi: 10.1111/j.1365-2958.1992.tb02193.x

Fol, M., Chauhan, A., Nair, N. K., Maloney, E., Moomey, M., Jagannath, C., et al. (2006). Modulation of *Mycobacterium tuberculosis* proliferation by MtrA, an essential two-component response regulator. *Mol. Microbiol.* 60, 643–657. doi: 10.1111/j.1365-2958.2006.05137.x

Fujikawa, N., Kurumizaka, H., Nureki, O., Terada, T., Shirouzu, M., Katayama, T., et al. (2003). Structural basis of replication origin recognition by the DnaA protein. *Nucleic Acids Res.* 31, 2077–2086. doi: 10.1093/nar/gkg309

Fujita, M., González-Pastor, J. E., and Losick, R. (2005). High- and low-threshold genes in the Spo0A regulon of *Bacillus subtilis*. *J. Bacteriol.* 187, 1357–1368. doi: 10.1128/JB.187.4.1357-1368.2005

Fujita, M., and Losick, R. (2005). Evidence that entry into sporulation in *Bacillus subtilis* is governed by a gradual increase in the level and activity of the master regulator Spo0A. *Genes Dev.* 19, 2236–2244. doi: 10.1101/gad.1335705

Fukuoka, T., Moriya, S., Yoshikawa, H., and Ogasawara, N. (1990). Purification and characterization of an initiation protein for chromosomal replication, DnaA, in *Bacillus subtilis*. *J. Biochem.* 107, 732–739.

Fuller, R. S., Funnell, B. E., and Kornberg, A. (1984). The dnaA protein complex with the *E. coli* chromosomal replication origin (oriC) and other DNA sites. *Cell* 38, 889–900. doi: 10.1016/0092-8674(84)90284-8

Gao, F., Luo, H., and Zhang, C.-T. (2013). DoriC 5.0: an updated database of oriC regions in both bacterial and archaeal genomes. *Nucleic Acids Res.* 41, D90–93. doi: 10.1093/nar/gks990

Gil, R., Silva, F. J., Zientz, E., Delmotte, F., González-Candelas, F., Latorre, A., et al. (2003). The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *Proc. Natl. Acad. Sci. U.S.A.* 100, 9388–9393. doi: 10.1073/pnas.1533499100

Gille, H., Egan, J. B., Roth, A., and Messer, W. (1991). The FIS protein binds and bends the origin of chromosomal DNA replication, oriC, of *Escherichia coli*. *Nucleic Acids Res.* 19, 4167–4172. doi: 10.1093/nar/19.15.4167

Grimwade, J. E., Ryan, V. T., and Leonard, A. C. (2000). IHF redistributes bound initiator protein, DnaA, on supercoiled oriC of *Escherichia coli*. *Mol. Microbiol.* 35, 835–844. doi: 10.1046/j.1365-2958.2000.01755.x

Grimwade, J. E., Torgue, J. J.-C., McGarry, K. C., Rozgaja, T., Enloe, S. T., and Leonard, A. C. (2007). Mutational analysis reveals *Escherichia coli* oriC interacts with both DnaA-ATP and DnaA-ADP during pre-RC assembly. *Mol. Microbiol.* 66, 428–439. doi: 10.1111/j.1365-2958.2007.05930.x

Harding, N. E., Cleary, J. M., Smith, D. W., Michon, J. J., Brusilow, W. S., and Zyskind, J. W. (1982). Chromosomal replication origins (oriC) of *Enterobacter aerogenes* and *Klebsiella pneumoniae* are functional in *Escherichia coli*. *J. Bacteriol.* 152, 983–993.

Hsu, J., Bramhill, D., and Thompson, C. M. (1994). Open complex formation by DnaA initiation protein at the *Escherichia coli* chromosomal origin requires the 13-mers precisely spaced relative to the 9-mers. *Mol. Microbiol.* 11, 903–911. doi: 10.1111/j.1365-2958.1994.tb00369.x

Hwang, D. S., and Kornberg, A. (1990). A novel protein binds a key origin sequence to block replication of an *E. coli* minichromosome. *Cell* 63, 325–331. doi: 10.1016/0092-8674(90)90165-B

Hwang, D. S., and Kornberg, A. (1992). Opening of the replication origin of *Escherichia coli* by DnaA protein with protein HU or IHF. *J. Biol. Chem.* 267, 23083–23086.

Hwang, D. S., Thöny, B., and Kornberg, A. (1992). IciA protein, a specific inhibitor of initiation of *Escherichia coli* chromosomal replication. *J. Biol. Chem.* 267, 2209–2213.

Jakimowicz, D., Majkadagger, J., Konopa, G., Wegrzyn, G., Messer, W., Schrempf, H., et al. (2000). Architecture of the *Streptomyces lividans* DnaA protein-replication origin complexes. *J. Mol. Biol.* 298, 351–364. doi: 10.1006/jmbi.2000.3686

Jakimowicz, D., Majka, J., Messer, W., Speck, C., Fernandez, M., Martin, M. C., et al. (1998). Structural elements of the Streptomyces oriC region and their interactions with the DnaA protein. *Microbiology* 144(Pt 5), 1281–1290. doi: 10.1099/00221287-144-5-1281

Jha, J. K., Baek, J., Venkova-Canova, T., and Chattoraj, D. K. (2012). Chromosome dynamics in multichromosome bacteria. *Biochim. Biophys. Acta* 1819, 826–829. doi: 10.1016/j.bbagrm.2012.01.012

Jiang, Y., Yao, S., Helinski, D., and Toukdarian, A. (2006). Functional analysis of two putative chromosomal replication origins from *Pseudomonas aeruginosa*. *Plasmid* 55, 194–200. doi: 10.1016/j.plasmid.2005.11.001

Kaguni, J. M. (2006). DnaA: controlling the initiation of bacterial DNA replication and more. *Annu. Rev. Microbiol.* 60, 351–375. doi: 10.1146/annurev.micro.60.080805.142111

Kaguni, J. M. (2011). Replication initiation at the *Escherichia coli* chromosomal origin. *Curr. Opin. Chem. Biol.* 15, 606–613. doi: 10.1016/j.cbpa.2011.07.016

Kasho, K., Fujimitsu, K., Matoba, T., Oshima, T., and Katayama, T. (2014). Timely binding of IHF and Fis to DARS2 regulates ATP-DnaA production and replication initiation. *Nucleic Acids Res.* 42, 13134–13149. doi: 10.1093/nar/gku1051

Katayama, T., Kubota, T., Kurokawa, K., Crooke, E., and Sekimizu, K. (1998). The initiator function of DnaA protein is negatively regulated by the sliding clamp of the *E. coli* chromosomal replicase. *Cell* 94, 61–71. doi: 10.1016/S0092-8674(00)81222-2

Katayama, T., Ozaki, S., Keyamura, K., and Fujimitsu, K. (2010). Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC. *Nat. Rev. Microbiol.* 8, 163–170. doi: 10.1038/nrmicro2314

Kato, J., and Katayama, T. (2001). Hda, a novel DnaA-related protein, regulates the replication cycle in *Escherichia coli*. *EMBO J.* 20, 4253–4262. doi: 10.1093/emboj/20.15.4253

Keyamura, K., Abe, Y., Higashi, M., Ueda, T., and Katayama, T. (2009). DiaA dynamics are coupled with changes in initial origin complexes leading to helicase loading. *J. Biol. Chem.* 284, 25038–25050. doi: 10.1074/jbc.M109.002717

Keyamura, K., Fujikawa, N., Ishida, T., Ozaki, S., Su'etsugu, M., Fujimitsu, K., et al. (2007). The interaction of DiaA and DnaA regulates the replication cycle in *E. coli* by directly promoting ATP DnaA-specific initiation complexes. *Genes Dev.* 21, 2083–2099. doi: 10.1101/gad.1561207

Kim, M. S., Bae, S.-H., Yun, S. H., Lee, H. J., Ji, S. C., Lee, J. H., et al. (2005). Cnu, a novel oriC-binding protein of *Escherichia coli*. *J. Bacteriol.* 187, 6998–7008. doi: 10.1128/JB.187.20.6998-7008.2005

Koch, B., Ma, X., and Løbner-Olesen, A. (2010). Replication of *Vibrio cholerae* chromosome I in *Escherichia coli*: dependence on dam methylation. *J. Bacteriol.* 192, 3903–3914. doi: 10.1128/JB.00311-10

Krause, M., Rückert, B., Lurz, R., and Messer, W. (1997). Complexes at the replication origin of *Bacillus subtilis* with homologous and heterologous DnaA protein. *J. Mol. Biol.* 274, 365–380. doi: 10.1006/jmbi.1997.1404

Kumar, S., Farhana, A., and Hasnain, S. E. (2009). *In-vitro* helix opening of M. *tuberculosis* oriC by DnaA occurs at precise location and is inhibited by IciA like protein. *PloS ONE* 4:e4139. doi: 10.1371/journal.pone.0004139

Kwan, J. C., and Schmidt, E. W. (2013). Bacterial endosymbiosis in a chordate host: long-term co-evolution and conservation of secondary metabolism. *PloS ONE* 8:e80822. doi: 10.1371/journal.pone.0080822

Lartigue, C., Blanchard, A., Renaudin, J., Thiaucourt, F., and Sirand-Pugnet, P. (2003). Host specificity of mollicutes oriC plasmids: functional analysis of replication origin. *Nucleic Acids Res.* 31, 6610–6618. doi: 10.1093/nar/gkg848

Laub, M. T., Chen, S. L., Shapiro, L., and McAdams, H. H. (2002). Genes directly controlled by CtrA, a master regulator of the Caulobacter cell cycle. *Proc. Natl. Acad. Sci. U.S.A.* 99, 4632–4637. doi: 10.1073/pnas.062065699

Laub, M. T., McAdams, H. H., Feldblyum, T., Fraser, C. M., and Shapiro, L. (2000). Global analysis of the genetic network controlling a bacterial cell cycle. *Science* 290, 2144–2148. doi: 10.1126/science.290.5499.2144

Lee, S.-W., Browning, G. F., and Markham, P. F. (2008). Development of a replicable oriC plasmid for *Mycoplasma gallisepticum* and *Mycoplasma imitans*, and gene disruption through homologous recombination in *M. gallisepticum*. *Microbiology* 154, 2571–2580. doi: 10.1099/mic.0.2008/019208-0

Lee, Y. S., Han, J. S., Jeon, Y., and Hwang, D. S. (2001). The arc two-component signal transduction system inhibits *in vitro Escherichia coli* chromosomal initiation. *J. Biol. Chem.* 276, 9917–9923. doi: 10.1074/jbc.M008629200

Lee, Y. S., Kim, H., and Hwang, D. S. (1996). Transcriptional activation of the dnaA gene encoding the initiator for oriC replication by IciA protein, an inhibitor of *in vitro* oriC replication in *Escherichia coli*. *Mol. Microbiol.* 19, 389–396. doi: 10.1046/j.1365-2958.1996.485902.x

Leonard, A. C., and Grimwade, J. E. (2010). Regulating DnaA complex assembly: it is time to fill the gaps. *Curr. Opin. Microbiol.* 13, 766–772. doi: 10.1016/j.mib.2010.10.001

Leonard, A. C., and Grimwade, J. E. (2011). Regulation of DnaA assembly and activity: taking directions from the genome. *Annu. Rev. Microbiol.* 65, 19–35. doi: 10.1146/annurev-micro-090110-102934

Leonard, A. C., and Méchali, M. (2013). DNA replication origins. *Cold Spring Harb. Perspect. Biol.* 5:a010116. doi: 10.1101/cshperspect.a010116

Lu, M., Campbell, J. L., Boye, E., and Kleckner, N. (1994). SeqA: a negative modulator of replication initiation in *E. coli*. *Cell* 77, 413–426. doi: 10.1016/0092-8674(94)90156-2

Mackiewicz, P., Zakrzewska-Czerwinska, J., Zawilak, A., Dudek, M. R., and Cebrat, S. (2004). Where does bacterial replication start? Rules for predicting the oriC region. *Nucleic Acids Res.* 32, 3781–3791. doi: 10.1093/nar/gkh699

Martin, R. G., Gillette, W. K., Rhee, S., and Rosner, J. L. (1999). Structural requirements for marbox function in transcriptional activation of mar/sox/rob regulon promoters in *Escherichia coli*: sequence, orientation and spatial relationship to the core promoter. *Mol. Microbiol.* 34, 431–441. doi: 10.1046/j.1365-2958.1999.01599.x

Messer, W. (2002). The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiol. Rev.* 26, 355–374. doi: 10.1111/j.1574-6976.2002.tb00620.x

Miller, C., Ingmer, H., Thomsen, L. E., Skarstad, K., and Cohen, S. N. (2003). DpiA binding to the replication origin of *Escherichia coli* plasmids and chromosomes destabilizes plasmid inheritance and induces the bacterial SOS response. *J. Bacteriol.* 185, 6025–6031. doi: 10.1128/JB.185.20.6025-6031.2003

Miller, D. T., Grimwade, J. E., Betteridge, T., Rozgaja, T., Torgue, J. J.-C., and Leonard, A. C. (2009). Bacterial origin recognition complexes direct assembly of higher-order DnaA oligomeric structures. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18479–18484. doi: 10.1073/pnas.0909472106

Molle, V., Fujita, M., Jensen, S. T., Eichenberger, P., González-Pastor, J. E., Liu, J. S., et al. (2003). The Spo0A regulon of *Bacillus subtilis*. *Mol. Microbiol.* 50, 1683–1701. doi: 10.1046/j.1365-2958.2003.03818.x

Moriya, S., Atlung, T., Hansen, F. G., Yoshikawa, H., and Ogasawara, N. (1992). Cloning of an autonomously replicating sequence (ars) from the *Bacillus subtilis* chromosome. *Mol. Microbiol.* 6, 309–315. doi: 10.1111/j.1365-2958.1992.tb01473.x

Moriya, S., Imai, Y., Hassan, A. K., and Ogasawara, N. (1999). Regulation of initiation of *Bacillus subtilis* chromosome replication. *Plasmid* 41, 17–29. doi: 10.1006/plas.1998.1381

Moriya, S., Kawai, Y., Kaji, S., Smith, A., Harry, E. J., and Errington, J. (2009). Effects of oriC relocation on control of replication initiation in *Bacillus subtilis*. *Microbiology* 155, 3070–3082. doi: 10.1099/mic.0.030080-0

Mott, M. L., and Berger, J. M. (2007). DNA replication initiation: mechanisms and regulation in bacteria. *Nat. Rev. Microbiol.* 5, 343–354. doi: 10.1038/nrmicro1640

Mott, M. L., Erzberger, J. P., Coons, M. M., and Berger, J. M. (2008). Structural synergy and molecular crosstalk between bacterial helicase loaders and replication initiators. *Cell* 135, 623–634. doi: 10.1016/j.cell.2008.09.058

Nandineni, M. R., and Gowrishankar, J. (2004). Evidence for an arginine exporter encoded by yggA (argO) that is regulated by the LysR-type transcriptional regulator ArgP in *Escherichia coli*. *J. Bacteriol.* 186, 3539–3546. doi: 10.1128/JB.186.11.3539-3546.2004

Nievera, C., Torgue, J. J.-C., Grimwade, J. E., and Leonard, A. C. (2006). SeqA blocking of DnaA-oriC interactions ensures staged assembly of the *E. coli* pre-RC. *Mol. Cell* 24, 581–592. doi: 10.1016/j.molcel.2006.09.016

O'Neill, E. A., and Bender, R. A. (1988). Klebsiella pneumoniae origin of replication (oriC) is not active in *Caulobacter crescentus*, *Pseudomonas putida*, and *Rhodobacter sphaeroides*. *J. Bacteriol.* 170, 3774–3777.

Ogasawara, N., Moriya, S., and Yoshikawa, H. (1991). Initiation of chromosome replication: structure and function of oriC and DnaA protein in eubacteria. *Res. Microbiol.* 142, 851–859. doi: 10.1016/0923-2508(91)90065-I

Ohnishi, Y., Yamazaki, H., Kato, J.-Y., Tomono, A., and Horinouchi, S. (2005). AdpA, a central transcriptional regulator in the A-factor regulatory cascade that leads to morphological development and secondary metabolism in *Streptomyces griseus*. *Biosci. Biotechnol. Biochem.* 69, 431–439. doi: 10.1271/bbb.69.431

Ozaki, S., Fujimitsu, K., Kurumizaka, H., and Katayama, T. (2006). The DnaA homolog of the hyperthermophilic eubacterium *Thermotoga maritima* forms an open complex with a minimal 149-bp origin region in an ATP-dependent manner. *Genes Cells* 11, 425–438. doi: 10.1111/j.1365-2443.2006.00950.x

Ozaki, S., and Katayama, T. (2012). Highly organized DnaA-oriC complexes recruit the single-stranded DNA for replication initiation. *Nucleic Acids Res.* 40, 1648–1665. doi: 10.1093/nar/gkr832

Pflock, M., Bathon, M., Schär, J., Müller, S., Mollenkopf, H., Meyer, T. F., et al. (2007). The orphan response regulator HP1021 of *Helicobacter pylori* regulates transcription of a gene cluster presumably involved in acetone metabolism. *J. Bacteriol.* 189, 2339–2349. doi: 10.1128/JB.01827-06

Polaczek, P. (1990). Bending of the origin of replication of *E. coli* by binding of IHF at a specific site. *New Biol.* 2, 265–271.

Rajagopalan, M., Dziedzic, R., Al Zayer, M., Stankowska, D., Ouimet, M.-C., Bastedo, D. P., et al. (2010). *Mycobacterium tuberculosis* origin of replication and the promoter for immunodominant secreted antigen 85B are the targets of MtrA, the essential response regulator. *J. Biol. Chem.* 285, 15816–15827. doi: 10.1074/jbc.M109.040097

Rajewska, M., Wegrzyn, K., and Konieczny, I. (2012). AT-rich region and repeated sequences – the essential elements of replication origins of bacterial replicons. *FEMS Microbiol. Rev.* 36, 408–434. doi: 10.1111/j.1574-6976.2011.00300.x

Rice, P. A., Yang, S., Mizuuchi, K., and Nash, H. A. (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* 87, 1295–1306. doi: 10.1016/S0092-8674(00)81824-3

Roggenkamp, A. (2007). Phylogenetic analysis of enteric species of the family Enterobacteriaceae using the oriC-locus. *Syst. Appl. Microbiol.* 30, 180–188. doi: 10.1016/j.syapm.2006.06.004

Rozgaja, T. A., Grimwade, J. E., Iqbal, M., Czerwonka, C., Vora, M., and Leonard, A. C. (2011). Two oppositely oriented arrays of low-affinity recognition sites in oriC guide progressive binding of DnaA during *Escherichia coli* pre-RC assembly. *Mol. Microbiol.* 82, 475–488. doi: 10.1111/j.1365-2958.2011.07827.x

Ryan, V. T., Grimwade, J. E., Camara, J. E., Crooke, E., and Leonard, A. C. (2004). *Escherichia coli* prereplication complex assembly is regulated by dynamic interplay among Fis, IHF and DnaA. *Mol. Microbiol.* 51, 1347–1359. doi: 10.1046/j.1365-2958.2003.03906.x

Ryan, V. T., Grimwade, J. E., Nievera, C. J., and Leonard, A. C. (2002). IHF and HU stimulate assembly of pre-replication complexes at *Escherichia coli* oriC by two different mechanisms. *Mol. Microbiol.* 46, 113–124. doi: 10.1046/j.1365-2958.2002.03129.x

Samitt, C. E., Hansen, F. G., Miller, J. F., and Schaechter, M. (1989). *In vivo* studies of DnaA binding to the origin of replication of *Escherichia coli*. *EMBO J.* 8, 989–993.

Scholefield, G., Errington, J., and Murray, H. (2012). Soj/ParA stalls DNA replication by inhibiting helix formation of the initiator protein DnaA. *EMBO J.* 31, 1542–1555. doi: 10.1038/emboj.2012.6

Shaheen, S. M., Ouimet, M.-C., and Marczynski, G. T. (2009). Comparative analysis of Caulobacter chromosome replication origins. *Microbiology* 155, 1215–1225. doi: 10.1099/mic.0.025528-0

Siam, R., Brassinga, A. K. C., and Marczynski, G. T. (2003). A dual binding site for integration host factor and the response regulator CtrA inside the *Caulobacter crescentus* replication origin. *J. Bacteriol.* 185, 5563–5572. doi: 10.1128/JB.185.18.5563-5572.2003

Siam, R., and Marczynski, G. T. (2000). Cell cycle regulator phosphorylation stimulates two distinct modes of binding at a chromosome replication origin. *EMBO J.* 19, 1138–1147. doi: 10.1093/emboj/19.5.1138

Skarstad, K., and Katayama, T. (2013). Regulating DNA replication in bacteria. *Cold Spring Harb. Perspect. Biol.* 5:a012922. doi: 10.1101/cshperspect.a012922

Skarstad, K., Thöny, B., Hwang, D. S., and Kornberg, A. (1993). A novel binding protein of the origin of the *Escherichia coli* chromosome. *J. Biol. Chem.* 268, 5365–5370.

Slager, J., Kjos, M., Attaiech, L., and Veening, J.-W. (2014). Antibiotic-induced replication stress triggers bacterial competence by increasing gene dosage near the origin. *Cell* 157, 395–406. doi: 10.1016/j.cell.2014.01.068

Slater, S., Wold, S., Lu, M., Boye, E., Skarstad, K., and Kleckner, N. (1995). *E. coli* SeqA protein binds oriC in two different methyl-modulated reactions appropriate to its roles in DNA replication initiation and origin sequestration. *Cell* 82, 927–936. doi: 10.1016/0092-8674(95)90272-4

Speck, C., and Messer, W. (2001). Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA. *EMBO J.* 20, 1469–1476. doi: 10.1093/emboj/20.6.1469

Strauch, M., Webb, V., Spiegelman, G., and Hoch, J. A. (1990). The SpoOA protein of *Bacillus* subtilis is a repressor of the abrB gene. *Proc. Natl. Acad. Sci. U.S.A.* 87, 1801–1805. doi: 10.1073/pnas.87.5.1801

Sutton, M. D., Carr, K. M., Vicente, M., and Kaguni, J. M. (1998). *Escherichia coli* DnaA protein. The N-terminal domain and loading of DnaB helicase at the E. coli chromosomal origin. *J. Biol. Chem.* 273, 34255–34262. doi: 10.1074/jbc.273.51.34255

Suvorov, A. N., and Ferretti, J. J. (2000). Replication origin of *Streptococcus pyogenes*, organization and cloning in heterologous systems. *FEMS Microbiol. Lett.* 189, 293–297. doi: 10.1111/j.1574-6968.2000.tb09246.x

Swinger, K. K., and Rice, P. A. (2004). IHF and HU: flexible architects of bent DNA. *Curr. Opin. Struct. Biol.* 14, 28–35. doi: 10.1016/j.sbi.2003.12.003

Taghbalout, A., Landoulsi, A., Kern, R., Yamazoe, M., Hiraga, S., Holland, B., et al. (2000). Competition between the replication initiator DnaA and the sequestration factor SeqA for binding to the hemimethylated chromosomal origin of *E. coli* in vitro. *Genes Cells* 5, 873–884. doi: 10.1046/j.1365-2443.2000.00380.x

Takeda, Y., Harding, N. E., Smith, D. W., and Zyskind, J. W. (1982). The chromosomal origin of replication (oriC) of *Erwinia carotovora*. *Nucleic Acids Res.* 10, 2639–2650. doi: 10.1093/nar/10.8.2639

Taylor, J. A., Ouimet, M.-C., Wargachuk, R., and Marczynski, G. T. (2011). The *Caulobacter crescentus* chromosome replication origin evolved two classes of weak DnaA binding sites. *Mol. Microbiol.* 82, 312–326. doi: 10.1111/j.1365-2958.2011.07785.x

Theisen, P. W., Grimwade, J. E., Leonard, A. C., Bogan, J. A., and Helmstetter, C. E. (1993). Correlation of gene transcription with the time of initiation of chromosome replication in *Escherichia coli*. *Mol. Microbiol.* 10, 575–584. doi: 10.1111/j.1365-2958.1993.tb00929.x

Thöny, B., Hwang, D. S., Fradkin, L., and Kornberg, A. (1991). iciA, an *Escherichia coli* gene encoding a specific inhibitor of chromosomal initiation of replication *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* 88, 4066–4070. doi: 10.1073/pnas.88.10.4066

Tsodikov, O. V., and Biswas, T. (2011). Structural and thermodynamic signatures of DNA recognition by *Mycobacterium tuberculosis* DnaA. *J. Mol. Biol.* 410, 461–476. doi: 10.1016/j.jmb.2011.05.007

Via, L. E., Curcic, R., Mudd, M. H., Dhandayuthapani, S., Ulmer, R. J., and Deretic, V. (1996). Elements of signal transduction in *Mycobacterium tuberculosis*: *in vitro* phosphorylation and *in vivo* expression of the response regulator MtrA. *J. Bacteriol.* 178, 3314–3321.

Weigel, C., Schmidt, A., Rückert, B., Lurz, R., and Messer, W. (1997). DnaA protein binding to individual DnaA boxes in the *Escherichia coli* replication origin, oriC. *EMBO J.* 16, 6574–6583. doi: 10.1093/emboj/16.21.6574

Weisberg, R. A., Freundlich, M., Friedman, D., Gardner, J., Goosen, N., Nash, H., et al. (1996). Nomenclature of the genes encoding IHF. *Mol. Microbiol.* 19, 642. doi: 10.1046/j.1365-2958.1996.t01-2-442924.x

Woelker, B., and Messer, W. (1993). The structure of the initiation complex at the replication origin, oriC, of *Escherichia coli*. *Nucleic Acids Res.* 21, 5025–5033. doi: 10.1093/nar/21.22.5025

Wolański, M., Donczew, R., Kois-Ostrowska, A., Masiewicz, P., Jakimowicz, D., and Zakrzewska-Czerwinska, J. (2011). The level of AdpA directly affects expression of developmental genes in *Streptomyces coelicolor*. *J. Bacteriol.* 193, 6358–6365. doi: 10.1128/JB.05734-11

Wolański, M., Jakimowicz, D., and Zakrzewska-Czerwińska, J. (2012). AdpA, key regulator for morphological differentiation regulates bacterial chromosome replication. *Open Biol.* 2, 120097. doi: 10.1098/rsob.120097

Wolański, M., Jakimowicz, D., and Zakrzewska-Czerwińska, J. (2014). Fifty years after the replicon hypothesis: cell-specific master regulators as new players in chromosome replication control. *J. Bacteriol.* 196, 2901–2911. doi: 10.1128/JB.01706-14

Yamanaka, K., Zheng, W., Crooke, E., Wang, Y. H., and Inouye, M. (2001). CspD, a novel DNA replication inhibitor induced during the stationary phase in *Escherichia coli*. *Mol. Microbiol.* 39, 1572–1584. doi: 10.1046/j.1365-2958.2001.02345.x

Yee, T. W., and Smith, D. W. (1990). Pseudomonas chromosomal replication origins: a bacterial class distinct from *Escherichia coli*-type origins. *Proc. Natl. Acad. Sci. U.S.A.* 87, 1278–1282. doi: 10.1073/pnas.87.4.1278

Yun, S. H., Ji, S. C., Jeon, H. J., Wang, X., Kim, S. W., Bak, G., et al. (2012a). The CnuK9E H-NS complex antagonizes DNA binding of DicA and leads to temperature-dependent filamentous growth in *E. coli*. *PloS ONE* 7:e45236. doi: 10.1371/journal.pone.0045236

Yun, S. H., Ji, S. C., Jeon, H. J., Wang, X., Lee, Y., Choi, B.-S., et al. (2012b). A mutational study of Cnu reveals attractive forces between Cnu and H-NS. *Mol. Cells* 33, 211–216. doi: 10.1007/s10059-012-0006-5

Zakrzewska-Czerwińska, J., Jakimowicz, D., Zawilak-Pawlik, A., and Messer, W. (2007). Regulation of the initiation of chromosomal replication in bacteria. *FEMS Microbiol. Rev.* 31, 378–387. doi: 10.1111/j.1574-6976.2007.00070.x

Zakrzewska-Czerwińska, J., and Schrempf, H. (1992). Characterization of an autonomously replicating region from the *Streptomyces lividans* chromosome. *J. Bacteriol.* 174, 2688–2693.

Zawilak, A., Cebrat, S., Mackiewicz, P., Król-Hulewicz, A., Jakimowicz, D., Messer, W., et al. (2001). Identification of a putative chromosomal replication origin from *Helicobacter pylori* and its interaction with the initiator protein DnaA. *Nucleic Acids Res.* 29, 2251–2259. doi: 10.1093/nar/29.11.2251

Zawilak, A., Durrant, M. C., Jakimowicz, P., Backert, S., and Zakrzewska-Czerwińska, J. (2003). DNA binding specificity of the replication initiator protein, DnaA from *Helicobacter pylori*. *J. Mol. Biol.* 334, 933–947. doi: 10.1016/j.jmb.2003.10.018

Zawilak, A., Kois, A., Konopa, G., Smulczyk-Krawczyszyn, A., and Zakrzewska-Czerwińska, J. (2004). *Mycobacterium tuberculosis* DnaA initiator protein: purification and DNA-binding requirements. *Biochem. J.* 382, 247–252. doi: 10.1042/BJ20040338

Zawilak-Pawlik, A., Kois, A., Majka, J., Jakimowicz, D., Smulczyk-Krawczyszyn, A., Messer, W., et al. (2005). Architecture of bacterial replication initiation complexes: orisomes from four unrelated bacteria. *Biochem. J.* 389, 471–481. doi: 10.1042/BJ20050143

Zorman, S., Seitz, H., Sclavi, B., and Strick, T. R. (2012). Topological characterization of the DnaA-oriC complex using single-molecule nanomanipuation. *Nucleic Acids Res.* 40, 7375–7383. doi: 10.1093/nar/gks371

Zyskind, J. W., Cleary, J. M., Brusilow, W. S., Harding, N. E., and Smith, D. W. (1983). Chromosomal replication origin from the marine bacterium *Vibrio harveyi* functions in *Escherichia coli*: oriC consensus sequence. *Proc. Natl. Acad. Sci. U.S.A.* 80, 1164–1168. doi: 10.1073/pnas.80.5.1164

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 October 2014; paper pending published: 06 November 2014; accepted: 05 December 2014; published online: 06 January 2015.

Citation: Wolański M, Donczew R, Zawilak-Pawlik A and Zakrzewska-Czerwińska J (2015) oriC-encoded instructions for the initiation of bacterial chromosome replication. Front. Microbiol. ***5****:735. doi: 10.3389/fmicb.2014.00735*

This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology.

Copyright © 2015 Wolański, Donczew, Zawilak-Pawlik and Zakrzewska-Czerwińska. 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.

frontiers in Microbiology

REVIEW
published: 16 June 2015
doi: 10.3389/fmicb.2015.00610

Redefining bacterial origins of replication as centralized information processors

Gregory T. Marczynski, Thomas Rolain† and James A. Taylor†*

Department of Microbiology and Immunology, McGill University, Montreal, QC, Canada

In this review we stress the differences between eukaryotes and bacteria with respect to their different cell cycles, replication mechanisms and genome organizations. One of the most basic and underappreciated differences is that a bacterial chromosome uses only one *ori* while eukaryotic chromosome uses multiple *oris*. Consequently, eukaryotic *ori*s work redundantly in a cell cycle divided into separate phases: First inactive replication proteins assemble on eukaryotic *ori*s, and then they await conditions (in the separate "S-phase") that activate only the *ori*-bound and pre-assembled replication proteins. S-phase activation (without re-assembly) ensures that a eukaryotic *ori* "fires" (starts replication) only once and that each chromosome consistently duplicates only once per cell cycle. This precise chromosome duplication does not require precise multiple *ori* firing in S-phase. A eukaryotic *ori* can fire early, late or not at all. The single bacterial *ori* has no such margin for error and a comparable imprecision is lethal. Single *ori* usage is not more primitive; it is a totally different strategy that distinguishes bacteria. We further argue that strong evolutionary pressures created more sophisticated single *ori* systems because bacteria experience extreme and rapidly changing conditions. A bacterial *ori* must rapidly receive and process much information in "real-time" and not just in "cell cycle time." This redefinition of bacterial *oris* as centralized information processors makes at least two important predictions: First that bacterial *oris* use many and yet to be discovered control mechanisms and second that evolutionarily distinct bacteria will use many very distinct control mechanisms. We review recent literature that supports both predictions. We will highlight three key examples and describe how negative-feedback, phospho-relay, and chromosome-partitioning systems act to regulate chromosome replication. We also suggest future studies and discuss using replication proteins as novel antibiotic targets.

Keywords: *oriC*, DnaA, chromosome replication, partitioning, cell-cycle, regulators

OPEN ACCESS

Edited by:
Feng Gao,
Tianjin University, China

Reviewed by:
Murty V. Madiraju,
University of Texas Health Center at Tyler, USA
Justine Collier,
University of Lausanne, Switzerland

***Correspondence:**
Gregory T. Marczynski,
Department of Microbiology and Immunology, McGill University,
3775 University Street,
Montreal, QC H3A 2B4, Canada
gregory.marczynski@mcgill.ca

†*These authors have contributed equally to this work.*

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *15 April 2015*
Accepted: *02 June 2015*
Published: *16 June 2015*

Citation:
Marczynski GT, Rolain T and Taylor JA (2015) Redefining bacterial origins of replication as centralized information processors. Front. Microbiol. 6:610. doi: 10.3389/fmicb.2015.00610

Introduction

This short review emphasizes the bacterial point of view for replication control and argues that bacterial chromosome origins (*oris*) of replication have an underappreciated importance for cell cycle control not shared by eukaryotic *oris*. If this view seems controversial, it is not because the data and literature are contradictory. Instead, our view only seems controversial because reviews typically over-emphasize the similarities among organisms. Our presentation aims to restore a balance that respects the complexities of bacteria and eukaryotes. We develop our argument from a historical

perspective and then, because space is limited, we give a few specific examples of uniquely bacterial control. Our literature review is therefore incomplete. However, the bacterial cell cycle field is growing and excellent reviews are available to fill the gaps. For example, a very recent review has covered *oris* in diverse model bacteria and it systematically surveyed the many different regulators of replication (Wolanski et al., 2015). The *Escherichia coli oriC* model and the DnaA mechanism for initiating chromosome replication have provided the most detailed molecular mechanisms that operate inside *oris* and recent reviews also provide new insights (Kaguni, 2011; Leonard and Grimwade, 2011; Skarstad and Katayama, 2013; Kaur et al., 2014). An especially lucid review with fine graphic summaries of bacterial cell cycle mechanisms was provided by Katayama and coworkers (Katayama et al., 2010). Our review aims to complement such reviews with a fresh perspective.

Historical and Theoretical Background

Bacteria were first studied as medical problems and later as simple models or substitutes for complex organisms. Today, bacteria are also studied as interesting organisms in their own right. The three kingdoms view of biology gives bacteria a separate and potentially unique place. Regarding replication genes, we now know that the other two kingdoms, the archaea and eukarya share homologous replication components and it is the bacteria that stand out (Makarova and Koonin, 2013). However, when the replicon hypothesis was first formulated to explain chromosome replication, *E. coli* replication was viewed as a valid and accurate representation for all organisms. This bold assertion reflected the basically valid conviction that all life is united by evolution. However, a unity at the biochemical level does not necessarily imply a unity at higher organizational levels. So while biosynthetic and polymerization reactions may all have common mechanisms, it does not follow that assembly and regulatory reactions should be similarly conserved. How proteins and other cell components bind and sequentially assemble, how these form dynamic cellular structures and how these communicate to regulate cellular functions, are all major themes of contemporary cell biology. We now know that regulatory systems are evolutionarily very flexible and this insight is also expressed in recent bacterial cell cycle reviews (Katayama et al., 2010; Collier, 2012; Jonas, 2014, Wolanski et al., 2015).

Chromosome replication is an especially sophisticated assembly reaction that communicates with many cellular processes. We will argue that bacteria present special challenges and that our studies are far from complete. However, before presenting some contemporary studies, we need to quickly review the original replicon hypothesis, because it has guided and unfortunately also misguided so much of what we know or think that we know.

The replicon hypothesis is now 50 years old (Wolanski et al., 2014). When this hypothesis was first proposed to explain chromosome replication, the operon hypothesis was simultaneously proposed to explain genetic transcription. Both hypotheses were viewed as parallel and complementing explanations for these fundamental processes. For example, while both hypotheses proposed specific DNA targets for proteins, the replicon hypothesis proposed proteins that only acted positively to stimulate DNA synthesis, while the operon hypothesis proposed exclusive negative regulation using the *lac* repressor as the model. In retrospect, it is hard to see why both positive and negative regulators should not have been considered, but this realization would require further studies of the *lac* and other operons as well as studies of RNA polymerase interacting with its promoter DNA sequences. By analogy to transcription promoters, bacterial origins of replication (*oris*) became viewed as places for assembling replication proteins (Kornberg and Baker, 1992). In rough outline, a bacterial *ori* is now viewed as a specific place where the DnaA protein binds multiple DnaA boxes to self-assemble and then to promote the assembly of the downstream replication proteins (Kaguni, 2011; Leonard and Grimwade, 2011; Bell and Kaguni, 2013; Kaur et al., 2014).

What is the Correct Definition of an Origin of Replication?

Most importantly for this review, the replicon hypothesis gave us the basic concept of "origins (*oris*) of replication." In other words, an *ori* is a fixed and dedicated place on the chromosome where replication always starts and by analogy to promoters, where most regulators act. While we all take this basic concept for granted, there is in fact no theoretical need for origins of replications as there is for transcriptional promoters. Genetic transcription requires fixed and dedicated promoters to selectively transcribe specific genes so that some genes are "on" while others are "off." However, if all genes required uniform transcription then specific start and stop sites would be optional and even wasteful. Therefore, to duplicate a whole chromosome the cell does not require that replication always initiates from one fixed place. Instead, what is required is that the chromosome is picked only once for each replication cycle. In fact, this is exactly what eukaryotic cells do in S-phase (Prasanth et al., 2004; Masai et al., 2010). So why do we conventionally say that eukaryotic chromosomes use specific *oris* if they are apparently not needed? This view is primarily a presumption from the earlier bacterial literature. Today, it is more accurate to say that eukaryotic chromosomes use preferential *oris*, including optional and conditional *oris* (Chang et al., 2011) but that they lack the fixed and dedicated *oris* of bacterial chromosomes (Gao et al., 2013). As we will explain further below, eukaryotic chromosomes have preferential *oris* only because the proteins that recognize them (the ORCs, origin recognition complex proteins) have preferential binding sites (Chang et al., 2011). However, the main role of eukaryotic ORC proteins is not to pick the place but the time (S-phase) for replication (Prasanth et al., 2004). ORCs mark the chromosome for replication and ORC placement is much less important. In contrast, the bacterial DnaA protein picks both the time and place to start chromosome replication. This distinction and the special regulatory functions of bacterial *oris* will be more apparent when we next consider the eukaryotic and the bacterial cell cycles.

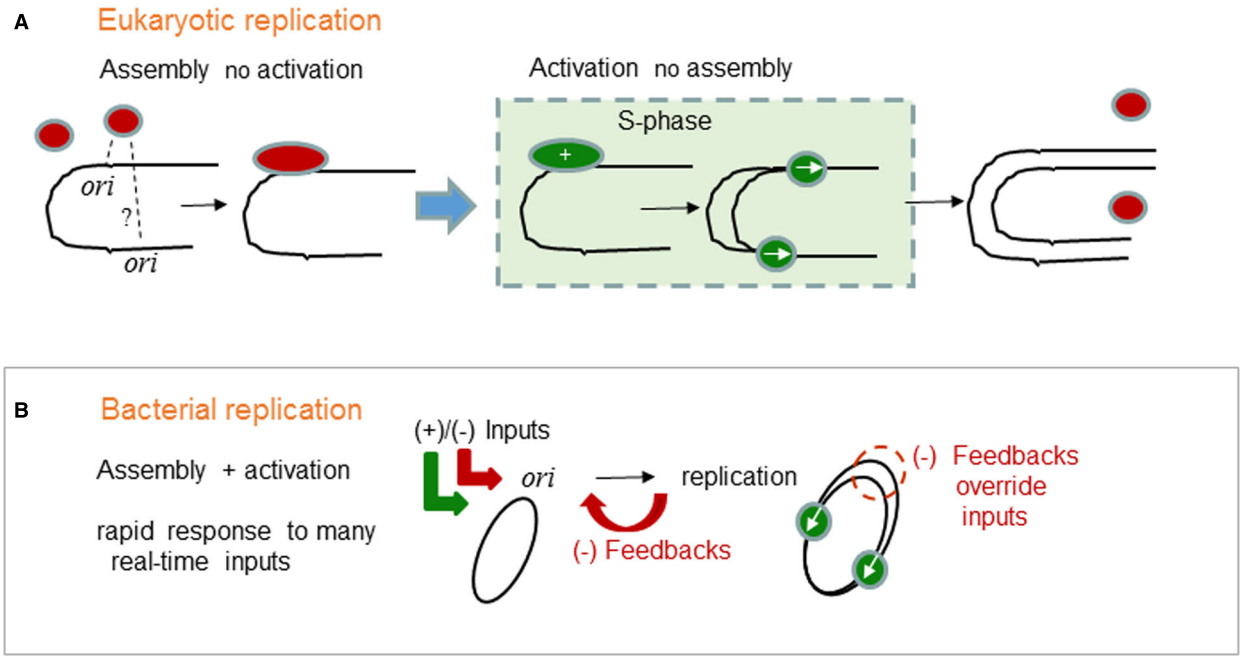


FIGURE 1 | Generalized logic of (A) eukaryotic and (B) bacterial chromosome replication control. (A) In eukaryotes, the commitment to chromosome replication occurs at the cellular-level. The whole cell moves into S-phase. Individual eukaryotic *ori*s do not participate in the commitment to S-phase. Instead, eukaryotic *ori*s passively respond to S-phase. Assembly of replication proteins on *ori*s is temporally separated from the activation of replication which can only occur once in S-phase. Red ovals are licensed ORC complexes, green ovals are initiation complexes and replication forks started only from those same pre-bound complexes. **(B)** In bacteria, the commitment to chromosome replication occurs at the single *ori*- level. Replication protein assembly and activation are integrated and subjected to many positive and negative (+)/(−) inputs. Precise chromosome duplication, without over-replication, also needs negative (−) feedback mechanisms that transiently override the (+) inputs and block assembly. The green ovals represent active replisomes. Integrated assembly and activation permit rapid and real-time responses that characterize bacterial physiology and permit survival in extreme and in rapidly changing environments.

Contrasts between Eukaryotic and Bacterial Replication Controls

Eukaryotes and bacteria have very different replication control strategies. In many respects, eukaryotic cell cycle controls are very sophisticated but at the DNA-binding level it is the bacteria that show the sophistication. In eukaryotes, the commitment to chromosome replication occurs at the cellular-level. The whole cell moves into S-phase (**Figure 1A**). Individual eukaryotic *ori*s do not participate in this commitment, instead they wait and passively respond to global changes such as threshold levels of cyclin-dependent kinases. First, replication proteins assemble on *ori*s and become primed for replication. Another important distinction is the "licensing" concept (Lygerou and Nurse, 2000; Nishitani et al., 2000), because it applies to eukaryotic and not to bacterial chromosomes. Licensing is a protein assembly reaction that occurs in G1 phase. In the separate S-phase only the *ori*-bound "licensed" assemblies can start replication. Assembly of replication proteins on *ori*s and their activation occur in separate phases of the cell cycle. It is this temporal separation that ensures that a chromosome will replicate only once per cell cycle. Precise duplication does not require a precise *ori* response. ORC and licensing proteins need not assemble at every *ori* and every *ori* need not fire (Woodward et al., 2006).

In contrast, bacteria absolutely need a precise *ori* response, because the chromosome has just one *ori*. This fact is unusually misinterpreted as a primitive state compared to eukaryotes. However, bacterial chromosomes are in fact well organized, e.g., the functional unity of operons, and highly evolved compared to those of eukaryotes. The single *ori* is not an accident but an evolved advantage. What advantages does a single *ori* provide? We argue that a single *ori* centralizes information processing. As we summarize for the *ori* in (**Figure 1B**), bacterial cell cycles do not have well defined phases. Instead, replication protein assembly and activation are integrated and subjected to many positive and negative (++)/(−) inputs (Wolanski et al., 2015). Precise chromosome duplication, without over-replication, also needs negative (−) feedback mechanisms that transiently override the (+) inputs and block assembly (Katayama et al., 2010). Integrated assembly and activation also permits rapid real-time responses that characterize bacterial physiology and permit survival in extreme and in rapidly changing environments.

Bacterial DnaA Replication Control

The DnaA protein is used by most and possibly all bacteria to initiate chromosome replication (Wolanski et al., 2014) and therefore DnaA is a major target for the positive and negative

(+)/(−) *ori* inputs implied schematically in **Figure 1B** (Wolanski et al., 2015). In *E. coli*, replication begins from a single *oriC* when a critical level of activated DnaA (ATP bound ATP-DnaA) is reached (Katayama et al., 2010; Kaguni, 2011). Both the activated ATP-DnaA and the inactive ADP-DnaA proteins bind to the main DnaA boxes in *oriC*, but only the activated ATP-DnaA proteins will bind and oligomerize at *oriC* using interactions between neighboring AAA^+ domains (Erzberger and Berger, 2006). Such DnaA assembly causes DNA unwinding and the recruitment of downstream replicative proteins. Specifically, *oriC* DNA unwinding allows DnaA to recruit DnaB, the replicative DNA helicase and DnaC, the helicase loader, onto the single-stranded DNA (Mott and Berger, 2007). Movement of two DnaB hexamers away from *oriC* results in the further recruitment of primase DnaG and the dissociation of the helicase loader DnaC. Next, the DNA polymerase III holoenzyme composed of the Pol III and the β-clamp (DnaN) are recruited to form the "replisome" that synthesizes the complementary DNA strands (Kaguni, 2011; Leonard and Grimwade, 2011; Skarstad and Katayama, 2013; Kaur et al., 2014).

This bacterial initiation process is often compared to eukaryotic entry into S-phase, especially since both DnaA and the ORC proteins use AAA^+ domains and ATP to facilitate assembly reactions (Erzberger and Berger, 2006). However, there are significant differences with major consequences for replication control. *First*, *E. coli* DnaA assembly at *oriC* is dynamic and *in vivo* there is probably both back and forth assembly and dis-assembly of DnaA until the critical amount of DnaA oligomerization is reached (Leonard and Grimwade, 2011). This is very different than the static licensing factor assemblies that attach to ORC-bound DNA (the eukaryotic *oris*) during G1 and await activation in S-phase. *Second*, the *E. coli* DnaB replicative helicase is loaded during the initiation process that is driven forward by DnaA oligomerization (Bell and Kaguni, 2013). This dynamic loading is also very different than the static replicative helicases (MCM proteins) that pre-loaded on ORC-bound DNA (eukaryotic *oris*) during G1 and await activation in S-phase.

Both dynamic features of *E. coli* replication initiation imply that there are many ways to shift the dynamics of DnaA and DnaB assembly and therefore bacterial initiation has the potential for a rapid response to many regulatory inputs (**Figure 1B**). In other words, unlike eukaryotic *oris*, the bacterial *oris* have the potential to process many regulatory signals before firing and committing to replication. Also, this processing can happen in real-time, because cell growth is not divided into cell cycle phases. Such regulation is very advantageous, because the conditions for growth and replication can change very rapidly for bacteria. In support of this dynamic view of *ori* signal processing, many regulators have been found and this is a rapidly expanding field of research. However, since recent reviews have covered the many proposed and established regulators of replication (Katayama et al., 2010, Wolanski et al., 2015), we will only present below the control mechanisms that have interested our lab the most. These include the next three topics on negative-feedback control, inputs from two-component systems and the co-regulation of replication with chromosome partitioning.

Bacterial Negative-feedback Control

The more dynamic bacterial initiation process also creates a greater reliance on negative-feedback controls. In eukaryotes, the licensing mechanisms automatically quench extra replication from the same *ori* in S-phase. In bacteria, as implied schematically in **Figure 1B**, to avoid potentially lethal over-replication, negative feedbacks must quench the forward replication potential created by high levels of active ATP-DnaA. *E. coli* has several negative-feedback mechanisms but the dominant one uses DnaN as a key regulatory component (Camara et al., 2003). DnaN forms a ring around the DNA to hold Pol III at the replication forks and a new DnaN ring is formed at each Okazaki fragment. Once replication starts, surplus DnaN rings accumulate and provide a platform for negative feedback regulators that limit replication. In *E. coli* this major regulatory mechanism of inhibiting replication is called RIDA for regulatory inactivation of DnaA. RIDA promotes ATP hydrolysis of ATP-DnaA and thus increases the ratio between inactive ADP-DnaA and active ATP-DnaA in the cell. Hda binds the DnaN ring which slides on the DNA to bring Hda into contact with DNA-bound DnaA protein. Hda has an AAA^+ domain that contacts the homologous AAA^+ oligomerization domain on DnaA and this is the specific interaction that stimulates the hydrolysis of DnaA-bound ATP (Kato and Katayama, 2001; Katayama et al., 2010; Nakamura and Katayama, 2010). Since ADP-DnaA cannot oligomerize, Hda can be regarded as an anti-oligomerization or as an anti-DnaA assembly factor.

If the *E. coli oriC* model applies to most bacteria and if surplus DnaN rings are deposited when replication starts, then do other bacteria also use RIDA? Yes, there is good evidence that the distantly related Gram-negative *Caulobacter crescentus* also uses a RIDA-like system. The *C. crescentus* homolog HdaA is very similar to *E. coli* Hda, and as expected down-regulation of HdaA causes chromosome over-replication (Collier and Shapiro, 2009). Also, fluorescence resonance energy transfer experiments demonstrate that *C. crescentus* HdaA interacts with DnaN in live cells (Fernandez-Fernandez et al., 2013). However, unlike *E. coli* DnaA protein, the *C. crescentus* DnaA protein is also regulated by cell cycle proteolysis (Gorbatyuk and Marczynski, 2005; Jonas et al., 2013). Therefore, it is important to consider that HdaA may regulate DnaA through both of these mechanisms and thereby fine-tuning DnaA activity more precisely for a cell cycle program which under natural conditions will experience sudden changes of nutrients, antibiotics and other growth challenges.

In distantly related Gram-positive *Bacillus subtilis*, a negative feedback system similar to RIDA is also present but it certainly evolved independently (Noirot-Gros et al., 2002, 2006). In this system, Hda is replaced by YabA. Interestingly, despite the lack of homology, YabA still forms a stable complex with DnaA as well as with DnaN. Deletion or mutations in *yabA* cause severe over-initiation of chromosome replication and *yabA* over-expression inhibits replication (Noirot-Gros et al., 2002; Goranov et al., 2009). Localization experiments also shown that YabA is associated with the replisome during chromosome replication through its interactions with DnaN (Goranov et al., 2009). Both YabA and Hda have been interpreted as anti-cooperativity or anti-assembly

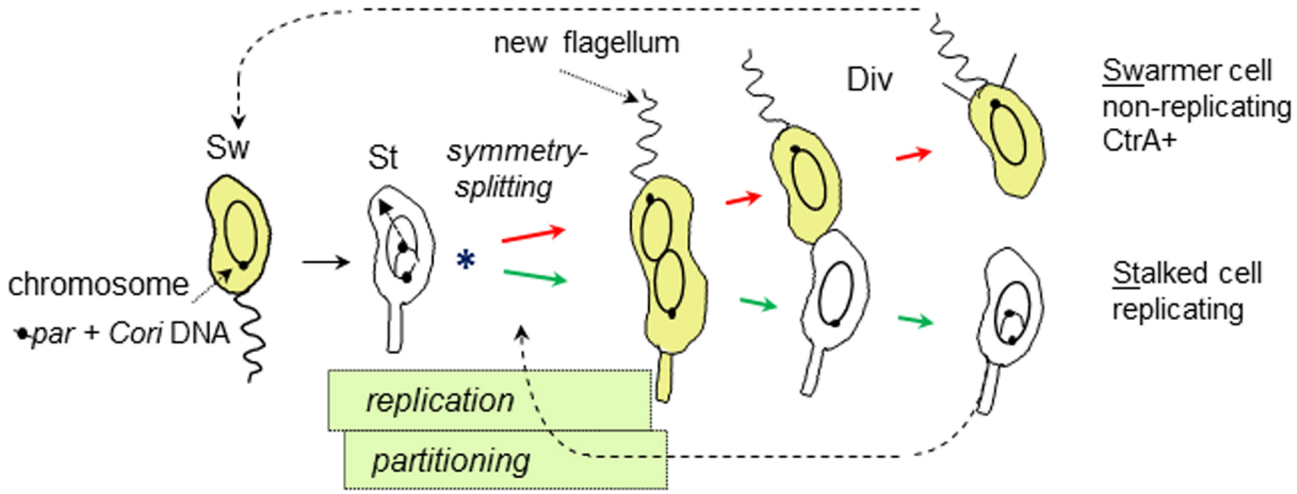


FIGURE 2 | Asymmetric cell division of *C. crescentus*, emphasizing key events and overlapping chromosome replication and partitioning periods. Swarmer cells (Sw) differentiate into stalked cells (St) and start chromosome replication with asymmetric (Sw and St-polar) division (Div). The chromosome origin of replication (*Cori*) initiates replication only once in the St cells. Linkage to the partition operon (*par,* containing *parABS*) ensures *Cori* placement at opposite cell poles. The asterisk (*) marks the chromosome symmetry-splitting stage of chromosome partitioning that is described in the text. CtrA protein (yellow) tracks the Sw cell-type due to its cell cycle synthesis and proteolysis.

factors that block the critical DnaA oligomerization step on *oriC* (Merrikh and Grossman, 2011).

Bacterial *ori* Regulation by Two-component Systems

The two-component systems proteins are an especially important class of regulators. These proteins dominate bacteria adaptive responses probably because they have a modular organization that aids the rapid evolution of paralogs that are easily altered to transduce many different signals (Garcia Vescovi et al., 2010; Capra and Laub, 2012). A conserved histidine kinase (HK) module and a conserved a response regulator (RR) module form the basis of a two-component signaling system. Although there is much variety, in many systems the HK is linked to a receptor while the RR is linked to a DNA-binding domain and the HK phosphorylates its cognate RR thereby sending the signal for activating the RR protein.

The *C. crescentus* RR protein called CtrA was the first example of bacterial *ori* regulation by a two-component system (Quon et al., 1996, 1998). Given the ubiquity and adaptive value of two-component systems, their regulatory inputs should be both common and varied. Since the first reports on CtrA, other RR proteins have been reported to regulate or at least to bind inside bacterial *oris*. Such examples include ArcA in *E. coli* (Lee et al., 2001), MtrA in *Mycobacterium tuberculosis* (Rajagopalan et al., 2010), Spo0A in *B. subtilis* (Boonstra et al., 2013), and most recently HP1021 in *Helicobacter pylori* (Donczew et al., 2015). In each case, the RR probably co-regulates replication with global cell activities, because each regulates many genes and the targets inside the *ori* are few compared to the many targets in the whole genome. Regarding the global cell activities, these probably include co-regulation with anaerobic growth by ArcA, macrophage invasion by MtrA, starvation-induced sporulation by Spo0A and stomach colonization by HP1021. Therefore, in each of these cases, environmental signals that drastically affect cell physiology are shunted into the *ori* for information processing, i.e., interactions with other replication proteins. In most cases these inputs are negative. For example, *E. coli* ArcA binds and blocks *ori* unwinding while *H. pylori* HP1021 probably binds to exclude DnaA from *ori*. However, these mechanisms of action are inferred primarily from *in vitro* studies and the *in vivo* activities are probably more complex.

CtrA remains the best studied example of bacterial *ori* regulation by two-component systems. CtrA (cell cycle transcription regulator) as the name implies regulates many cell cycle processes including DNA methylation and cell division (Quon et al., 1996; Kelly et al., 1998). CtrA is an essential master regulator of the dimorphic cell cycle that characterizes *C. crescentus* and therefore CtrA links chromosome replication with a series of intrinsic cell cycle programs that direct cell development.

Understanding CtrA regulation requires the following outline of the *C. crescentus* cell cycle (**Figure 2**): The non-replicating swarmer cell-type swims until it differentiates into the replicating stalked cell-type. Chromosome replication initiates only once in the stalked cell-type (Marczynski, 1999) which proceeds to grow and divide asymmetrically such that a new swarmer cell-pole is built opposite to the stalked cell-pole. Once replication initiates, the newly replicated DNA is partitioned into these emerging cell compartments that upon cell division will become

distinct replicating (stalked) and non-replicating (swarmer) cell-types. CtrA activity is associated with the swarmer cell-type and although CtrA has multiple roles, a major role is to bind and repress the *C. crescentus* origin of replication (*Cori*) in the non-replicating swarmer cells (Quon et al., 1998; Siam et al., 2003; Bastedo and Marczynski, 2009).

How is CtrA activity regulated? This complex topic itself requires a separate review (Tsokos and Laub, 2012). For our purposes, we note that synthesis and proteolysis adjust CtrA protein concentrations so that they are high in swarmer but low in stalked cells. However, protein turn-over is a secondary layer of regulation and as expected, CtrA activity is primarily adjusted by phosphorylation of its cognate RR domain (Domian et al., 1997; Spencer et al., 2009). The dimorphic and asymmetric mode of cell division directs CtrA phosphorylation through kinases and phosphatases that are localized at the swarmer and stalked cell poles (Tsokos and Laub, 2012). It is misleading to call this a "two-component" system, because like Spo0A of *B. subtilis*, CtrA activity is the final readout of a phopho-relay system that integrates many signals with multiple HK and RR modules. Such phosphor-relays do not just pass the signal, they in effect "decide" whether or not to pass the signal by in effect "consulting" many lateral inputs. One interesting aspect of the *C. crescentus* phopho-relay is that it creates a spatial gradient of CtrA activity during asymmetric cell division from high CtrA activity at the emerging swarmer cell-pole to low CtrA activity at the stalked pole (Chen et al., 2011). Another, very interesting aspect of the CtrA phospho-relay is a novel compartment sensing mechanism, so that once the compartments seal, the communication between the opposite poles is cut and this in turn strongly increases CtrA activity in the swarmer compartment while CtrA activity is quenched in the stalked cell compartment (Childers et al., 2014).

How does the *C. crescentus* origin of replication (*Cori*) use CtrA? *Cori* has five high-affinity binding sites for CtrA (Siam and Marczynski, 2000) and four of these sites are evolutionarily conserved among freshwater *Caulobacter* species (Shaheen et al., 2009). Interestingly, the *oris* of some marine *Caulobacter* species also use CtrA but unexpectedly, this usage probably evolved independently. *Caulobacters* belong to the alpha-proteobacteria and while CtrA seems to be a master regulator in this whole group of bacteria (Brilli et al., 2010), except possibly for *Rickettsia prowazekii* (Brassinga et al., 2002), CtrA binding sites are not seen in other *oris*. Therefore, CtrA also illustrates the principle that regulatory systems are evolutionarily very flexible.

What mechanisms does CtrA use to regulate *Cori*? One mechanism may involve transcriptional promoter activation in the stalked cells (Siam and Marczynski, 2000), but how new RNA synthesis promotes replication is not yet clear. The simplest mechanism seems to be a steric exclusion of DnaA protein from *Cori* (Taylor et al., 2011). Therefore, when CtrA activity rises in swarmer cells it binds and blocks replication in the swarmer cells by excluding DnaA. Interestingly, *Cori* has two classes of DnaA binding sites: A moderate affinity class termed G-boxes and a very weak class termed W-boxes (Taylor et al., 2011). The G-boxes have a conserved T to G substitution that reduces the otherwise high affinity of typical DnaA boxes present in other bacterial *oris*. *Cori* has only two G-boxes and both are targeted by their proximity or overlap with CtrA binding sites. The W-boxes are very weak and require cooperative binding with G-boxes for occupancy. The relatively weak G-box and W-box binding sites seem to have a precisely tuned low affinity for DnaA, because mutations that increase their affinity for DnaA can unexpectedly decrease replication (Taylor et al., 2011).

Therefore, *Cori* presents what seems to be a contradiction. *Cori* has a high affinity for CtrA (a protein not typically associated with *oris*) and yet a relatively low affinity for DnaA (the protein that is always required for bacterial *ori* function). In fact *Cori* is the highest affinity target for CtrA in the whole genome (Laub et al., 2002; Taylor et al., 2011). In contrast, since DnaA is also a transcription regulator, many *C. crescentus* promoters have DnaA boxes and some have higher affinity DnaA boxes than those in *Cori* (Hottes et al., 2005; Taylor et al., 2011). To better understand how CtrA binding regulates *Cori*, we systematically removed the CtrA binding sites from *Cori* at its natural locus on the chromosome (Bastedo and Marczynski, 2009). By combining site-directed mutations with homologous recombination, we created strains with substantially lower CtrA affinity in all five binding sites. To our surprise, the normal cell cycle program of chromosome replication was only mildly perturbed. Our interpretation of this result is that under constant laboratory culture conditions, the cell cycle runs like a clock. Most likely DnaA regulators and particularly RIDA (as discussed above) drive the replication cycle with only small adjustments form CtrA (Jonas et al., 2011). Such results forced us to reconsider *Cori* regulation, because obviously *C. crescentus* did not evolve in laboratory cultures but faced many environmental stresses that required constant monitoring. Typical environmental stresses for *C. crescentus* might be starvation and antibiotics. To support this view, we noticed that *C. crescentus* strains lacking CtrA binding at *Cori* became very sensitive to otherwise sub-lethal pulses of antibiotics (Bastedo and Marczynski, 2009).

Most significantly, *Cori* CtrA binding sites become essential when cells encounter both nutrients and antibiotics, a situation that presumably simulates natural bacterial competition and evolutionary pressures (Bastedo and Marczynski, 2009). Therefore, CtrA has at least two major roles in *Cori*: *First*, to help maintain or reinforce the cell cycle pattern of replication, so that replication is "off" in swarmer cells and "on" in the stalked cells (**Figure 2**). *Second*, to coordinate replication with cell growth in stressful and rapidly changing environments (nutrient up-shifts and antibiotic pulses). We argue that it is this second role for CtrA that provided the main selective pressure for evolving control by CtrA. This second role also presumes rapid real-time inputs into *Cori* that target DnaA. We tentatively interpret the G-box and W-box distribution in *Cori* (Taylor et al., 2011) as a variation of the DnaA box distribution in *E. coli oriC* that permits dynamic back and forth assembly and dis-assembly of DnaA (Leonard and Grimwade, 2011) until regulatory inputs, from CtrA and probably other regulators, drive the DnaA oligomerization toward critical initiation levels. Our search for additional *Cori* regulators identified a novel protein termed OpaA that we describe below, because it participates in both chromosome replication and partitioning. In addition to real-time inputs, environmental signals, such as sudden starvation, are especially important to

arrest the normal clockwork cell cycle pattern. For example, such arrests happen when *C. crescentus* is starved and DnaA is removed by targeted proteolysis (Gorbatyuk and Marczynski, 2005; Lesley and Shapiro, 2008; Jonas et al., 2013). Limited space does not allow us to expand on this topic, but the importance of environmental signals for bacterial cell cycle regulation as well as some recent developments have also received a fine review (Jonas, 2014).

Co-regulation of Chromosome Replication and Chromosome Partitioning

The initiation of chromosome replication immediately precedes the initiation of chromosome partitioning into the daughter cell compartments that will eventually form the daughter cells at cell division (Toro and Shapiro, 2010; **Figure 2**). This close temporal link suggests that it would be advantageous to co-regulate replication and partitioning. In many bacteria, chromosome partitioning employs a tripartite Par system consisting of a chromosomal centromere site (*parS*), a DNA binding protein (ParB) that binds *parS* DNA and a Walker-type ATPase protein (ParA) that probably uses non-specific DNA sequence affinity and ATP hydrolysis to pull the ParB-*parS* complex into opposite daughter compartments (Vecchiarelli et al., 2010). Interestingly, the *parS* site is usually located close to the *ori*, presumably to minimize the delay between replication and the onset of chromosome partitioning. For example, in *C. crescentus* the *parS* site is located within 8 kb of *Cori*, and in *B. subtilis* the three primary *parS* sites are located within 10 kb of *oriC*. In a survey of over 1,000 genomes, 92% of the *parS* sites were found to be located in the 15% of the chromosome closest to the *ori* (Livny et al., 2007).

Given these close temporal and spatial links, what is the evidence for co-regulation and communication between the replication and partitioning systems? In *B. subtilis*, Soj (a ParA homologue) directly interacts with DnaA protein to regulate replication both positively and negatively at *oriC*, depending on the quaternary state of Soj protein (Murray and Errington, 2008). In turn, Spo0J (ParB homologue) regulates this quaternary state, thus controlling replication through Soj (Scholefield et al., 2011). An innovative study employing recombinant DnaA to allow specific crosslinking of DnaA molecules during their helical oligomerization showed that monomeric-Soj/DnaA interaction blocks the formation of helical DnaA oligomers both *in vivo* and *in vitro* (Scholefield et al., 2012). The mechanism by which dimeric Soj positively influences replication remains unclear but these studies clearly establish co-regulation.

Vibrio cholera provides more insights from a very different evolutionary perspective. Unlike most bacteria, *V. cholera* has two chromosomes that use different replication-initiation mechanisms. Chromosome I (chrI) encodes and employs the canonical DnaA mediated replication mechanism while chromosome II (chrII) encodes and employs a different protein, RctB, which performs the analogous initiation function (Egan and Waldor, 2003). Both chromosomes also encode their own Par systems, which act specifically on the chromosome that encodes them. Most interestingly, both Par systems also regulate the replication of their respective chromosomes. ChrI replication is stimulated by ParA1, apparently through direct interactions with DnaA, while ParB1 plays an inhibitory role (Kadoya et al., 2011). On chrII, where replication is initiated by the RctB protein, titration of RctB by the *rctA* site, adjacent to the *ori*, inhibits replication (Venkova-Canova et al., 2006). Yamaichi and colleagues showed that this inhibition is counteracted by ParB2 binding to a *parS2* site within the *rctA* site (Yamaichi et al., 2011). In addition, ParB can directly compete for a strong RctB binding site that inhibits replication within *oriCII* (Venkova-Canova et al., 2013). Thus two ParB2 activities promote replication by reducing RctB binding to inhibitory DNA sequences. These results suggest co-regulation whereby replication is promoted only when ParB2 levels become sufficient for chromosome partitioning.

The previous examples show how partitioning systems can signal replication initiation but logically the signals could flow both ways. Accordingly, a recent study by Mera and colleagues implicated DnaA in controlling ParA dependent chromosome partitioning in *C. crescentus* (Mera et al., 2014). A conditional DnaA expression strain, in which DnaA was shut off failed to initiate chromosome replication, as expected (Gorbatyuk and Marczynski, 2001), and kept the single ParB/*parS* centromere complex at the old cell pole. However, when DnaA was expressed at a low concentration that was insufficient to initiate replication, some cells "partitioned," i.e., moved the single un-replicated ParB/*parS* centromere complex to the new cell pole using the ParA mechanism. This faulty partitioning requires a DnaA binding site located within *parS*, suggesting that DnaA binding at *parS* directly controls partitioning.

Closer examination of *C. crescentus* chromosome partitioning suggests a need for novel components and perhaps novel mechanisms at the earliest stage of chromosome partitioning. This is a key chromosome symmetry-splitting stage (**Figure 2**), because immediately following the start of chromosome replication one *parS* locus will stay at the staked pole while the other *parS* locus will partition to the swarmer pole. Subsequent replication will eventually yield polarized chromosomes in their respective stalked cell (replicating) and swarmer cell (non-replicating) compartments (**Figure 2**). Time-lapse microscopy showed that this partitioning is a multi-step process involving *parS* separation, *parS* discrimination, *parS* slow-movement away from the stalked pole and finally *parS* fast-movement toward the swarmer pole (Shebelut et al., 2010). Further genetic analysis showed that only the final *parS* fast-movement step requires ParA (Shebelut et al., 2010). Therefore, neither the regulators nor the motors of the preceding early steps are known. However, we can speculate that as for DnaA (described above) novel partitioning components might be found among the proteins that first interact with the origins of chromosome replication. These considerations also provide a further motivation for seeking novel replication proteins.

Therefore, co-regulation of partitioning and replication control systems is both phylogenetically widespread and diverse in terms of the molecular interactions involved. Such co-regulation may be advantageous as it ensures that protein concentrations or activity levels required for each process are achieved simultaneously. To our knowledge, no studies have systematically addressed whether the proximity of *par* and *ori* loci is also important for their

co-regulation. However, the conservation of this proximity among so many bacterial chromosomes argues very strongly that *par* and *ori* communication is an important part of uniquely bacterial cell cycle strategies.

Implications for Novel Antibiotic Targets

We are running out of antibiotics and options for treating antibiotic-resistant infections. This fact is well known but if history is any guide, then new treatments will probably not come from established studies but from unexpected sources revealed by new basic research. Chromosome replication studies will contribute toward finding new antibiotics for at least two major reasons: *First*, because replication is essential and it predisposes cells to lethal damage; *Second*, as we argued in this review, because replication must communicate with essential cell cycle processes including for example chromosome partitioning. The first reason suggests finding new direct targets for antibiotics that might disrupt replication regulators. While the second reason suggests that indirect targets may be equally valuable. Such targets may not be directly lethal but they could nonetheless be very effective as *in vivo* antimicrobials.

This short review cannot begin to address this question but it again raises our main issue of bacterial molecular communication and our reinterpretation of *oris* as centralized information processors. From the microbe's point of view, an infection requires complex navigation and communication in an ever-changing, alternatively hostile and benign tissue environment. As we argue, such communication must ultimately connect with *ori* which must process much information in real-time to determine the life or death of the cell. Therefore, an effective *in vivo* antimicrobial may be one that confuses bacteria so that they make mistakes and fall prey to the natural and overwhelming antimicrobial activities of the immune system. Finding such targeted antimicrobials requires much better knowledge of bacterial communication. Given the varieties of bacterial communication, it is also likely that future antibiotics may be customized for the specific regulators of specific species. We normally think of personalized medicine as a match between a specific human genotype and a specific medication. In the future, considering the ease of identifying bacteria by deep-sequencing techniques, another form of personalized medicine may be a matching between a microbial genotype and specific replication-disrupting antibiotics.

Acknowledgments

This work was funded by the Canadian Institutes for Health Research (CIHR) operating grant MOP-12599 and by the Natural Sciences and Engineering Research Council of Canada (NSERC, Rgpin 184894-09).

References

Bastedo, D. P., and Marczynski, G. T. (2009). CtrA response regulator binding to the *Caulobacter chromosome* replication origin is required during nutrient and antibiotic stress as well as during cell cycle progression. *Mol. Microbiol.* 72, 139–154. doi: 10.1111/j.1365-2958.2009.06630.x

Bell, S. P., and Kaguni, J. M. (2013). Helicase loading at chromosomal origins of replication. *Cold Spring Harb. Perspect. Biol.* 5, a010124. doi: 10.1101/cshperspect.a010124

Boonstra, M., de Jong, I. G., Scholefield, G., Murray, H., Kuipers, O. P., and Veening, J. W. (2013). Spo0A regulates chromosome copy number during sporulation by directly binding to the origin of replication in *Bacillus subtilis*. *Mol. Microbiol.* 87, 925–938. doi: 10.1111/mmi.12141

Brassinga, A. K., Siam, R., McSween, W., Winkler, H., Wood, D., and Marczynski, G. T. (2002). Conserved response regulator CtrA and IHF binding sites in the alpha-proteobacteria *Caulobacter crescentus* and *Rickettsia prowazekii* chromosomal replication origins. *J. Bacteriol.* 184, 5789–5799. doi: 10.1128/JB.184.20.5789-5799.2002

Brilli, M., Fondi, M., Fani, R., Mengoni, A., Ferri, L., Bazzicalupo, M., et al. (2010). The diversity and evolution of cell cycle regulation in alpha-proteobacteria: a comparative genomic analysis. *BMC Syst. Biol.* 4:52. doi: 10.1186/1752-0509-4-52

Camara, J. E., Skarstad, K., and Crooke, E. (2003). Controlled initiation of chromosomal replication in *Escherichia coli* requires functional Hda protein. *J. Bacteriol.* 185, 3244–3248. doi: 10.1128/JB.185.10.3244-3248.2003

Capra, E. J., and Laub, M. T. (2012). Evolution of two-component signal transduction systems. *Annu. Rev. Microbiol.* 66, 325–347. doi: 10.1146/annurev-micro-092611-150039

Chang, F., May, C. D., Hoggard, T., Miller, J., Fox, C. A., and Weinreich, M. (2011). High-resolution analysis of four efficient yeast replication origins reveals new insights into the ORC and putative MCM binding elements. *Nucleic Acids Res.* 39, 6523–6535. doi: 10.1093/nar/gkr301

Chen, Y. E., Tropini, C., Jonas, K., Tsokos, C. G., Huang, K. C., and Laub, M. T. (2011). Spatial gradient of protein phosphorylation underlies replicative asymmetry in a bacterium. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1052–1057. doi: 10.1073/pnas.1015397108

Childers, W. S., Xu, Q., Mann, T. H., Mathews, I. I., Blair, J. A., Deacon, A. M., et al. (2014). Cell fate regulation governed by a repurposed bacterial histidine kinase. *PLoS Biol.* 12:e1001979. doi: 10.1371/journal.pbio.1001979

Collier, J. (2012). Regulation of chromosomal replication in *Caulobacter crescentus*. *Plasmid* 67, 76–87. doi: 10.1016/j.plasmid.2011.12.007

Collier, J., and Shapiro, L. (2009). Feedback control of DnaA-mediated replication initiation by replisome-associated HdaA protein in *Caulobacter*. *J. Bacteriol.* 191, 5706–5716. doi: 10.1128/JB.00525-09

Domian, I. J., Quon, K. C., and Shapiro, L. (1997). Cell type-specific phosphorylation and proteolysis of a transcriptional regulator controls the G1-to-S transition in a bacterial cell cycle. *Cell* 90, 415–424. doi: 10.1016/S0092-8674(00)80502-4

Donczew, R., Makowski, Ł., Jaworski, P., Bezulska, M., Nowaczyk, M., Zakrzewska-Czerwińska, J., et al. (2015). The atypical response regulator HP1021 controls formation of the *Helicobacter pylori* replication initiation complex. *Mol. Microbiol.* 95, 297–312. doi: 10.1111/mmi.12866

Egan, E. S., and Waldor, M. K. (2003). Distinct replication requirements for the two *Vibrio cholerae* chromosomes. *Cell* 114, 521–530. doi: 10.1016/S0092-8674(03)00611-1

Erzberger, J. P., and Berger, J. M. (2006). Evolutionary relationships and structural mechanisms of AAA^+ proteins. *Annu. Rev. Biophys. Biomol. Struct.* 35, 93–114. doi: 10.1146/annurev.biophys.35.040405.101933

Fernandez-Fernandez, C., Grosse, K., Sourjik, V., and Collier, J. (2013). The beta-sliding clamp directs the localization of HdaA to the replisome in *Caulobacter crescentus*. *Microbiology* 159, 2237–2248. doi: 10.1099/mic.0.068577-0

Gao, F., Luo, H., and Zhang, C. T. (2013). DoriC 5.0: an updated database of oriC regions in both bacterial and archaeal genomes. *Nucleic Acids Res.* 41, D90–D93. doi: 10.1093/nar/gks990

Garcia Vescovi, E., Sciara, M. I., and Castelli, M. E. (2010). Two component systems in the spatial program of bacteria. *Curr. Opin. Microbiol.* 13, 210–218. doi: 10.1016/j.mib.2009.12.012

Goranov, A. I., Breier, A. M., Merrikh, H., and Grossman, A. D. (2009). YabA of *Bacillus subtilis* controls DnaA-mediated replication initiation but not the transcriptional response to replication stress. *Mol. Microbiol.* 74, 454–466. doi: 10.1111/j.1365-2958.2009.06876.x

Gorbatyuk, B., and Marczynski, G. T. (2001). Physiological consequences of blocked *Caulobacter crescentus* dnaA expression, an essential DNA

replication gene.*Mol. Microbiol.* 40, 485–497. doi: 10.1046/j.1365-2958.2001.02404.x

Gorbatyuk, B., and Marczynski, G. T. (2005). Regulated degradation of chromosome replication proteins DnaA and CtrA in *Caulobacter crescentus*. *Mol. Microbiol.* 55, 1233–1245. doi: 10.1111/j.1365-2958.2004.04459.x

Hottes, A. K., Shapiro, L., and McAdams, H. H. (2005). DnaA coordinates replication initiation and cell cycle transcription in *Caulobacter crescentus*. *Mol. Microbiol.* 58, 1340–1353. doi: 10.1111/j.1365-2958.2005.04912.x

Jonas, K. (2014). To divide or not to divide: control of the bacterial cell cycle by environmental cues. *Curr. Opin. Microbiol.* 18, 54–60. doi: 10.1016/j.mib.2014.02.006

Jonas, K., Chen, Y. E., and Laub, M. T. (2011). Modularity of the bacterial cell cycle enables independent spatial and temporal control of DNA replication. *Curr. Biol.* 21, 1092–1101. doi: 10.1016/j.cub.2011.05.040

Jonas, K., Liu, J., Chien, P., and Laub, M. T. (2013). Proteotoxic stress induces a cell-cycle arrest by stimulating Lon to degrade the replication initiator DnaA. *Cell* 154, 623–636. doi: 10.1016/j.cell.2013.06.034

Kadoya, R., Baek, J. H., Sarker, A., and Chattoraj, D. K. (2011). Participation of chromosome segregation protein ParAI of *Vibrio cholerae* in chromosome replication. *J. Bacteriol.* 193, 1504–1514. doi: 10.1128/JB.01067-10

Kaguni, J. M. (2011). Replication initiation at the *Escherichia coli* chromosomal origin. *Curr. Opin. Chem. Biol.* 15, 606–613. doi: 10.1016/j.cbpa.2011.07.016

Katayama, T., Ozaki, S., Keyamura, K., and Fujimitsu, K. (2010). Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC. *Nat. Rev. Microbiol.* 8, 163–170. doi: 10.1038/nrmicro2314

Kato, J., and Katayama, T. (2001). Hda, a novel DnaA-related protein, regulates the replication cycle in *Escherichia coli*. *EMBO J.* 20, 4253–4262. doi: 10.1093/emboj/20.15.4253

Kaur, G., Vora, M. P., Czerwonka, C. A., Rozgaja, T. A., Grimwade, J. E., and Leonard, A. C. (2014). Building the bacterial orisome: high-affinity DnaA recognition plays a role in setting the conformation of oriC DNA. *Mol. Microbiol.* 91, 1148–1163. doi: 10.1111/mmi.12525

Kelly, A. J., Sackett, M. J., Din, N., Quardokus, E., and Brun, Y. V. (1998). Cell cycle-dependent transcriptional and proteolytic regulation of FtsZ in *Caulobacter*. *Genes Dev.* 12, 880–893. doi: 10.1101/gad.12.6.880

Kornberg, A., and Baker, T. A. (1992). *DNA Replication*. New York: W.H. Freeman.

Laub, M. T., Chen, S. L., Shapiro, L., and McAdams, H. H. (2002). Genes directly controlled by CtrA, a master regulator of the *Caulobacter* cell cycle. *Proc. Natl. Acad. Sci. U.S.A.* 99, 4632–4637. doi: 10.1073/pnas.062065699

Lee, Y. S., Han, J. S., Jeon, Y., and Hwang, D. S. (2001). The arc two-component signal transduction system inhibits *in vitro Escherichia coli* chromosomal initiation. *J. Biol. Chem.* 276, 9917–9923. doi: 10.1074/jbc.M008629200

Leonard, A. C., and Grimwade, J. E. (2011). Regulation of DnaA assembly and activity: taking directions from the genome. *Annu. Rev. Microbiol.* 65, 19–35. doi: 10.1146/annurev-micro-090110-102934

Lesley, J. A., and Shapiro, L. (2008). SpoT regulates DnaA stability and initiation of DNA replication in carbon-starved *Caulobacter crescentus*. *J. Bacteriol.* 190, 6867–6880. doi: 10.1128/JB.00700-08

Livny, J., Yamaichi, Y., and Waldor, M. K. (2007). Distribution of centromere-like parS sites in bacteria: insights from comparative genomics. *J. Bacteriol.* 189, 8693–8703. doi: 10.1128/JB.01239-07

Lygerou, Z., and Nurse, P. (2000). Cell cycle. License withheld–geminin blocks DNA replication. *Science* 290, 2271–2273. doi: 10.1126/science.290.5500.2271

Makarova, K. S., and Koonin, E. V. (2013). Archaeology of eukaryotic DNA replication. *Cold Spring Harb. Perspect. Med.* 3, a012963. doi: 10.1101/cshperspect.a012963

Marczynski, G. T. (1999). Chromosome methylation and measurement of faithful, once and only once per cell cycle chromosome replication in *Caulobacter crescentus*. *J. Bacteriol.* 181, 1984–1993.

Masai, H., Matsumoto, S., You, Z., Yoshizawa-Sugata, N., and Oda, M. (2010). Eukaryotic chromosome DNA replication: where, when, and how? *Annu. Rev. Biochem.* 79, 89–130. doi: 10.1146/annurev.biochem.052308.103205

Mera, P. E., Kalogeraki, V. S., and Shapiro, L. (2014). Replication initiator DnaA binds at the *Caulobacter* centromere and enables chromosome segregation. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16100–16105. doi: 10.1073/pnas.1418989111

Merrikh, H., and Grossman, A. D. (2011). Control of the replication initiator DnaA by an anti-cooperativity factor. *Mol. Microbiol.* 82, 434–446. doi: 10.1111/j.1365-2958.2011.07821.x

Mott, M. L., and Berger, J. M. (2007). DNA replication initiation: mechanisms and regulation in bacteria. *Nat. Rev. Microbiol.* 5, 343–354. doi: 10.1038/nrmicro1640

Murray, H., and Errington, J. (2008). Dynamic control of the DNA replication initiation protein DnaA by Soj/ParA. *Cell* 135, 74–84. doi: 10.1016/j.cell.2008.07.044

Nakamura, K., and Katayama, T. (2010). Novel essential residues of Hda for interaction with DnaA in the regulatory inactivation of DnaA: unique roles for Hda AAA Box VI and VII motifs. *Mol. Microbiol.* 76, 302–317. doi: 10.1111/j.1365-2958.2010.07074.x

Nishitani, H., Lygerou, Z., Nishimoto, T., and Nurse, P. (2000). The Cdt1 protein is required to license DNA for replication in fission yeast. *Nature* 404, 625–628. doi: 10.1038/35007110

Noirot-Gros, M. F., Dervyn, E., Wu, L. J., Mervelet, P., Errington, J., Ehrlich, S. D., et al. (2002). An expanded view of bacterial DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* 99, 8342–8347. doi: 10.1073/pnas.122040799

Noirot-Gros, M. F., Velten, M., Yoshimura, M., McGovern, S., Morimoto, T., Ehrlich, S. D., et al. (2006). Functional dissection of YabA, a negative regulator of DNA replication initiation in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2368–2373. doi: 10.1073/pnas.0506914103

Prasanth, S. G., Méndez, J., Prasanth, K. V., and Stillman, B. (2004). Dynamics of pre-replication complex proteins during the cell division cycle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 7–16. doi: 10.1098/rstb.2003.1360

Quon, K. C., Marczynski, G. T., and Shapiro, L. (1996). Cell cycle control by an essential bacterial two-component signal transduction protein. *Cell* 84, 83–93. doi: 10.1016/S0092-8674(00)80995-2

Quon, K. C., Yang, B., Domian, I. J., Shapiro, L., and Marczynski, G. T. (1998). Negative control of bacterial DNA replication by a cell cycle regulatory protein that binds at the chromosome origin. *Proc. Natl. Acad. Sci. U.S.A.* 95, 120–125.

Rajagopalan, M., Dziedzic, R., Al Zayer, M., Stankowska, D., Ouimet, M. C., Bastedo, D. P., et al. (2010). Mycobacterium tuberculosis origin of replication and the promoter for immunodominant secreted antigen 85B are the targets of MtrA, the essential response regulator. *J. Biol. Chem.* 285, 15816–15827. doi: 10.1074/jbc.M109.040097

Scholefield, G., Errington, J., and Murray, H. (2012). Soj/ParA stalls DNA replication by inhibiting helix formation of the initiator protein DnaA. *EMBO J.* 31, 1542–1555. doi: 10.1038/emboj.2012.6

Scholefield, G., Whiting, R., Errington, J., and Murray, H. (2011). Spo0J regulates the oligomeric state of Soj to trigger its switch from an activator to an inhibitor of DNA replication initiation. *Mol. Microbiol.* 79, 1089–1100. doi: 10.1111/j.1365-2958.2010.07507.x

Shaheen, S. M., Ouimet, M. C., and Marczynski, G. T. (2009). Comparative analysis of *Caulobacter chromosom*e replication origins. *Microbiology* 155, 1215–1225. doi: 10.1099/mic.0.025528-0

Shebelut, C. W., Guberman, J. M., van Teeffelen, S., Yakhnina, A. A., and Gitai, Z. (2010). *Caulobacter chromosom*e segregation is an ordered multistep process. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14194–14198. doi: 10.1073/pnas.1005274107

Siam, R., Brassinga, A. K., and Marczynski, G. T. (2003). A dual binding site for integration host factor and the response regulator CtrA inside the *Caulobacter crescentus* replication origin. *J. Bacteriol.* 185, 5563–5572. doi: 10.1128/JB.185.18.5563-5572.2003

Siam, R., and Marczynski, G. T. (2000). Cell cycle regulator phosphorylation stimulates two distinct modes of binding at a chromosome replication origin. *EMBO J.* 19, 1138–1147. doi: 10.1093/emboj/19.5.1138

Skarstad, K., and Katayama, T. (2013). Regulating DNA replication in bacteria. *Cold Spring Harb. Perspect. Biol.* 5, a012922. doi: 10.1101/cshperspect.a012922

Spencer, W., Siam, R., Ouimet, M. C., Bastedo, D. P., and Marczynski, G. T. (2009). CtrA, a global response regulator, uses a distinct second category of weak DNA binding sites for cell cycle transcription control in *Caulobacter crescentus*. *J. Bacteriol.* 191, 5458–5470. doi: 10.1128/JB.00355-09

Taylor, J. A., Ouimet, M. C., Wargachuk, R., and Marczynski, G. T. (2011). The *Caulobacter crescentus* chromosome replication origin evolved two classes of weak DnaA binding sites. *Mol. Microbiol.* 82, 312–326. doi: 10.1111/j.1365-2958.2011.07785.x

Toro, E., and Shapiro, L. (2010). Bacterial chromosome organization and segregation. *Cold Spring Harb. Perspect. Biol.* 2, a000349. doi: 10.1101/cshperspect.a000349

Tsokos, C. G., and Laub, M. T. (2012). Polarity and cell fate asymmetry in *Caulobacter crescentus*. *Curr. Opin. Microbiol.* 15, 744–750. doi: 10.1016/j.mib.2012.10.011

Vecchiarelli, A. G., Han, Y. W., Tan, X., Mizuuchi, M., Ghirlando, R., Biertümpfel, C., et al. (2010). ATP control of dynamic P1 ParA-DNA interactions: a key role for the nucleoid in plasmid partition. *Mol. Microbiol.* 78, 78–91. doi: 10.1111/j.1365-2958.2010.07314.x

Venkova-Canova, T., Baek, J. H., Fitzgerald, P. C., Blokesch, M., and Chattoraj, D. K. (2013). Evidence for two different regulatory mechanisms linking replication and segregation of *Vibrio cholerae* chromosome II. *PLoS Genet.* 9:e1003579. doi: 10.1371/journal.pgen.1003579

Venkova-Canova, T., Srivastava, P., and Chattoraj, D. K. (2006). Transcriptional inactivation of a regulatory site for replication of *Vibrio cholerae* chromosome II. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12051–12056. doi: 10.1073/pnas.0605120103

Wolanski, M., Donczew, R., Zawilak-Pawlik, A., and Zakrzewska-Czerwińska, J. (2015). oriC-encoded instructions for the initiation of bacterial chromosome replication. *Front. Microbiol.* 5:735. doi: 10.3389/fmicb.2014.00735

Wolanski, M., Jakimowicz, D., and Zakrzewska-Czerwińska, J. (2014). Fifty years after the replicon hypothesis: cell-specific master regulators as new players in chromosome replication control. *J. Bacteriol.* 196, 2901–2911. doi: 10.1128/JB.01706-14

Woodward, A. M., Göhler, T., Luciani, M. G., Oehlmann, M., Ge, X., Gartner, A., et al. (2006). Excess Mcm2-7 license dormant origins of replication that can be used under conditions of replicative stress. *J. Cell Biol.* 173, 673–683. doi: 10.1083/jcb.200602108

Yamaichi, Y., Gerding, M. A., Davis, B. M., and Waldor, M. K. (2011). Regulatory cross-talk links *Vibrio cholerae* chromosome II replication and segregation. *PLoS Genet.* 7:e1002189. doi: 10.1371/journal.pgen.1002189

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.

Copyright © 2015 Marczynski, Rolain and Taylor. 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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 24 September 2015
doi: 10.3389/fmicb.2015.01011

Control regions for chromosome replication are conserved with respect to sequence and location among *Escherichia coli* strains

Jakob Frimodt-Møller[1,2], *Godefroid Charbon*[1], *Karen A. Krogfelt*[2] and *Anders Løbner-Olesen*[1]*

[1] Department of Biology, Section for Functional Genomics and Center for Bacterial Stress Response and Persistence, University of Copenhagen, Copenhagen, Denmark, [2] Department of Microbiology and Infection Control, Statens Serum Institut, Copenhagen, Denmark

OPEN ACCESS

Edited by:
Feng Gao,
Tianjin University, China

Reviewed by:
Dhruba Chattoraj,
National Institutes of Health, USA
Tsutomu Katayama,
Kyushu University, Japan

***Correspondence:**
Anders Løbner-Olesen,
Department of Biology, Section for Functional Genomics and Center for Bacterial Stress Response and Persistence, University of Copenhagen, Ole Maaløe's Vej 5, 2200 Copenhagen, Denmark
lobner@bio.ku.dk

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: 06 August 2015
Accepted: 07 September 2015
Published: 24 September 2015

Citation:
Frimodt-Møller J, Charbon G, Krogfelt KA and Løbner-Olesen A (2015) Control regions for chromosome replication are conserved with respect to sequence and location among Escherichia coli strains. Front. Microbiol. 6:1011. doi: 10.3389/fmicb.2015.01011

In *Escherichia coli*, chromosome replication is initiated from *oriC* by the DnaA initiator protein associated with ATP. Three non-coding regions contribute to the activity of DnaA. The *datA* locus is instrumental in conversion of $DnaA^{ATP}$ to $DnaA^{ADP}$ (*datA* dependent $DnaA^{ATP}$ hydrolysis) whereas DnaA rejuvenation sequences 1 and 2 (*DARS1* and *DARS2*) reactivate $DnaA^{ADP}$ to $DnaA^{ATP}$. The structural organization of *oriC*, *datA*, *DARS1,* and *DARS2* were found conserved among 59 fully sequenced *E. coli* genomes, with differences primarily in the non-functional spacer regions between key protein binding sites. The relative distances from *oriC* to *datA*, *DARS1,* and *DARS2*, respectively, was also conserved despite of large variations in genome size, suggesting that the gene dosage of either region is important for bacterial growth. Yet all three regions could be deleted alone or in combination without loss of viability. Competition experiments during balanced growth in rich medium and during mouse colonization indicated roles of *datA*, *DARS1,* and *DARS2* for bacterial fitness although the relative contribution of each region differed between growth conditions. We suggest that this fitness advantage has contributed to conservation of both sequence and chromosomal location for *datA*, *DARS1*, and *DARS2*.

Keywords: *E. coli*, *oriC*, non-coding replication control regions, *datA*, *DARS1* and *DARS2*, fitness *in vitro* and *in vivo*

Introduction

In *Escherichia coli* chromosome replication is initiated from a single origin, *oriC*, and proceeds bi-directionally until the two replication forks meet at terminus of replication (*terC*). The initiator protein DnaA belongs to the AAA^+ (ATPases Associated with diverse Activities) proteins. DnaA can bind ATP and ADP with similar high affinities (Skarstad and Katayama, 2013), but only DnaA bound to ATP is able to initiate DNA replication (Sekimizu et al., 1987). Different recognition sites for DnaA has been identified in *oriC*; three high to medium affinity sites (R1, R4, and R2) that binds both $DnaA^{ATP}$ and $DnaA^{ADP}$ (Fuller et al., 1984), and multiple lower affinity sites (R3, R5/M, I1, I2, I3, C1, C2, C3, τ1, and τ2) (McGarry et al., 2004; Kawakami et al., 2005; Rozgaja et al., 2011) (**Figure 1**). Only $DnaA^{ATP}$ is capable to bind to low-affinity sites (McGarry et al., 2004; Kawakami et al., 2005; Rozgaja et al., 2011), and single stranded DnaA boxes (Speck and Messer, 2001). Binding of the Fis protein to *oriC* is reported to both inhibit initiation of

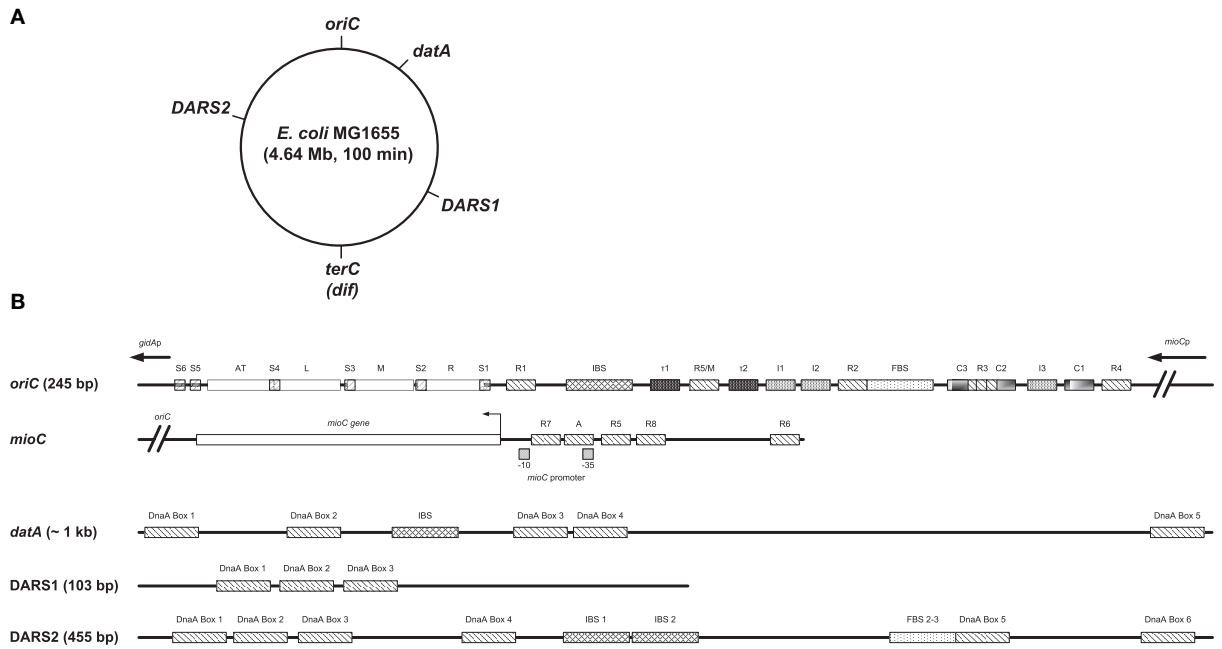


FIGURE 1 | Structures of *oriC*, *mioC* promoter, *datA*, *DARS1*, and *DARS2*. (A) The MG1655 chromosome is shown as a circle and the locations of *oriC* (at the genome map position of 84.6 min), *datA* (at 94.6 min), *DARS1* (at 17.5 min), *DARS2* (at 64.0 min), and *terC*/*dif* (around 36 min) are indicated. **(B)** Basic structures of *oriC*, the *mioC* promoter, *datA*, *DARS1*, and *DARS2* are schematically shown. Both high- to medium affinity sites (R1, R4, and R2) that binds both $DnaA^{ATP}$ and $DnaA^{ADP}$, lower affinity sites (R3, R5/M, I1, I2, I3, C1, C2, C3, τ1, and τ2), and Single-stranded DnaA-ATP box 1–6 (S1-6) that only binds $DnaA^{ATP}$ are indicated in *oriC* (see text for details). The minimal *oriC* sequence (245 bp) is defined to end just to the left of the 13-mer termed L and to the right of DnaA Box R4. Note that the R3 DnaA box overlaps with DnaA box C3 and C2 in *oriC*. DnaA boxes in the *mioC* promoter are indicated as described by Hansen et al. (2007), with DnaA Box R5 and R6 being high affinity sites, while DnaA Box R7, R8, and A are lower affinity sites. *DARS1* contains 3 DnaA boxes, *DARS2* contains 6 DnaA boxes an IBS and a FBS, and *datA* contains 5 DnaA boxes and an IBS. IBS, IHF-binding site; FBS, FIS-binding site. Figure is not to scale.

replication (Wold et al., 1996; Ryan et al., 2004; Riber et al., 2009), stimulate initiation (Flåtten and Skarstad, 2013), or have no effect on initiation (Margulies and Kaguni, 1998), while the binding of integration host factor (IHF) plays a central role in forming an optimal complex (Ryan et al., 2002; Keyamura et al., 2007; Ozaki and Katayama, 2012). Binding of $DnaA^{ATP}$ to both high- and low affinity DnaA boxes in *oriC* are proposed to result in a oligomeric DnaA structure, which assisted by IHF leads to duplex opening in the AT-rich region, i.e., open complex formation (Skarstad and Katayama, 2013). Following duplex opening the helicase DnaB is loaded onto the now single-stranded DNA by the help of DnaA, which leads to further duplex opening and assembly of the replisome (Skarstad and Katayama, 2013).

Initiation of replication is a highly regulated process in *E. coli*. Replication begins essentially simultaneously at all cellular origins (Skarstad et al., 1986), i.e., in synchrony and only once per cell cycle. The tight control is primarily ensured by the oscillation of $DnaA^{ATP}$ that has a temporal increase around the time of initiation, and decreases rapidly thereafter (Kurokawa et al., 1999). Following initiation, *oriC* is temporarily inactivated by the binding of SeqA to hemi-methylated GATC-sites (Campbell and Kleckner, 1990; Lu et al., 1994). This sequestration lasts for about 1/3 of the doubling time and provides a time period for RIDA (Regulatory Inactivation of DnaA) and DDAH (*datA*-dependent $DnaA^{ATP}$ hydrolysis) to hydrolyse $DnaA^{ATP}$ to $DnaA^{ADP}$. In RIDA, the Hda protein, in association with the DNA-loaded β-clamp (DnaN), activates the intrinsic ATPase activity of DnaA, which converts $DnaA^{ATP}$ into $DnaA^{ADP}$ (Kurokawa et al., 1999; Kato and Katayama, 2001). DDAH is an IHF dependent hydrolysis of $DnaA^{ATP}$ to $DnaA^{ADP}$, which takes place at the *datA* locus (Kasho and Katayama, 2013). *datA* contain five DnaA boxes as well as an IHF-binding site (Nozaki et al., 2009; Kasho and Katayama, 2013) (**Figure 1**). Common for both RIDA and DDAH is that both processes lower the $DnaA^{ATP}/DnaA^{ADP}$ ratio to counter unwanted re-initiation of replication. At later stages in the cell cycle the $DnaA^{ATP}$ level must increase past a critical level for a new round of initiation of replication. This is done by rejuvenation of $DnaA^{ADP}$ to $DnaA^{ATP}$ at the *DARS1* and *DARS2* loci, where rejuvenation at the *DARS2* locus is dependent on IHF and Fis (Kasho et al., 2014). In addition *de novo* synthesis of DnaA, which by and large will be ATP bound because ATP is more abundant than ADP within the cell, will also contribute to the increase in $DnaA^{ATP}$ (Kurokawa et al., 1999). *DARS1* and *DARS2* contain a core of three DnaA boxes (**Figure 1**). In addition, *DARS1* needs a specific DNA region flanking the core for stimulation of ADP dissociation from DnaA (Fujimitsu et al., 2009), while *DARS2* contains three additional DnaA boxes and requires both Fis binding sites (FBS) 2 and 3, and IHF binding

to IHF binding sites (IBS) 1 and 2 be active (Kasho et al., 2014) (**Figure 1**).

Termination of replication occurs in *terC*, a poorly-defined region approximately 180° away from *oriC* (Hill et al., 1987). If an uneven number of homologous recombination events between daughter chromosomes have taken place during replication, the end result will be a chromosome dimer (Sherratt et al., 2004). Resolution takes place at a 28 bp site *dif,* located in *terC* (Sherratt et al., 2004) in a process involving two tyrosine recombinases, XerC and XerD. The XerCD recombinase is activated and delivered at *dif* by the FtsK translocase (Bigot et al., 2005). Numerous forces seem to shape the organization of bacterial chromosomes, and the pattern of these forces on the chromosome is evident at different levels. In both the Gram-negative bacteria *E. coli* (Bergthorsson and Ochman, 1998) and *Salmonella enterica* (Liu and Sanderson, 1995a,b), as well as the Gram-positive bacterium *Lactococcus lactis* (Campo et al., 2004), selective pressure maintains a global architecture of the chromosome, which preserves two replication arms of nearly equal length. In addition to chromosome symmetry further chromosomal constrains are observed in *E. coli*. Four insulated macrodomains (MD) and two less constrained regions called non-structured (NS) regions has been uncovered (Niki et al., 2000; Valens et al., 2004). MDs are defined as regions where DNA interactions occur preferentially, while DNA interactions between the different MDs are highly restricted. NS regions can however interact with both its flaking MDs (Valens et al., 2004). *oriC* and *datA* are contained within the Ori MD, while the Ter MD contains *dif*. The Ori MD is flanked by NS^{Right} and NS^{Left} (where *DARS2* is found) whereas the Ter MD is flanked by the Left MD and the Right MD which contain *DARS1* (Valens et al., 2004) (**Figure 2**). Several observations indicates that the MDs and NS plays a part in the segregation of sister chromatids and the mobility of chromosomal DNA. The Ori MD is centered on a centromere-like 25 bp sequence designated *migS*, which affects *oriC* positioning during chromosome segregation (Yamaichi and Niki, 2004; Fekete and Chattoraj, 2005). Furthermore, movement of the Ter MD is maintained by several factors including the MatP/*matS* system (Mercier et al., 2008), and ZapA, ZapB, and FtsZ (Espeli et al., 2012; Buss et al., 2015).

Despite of the restraints on the *E. coli* K-12 chromosome, the size of genomes of other *E. coli* species varies from 4.6 to 5.7 Mb, indicating that horizontal gene transfer and genome reductions frequently takes place (Leimbach et al., 2013). A very dynamic genome structure underlies the metabolic and phenotypic diversity of *E. coli*. The genome of a bacterial species can be grouped into two categories. The core genome contains genes present in all strains, while the flexible genome comprises genes that are present in only a few strains or unique to single isolates (Medini et al., 2005). The pan-genome of a bacterial species is the combination of the core genome and the flexible genome (Medini et al., 2005). A typical *E. coli* genome has approximately 5000 genes, where roughly 2200 genes represent the core genome (Rasko et al., 2008). *E. coli* has a very large pan-genome (>18,000 genes), which grows for each new genome sequenced (Medini et al., 2005; Rasko et al., 2008). This indicates that there is a great diversity in gene content between *E. coli* species. Nevertheless, comparison of bacterial

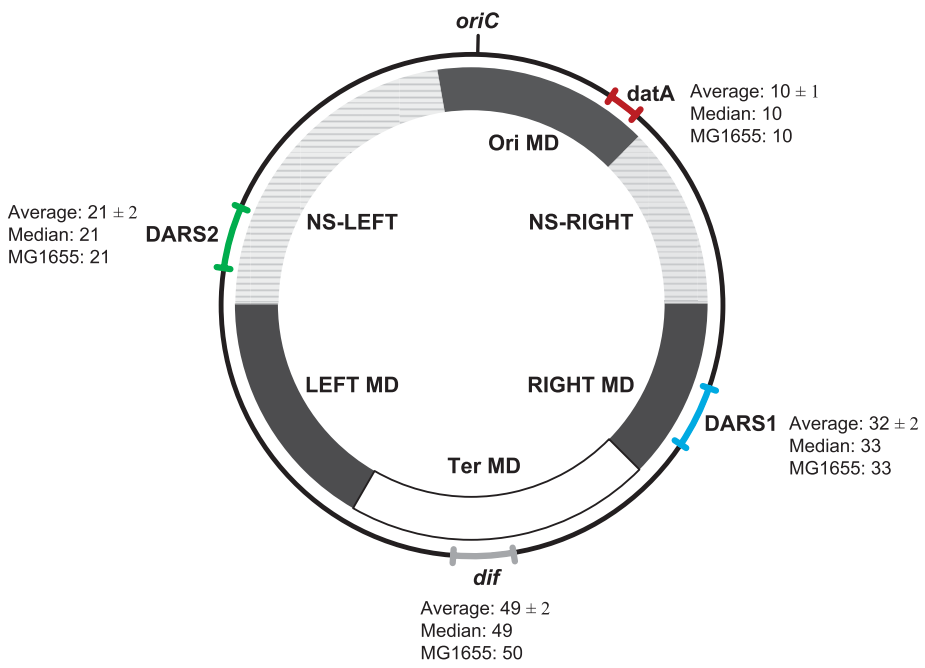


FIGURE 2 | Location of *datA*, *DARS1*, *DARS2*, and *dif regions on the E. coli* genomes. The range (the relative distance ± the standard deviation in centisomes) in the chromosomal position of *datA* (red), *DARS1* (blue), *DARS2* (green), and *dif* (gray) are shown in the outer circle. For each region the average relative distance with the standard deviation (see Materials and Methods), the median, as well as the chromosomal position in MG1655 (MG1655) are indicated. All distances are given in centisomes. The inner circle schematically shows the location of the different MD- and NS-regions as indicated by Esnault et al. (2007).

chromosomes from related genera revealed a conservation of organization (Eisen et al., 2000). For instance, even though *E. coli* and *Salmonella typhimurium* diverged from a common ancestor about 140 million years ago their genetic maps are extensively superimposable (Groisman and Ochman, 1997).

Here we report a conserved chromosomal position of the non-coding regions *datA*, *DARS1*, and *DARS2* relative to *oriC* in *E. coli*. In addition, we report that the structural organization of *oriC*, *datA*, *DARS1*, and *DARS2* regions are conserved in all *E. coli* strains analyzed. Furthermore, we demonstrate that even though the loss of *datA, DARS1*, or *DARS2* did not result in a measurable reduction in growth rate, the mutant cells had a lower fitness than wild-type when tested under laboratory conditions or in mice.

Materials and Methods

Growth Conditions

Cells were grown in Luria–Bertani Broth (LB) or AB minimal medium supplemented with 0.2% glucose, 10 μg/ml thiamine, and when indicated 0.5% casamino acids. MacConkey agar was used for mice experiments. All cells were cultured at 37°C. When necessary, antibiotics were added to the following concentrations: kanamycin, 50 μg/ml; streptomycin, 100 μg/ml; chloramphenicol, 20 μg/ml; ampicillin, 150 μg/ml.

Bacterial Strains and Plasmids

A spontaneous streptomycin resistance mutant of MG1655 (ALO1825) was obtained by plating an overnight culture on streptomycin-plates, resulting in MG1655 StrR (ALO4292) (see Table S1 for used strains).

The *DARS2* region containing DnaA box I-III was replaced with the cat gene in MG1655 by the lambda red procedure (Datsenko and Wanner, 2000), resulting in the Δ*DARS2*::*cat* mutant (ALO4254). Briefly, DNA fragments were PCR amplified using modified primers MutH-9 (5′-TCACAGTTATGTGCAGAGTTATAAACAGAGGAAGGGGTGGATAGCCGTTTCGATTTATTCAACAAAGCCACG-3′) and MutH-10 (5′-CTACGGAATTACTACGGGAAAACCCGGAGCATTCTGAATAAGCCCGATATGCCAGTGTTACAACCAATTAACC-3′), where the underlined sequence will anneal to pKD3 (Datsenko and Wanner, 2000). Each deletion was verified by PCR. The *DARS2* deletion was moved from ALO4254 to ALO4292 by P1 transduction using established procedures (Miller, 1972) and by selection for chloramphenicol resistance, resulting in ALO4310. The *cat* gene was removed from ALO4310 by pCP20, according to a method described previously (Cherepanov and Wackernagel, 1995), resulting in ALO4312.

The *DARS1* region was replaced with the *cat* gene in ALO4292 harboring pKD46, as described above, resulting in the Δ*DARS1*::cat mutant (ALO4313). DNA fragments were PCR amplified using pKD3 as template and primers *DARS1*_pKD3_FW (5′-TACATAAACCTTGCCTTGTTGTAGCCATTCTGTATTCGATTTATTCAACAAAGCCACG-3′) and *DARS1*_pKD3_RV (5′-AAAACAGTTCATCACCATAATATTTCTGATACAGCGTAAAGCCAGTGTTACAACCAATTAACC-3′) using pKD3 as a template. Each deletion was verified by PCR. The double deletion of *DARS1* and *DARS2* was obtained in the same background by moving the *DARS1* deletion from ALO4313 into the *cat* sensitive ALO4312 by P1 transduction, resulting in ALO4315.

The Δ*datA*::*kan* allele was obtained from RSD428 (Kitagawa et al., 1998). The *datA* deletion was moved into ALO4292 and ALO4315 by P1 transduction, selecting for kanamycin, resulting in ALO4331 and ALO4511 respectively. Each deletion was verified by PCR.

lacZ::Tn*5*::*kan* was moved from MC1000 F' lacI^q, *lacZ*::Tn*5* (laboratory stock) to MG1655 by P1 transduction and selecting for kanamycin, resulting in ALO1257. *DARS1* and *DARS2* were deleted in ALO1257 (as described above) to give ALO4618 and ALO4619. The *cat* gene was removed from ALO4618 and ALO4619 by pCP20, before transformation of pALO75 (Løbner-Olesen et al., 1987) for investigation of β-galactosidase synthesis from the *mioC* promoter. Strain RB210 (MC1000 carries a *dnaA-lacZ* translational fusion on phage λRB1 integrated at *att*λ (Braun et al., 1985). λRB1 was transduced from RB210 to ALO1257, resulting in ALO1265. The deletion of *DARS2* or *datA* was done as described above and resulted in strains ALO4626 and ALO4627, respectively, for investigation of β-galactosidase synthesis from the *dnaA* promoter.

Flow Cytometry

Flow cytometry was performed as described previously (Løbner-Olesen et al., 1989) using an Apogee A10 instrument. For each sample, a minimum of 30,000 cells were analyzed. Numbers of origins per cell and relative cell mass were determined as described previously (Løbner-Olesen et al., 1989).

The distribution of origins per cell was measured after treating exponentially growing cells with rifampicin and cephalexin for 4 h. Rifampicin block initiation of replication, while cephalexin will block cell division. The average number of chromosomes per cell will therefore be equivalent to the number of *oriC's* present in the cell at the time the drugs were added.

Relative Distance

The relative distance between *oriC* and *DARS1*, *DARS2*, *datA,* and *dif* were calculated in centisomes (see equation below). Each *E. coli* chromosome is by definition 100 centisomes. The relative distance is set as the distance in base pairs between *oriC* and the region of interest divided by the size of the genome in base pairs of the investigated *E. coli* strain. There are two distances to *dif*, one for each replication arm. In this study only the shortest replication arm is presented, while the distance of the longest replication arm by default is the sum of the shortest replication arm substracted from 100.

$$\textit{Relative distance in centisomes} = \left(\frac{[\textit{Position of oriC}] - [\textit{Position of region of interest}]}{[\textit{Chromosome size}]}\right) * 100$$

To calculate the relative distance the chromosome needs to be fully assembled. For this the sequence of 70 fully assembled *E. coli* chromosomes from The European Nucleotide Archive (http://www.ebi.ac.uk/genomes/bacteria.html) were obtained and analyzed. Two times two strains were uploaded under the same

strain name, i.e., W and ST540. We denote them Wa (uploaded under sequence CP002185), Wb (uploaded under sequence CP002967), ST540a (uploaded under sequence CP002185), and ST540b (uploaded under sequence CP002967). Wa are identical to Wb, why Wa was used, while ST540b was used as ST540a was excluded (see below). BL21-DE3 and BL21-Gold were excluded for being deviates of B str. REL606, while KO11, KO11FL, and LY180 were excluded for being deviates of W.

Six *E. coli* genomes were found to have relative distances, which were two times the standard deviation or more away from the average (Table S2). Of these W3110, MC4100, and strain ST540a were excluded. ST540a and MC4100 has 20% or more imbalance between the length of the two replication arms, which have been shown to give abnormal cells that was dependent on the RecBC-dependent homologous recombination for viability (Esnault et al., 2007), why they were excluded. W3110 is disqualified due to a known inversion around *oriC* (Hayashi et al., 2006), which explains the altered relative distances compared to the *E. coli* average.

The final dataset comprised of 59 fully assembled *E. coli* genomes (Table S3). The position and sequence of *oriC*, *DARS1*, *DARS2*, *datA*, and *dif* are known in MG1655, but not annotated in the dataset. We therefore choose to annotate the regions in 58 remaining *E. coli* genomes. The sequence of each of the regions (see Supplementary Data) from MG1655 where therefore aligned with the chosen *E. coli* genomes to obtain the chromosomal position and sequence of the region in each individual *E. coli* genomes.

Mutation Frequency

The mutation frequency was estimated for intergenic regions. Regions between protein-coding genes of more than 300 bp were selected in MG1655. These regions were trimmed for 100 bps on each side, to avoid conserved promoter-regions, and blasted against the 58 remaining *E. coli* genomes. If not present in the entire dataset the intergenic region was discharged, resulting in 109 regions. Of these 13 intergenic regions contained known conserved sRNA or tRNA's, why they were removed, resulting in 96 intergenic regions. Each intergenic region was aligned and number of nucleotides that were not present in every genome was calculated to give a mutation frequency.

Competition Experiment in LB

The fitness of Δ*DARS1*, Δ*DARS2*, Δ*DARS1* Δ*DARS2*, and Δ*datA* compared to the wild-type were investigated during direct competition in LB medium. The competing strains were inoculated pairwise at an approximate concentration of (10^7 CFU/mL) each. The populations were propagated by continuously transfers in LB medium. Samples from each population were taken at 10-generation intervals. Each sample was diluted in 0.9% NaCl and plated on LB plates with appropriate antibiotics. To distinguish the various *E. coli* strains, dilutions were plated on LB plates containing no antibiotic, kanamycin, or chloramphenicol. All plates were incubated for 18–24 h at 37°C prior to counting. When necessary to distinguish strains, 100 colonies from plates containing no antibiotic were toothpicked onto LB plates containing kanamycin or LB plates containing chloramphenicol.

Mouse Colonization Experiments

The specifics of the streptomycin-treated mouse model used to compare the large intestine colonizing abilities of *E. coli* strains in mice have been described previously (Leatham et al., 2005; Leatham-Jensen et al., 2012). Briefly, Six-to-eight-week-old, outbreed female CD-1 (Charles River Laboratories, Netherlands) mice were given drinking water containing streptomycin sulfate (5 g/l) for 24 h to eliminate resident facultative anaerobic bacteria (Miller and Bohnhoff, 1963). Mice were orally fed 100 μL of 20% (wt/vol) sucrose containing 10^6 CFU LB grown *E. coli* strains. The number of *E. coli* colonizing the mouse large intestine is reflected in the mouse feces, which is why fecal counts are used to estimate the various *E. coli* strains' ability to colonize the mouse intestine (Leatham-Jensen et al., 2012). After ingesting the bacterial suspension; feces was collected after 24 h, and as indicated. The mice were caged in groups of three mice, and cages were changed weekly. Mice were marked so they could be isolated and fecal pellets could be collected from each individual mouse. Mice were given fresh drinking water containing streptomycin sulfate (5 g/l) each day. Each fecal sample was homogenized in 1% Bacto tryptone (Difco Laboratories, NJ, USA), diluted in the same medium, and plated on MacConkey agar plates with appropriate antibiotics. When appropriate, 1 ml of a fecal homogenate (sampled after the feces had settled) was centrifuged at 12,000 X *g*, resuspended in 100 μL of 1% Bacto tryptone, and plated on a MacConkey agar plate with the appropriate antibiotics. This procedure increases the sensitivity of the assay from 10^2 CFU/gram of feces to 10 CFU/per g of feces. To distinguish the various *E. coli* strains in feces, dilutions were plated on lactose MacConkey agar containing either streptomycin, streptomycin and kanamycin, or streptomycin and chloramphenicol. All plates were incubated for 18–24 h at 37°C prior to counting. When necessary to distinguish strains, 100 colonies from plates containing streptomycin were toothpicked onto MacConkey agar plates containing streptomycin and kanamycin or onto MacConkey agar plates containing streptomycin and chloramphenicol. Ethics approval statement; 2007/561-1430.

β-Galactosidase Assays

Cells were grown exponentially at 37°C in AB minimal medium supplemented with casamino acids, and β-galactosidase activities were measured as described by Miller (1972).

Results

Conserved Relative Distance from *oriC* to *datA*, *DARS1*, *DARS2*, and *dif* in *E. coli*

Bergthorsson and Ochman (1998) suggested that there is an evolutionary pressure on keeping the *E. coli* chromosome symmetric, so an approximately equal length of the two replication arms are maintained. The non-coding regions *DARS1*, *DARS2*, and *datA* are all indirectly involved in initiation of replication at *oriC* as they modulate the activity and for *datA*

also the amount of DnaA available for initiation in a dosage dependent manner. The replication-associated gene dosage of each region relative to *oriC* changes with growth rate and is given by the formula $N_x/N_{oriC} = 2^{([C \times (1-x) + D]/\tau)}$ where x is the relative distance from *oriC*, C is the replication period, D is the time following termination of replication until cell division, and τ is the doubling time (Bremer and Churchward, 1977). We therefore decided to investigate if there was any evolutionary pressure on their chromosomal position relative to *oriC*. The genome size of *E. coli* varies from 4.6 to 5.7 Mb (Leimbach et al., 2013). Thus, to compare chromosomal positions between genomes with up to 1 Mb difference we calculated a relative distance from *oriC* (see Materials and Methods) while using MG1655 as reference strain. The replication terminus is not as well defined as the origin of replication, this is why *dif* was chosen to represent *terC* (Hendrickson and Lawrence, 2007).

Although the study was limited to 59 "closed" *E. coli* genomes, the dataset includes a wide variety of different *E. coli* (see Table S3). Pathogenic *E. coli* strains are categorized into pathotypes (Kaper et al., 2004). The dataset includes four pathotypes, which are associated with diarrhea, namely shiga toxin-producing *E. coli* (STEC)/enterohemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteropathogenic *E. coli* (EPEC), and enteroaggregative *E. coli* (EAEC). In addition to the intestinal pathogens two *E. coli* associated with the inflammatory bowel disease Crohn's disease were also included. In contrast to intestinal pathogenic *E. coli* (IPEC), which are obligate pathogens, extraintestinal pathogenic *E. coli* (ExPEC) are facultative pathogens which belong to the normal gut flora of a certain fraction of the healthy population where they live as commensals (Köhler and Dobrindt, 2011). The dataset contains ExPEC associated with neonatal-meningitis, asymptomatic bacteriuria, acute cystitis, the multidrug resistant ST131, as well as several uropathogenic *E. coli* (UPEC). In addition to human pathogenic *E. coli* strains, several *E. coli* strains isolated from the feces of healthy individuals (human commensals) are included. Apart from numerous common *E. coli* laboratory strains the dataset is concluded by three *E. coli* strains shown to be pathogenic in animals (avian pathogenic *E. coli* (APEC), and porcine enterotoxigenic *E. coli*) as well as an *E. coli* isolated from a toxic-metal contaminated site (for references see Table S3).

The chosen *E. coli* genome dataset had a median genome size of 5,095,204 bp, spanning from 3,976,195 bp (MDS42) to 5,697,240 bp (O26:H11 str. 11368). MDS42 is a "man-made" reduced *E. coli* K-12 genome derived from MG1655, which was constructed to identify non-essential genes (Pósfai et al., 2006). The smallest non-lab constructed *E. coli* chromosome was BW2952 with 4,578,159 bp. Due to the great diversity in both origin of isolation and genome size we believe that the dataset will be representative of *E. coli* as a whole.

Despite the large differences in genome size between *E. coli* strains, we found approximately the same relative distance from *oriC* to *dif, datA*, *DARS1*, and *DARS2*, respectively (**Figure 2**). This observation points to a conserved chromosomal organization. This organization is further conserved at the replichore level as, *DARS2* was always found on one replichore, while *datA* and *DARS1* were always found on the other replichore. *dif* was found at the chromosomal position opposite of *oriC*, which indicates that both replications arms were of approximately equal length in accordance with data from Bergthorsson and Ochman (1998).

E. coli Chromosome Symmetry

The conserved position of *DARS1*, *DARS2*, *datA*, and *dif* relative to *oriC* in the tested 59 *E. coli* genomes, suggests that new DNA obtained by horizontal gene transfer has been equally distributed between the two replication arms, but also between the different regions on each of the replication arms. Strain O157:H7 EDL933 that has a genome size of 5.53 Mb, i.e., about 0.9 Mb bigger than the laboratory strain MG1655, exemplifies this (**Figure 3**). MG1655 and O157:H7 EDL933 shares a common 4.1 Mb backbone, which is co-linear except for one 422-kilobase inversion spanning the replication terminus (Perna et al., 2001). The differences between the two genomes are reflected in K-islands (0.53 Mb), which is the DNA present only in MG1655 and O-islands (1.34 Mb), which is unique to O157:H7 EDL933 (Perna et al., 2001). When a circular genome map of O157:H7 EDL933 is compared to MG1655 the 1.34 Mb DNA unique to O157:H7 EDL933 is not only distributed between the two replication arms but as expected also among the *cis*-acting regions for regulation of initiation of replication (gray boxes; see **Figure 3**).

The *E. coli* strain MDS42 (Pósfai et al., 2006) contains a 14.0% reduced genome relative to it's the parental MG1655. However, it maintained a similar relative distance from *oriC* to *DARS1*, *DARS2*, *datA*, and *dif* as the parental strain (see Table S3), i.e., the non-essential DNA lost from MG1655 was distributed between the different non-coding *cis*-acting regions.

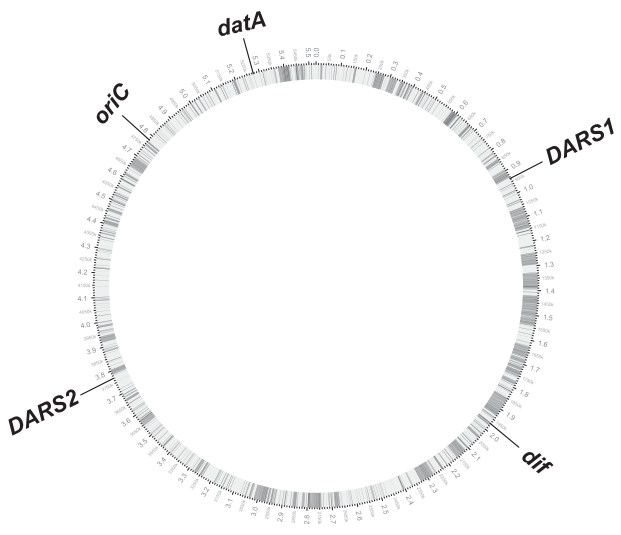


FIGURE 3 | Circular genome map of O157:H7 EDL933 compared with MG1655. The outer circle shows the genomic backbone of the O157:H7 EDL933 genome with the chromosomal position of *oriC, datA*, *DARS1*, *DARS2*, and *dif* annotated. The inner circle displays in white the common 4.1 Mb backbone between the two *E. coli* genomes, while the gray bars illustrates the distribution of the 1.34 Mb DNA unique to O157:H7.

Conservation of *oriC, DARS1, DARS2* and *datA* regions

Only a few genomes showed 100% sequence identity of the *oriC*, *DARS1*, *DARS2*, and *datA*-regions to those of MG1655. The comparison between the nucleotide sequences from the 59 different *E. coli* genomes is found in the Supplementary Material (Supplementary Figures S12–S15).

In order to estimate the mutation pressure on *oriC*, *datA*, *DARS1*, and *DARS2* we calculated the mutation frequency for intergenic regions in *E. coli* (see Materials and Methods). It was found to be 6 mutations per 100 ± 3 bp. Neither of the *oriC, datA*, *DARS1,* or *DARS2* regions differed significantly from this average frequency (Not shown). However, the vast majority of changes observed were found in spacer regions whereas binding sites for key proteins were conserved among all genomes (**Table 1**) which underlines their important role for cell cycle control.

oriC

In *oriC* both of the AT-rich 13-mer regions L and M were identical among the 59 strains, whereas the R 13-mer varied in the two outer positions (Supplementary Figure S1). Three of the six 6-mer sites present in the AT-rich region (Supplementary Figures S3, S4, S6) that specifies binding of $DnaA^{ATP}$ when in the single stranded configuration (**Figure 1**) were identical to the same regions in MG1655. However, Supplementary Figures S1, S2, S5 carries single nucleotide changes relative to MG1655 in a subset of strains. In five strains a nucleotide alteration was found in the single stranded DnaA-ATP box 2 that demolished a GATC-site, which is the substrate for Dam methyltransferase (Supplementary Figure S2).

The majority of DnaA binding sites in *oriC* (R1, R2, R5, τ1, τ2, I3, C1, and C2) were completely conserved, whereas R3, C3, I1, I2, and R4 binding sites carried differences to the corresponding MG1655 sequences in some strains (Supplementary Figure S12). Controversy about which of DnaA Box R3 and DnaA Box C3 are functional during unwinding of *oriC* exists (Kaur et al., 2014). The DnaA Box R3 overlaps with both DnaA Box C2 and C3 (**Figure 1**; Supplementary Figure S12). Nonetheless, all three DnaA boxes are included in the present study.

The consensus sequence for the R-box is TTWTNCACA (W is dA or dT and N is any nucleotide) (Schaper and Messer, 1995). In MG1655 DnaA Box I1 differs from the R-box consensus sequence by three nucleotides, while DnaA Box C3 and I2 differ by four nucleotides (Grimwade et al., 2000; Ryan et al., 2002; Rozgaja et al., 2011). We only identified sequence alterations that resulted in an altered identity to the R-box consensus sequences for DnaA binding sites, I2, R3, and R4. The I2 binding site from strain ED1a was found to fit better to the R-box consensus sequence compared to the DnaA box I2 sequence from MG1655 (Supplementary Figure S4). The MG1655 DnaA Box R3 differs from the R-box consensus sequence by one nucleotide, while the 10 *E. coli* strains deviating from the MG1655 DnaA Box R3 sequence deviate from the R-box consensus sequence by two nucleotides (Supplementary Figure S5). Strain 0127:H6 E2348/69 deviates from the R-box consensus sequence by a nucleotide in DnaA Box R4 (Supplementary Figure S6). The change from a dT to a dC in nucleotide position number 4 diminishes the identity to the R-box consensus sequence.

The IHF binding site consensus sequence is WATCAANNNNTTR [W is dA or dT, R is dA or dG, and N is any nucleotide (Hales et al., 1994)]. Three strains were found to differ with respect to the *oriC* IBS (Supplementary Figure S7).

TABLE 1 | Sequence deviations in conserved regions of *oriC*, *DARS1*, *DARS2* and *datA*.

Region	Identity	Functional site[a]	No. of sequences deviating from MG1655[b]	See Figures
oriC	Conserved	13-mer L 13-mer M DnaA Box R1 DnaA Box R2 DnaA Box R5 DnaA Box τ1 DnaA Box τ2 DnaA Box I3 DnaA Box C1 DnaA Box C2 SSDA Box 3 SSDA Box 4 SSDA Box 6 AT		
	Not conserved	13-mer R	6/59	S1
		SSDA Box 1	1/59	S2
		SSDA Box 2	5/59	S2
		SSDA Box 5	1/59	S2
		DnaA Box I1	6/59	S3
		DnaA Box I2	1/59	S4
		DnaA Box C3/R3	10/59	S5
		DnaA Box R4	1/59	S6
		IBS	3/59	S7
DARS1	Conserved	DnaA Box 1 DnaA Box 2 DnaA Box 3		
DARS2	Conserved	DnaA Box 1 DnaA Box 2 DnaA Box 3 DnaA Box 4 DnaA Box 5 DnaA Box 6 IBS-1		
	Not conserved	FBS-2	3/59	S8
		FBS-3	9/59	S9
		IBS-2	4/59	S10
datA	Conserved	DnaA Box 1 DnaA Box 2 DnaA Box 3 DnaA Box 4 DnaA Box 5		
	Not conserved	IBS	12/59	S11

[a] *FBS, Fis Binding Site; IBS, IHF Binding Site; SSDA, Single stranded binding site for* $DnaA^{ATP}$.

[b] *Number of isolates within the dataset which deviates from the region in MG1655.*

O26:H11 str. 11368 was found to have a diminished identity to the IHF consensus sequence compared to MG1655, while both O145:H28 str. RM12761 and O145:H28 str. RM13516 was found to have a better fit. The Fis binding site in *oriC* was found completely conserved in all strains (Supplementary Figure S12).

The nucleotide distances between the protein binding regions were highly conserved between strains. Only O26:H11 str. 11368 and O111:H- str. 11128 lacked a nucleotide in the spacer region between the AT-rich 13-mer termed R and the DnaA Box R1 in *oriC* compared to MG1655 (Supplementary Figure S12).

Based on this analysis it is hard to deduce a hierarchy of the importance of the different DnaA binding sites in *oriC*. Ten strains had changes in the R3/C3 boxes. Whereas the alterations resulted in a R3 box with poorer resemblance to the R-box consensus, this was not the case for C3. Therefore, it is likely that C3 represents the functional DnaA binding site in the replication origins.

DARS1, DARS2, and *datA*

All DnaA binding sites in *datA, DARS1,* and *DARS2* were completely conserved between the strains analyzed (**Table 1**) (Supplementary Figures S13–S15). FBS-2 of *DARS2* differed from that of MG1655 in three strains. However, since Fis has the consensus sequences GNNYANNNNNTRNNC (Y is dC or dT, R is dA or dG, and N is any nucleotide) (Finkel and Johnson, 1992) none of the observed differences resulted in a reduced similarity to the Fis consensus sequence (Supplementary Figure S8). For FBS-3 of *DARS2*, four *E. coli* strains differed from MG1655 and had a reduced identity the Fis consensus sequence (Supplementary Figure S9).

The IHF binding site IBS-1 of *DARS2* was identical in all 59 genomes. The sequence variations of IBS-2 of *DARS2* (Supplementary Figure S10) or the IBS in *datA* (Supplementary Figure S11), relative to MG1655 did not change the identity to the IHF consensus sequences. The *datA* region of strains HS and O103:H2 str. 12009 DNA lacked four nucleotides between DnaA Box 1 and DnaA Box 2, while strain 536 lacked a nucleotide between the IBS and DnaA Box 3.

Altogether, these observations suggest that there is a strong selection pressure on maintaining the sequence and spacing of protein binding sites, and thereby functionality of *oriC*, *DARS1*, *DARS2*, and *datA*. The majority of the nucleotide differences observed was located in the non-functional spacer regions between the different protein binding sites. The majority of differences found within protein binding sites, did not reduce the identity to the investigated consensus sequence.

Importance of *datA, DARS1*, and *DARS2* for Cell Cycle Control

Despite of the conservation of the chromosomal positions of *DARS1*, *DARS2,* and *datA* relative to *oriC*, neither is essential (Kitagawa et al., 1998; Fujimitsu et al., 2009). Cells with and without *datA* were also previously found to have similar doubling times (Kitagawa et al., 1998). We created cells with deletions of *datA*, *DARS1,* and *DARS2* individually and in various combinations. These cells were viable no matter which combination of *DARS1*, *DARS2*, and *datA* we deleted. The cellular doubling time was not affected by individual deletions but increased when combinations of *DARS1*, *DARS2*, and *datA* were deleted (**Table 2**). Cells carrying Δ*DARS1* Δ*DARS2* and Δ*DARS1* Δ*DARS2* Δ*datA* were found to have the longest doubling time in minimal medium supplemented with glucose and casamino acids, while Δ*DARS1* Δ*datA* and Δ*DARS2* Δ*datA* cells were found to have the longest doubling time in the same medium without casamino acids (**Table 2**).

We used *lacZ* fusions of *dnaA* and *mioC* promoters to assess the effect of *datA*, *DARS1,* and *DARS2* loss on the cellular $DnaA^{ATP}/DnaA^{ADP}$ ratio. The *dnaA* gene is transcribed from two upstream promoters, termed *dnaA1p* and *dnaA2p*. Four DnaA boxes are located between the two promoters, with only one of them containing the stringent consensus sequence (Hansen et al., 1982, 2007; Armengod et al., 1988). Both *dnaA* promoters are negatively regulated by the DnaA protein (Hansen et al., 2007), with $DnaA^{ATP}$ being most efficient in repressing *dnaA* expression (Speck et al., 1999). $DnaA^{ATP}$ also repress the *mioC* promoter located upstream of *oriC* prior to initiation by binding to five DnaA boxes located within and/or close the promoter. Of the five DnaA boxes only one contains the stringent consensus sequence (**Figure 1**) (Ogawa and Okazaki, 1994; Bogan and Helmstetter, 1997; Hansen et al., 2007). Loss of *datA* resulted in a slight repression of *dnaA* (**Table 3**). This is in agreement with an increase in the $DnaA^{ATP}/DnaA^{ADP}$ ratio,

TABLE 2 | Cell cycle parameters of mutant strains.

Strain/Genotype[a]	ABTG + CAA[b] (Min)	Origin/ Cell[c]	Cell mass[d]	Origins/ Mass[e]
Wild-type	40	5.2	1	1
Δ*datA*	41	7.9	1	1.5
Δ*DARS1*	40	4.7	0.9	0.8
Δ*DARS2*	41	4.4	0.9	0.8
Δ*DARS1* Δ*DARS2*	45	3.1	1.1	0.6
Δ*DARS1* Δ*datA*	41	7.2	1	1.4
Δ*DARS2* Δ*datA*	42	6.1	1.1	1.2
Δ*DARS1* Δ*DARS2* Δ*datA*	49	4	1	0.8
	ABTG[b]			
Wild-type	88	1.4	1	1
Δ*datA*	86	1.7	0.8	1.2
Δ*DARS1*	85	1.3	1.2	0.95
Δ*DARS2*	87	1.3	1.1	0.95
Δ*DARS1* Δ*DARS2*	91	1.2	1.5	0.9
Δ*DARS1* Δ*datA*	96	1.4	0.9	1.1
Δ*DARS2* Δ*datA*	103	1.4	0.9	1.1
Δ*DARS1* Δ*DARS2* Δ*datA*	91	1.2	1.1	0.9

[a] *Wild-type is MG1655.*

[b] *Doubling time in minimal medium supplemented with glucose and casamino acids or minimal medium supplemented with glucose. Each strain has a standard deviation of ± 2 min for growth in ABTG + CAA and ± 3 min for growth in ABTG.*

[c] *Determined from flow cytometric analysis.*

[d] *Determined as average light scatter from flow cytometric analysis. Numbers are normalized to 1 for wild-type.*

[e] *Average fluorescence/average light scatter. Numbers are normalized to 1 for wild-type.*

TABLE 3 | Expression of the *dnaA* and *mioC* genes.

Strain	*dnaA* expression (SD)[a]	*mioC* expression (SD)[b]
Wild-type	100	100
Δ*DARS1*	ND	108 (± 5)
Δ*DARS2*	116 (± 1)	122 (± 4)
Δ*datA*	95 (± 3)	ND

[a]*Measured in MG1655 using the dnaA-lacZ translational fusion carried on λRB1 (Braun et al., 1985) in strain MG1655 lacZ::Tn5. Numbers are given relative to wild-type expression of 100% corresponding to 46 Miller units. ND, Not determined; SD, Standard deviation.*

[b]*Measured in MG1655 lacZ::Tn5 using the mioC-lacZ transcriptional fusion carried on plasmid pALO75 (Løbner-Olesen et al., 1987). Numbers are given relative to wild-type expression of 100% corresponding to 302 Miller units. ND, Not determined; SD, Standard deviation.*

and the *dnaA* promoter being repressed by $DnaA^{ATP}$ (Kitagawa et al., 1998; Speck et al., 1999; Kasho and Katayama, 2013). Loss of *DARS1* led to an increased expression of the *mioC* gene while loss of *DARS2* led to an increased expression of both the *dnaA* and *mioC* genes (**Table 3**), since both promoters are subject to negative transcriptional control by $DnaA^{ATP}$ (Speck et al., 1999; Hansen et al., 2007). This agrees with *DARS1* and *DARS2* being instrumental in increasing the cellular $DnaA^{ATP}$ level, and that *DARS2* is more efficient than *DARS1* (Fujimitsu et al., 2009).

We proceeded to analyze the cell cycle characteristics by flow cytometry (**Table 2**; **Figure 4**). Wild-type cells exhibited the expected synchronous initiation pattern with the majority of cells containing 2, 4, or 8 replication origins (**Figure 4A**) (Skarstad et al., 1986). *datA* deficient cells had an increased origin concentration (origins/mass) (Kitagawa et al., 1998), which resulted both from an increase number of origins per cell and a decreased cell mass (during slow growth only) (**Table 2**). A high degree of initiation asynchrony was observed for Δ*datA* cells (**Figure 4E**) (Kitagawa et al., 1998). Cells deficient in *DARS1*, *DARS2* or both regions had a reduced origin concentration relative to wild-type cells (**Table 2**; **Figures 4B–D**) (Fujimitsu et al., 2009). Compared to wild-type, all DARS mutant cells had an increased cell mass during slow growth whereas only the Δ*DARS1* Δ*DARS2* double mutant had increased cell mass during fast growth. Asynchrony of initiation was observed for Δ*DARS2* (**Figure 4C**) and Δ*DARS1* Δ*DARS2* (**Figure 4D**) cells, but not for cells carrying the Δ*DARS1* mutation alone (**Figure 4B**) (Fujimitsu et al., 2009).

Because the *datA* region promotes inactivation of $DnaA^{ATP}$ to $DnaA^{ADP}$, and the DARS regions promote the opposite, i.e., DnaA reactivation, we decided to see whether loss of *DARS1*, *DARS2* or both could suppress the initiation defect of *datA* cells. Deletion of *DARS1* in Δ*datA* cells only marginally lowered the origin per mass (from 1.5 to 1.4; **Table 2**). A similar but larger effect was observed when *DARS2* was deleted suggesting that *DARS2* is more efficient than *DARS1* for DnaA rejuvenation. Deleting both *DARS1* and *DARS2* in Δ*datA* cells lowered the origin concentration below wild-type level (**Table 2**) and also partly restored initiation synchrony (**Figure 4H**). Overall, these experiments show that loss of rejuvenation activity overcompensates for loss of DDAH. This may be explained by the RIDA process, which being active in the triple mutant so that $DnaA^{ATP}$ to $DnaA^{ADP}$ conversion is still ongoing.

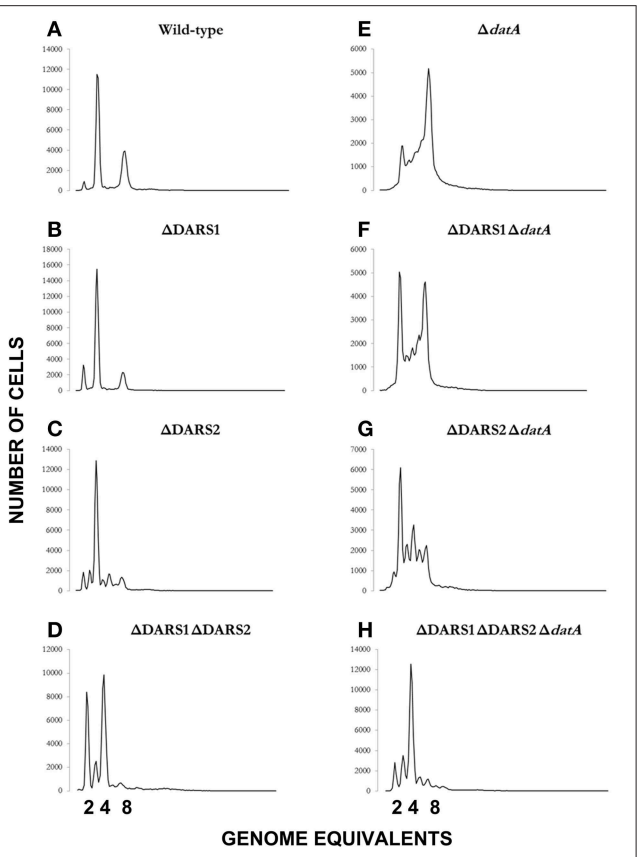


FIGURE 4 | Cellular origin distribution. Prior to flow cytometric analysis exponentially growing wild-type and mutant cells was treated with rifampicin and cephalexin. Cells were grown in AB minimal medium supplemented with 0.2% glucose, 10 μg/ml thiamine, and 0.5% casamino acids at 37°C. Wild-type is MG1655 **(A)**; relevant mutations are indicated in individual panels **(B–H)**.

DARS1 and *DARS2* are Required for Mouse Colonization

In order to examine the fitness cost of losing DARS or *datA* activity we performed two different competition experiments: continued growth in LB medium and during mouse colonization, where the streptomycin-treated mouse was chosen as the *in vivo* model. For both competition experiments strains were introduced pairwise at approximately equal numbers (**Figure 5**). If they have the same fitness they would also be recovered in equal numbers.

During growth in LB the wild-type was more fit than the cells deficient in either *DARS1*, *DARS2*, both *DARS1* and *DARS2*, or *datA* (**Figure 5**). The biggest fitness cost resulted from loss of both *DARS1* and *DARS2* (**Figure 5C**) followed by loss of *DARS2* (**Figure 5B**) loss of *DARS1* (**Figure 5A**) which was similar to loss of *datA* (**Figure 5D**).

The same order of fitness was not observed when evaluated in mice. Following colonization, the number of wild-type *E. coli* increased for about 3 days until stabilizing around 10^9 cfu

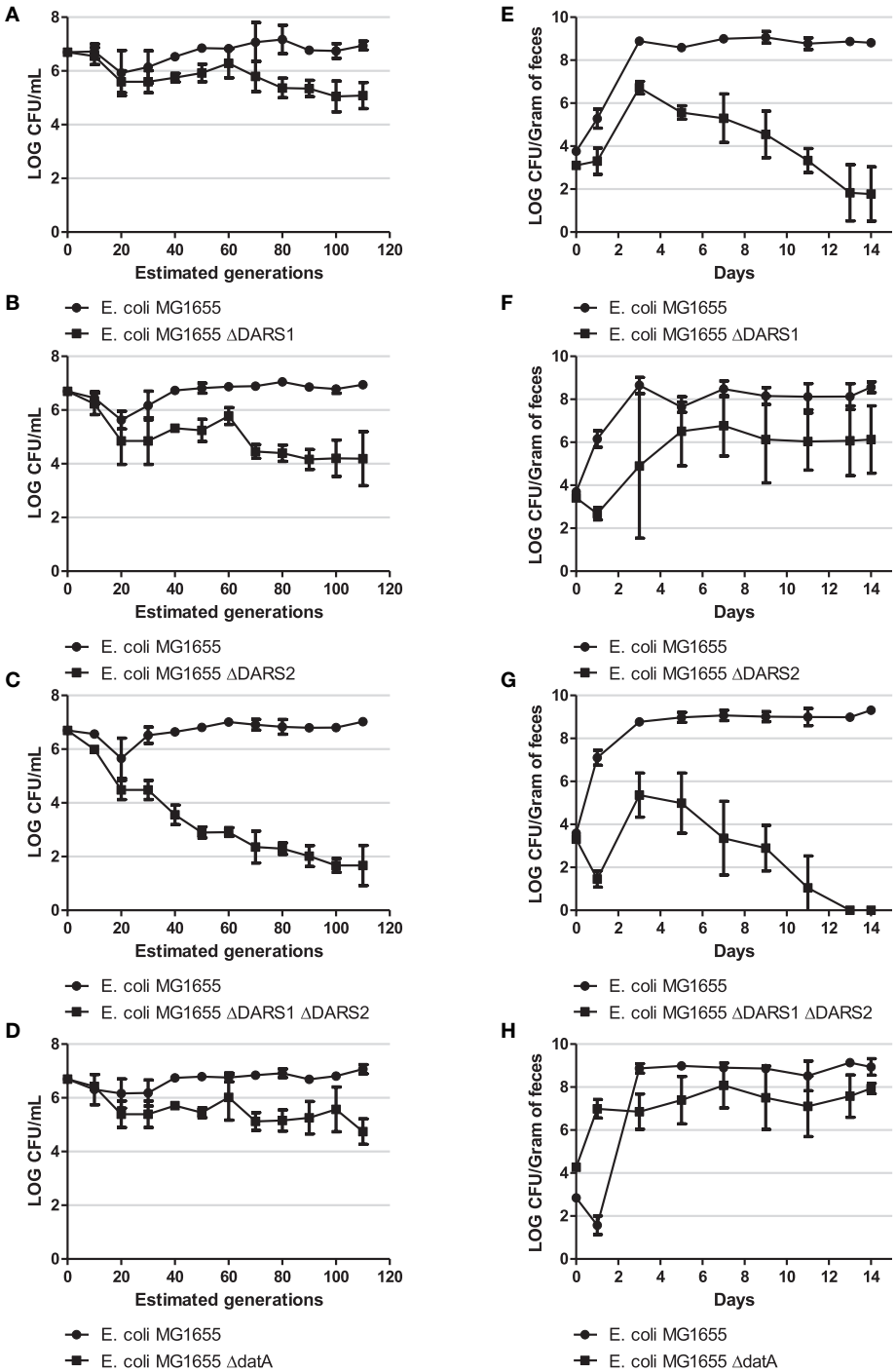


FIGURE 5 | Competition experiment. Competition experiment in LB medium **(A–D)** and during colonization of the mouse large intestine **(E–H)** for MG1655 and MG1655 Δ*DARS1* (**A,E**, respectively), MG1655 and MG1655 Δ*DARS2* (**B,F**, respectively), MG1655 and MG1655 Δ*DARS1* Δ*DARS2* (**C,G**, respectively), and MG1655 and MG1655 Δ*datA* (**D,H**, respectively). At the indicated times, samples were homogenized, diluted, and plated as described in Materials and Methods. Bars represent the standard error of the log10 mean number of CFU per mL (competition experiment in LB medium) or CFU per gram of feces (mice experiment).

per gram of mouse feces (**Figures 5E–H**). Cells deficient in *DARS1* increased in number to peak at about 10^7 cfu/gram feces at day 3 followed by a rapid decline in number over the next days to end around 10^2 cfu/gram feces at day 14 (**Figure 5E**) suggesting that these cells were rapidly out-competed by wild-type cells. Cells deficient in both *DARS1* and *DARS2* (**Figure 5G**) were outcompeted at a slightly faster rate than cells deficient in only *DARS1,* suggesting that DARS2 plays a

minor role to DARS1 in fitness during mouse colonization. In agreement with this, *DARS2* mutant cells were able to coexist in the mouse along with wild-type cells albeit at a lower number (**Figure 5F**). Loss of *datA* was similar to the loss of *DARS2*. Following co-infection in mice both wild-type and *datA* cells increased in numbers to level at 10^9 and 10^7 cfu per gram feces, respectively, and remained at these levels for the duration of the experiment (**Figure 5H**). Therefore, *datA* and *DARS2* deficient cells were poor at establishing colonization relative to the wild-type, but once established cells were not outcompeted with time.

On day 14 post-feeding, wild-type and a Δ*DARS1*, Δ*DARS2*, Δ*DARS1* Δ*DARS2*, or Δ*datA* cells (depending on the competition experiment) were isolated from the feces of each mouse for further study. The origin per mass and asynchrony index score were determined for each strain isolated post-infection and found to be similar to the initial strains fed to each mouse [data not shown] showing that secondary mutations were not likely to have been selected during growth in the mouse.

Overall these experiments indicate that different factors determine fitness of cells dependent on growth conditions. During continued growth in LB medium, both promotion and prevention of $DnaA^{ATP}$ to $DnaA^{ADP}$ conversion resulted in a fitness cost. On the other hand, overinitiation resulting from $DnaA^{ATP}$ accumulation in *datA* cells did not seem to affect mouse colonization to the same extent as loss of rejuvenation ability, especially promoted by *DARS1*.

Discussion

In this study we found conservation in distances from *oriC* to the non-coding regions *DARS1*, *DARS2*, *datA*, and *dif* in *E. coli*. *DARS1* and *datA* were always found on the same replichore, while *DARS2* were found on the other replichore. The *oriC*, *DARS1*, *DARS2*, and *datA* regions were found to be structurally similar among the tested *E. coli*, with most of the sequence differences found to be in the non-functional spacer regions between key protein binding sites. Cells deficient in *DARS1*, *DARS2*, or *datA* were viable and had doubling times similar to wild-type. However, replication initiation was perturbed. Cells deficient in *datA* were found to initiate asynchronously, and this could not be counteracted by further deletions of either *DARS1* or *DARS2*. Cells deficient in *DARS1*, *DARS2*, *DARS1,* and *DARS2*, or *datA* were found to be less fit than the wild-type in both LB medium and during mouse colonization.

Conservation of *oriC*

In the chromosomal context initiation of replication can be initiated from a mutant *oriC* without DnaA box R2, R3, R4, or R5, the IBS or the FBS (Weigel et al., 2001), as well as DnaA boxes I1, I2, or I3 (Riber et al., 2009). It is also possible to invert the direction of R4, add 14 bp between DnaA Box R3 and DnaA Box R4, or delete the right half of *oriC* (from position 275 to 352) (Weigel et al., 2001). Although DnaA Box R1 was originally found to be essential (Weigel et al., 2001), a more efficient recombining technique demonstrated that DnaA Box R1 is also dispensable (Kaur et al., 2014). Surprisingly, only deletion of DnaA Box R3, R4, and the right half of *oriC* (from position 275 to 352) was reported to result in slow growth relative to wild-type cells (Weigel et al., 2001). Asynchrony, a sensitive measure for perturbations of the initiation process, was observed with the deletion of the IBS, DnaA Box R2, R4, R5, extending the spacer region between R3 and R4 (Weigel et al., 2001), and the deletion of DnaA Box R1 (Kaur et al., 2014). These studies demonstrate that initiation from *oriC* is very robust and that only major changes in the origin results in loss of function altogether. On the other hand mutant origins fail to compete with their wild-type counterparts as shown by the inability to establish minichromosomes carrying *oriC* mutations in cells with a wild-type chromosomal copy of *oriC* (Weigel et al., 2001). It is conceivable that a similar competition takes place between cells in a population and that even small changes in important regions of *oriC*, that does not affect viability, may result in replication perturbation, loss of fitness and inability to co-exist with wild-type cells and that this explain the high degree of *oriC* conservation observed.

Chromosomal Position of *datA, DARS1*, and *DARS2*

The relative chromosomal locations of the *datA*, *DARS1*, and *DARS2* regions are conserved among *E. coli* strains. This is somewhat surprising as *E. coli* genomes are highly fluidic, i.e., they frequently mutate, change size, and rearrange. The frequency of genome rearrangement, measured between *rrn* sites, is about 10^3–10^4 changes/(generation.genome) (Hill and Gray, 1988). For example, the 1.34 Mb DNA unique to O157:H7 EDL933, is inserted compared to MG1655 in such a way that the chromosomal location of *datA*, *DARS1*, and *DARS2* relative to *oriC*, remain unchanged. Similarly, the non-essential DNA that was removed from MG1655 to create MSD42 (Pósfai et al., 2006), was also dispersed between regions so that MSD42 has the same relative chromosomal location of *datA*, *DARS1*, and *DARS2*. There may be at least three reasons for the conserved location of the three regions. First, chromosome asymmetry, i.e., different lengths of the two replication arms leads to slow growth (Hill and Gray, 1988). Second, chromosomal rearrangements resulting in a mixture of different macrodomains have deleterious effects of cell growth (Esnault et al., 2007) which may explain why *datA*, *DARS1* and *DARS2* regions located within the Ori MD, the Right MD and NS^{Left}, respectively (Valens et al., 2004), are always found on the same replichore. Third, the correct chromosomal location of *DARS1*, *DARS2*, and *datA* may be important for proper function and cell cycle progression (see below). As the activity of these regions in modulating DnaA binding to ATP or ADP is dependent on their copy number the proper distance to *oriC* becomes important for function. The relative copy number of *DARS1*, *DARS2*, and *datA* (replication-associated gene dosage) decreases with distance from *oriC* (Bremer and Churchward, 1977; Couturier and Rocha, 2006). *datA* was always found close to *oriC*. It is therefore conceivable that, *datA* is duplicated while *oriC* is still sequestered (Kitagawa et al., 1998; Kasho and Katayama, 2013), in all strains. The *datA*

site promotes $DnaA^{ATP}$ to $DnaA^{ADP}$ conversion, to prevent re-initiation when the concentration of $DnaA^{ATP}$ is high, i.e., just after sequestration ends and this may provide the evolutionary pressure that has resulted in a conserved chromosomal location. In agreement with this relocation of *datA* to a chromosomal position close to *terC* resulted in high asynchrony in initiation of replication while other positions closer to *oriC* resulted in a near wild-type phenotype (Kitagawa et al., 1998). It is likely that the chromosomal positions of *DARS1* and *DARS2* are also important for cell cycle control as they serve to re-activate the DnaA initiator protein in time for the next initiation. Relocation of DARS sequences has not been experimentally pursued. It is however tempting to speculate that the genomic arrangement of *DARS1* and *DARS2* will ensure that rejuvenation of $DnaA^{ADP}$ to $DnaA^{ATP}$ will be accelerated during later stages of the replication cycle and following duplication of these regions. This rejuvenation is important for increasing the $DnaA^{ATP}$ level for the following round of initiations.

The *datA, DARS1*, and *DARS2* Regions Are important for Fitness

In the *DARS1* and *DARS2* regions, all DnaA binding boxes as well as spacer distances were conserved among *E. coli* species. Especially the *DARS1* region had a very low mutation frequency which is consistent with reports that all three DnaA boxes along with the region flanking the last DnaA Box (42 bp spanning from base number 198–239) is required for full ADP-releasing activity of *in vitro* (Fujimitsu et al., 2009). Similarly DnaA Box 1 and DnaA Box 2 are crucial for ADP-releasing activity of *DARS2*, while DnaA Box 3 is required for full ADP-releasing activity *in vitro* (Fujimitsu et al., 2009). IBS1–2 and FBS2–3 are required for $DnaA^{ATP}$ regeneration *in vivo* (Kasho et al., 2014). In four *E. coli* isolates, FBS-3 has a weaker sequence identity to the Fis consensus sequence than in MG1655. Fis-binding sites are difficult to define due to the lack of an obvious consensus sequence (Finkel and Johnson, 1992), and the effect of the observed weaker identity in FBS-3 is hard to interpret.

DARS2 was previously reported to be more efficient in rejuvenation of $DnaA^{ADP}$ to $DnaA^{ATP}$ than *DARS1* (Fujimitsu et al., 2009) and both the fitness experiment performed in LB medium (compare **Figure 5A** and **Figure 5B**) and the *dnaA* and *mioC* expression studies agreed with this. The situation was reversed in the mouse model where loss of *DARS1* was associated with the biggest fitness cost (compare **Figure 5E** and **Figure 5F**). *DARS2* is activated by the binding of both IHF and Fis, whereas no protein factors are required for the function of *DARS1* (Fujimitsu et al., 2009; Kasho et al., 2014). While IHF is abundant in the cell during every growth phase, although the concentration is highest in the stationary phase (Azam and Ishihama, 1999), the concentration of Fis is dependent on the growth phase; i.e., it is highly abundant (10,000–50,000 molecules/cell) in early exponential phase, but decrease to <100 molecules/cell from late exponential phase to stationary phase. The level of Fis also varies during steady state growth; i.e., it is low during slow growth and high during fast growth (Nilsson et al., 1992; Flåtten and Skarstad, 2013). Limited data are available on the growth of *E. coli* in mouse intestines but overall slow growth was reported with doubling times between 80 and 125 min (Rang et al., 1999). It also seems reasonable that cells under these conditions never reaches exponential growth but grows whenever food becomes available, i.e., with relative short growth phases and frequent entries into stationary phase. Therefore, the bacterial Fis level during intestinal colonization may be significantly lower than during fast exponential growth in rich medium. The relative contribution of *DARS2* to DnaA rejuvenation may therefore be low in the mouse, and explain the bigger fitness cost associated with loss of *DARS1* under these conditions. In agreement with this we only observed a minor further fitness cost associated with deletion of *DARS2* in *DARS1* deficient cells. Such Δ*DARS1* Δ*DARS2* cells rely on *de novo* synthesis of DnaA or the speculated DARS3 to produce $DnaA^{ATP}$ during colonization of a mouse (Kasho et al., 2014). The fitness cost associated with loss of *DARS1* or *DARS2* may readily explain why mutations are rarely observed in these regions, but not the conserved distance to *oriC*. This needs to further elucidated by relocation to other chromosomal positions.

The activity of *datA* is absolutely dependent on DnaA Boxes 2 and 3 along with the IBS (Nozaki et al., 2009; Kasho and Katayama, 2013). Also the spacing between DnaA Box 2 and IBS, as well as the spacing between the IBS and DnaA Box 3 has been shown to be important for *datA* function (Nozaki et al., 2009; Kasho and Katayama, 2013). In accordance with this we found all DnaA boxes as well as the identities to the IHF consensus sequence were conserved. Changes were only observed in the length of the spacer region between DnaA Box 1 and DnaA Box 2 in strain HS and O103:H2 str. 12009 and between the IBS and DnaA Box 3 in strain 536. The effect of the altered spacing is hard to interpret although the latter may lead to a lower efficiency in converting $DnaA^{ATP}$ to $DnaA^{ADP}$ compared to MG1665 (Nozaki et al., 2009). Cells deficient in *datA* only had a 5% decrease in *dnaA* expression correlating with previous reports showing that a *datA* deletion slightly (i.e., 5–10%) increased the $DnaA^{ATP}$ level (Katayama et al., 2001). Loss of *datA* accompanied overinitiation only resulted in modest fitness cost during fast growth in LB and during colonization. The DDAH and RIDA (Hda dependent) pathways both contribute to convert $DnaA^{ATP}$ to $DnaA^{ADP}$ in *E. coli*, but where loss of RIDA is associated with severe overinitiation and inviability unless second site suppressor mutations arise (Riber et al., 2006); loss of DDAH is tolerated. Therefore, DDAH plays a minor role to RIDA and this may explain the limited fitness cost of *datA* cells. The limited fitness cost of losing *datA* relates poorly to the high degree of conservation observed between species. We do not have a good explanation for this observation but it may relate to the DDAH process being important during growth conditions other than those employed by us.

Of interest the Gram-positive bacteria *Bacillus subtilis* and *Streptomyces coelicolor* contain DnaA box clusters close to *oriC* that can repress untimely initiation (Smulczyk-Krawczyszyn et al., 2006; Okumura et al., 2012), i.e., a function similar to that of *datA* in *E. coli*. In addition, several *E. coli* related bacterial species contains *DARS1*-like sequence and *DARS2*-like sequences in a genomic position similar to that of *E. coli* (Fujimitsu et al., 2009; Kasho et al., 2014). These observations indicates that both *datA* and DARSs mechanism, and genomic positions, maybe common

to many bacterial species whose genomes contain DnaA box clusters.

Author Contributions

JF and AL planned the experiments. JF performed the experiments. JF, GC, KK, and AL analyzed data. JF and AL wrote the manuscript.

Funding

This work was supported by grant PIRG05-GA-2009-247241 from the European Union, by grant 09-064250/FNU from the Danish Research Council for Natural sciences, by grant 09-067075 from the Danish Strategic Research Council and by grants from the Lundbeck Foundation and the Novo Nordisk Foundation.

Acknowledgments

We thank Shiraz Shah for help with the bioinformatics analysis and Michaela Lederer for technical assistance.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2015.01011

References

Armengod, M. E., García-Sogo, M., and Lambíes, E. (1988). Transcriptional organization of the dnaN and recF genes of *Escherichia coli* K-12. *J. Biol. Chem.* 263, 12109–12114.

Azam, T. A., and Ishihama, A. (1999). Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity. *J. Biol. Chem.* 274, 33105–33113. doi: 10.1074/jbc.274.46.33105

Bergthorsson, U., and Ochman, H. (1998). Distribution of chromosome length variation in natural isolates of *Escherichia coli*. *Mol. Biol. Evol.* 15, 6–16. doi: 10.1093/oxfordjournals.molbev.a025847

Bigot, S., Saleh, O. A., Lesterlin, C., Pages, C., El Karoui, M., Dennis, C., et al. (2005). KOPS: DNA motifs that control *E. coli* chromosome segregation by orienting the FtsK translocase. *EMBO J.* 24, 3770–3780. doi: 10.1038/sj.emboj.7600835

Bogan, J. A., and Helmstetter, C. E. (1997). DNA sequestration and transcription in the oriC region of *Escherichia coli*. *Mol. Microbiol.* 26, 889–896. doi: 10.1046/j.1365-2958.1997.6221989.x

Braun, R. E., O'Day, K., and Wright, A. (1985). Autoregulation of the DNA replication gene dnaA in *E. coli* K-12. *Cell* 40, 159–169. doi: 10.1016/0092-8674(85)90319-8

Bremer, H., and Churchward, G. (1977). An examination of the Cooper-Helmstetter theory of DNA replication in bacteria and its underlying assumptions. *J. Theor. Biol.* 69, 645–654. doi: 10.1016/0022-5193(77)90373-3

Buss, J., Coltharp, C., Shtengel, G., Yang, X., Hess, H., and Xiao, J. (2015). A multi-layered protein network stabilizes the *Escherichia coli* FtsZ-ring and modulates constriction dynamics. *PLoS Genet.* 11:e1005128. doi: 10.1371/journal.pgen.1005128

Campbell, J. L., and Kleckner, N. (1990). *E. coli* oriC and the dnaA gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. *Cell* 62, 967–979. doi: 10.1016/0092-8674(90)90271-F

Campo, N., Dias, M. J., Daveran-Mingot, M. L., Ritzenthaler, P., and Le Bourgeois, P. (2004). Chromosomal constraints in Gram-positive bacteria revealed by artificial inversions. *Mol. Microbiol.* 51, 511–522. doi: 10.1046/j.1365-2958.2003.03847.x

Cherepanov, P. P., and Wackernagel, W. (1995). Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* 158, 9–14. doi: 10.1016/0378-1119(95)00193-A

Couturier, E., and Rocha, E. P. (2006). Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Mol. Microbiol.* 59, 1506–1518. doi: 10.1111/j.1365-2958.2006.05046.x

Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6640–6645. doi: 10.1073/pnas.120163297

Eisen, J. A., Heidelberg, J. F., White, O., and Salzberg, S. L. (2000). Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol.* 1:RESEARCH0011. doi: 10.1186/gb-2000-1-6-research0011

Esnault, E., Valens, M., Espéli, O., and Boccard, F. (2007). Chromosome structuring limits genome plasticity in *Escherichia coli*. *PLoS Genet.* 3:e226. doi: 10.1371/journal.pgen.0030226

Espeli, O., Borne, R., Dupaigne, P., Thiel, A., Gigant, E., Mercier, R., et al. (2012). A MatP-divisome interaction coordinates chromosome segregation with cell division in *E. coli*. *EMBO J.* 31, 3198–3211. doi: 10.1038/emboj.2012.128

Fekete, R. A., and Chattoraj, D. K. (2005). A cis-acting sequence involved in chromosome segregation in *Escherichia coli*. *Mol. Microbiol.* 55, 175–183. doi: 10.1111/j.1365-2958.2004.04392.x

Finkel, S. E., and Johnson, R. C. (1992). The Fis protein: it's not just for DNA inversion anymore. *Mol. Microbiol.* 6, 3257–3265. doi: 10.1111/j.1365-2958.1992.tb02193.x

Flåtten, I., and Skarstad, K. (2013). The Fis protein has a stimulating role in initiation of replication in *Escherichia coli in vivo*. *PLoS ONE* 8:e83562. doi: 10.1371/journal.pone.0083562

Fujimitsu, K., Senriuchi, T., and Katayama, T. (2009). Specific genomic sequences of *E. coli* promote replicational initiation by directly reactivating ADP-DnaA. *Genes Dev.* 23, 1221–1233. doi: 10.1101/gad.1775809

Fuller, R. S., Funnell, B. E., and Kornberg, A. (1984). The dnaA protein complex with the *E. coli* chromosomal replication origin (oriC) and other DNA sites. *Cell* 38, 889–900. doi: 10.1016/0092-8674(84)90284-8

Grimwade, J. E., Ryan, V. T., and Leonard, A. C. (2000). IHF redistributes bound initiator protein, DnaA, on supercoiled oriC of *Escherichia coli*. *Mol. Microbiol.* 35, 835–844. doi: 10.1046/j.1365-2958.2000.01755.x

Groisman, E. A., and Ochman, H. (1997). How Salmonella became a pathogen. *Trends Microbiol.* 5, 343–349. doi: 10.1016/S0966-842X(97)01099-8

Hales, L. M., Gumport, R. I., and Gardner, J. F. (1994). Determining the DNA sequence elements required for binding integration host factor to two different target sites. *J. Bacteriol.* 176, 2999–3006.

Hansen, E. B., Hansen, F. G., and Von Meyenburg, K. (1982). The nucleotide sequence of the dnaA gene and the first part of the dnaN gene of *Escherichia coli* K-12. *Nucleic Acids Res.* 10, 7373–7385. doi: 10.1093/nar/10.22.7373

Hansen, F. G., Christensen, B. B., and Atlung, T. (2007). Sequence characteristics required for cooperative binding and efficient *in vivo* titration of the replication initiator protein DnaA in *E. coli*. *J. Mol. Biol.* 367, 942–952. doi: 10.1016/j.jmb.2007.01.056

Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., Isono, K., Choi, S., et al. (2006). Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol. Syst. Biol.* 2:2006.0007. doi: 10.1038/msb4100049

Hendrickson, H., and Lawrence, J. G. (2007). Mutational bias suggests that replication termination occurs near the dif site, not at Ter sites. *Mol. Microbiol.* 64, 42–56. doi: 10.1111/j.1365-2958.2007.05596.x

Hill, C. W., and Gray, J. A. (1988). Effects of chromosomal inversion on cell fitness in *Escherichia coli* K-12. *Genetics* 119, 771–778.

Hill, T. M., Henson, J. M., and Kuempel, P. L. (1987). The terminus region of the *Escherichia coli* chromosome contains two separate loci that exhibit polar inhibition of replication. *Proc. Natl. Acad. Sci. U.S.A.* 84, 1754–1758. doi: 10.1073/pnas.84.7.1754

Kaper, J. B., Nataro, J. P., and Mobley, H. L. (2004). Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* 2, 123–140. doi: 10.1038/nrmicro818

Kasho, K., Fujimitsu, K., Matoba, T., Oshima, T., and Katayama, T. (2014). Timely binding of IHF and Fis to DARS2 regulates ATP-DnaA production and replication initiation. *Nucleic Acids Res.* 42, 13134–13149. doi: 10.1093/nar/gku1051

Kasho, K., and Katayama, T. (2013). DnaA binding locus datA promotes DnaA-ATP hydrolysis to enable cell cycle-coordinated replication initiation. *Proc. Natl. Acad. Sci. U.S.A.* 110, 936–941. doi: 10.1073/pnas.1212070110

Katayama, T., Fujimitsu, K., and Ogawa, T. (2001). Multiple pathways regulating DnaA function in *Escherichia coli*: distinct roles for DnaA titration by the datA locus and the regulatory inactivation of DnaA. *Biochimie* 83, 13–17. doi: 10.1016/S0300-9084(00)01206-2

Kato, J., and Katayama, T. (2001). Hda, a novel DnaA-related protein, regulates the replication cycle in *Escherichia coli*. *EMBO J.* 20, 4253–4262. doi: 10.1093/emboj/20.15.4253

Kaur, G., Vora, M. P., Czerwonka, C. A., Rozgaja, T. A., Grimwade, J. E., and Leonard, A. C. (2014). Building the bacterial orisome: high-affinity DnaA recognition plays a role in setting the conformation of oriC DNA. *Mol. Microbiol.* 91, 1148–1163. doi: 10.1111/mmi.12525

Kawakami, H., Keyamura, K., and Katayama, T. (2005). Formation of an ATP-DnaA-specific initiation complex requires DnaA Arginine 285, a conserved motif in the AAA+ protein family. *J. Biol. Chem.* 280, 27420–27430. doi: 10.1074/jbc.M502764200

Keyamura, K., Fujikawa, N., Ishida, T., Ozaki, S., Su'Etsugu, M., Fujimitsu, K., et al. (2007). The interaction of DiaA and DnaA regulates the replication cycle in *E. coli* by directly promoting ATP DnaA-specific initiation complexes. *Genes Dev.* 21, 2083–2099. doi: 10.1101/gad.1561207

Kitagawa, R., Ozaki, T., Moriya, S., and Ogawa, T. (1998). Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for *Escherichia coli* DnaA protein. *Genes Dev.* 12, 3032–3043. doi: 10.1101/gad.12.19.3032

Köhler, C. D., and Dobrindt, U. (2011). What defines extraintestinal pathogenic *Escherichia coli*? *Int. J. Med. Microbiol.* 301, 642–647. doi: 10.1016/j.ijmm.2011.09.006

Kurokawa, K., Nishida, S., Emoto, A., Sekimizu, K., and Katayama, T. (1999). Replication cycle-coordinated change of the adenine nucleotide-bound forms of DnaA protein in *Escherichia coli*. *EMBO J.* 18, 6642–6652. doi: 10.1093/emboj/18.23.6642

Leatham, M. P., Stevenson, S. J., Gauger, E. J., Krogfelt, K. A., Lins, J. J., Haddock, T. L., et al. (2005). Mouse intestine selects nonmotile flhDC mutants of *Escherichia coli* MG1655 with increased colonizing ability and better utilization of carbon sources. *Infect. Immun.* 73, 8039–8049. doi: 10.1128/IAI.73.12.8039-8049.2005

Leatham-Jensen, M. P., Frimodt-Møller, J., Adediran, J., Mokszycki, M. E., Banner, M. E., Caughron, J. E., et al. (2012). The streptomycin-treated mouse intestine selects *Escherichia coli* envZ missense mutants that interact with dense and diverse intestinal microbiota. *Infect. Immun.* 80, 1716–1727. doi: 10.1128/IAI.06193-11

Leimbach, A., Hacker, J., and Dobrindt, U. (2013). *E. coli* as an all-rounder: the thin line between commensalism and pathogenicity. *Curr. Top. Microbiol. Immunol.* 358, 3–32. doi: 10.1007/978-3-662-45793-1_303

Liu, S. L., and Sanderson, K. E. (1995a). The chromosome of Salmonella paratyphi A is inverted by recombination between rrnH and rrnG. *J. Bacteriol.* 177, 6585–6592.

Liu, S. L., and Sanderson, K. E. (1995b). I-CeuI reveals conservation of the genome of independent strains of Salmonella typhimurium. *J. Bacteriol.* 177, 3355–3357.

Løbner-Olesen, A., Atlung, T., and Rasmussen, K. V. (1987). Stability and replication control of *Escherichia coli* minichromosomes. *J. Bacteriol.* 169, 2835–2842.

Løbner-Olesen, A., Skarstad, K., Hansen, F. G., von Meyenburg, K., and Boye, E. (1989). The DnaA protein determines the initiation mass of *Escherichia coli* K-12. *Cell* 57, 881–889. doi: 10.1016/0092-8674(89)90802-7

Lu, M., Campbell, J. L., Boye, E., and Kleckner, N. (1994). SeqA: a negative modulator of replication initiation in *E. coli*. *Cell* 77, 413–426. doi: 10.1016/0092-8674(94)90156-2

Margulies, C., and Kaguni, J. M. (1998). The FIS protein fails to block the binding of DnaA protein to oriC, the *Escherichia coli* chromosomal origin. *Nucleic Acids Res.* 26, 5170–5175. doi: 10.1093/nar/26.22.5170

McGarry, K. C., Ryan, V. T., Grimwade, J. E., and Leonard, A. C. (2004). Two discriminatory binding sites in the *Escherichia coli* replication origin are required for DNA strand opening by initiator DnaA-ATP. *Proc. Natl. Acad. Sci. U.S.A.* 101, 2811–2816. doi: 10.1073/pnas.0400340101

Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R. (2005). The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15, 589–594. doi: 10.1016/j.gde.2005.09.006

Mercier, R., Petit, M. A., Schbath, S., Robin, S., El Karoui, M., Boccard, F., et al. (2008). The MatP/matS site-specific system organizes the terminus region of the *E. coli* chromosome into a macrodomain. *Cell* 135, 475–485. doi: 10.1016/j.cell.2008.08.031

Miller, C. P., and Bohnhoff, M. (1963). Changes in the mouse's enteric microflora associated with enhanced susceptibility to salmonella infection following streptomycin treatment. *J. Infect. Dis.* 113, 59–66. doi: 10.1093/infdis/113.1.59

Miller, J. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Niki, H., Yamaichi, Y., and Hiraga, S. (2000). Dynamic organization of chromosomal DNA in *Escherichia coli*. *Genes Dev.* 14, 212–223. doi: 10.1101/gad.14.2.212

Nilsson, L., Verbeek, H., Vijgenboom, E., Van Drunen, C., Vanet, A., and Bosch, L. (1992). FIS-dependent trans activation of stable RNA operons of *Escherichia coli* under various growth conditions. *J. Bacteriol.* 174, 921–929.

Nozaki, S., Yamada, Y., and Ogawa, T. (2009). Initiator titration complex formed at datA with the aid of IHF regulates replication timing in *Escherichia coli*. *Genes Cells* 14, 329–341. doi: 10.1111/j.1365-2443.2008.01269.x

Ogawa, T., and Okazaki, T. (1994). Cell cycle-dependent transcription from the gid and mioC promoters of *Escherichia coli*. *J. Bacteriol.* 176, 1609–1615.

Okumura, H., Yoshimura, M., Ueki, M., Oshima, T., Ogasawara, N., and Ishikawa, S. (2012). Regulation of chromosomal replication initiation by oriC-proximal DnaA-box clusters in Bacillus subtilis. *Nucleic Acids Res.* 40, 220–234. doi: 10.1093/nar/gkr716

Ozaki, S., and Katayama, T. (2012). Highly organized DnaA-oriC complexes recruit the single-stranded DNA for replication initiation. *Nucleic Acids Res.* 40, 1648–1665. doi: 10.1093/nar/gkr832

Perna, N. T., Plunkett, G. III., Burland, V., Mau, B., Glasner, J. D., Rose, D. J., et al. (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409, 529–533. doi: 10.1038/35054089

Pósfai, G., Plunkett, G. III., Fehér, T., Frisch, D., Keil, G. M., Umenhoffer, K., et al. (2006). Emergent properties of reduced-genome *Escherichia coli*. *Science* 312, 1044–1046. doi: 10.1126/science.1126439

Rang, C. U., Licht, T. R., Midtvedt, T., Conway, P. L., Chao, L., Krogfelt, K. A., et al. (1999). Estimation of growth rates of *Escherichia coli* BJ4 in streptomycin-treated and previously germfree mice by *in situ* rRNA hybridization. *Clin. Diagn. Lab. Immunol.* 6, 434–436.

Rasko, D. A., Rosovitz, M. J., Myers, G. S., Mongodin, E. F., Fricke, W. F., Gajer, P., et al. (2008). The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J. Bacteriol.* 190, 6881–6893. doi: 10.1128/JB.00619-08

Riber, L., Fujimitsu, K., Katayama, T., and Lobner-Olesen, A. (2009). Loss of Hda activity stimulates replication initiation from I-box, but not R4 mutant origins in *Escherichia coli*. *Mol. Microbiol.* 71, 107–122. doi: 10.1111/j.1365-2958.2008.06516.x

Riber, L., Olsson, J. A., Jensen, R. B., Skovgaard, O., Dasgupta, S., Marinus, M. G., et al. (2006). Hda-mediated inactivation of the DnaA protein and dnaA gene autoregulation act in concert to ensure homeostatic maintenance of the *Escherichia coli* chromosome. *Genes Dev.* 20, 2121–2134. doi: 10.1101/gad.379506

Rozgaja, T. A., Grimwade, J. E., Iqbal, M., Czerwonka, C., Vora, M., and Leonard, A. C. (2011). Two oppositely oriented arrays of low-affinity recognition sites in oriC guide progressive binding of DnaA during *Escherichia coli* pre-RC assembly. *Mol. Microbiol.* 82, 475–488. doi: 10.1111/j.1365-2958.2011.07827.x

Ryan, V. T., Grimwade, J. E., Camara, J. E., Crooke, E., and Leonard, A. C. (2004). *Escherichia coli* prereplication complex assembly is regulated by dynamic interplay among Fis, IHF and DnaA. *Mol. Microbiol.* 51, 1347–1359. doi: 10.1046/j.1365-2958.2003.03906.x

Ryan, V. T., Grimwade, J. E., Nievera, C. J., and Leonard, A. C. (2002). IHF and HU stimulate assembly of pre-replication complexes at *Escherichia coli* oriC by two different mechanisms. *Mol. Microbiol.* 46, 113–124. doi: 10.1046/j.1365-2958.2002.03129.x

Schaper, S., and Messer, W. (1995). Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target. *J. Biol. Chem.* 270, 17622–17626. doi: 10.1074/jbc.270.29.17622

Sekimizu, K., Bramhill, D., and Kornberg, A. (1987). ATP activates dnaA protein in initiating replication of plasmids bearing the origin of the *E. coli* chromosome. *Cell* 50, 259–265. doi: 10.1016/0092-8674(87)90221-2

Sherratt, D. J., Søballe, B., Barre, F. X., Filipe, S., Lau, I., Massey, T., et al. (2004). Recombination and chromosome segregation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 61–69. doi: 10.1098/rstb.2003.1365

Skarstad, K., Boye, E., and Steen, H. B. (1986). Timing of initiation of chromosome replication in individual *Escherichia coli* cells. *EMBO J.* 5, 1711–1717.

Skarstad, K., and Katayama, T. (2013). Regulating DNA replication in bacteria. *Cold Spring Harb. Perspect. Biol.* 5:a012922. doi: 10.1101/cshperspect.a012922

Smulczyk-Krawczyszyn, A., Jakimowicz, D., Ruban-Osmialowska, B., Zawilak-Pawlik, A., Majka, J., Chater, K., et al. (2006). Cluster of DnaA boxes involved in regulation of Streptomyces chromosome replication: from *in silico* to *in vivo* studies. *J. Bacteriol.* 188, 6184–6194. doi: 10.1128/JB.00528-06

Speck, C., and Messer, W. (2001). Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA. *EMBO J.* 20, 1469–1476. doi: 10.1093/emboj/20.6.1469

Speck, C., Weigel, C., and Messer, W. (1999). ATP- and ADP-dnaA protein, a molecular switch in gene regulation. *EMBO J.* 18, 6169–6176. doi: 10.1093/emboj/18.21.6169

Valens, M., Penaud, S., Rossignol, M., Cornet, F., and Boccard, F. (2004). Macrodomain organization of the *Escherichia coli* chromosome. *EMBO J.* 23, 4330–4341. doi: 10.1038/sj.emboj.7600434

Weigel, C., Messer, W., Preiss, S., Welzeck, M., Morigen, and Boye, E. (2001). The sequence requirements for a functional *Escherichia coli* replication origin are different for the chromosome and a minichromosome. *Mol. Microbiol.* 40, 498–507. doi: 10.1046/j.1365-2958.2001.02409.x

Wold, S., Crooke, E., and Skarstad, K. (1996). The *Escherichia coli* Fis protein prevents initiation of DNA replication from oriC *in vitro*. *Nucleic Acids Res.* 24, 3527–3532. doi: 10.1093/nar/24.18.3527

Yamaichi, Y., and Niki, H. (2004). migS, a cis-acting site that affects bipolar positioning of oriC on the *Escherichia coli* chromosome. *EMBO J.* 23, 221–233. doi: 10.1038/sj.emboj.7600028

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.

Copyright © 2015 Frimodt-Møller, Charbon, Krogfelt and Løbner-Olesen. 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.

frontiers in Microbiology

OPINION
published: 20 April 2015
doi: 10.3389/fmicb.2015.00324

Bacteria may have multiple replication origins

Feng Gao [1, 2, 3]*

[1] Department of Physics, Tianjin University, Tianjin, China, [2] Key Laboratory of Systems Bioengineering (Ministry of Education), Tianjin University, Tianjin, China, [3] SynBio Research Platform, Collaborative Innovation Center of Chemical Science and Engineering, Tianjin, China

Keywords: bacteria, replication origin, Z-curve, DnaA box, DNA replication, synthetic biology, bipartite origin

Introduction

Since the pioneer work of Woese and Fox (1977), it has been known that life on the Earth is generally classified into three main evolutionary lineages: Archaea, Bacteria, and Eukarya. In terms of DNA replication origin per chromosome, bacteria typically have a single replication origin (*oriC*), and eukaryotic organisms have multiple replication origins, whereas archaea are in between, see a recent review paper for the details (Leonard and Mechali, 2013). Among bacteria, one replication origin is the norm and there is currently no evidence that two functional origins are ever used on the same chromosome. However, it seems that there are always exceptions to the rules of biological systems. For example, Wang et al. have constructed *Escherichia coli* cells with two identical functional replication origins separated by 1 Mb in their 4.64-Mb chromosome artificially. Consequently, synchronous initiation at both spatially separate origins is followed by productive replication, and this is the first study in which cells with more than one WT origin on a bacterial chromosome have been extensively characterized (Wang et al., 2011). Recent developments in synthetic biology methodologies make the synthesis of synthetic chromosomes a feasible goal. Liang et al. fragmented the *E. coli* chromosome of 4.64 Mb into two linear autonomous replicating units with the *E. coli oriC* on the chromosome of 3.27 Mb and the replication origin of chromosome II in *Vibrio cholerae* on the chromosome of 1.37 Mb (Liang et al., 2013). Subsequently, Messerschmidt et al. also constructed the synthetic secondary *E. coli* chromosomes successfully based on the replication origin of chromosome II in *V. cholerae* (Messerschmidt et al., 2015). Recently, there are also a growing number of cases confirmed by experiments where the replication origin exists in a bipartite configuration in both Gram-positive and Gram-negative bacteria (Wolanski et al., 2015), such as Gram-positive *Bacillus subtilis* (Moriya et al., 1992) and Gram-negative *Helicobacter pylori* (Donczew et al., 2012). In addition, two autonomously replicating elements isolated from *Pseudomonas aeruginosa* have been characterized *in vitro* for pre-priming complex formation using combinations of replication proteins from *P. aeruginosa* and *E. coli* (Yee and Smith, 1990; Smith et al., 1991).

Then, could multiple replication origins occur on a bacterial chromosome? This open question has even been raised by Prof. Pavel Pevzner in a popular online course "Bioinformatics Algorithms" on Coursera (http://coursera.org/course/bioinformatics) recently. Based on the summarization of the diverse patterns of strand asymmetry among different taxonomic groups, Xia suggested that the single-origin replication may not be universal among some bacterial species that exhibit strand asymmetry patterns consistent with the multiple origins of replication (Xia, 2012). However, the strand asymmetry patterns were caused not only by replication-associated mutational pressure, and many phenomena, such as genome rearrangements, could influence the strand asymmetry patterns. Consequently, the local minima in the skew diagram do not always correspond to the positions of functional replication origins (Mackiewicz et al., 2004). Therefore, more evidences are needed to support multiple replication origins on a bacterial chromosome.

OPEN ACCESS

Edited by:
Frank T. Robb, University of Maryland School of Medicine, USA

Reviewed by:
Alan Leonard, Florida Institute of Technology, USA
Frédéric Boccard, Centre National de la Recherche Scientifique, France

***Correspondence:**
Feng Gao, fgao@tju.edu.cn

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *28 January 2015*
Accepted: *31 March 2015*
Published: *20 April 2015*

Citation:
Gao F (2015) Bacteria may have multiple replication origins. Front. Microbiol. 6:324. doi: 10.3389/fmicb.2015.00324

Conserved Features for Typical Bacterial Replication Origins Identified by the Z-Curve Methodology

The Z-curve is a three-dimensional curve that constitutes a unique representation of a DNA sequence, whose components represent three independent distributions that completely describe the DNA sequence being studied. The components x_n, y_n, and z_n, display the distributions of purine versus pyrimidine (R vs. Y), amino versus keto (M vs. K) and strong H-bond versus weak H-bond (S vs. W) bases along the DNA sequence, respectively. Among them, the x_n and y_n components are termed RY and MK disparity curves, respectively. The AT and GC disparity curves are defined by $(x_n + y_n)/2$ and $(x_n - y_n)/2$, which show the excess of A over T and G over C along the DNA sequence, respectively. The RY and MK disparity curves, as well as the AT and GC disparity curves, could be used to predict replication origins, since Z-curves can display the asymmetrical nucleotide distributions around *oriC*s (Zhang and Zhang, 2005; Gao, 2014). For example, the Z-curve analysis suggested the existence of multiple replication origins in archaeal genome for the first time (Zhang and Zhang, 2003), and the locations of the three predicted replication origins in *Sulfolobus solfataricus* P2 are all consistent with the results of subsequent *in vivo* studies (Lundgren et al., 2004; Robinson et al., 2004).

Based on the Z-curve method, with the means of comparative genomics, a web-based system, Ori-Finder, has been developed to identify *oriC*s in bacterial and archaeal genomes with high accuracy and reliability (Gao and Zhang, 2008; Luo et al., 2014). The predicted *oriC* regions have been organized into a database of *oriC* regions in bacterial and archaeal genomes (DoriC) (Gao and Zhang, 2007; Gao et al., 2013). Based on the predicted *oriC* regions in DoriC, conserved features for typical bacterial *oriCs* could be summarized, such as the asymmetrical nucleotide distributions around *oriC*s, the occurrence of the replication related genes adjacent to *oriC*s and the clustered DnaA boxes within *oriC*s etc. In fact, it has been noted that Ori-Finder outputs several prediction results for some bacterial chromosomes. However, only the most probable origin was presented in DoriC based on the hypothesis that bacteria only have a single replication origin, although some others also have almost all the sequence hallmarks of bacterial *oriC*s summarized above. Here, we explore the thousands of bacterial chromosomes in DoriC again, in search of multiple replication origins that comply with the above criteria on a bacterial chromosome. That is, the candidate *oriC* regions should be closely next to the replication related genes as well as the switch of Z-curves (RY, MK, AT and GC disparity curves), and contain at least three DnaA boxes. Note that only the *E. coli* perfect DnaA box (TTATCCACA) was considered with no more than one mismatch currently.

Representative Bacteria with Putative Double Replication Origins

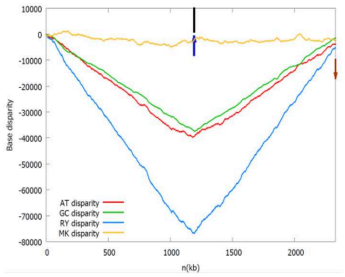
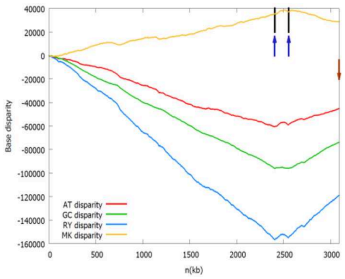
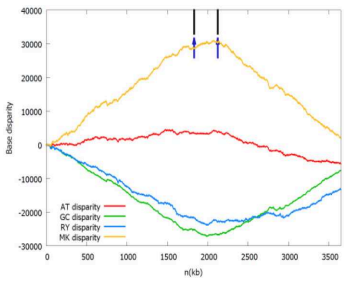
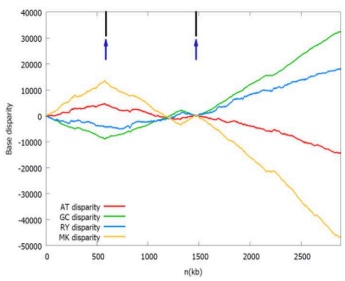
The *oriC* information of some representative bacterial chromosomes with putative double origins of replication in DoriC is listed in **Table 1**. Among them, some bacteria contain double replication origins, which are located very close to each other and exhibit bipartite configuration. For example, the *oriC* regions of *Acidaminococcus fermentans* DSM 20731 are located within the *rpmH-dnaA-dnaA-dnaN-recF-gyrB-gyrA* genes cluster, next to the *dnaA* genes encoding the chromosomal replication initiator proteins. The *oriC* region is frequently within the genes cluster *rpmH-dnaA-dnaN-recF-gyrB-gyrA* for a great number of bacteria, usually next to the *dnaA* gene. The only difference is that two *dnaA* genes are present in the genes cluster in *A. fermentans* DSM 2073, which is a unique configuration. The two identified *oriC*s are both putative bipartite origins that are composed of two sub-regions, each of which contains a cluster of DnaA boxes (Wolanski et al., 2015). Here, the bipartite origin is split into two sub-regions by the *dnaA* gene, and 13 DnaA boxes were identified in *oriC* 1 while 20 DnaA boxes were identified in *oriC* 2. The presence of the additional *dnaA* gene and *oriC* region may be due to the chromosomal duplication, which is especially typical for *Mycobacterium bovis* BCG str. Pasteur 1173P2. Two identical copies of the *rnpA-rpmH-dnaA-dnaN-recF-gyrB-gyrA* structure have been found in its *oriC* regions.

We also found *Dehalobacter* sp. CF chromosome may have two origins of replication separated by 150 kb. One is adjacent to the *dnaA* gene (*oriC* 1), and the other (*oriC* 2) is adjacent to the *parB* gene, which encodes the chromosome (plasmid) partitioning protein ParB. The *oriC* 2 is located within a putative genomic island carrying many horizontally transferred genes, such as transposase, phage integrase. Therefore, the putative *oriC* 2 may be introduced by an extrachromosomal element. These two replication origins are both located close to the local minima of the RY disparity curve as shown in the related Z-curves in **Table 1**.

In addition, on the chromosome 1, *Ralstonia pickettii* 12D and *Ochrobactrum anthropi* ATCC 49188 may have two separated origins of replication, which are adjacent to the *dnaA* gene and the *hemE* gene, respectively. The later condition is similar to the well-studied *oriC* of *Caulobacter crescentus* (Marczynski and Shapiro, 2002). The two replication origins of *R. pickettii* 12D and *O. anthropi* ATCC 49188 are separated by 291 and 882 kb, respectively. For *O. anthropi* ATCC 49188, the two replication origins are both located close to the local minima of the GC disparity curve, and are significantly more separated compared to the bipartite origins in *B. subtilis* and *H. pylori* that are usually close together.

As shown in the related Z-curves, the two putative replication origins in *A. fermentans* DSM 20731, *R. picketti* 12D and *Dehalobacter* sp. CF are located close to each other, which are around the global minima of the GC disparity curve. Therefore, the asymmetry pattern of replichores in these species is similar to that in most bacteria with single replication origin, and the asymmetric composition of the strands could be reflected by the V-shape of the Z-curves, where the minimum and maximum correspond to the origin and terminus of DNA replication. However, for *O. anthropi* ATCC 49188, the two putative replication origins are far apart, which are located at different local minima of the GC disparity curve. Consequently, the Z-curves exhibit strand asymmetry patterns consistent with the

TABLE 1 | The *oriC* information of some representative bacterial chromosomes with putative double origins of replication in DoriC.

Organism	Refseq	Location	Adjacent gene	No. of DnaA boxes[a]	Z-curves[b]
Acidaminococcus fermentans DSM 20731	NC_013740	2540..2750 nt and 4140..4386 nt (*oriC* 1) 2329473..33 nt and 1195..1915 nt (*oriC* 2)	*dnaA* (2751..4139 nt; *oriC* 1) *dnaA* (34..1194 nt; *oriC* 2)	13 (*oriC* 1) 20 (*oriC* 2)	
Dehalobacter sp. CF	NC_018867	3091418..3092048 nt (*oriC* 1) 150334..150923 nt (*oriC* 2)	*dnaA* (1..1338 nt; *oriC* 1) *parB* (149395..150333 nt; *oriC* 2)	4 (*oriC* 1) 3 (*oriC* 2)	
Ralstonia pickettii 12D chromosome 1	NC_012856	3356072..3356652 nt (*oriC* 1) 3647708..612 nt (*oriC* 2)	*hemE* (3356653..3357756 nt; *oriC* 1) *dnaA* (613..2202 nt; *oriC* 2)	3 (*oriC* 1) 3 (*oriC* 2)	
Ochrobactrum anthropi ATCC 49188 chromosome 1	NC_009667	544..1438 nt (*oriC* 1) 883808..884261 nt (*oriC* 2)	*dnaA* (1439..3001 nt; *oriC* 1) *hemE* (884262..885287 nt; *oriC* 2)	3 (*oriC* 1) 3 (*oriC* 2)	

[a] *Note that only the E. coli perfect DnaA box (TTATCCACA) was considered with no more than one mismatch.*

[b] *The Z-curves (that is, RY, MK, AT, and GC disparity curves) are plotted for the rotated sequence beginning and ending in dif site or the maximum of the GC disparity curve. Short vertical black line indicates the location of the adjacent gene listed in the table, while short up vertical dark blue arrow indicates the location of the identified oriC (note that the left arrow indicates oriC 1 and the right arrow indicates oriC 2) and short down vertical brown arrow indicates dif site location, if any. It should be noted that both the black lines and dark blue arrows in the first panel (Acidaminococcus fermentans) are located too close together to be drawn individually.*

multiple origins of replication in archaea (Zhang and Zhang, 2003).

The *in silico* analysis presented here shows that some bacteria, although very few, may have double origins of replication per bacterial chromosome. However, there is also a possibility that not both origins of replication are functional despite the finding of the evidences, such as the clustered DnaA boxes and *dnaA* gene duplications. For example, functional analysis of two autonomously replicating chromosomal replication origins from *P. aeruginosa* has shown that only one is essential for cell viability under typical laboratory growth conditions. An alternative and intriguing possibility is that the non-functional origin was once functional but no longer used as a result of structural changes (Jiang et al., 2006). This explanation may also apply to the cases

presented here, especially to the *oriC*2 of *Dehalobacter* sp. CF that may be introduced by an extrachromosomal element. Anyway, the experimental confirmation of them may provide the examples of the bacteria occurring in nature with double origins of replication and determine whether both origins of replication are functional or not, which would provide new insight into the understanding of replication mechanism of bacterial genomes and contribute to the design of synthetic bacterial genome finally.

Acknowledgments

The author would like to thank Prof. Chun-Ting Zhang for the invaluable assistance and inspiring discussions. The present work was supported in part by National Natural Science Foundation of China (Grant Nos. 31171238 and 30800642), Program for New Century Excellent Talents in University (No. NCET-12-0396), and the China National 863 High-Tech Program (2015AA020101).

References

Donczew, R., Weigel, C., Lurz, R., Zakrzewska-Czerwinska, J., and Zawilak-Pawlik, A. (2012). *Helicobacter pylori* oriC–the first bipartite origin of chromosome replication in Gram-negative bacteria. *Nucleic Acids Res.* 40, 9647–9660. doi: 10.1093/nar/gks742

Gao, F. (2014). Recent advances in the identification of replication origins based on the Z-curve method. *Curr. Genomics* 15, 104–112. doi: 10.2174/1389202915999140328162938

Gao, F., Luo, H., and Zhang, C. T. (2013). DoriC 5.0: an updated database of oriC regions in both bacterial and archaeal genomes. *Nucleic Acids Res.* 41, D90–D93. doi: 10.1093/nar/gks990

Gao, F., and Zhang, C. T. (2007). DoriC: a database of oriC regions in bacterial genomes. *Bioinformatics* 23, 1866–1867. doi: 10.1093/bioinformatics/btm255

Gao, F., and Zhang, C. T. (2008). Ori-Finder: a web-based system for finding oriCs in unannotated bacterial genomes. *BMC Bioinformatics* 9:79. doi: 10.1186/1471-2105-9-79

Jiang, Y., Yao, S., Helinski, D., and Toukdarian, A. (2006). Functional analysis of two putative chromosomal replication origins from *Pseudomonas aeruginosa*. *Plasmid* 55, 194–200. doi: 10.1016/j.plasmid.2005.11.001

Leonard, A. C., and Mechali, M. (2013). DNA replication origins. *Cold Spring Harb. Perspect. Biol.* 5:a010116. doi: 10.1101/cshperspect.a010116

Liang, X., Baek, C. H., and Katzen, F. (2013). *Escherichia coli* with two linear chromosomes. *ACS Synth. Biol.* 2, 734–740. doi: 10.1021/sb400079u

Lundgren, M., Andersson, A., Chen, L., Nilsson, P., and Bernander, R. (2004). Three replication origins in Sulfolobus species: synchronous initiation of chromosome replication and asynchronous termination. *Proc. Natl. Acad. Sci. U.S.A.* 101, 7046–7051. doi: 10.1073/pnas.0400656101

Luo, H., Zhang, C. T., and Gao, F. (2014). Ori-Finder 2, an integrated tool to predict replication origins in the archaeal genomes. *Front. Microbiol.* 5:482. doi: 10.3389/fmicb.2014.00482

Mackiewicz, P., Zakrzewska-Czerwinska, J., Zawilak, A., Dudek, M. R., and Cebrat, S. (2004). Where does bacterial replication start? Rules for predicting the oriC region. *Nucleic Acids Res.* 32, 3781–3791. doi: 10.1093/nar/gkh699

Marczynski, G. T., and Shapiro, L. (2002). Control of chromosome replication in caulobacter crescentus. *Annu. Rev. Microbiol.* 56, 625–656. doi: 10.1146/annurev.micro.56.012302.161103

Messerschmidt, S. J., Kemter, F. S., Schindler, D., and Waldminghaus, T. (2015). Synthetic secondary chromosomes in *Escherichia coli* based on the replication origin of chromosome II in Vibrio cholerae. *Biotechnol. J.* 10, 302–314. doi: 10.1002/biot.201400031

Moriya, S., Atlung, T., Hansen, F. G., Yoshikawa, H., and Ogasawara, N. (1992). Cloning of an autonomously replicating sequence (ars) from the *Bacillus subtilis* chromosome. *Mol. Microbiol.* 6, 309–315. doi: 10.1111/j.1365-2958.1992.tb01473.x

Robinson, N. P., Dionne, I., Lundgren, M., Marsh, V. L., Bernander, R., and Bell, S. D. (2004). Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell* 116, 25–38. doi: 10.1016/S0092-8674(03)01034-1

Smith, D. W., Yee, T. W., Baird, C., and Krishnapillai, V. (1991). Pseudomonad replication origins: a paradigm for bacterial origins? *Mol. Microbiol.* 5, 2581–2587. doi: 10.1111/j.1365-2958.1991.tb01966.x

Wang, X., Lesterlin, C., Reyes-Lamothe, R., Ball, G., and Sherratt, D. J. (2011). Replication and segregation of an *Escherichia coli* chromosome with two replication origins. *Proc. Natl. Acad. Sci. U.S.A.* 108, E243–E250. doi: 10.1073/pnas.1100874108

Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5088–5090. doi: 10.1073/pnas.74.11.5088

Wolanski, M., Donczew, R., Zawilak-Pawlik, A., and Zakrzewska-Czerwinska, J. (2015). oriC-encoded instructions for the initiation of bacterial chromosome replication. *Front. Microbiol.* 5:735. doi: 10.3389/fmicb.2014.00735

Xia, X. (2012). DNA replication and strand asymmetry in prokaryotic and mitochondrial genomes. *Curr. Genomics* 13, 16–27. doi: 10.2174/138920212799034776

Yee, T. W., and Smith, D. W. (1990). Pseudomonas chromosomal replication origins: a bacterial class distinct from *Escherichia coli*-type origins. *Proc. Natl. Acad. Sci. U.S.A.* 87, 1278–1282. doi: 10.1073/pnas.87.4.1278

Zhang, R., and Zhang, C. T. (2003). Multiple replication origins of the archaeon Halobacterium species NRC-1. *Biochem. Biophys. Res. Commun.* 302, 728–734. doi: 10.1016/S0006-291X(03)00252-3

Zhang, R., and Zhang, C. T. (2005). Identification of replication origins in archaeal genomes based on the Z-curve method. *Archaea* 1, 335–346. doi: 10.1155/2005/509646

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 Gao. 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.

frontiers in Microbiology

REVIEW
published: 30 September 2015
doi: 10.3389/fmicb.2015.01049

Choosing a suitable method for the identification of replication origins in microbial genomes

*Chengcheng Song[1,2,3], Shaocun Zhang[1,2,3] and He Huang[1,2,3]**

[1] Department of Biochemical Engineering, School of Chemical Engineering and Technology, Tianjin University, Tianjin, China, [2] Key Laboratory of Systems Bioengineering, Ministry of Education, Tianjin University, Tianjin, China, [3] Collaborative Innovation Center of Chemical Science and Engineering, Tianjin, China

As the replication of genomic DNA is arguably the most important task performed by a cell and given that it is controlled at the initiation stage, the events that occur at the replication origin play a central role in the cell cycle. Making sense of DNA replication origins is important for improving our capacity to study cellular processes and functions in the regulation of gene expression, genome integrity in much finer detail. Thus, clearly comprehending the positions and sequences of replication origins which are fundamental to chromosome organization and duplication is the first priority of all. In view of such important roles of replication origins, tremendous work has been aimed at identifying and testing the specificity of replication origins. A number of computational tools based on various skew types have been developed to predict replication origins. Using various *in silico* approaches such as Ori-Finder, and databases such as DoriC, researchers have predicted the locations of replication origins sites for thousands of bacterial chromosomes and archaeal genomes. Based on the predicted results, we should choose an effective method for identifying and confirming the interactions at origins of replication. Here we describe the main existing experimental methods that aimed to determine the replication origin regions and list some of the many the practical applications of these methods.

Keywords: replication origin, EMSA, Dnase I footprinting, SPR, RIP mapping, ITC, ChIP, ChIP-seq

OPEN ACCESS

Edited by:
Frank T. Robb,
University of Maryland, USA

Reviewed by:
Yoshizumi Ishino,
Kyushu University, Japan
Andrew F. Gardner,
New England Biolabs, USA

****Correspondence:***
He Huang,
Department of Biochemical Engineering, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China
huang@tju.edu.cn

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *03 April 2015*
Accepted: *14 September 2015*
Published: *30 September 2015*

Citation:
Song C, Zhang S and Huang H (2015) Choosing a suitable method for the identification of replication origins in microbial genomes. Front. Microbiol. 6:1049. doi: 10.3389/fmicb.2015.01049

Introduction

Genome duplication is essential for cellular life. Since the determination of complete genome sequences of many species, attention has been given to the understanding of DNA replication. There are important differences among bacteria, archaea, and eukaryotes in the process of DNA replication, but they all have the same core components of replication machines: DNA polymerases, circular sliding clamps, a pentameric clamp loader, helicase, primase, and single-strand binding protein (SSB) (Waga and Stillman, 1998; Garg and Burgers, 2005; Johnson and O'Donnell, 2005; Barry and Bell, 2006). The number of replication origins varies in terms of different evolutionary lineages (Aves, 2009). In bacteria, a single DNA replication origin is sufficient enough to ensure complete and opportune replication of the entire genome precisely once in each cell cycle. In the case of *Escherichia coli*, bacteria often contain only a single replication origin in one chromosome although not all bacteria follow this paradigm (**Figure 1**). Similarly, in archaea, single replication origins have been found in *Pyrococcus* and *Archaeoglobus* (Myllykallio et al., 2000;

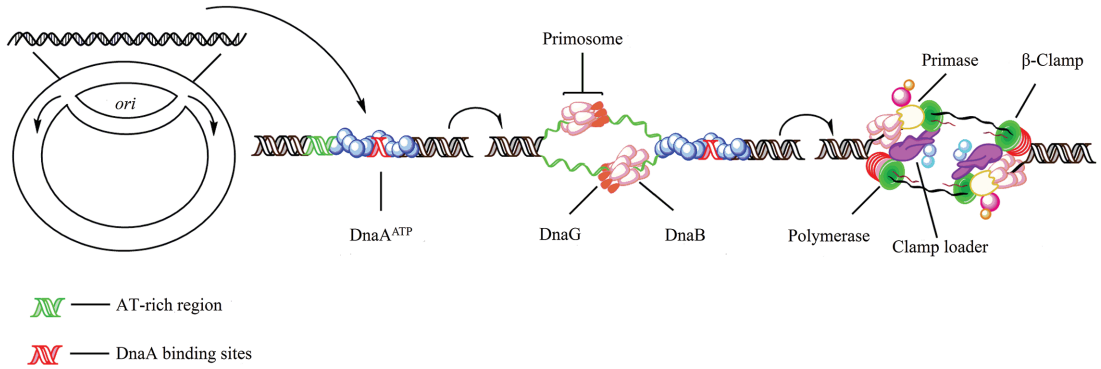


FIGURE 1 | Initiation of replication in bacteria. In *Escherichia coli*, replication initiation requires binding of the DNA-binding protein DnaA to DnaA-boxes at the chromosome origin *oriC* which is regulated by SeqA (Dame et al., 2011). Then, with the activation of ATP, two DnaB hexamers and the helicase loader DnaC, one double hexamer for each replication direction, are positioned by DnaA into the loop (Wahle et al., 1989; Skarstad and Katayama, 2013). Primase (DnaG) which can enter the complex and synthesize two leading strand primers, stimulates release of the regulatory protein DnaC from DnaB after transiently binding to the DnaB replicative helicase (Arias-Palomo et al., 2013). Also, DnaB binds to the sliding clamp loader, a ring-shaped dimer of the β-subunit which in turn binds the DNA polymerases III (Kelman and O'Donnell, 1995; O'Donnell et al., 2013).

Maisnier-Patin et al., 2002), two have been found in *Aeropyrum* (Robinson and Bell, 2007), three in *Sulfolobales*, four replication origins in the archaeon *Pyrobaculum calidifontis* (Pelve et al., 2012) and even multiple replication origins have been suggested in other genera, including *Methanocaldococcus* (Maisnier-Patin et al., 2002), *Halobacterium* (Berquist and DasSarma, 2003; Zhang and Zhang, 2003; Coker et al., 2009), and *Haloferax* (Norais et al., 2007). This illustrates how the events that occur at the DNA replication origins are predominant in the processes of DNA replication (Baker and Bell, 1998; **Figure 2**).

Initiator proteins were first proposed as the essential trans-acting factors for the initiation of DNA replication by Jacob et al. (1963). The initiator protein DnaA is the prerequisite protein in the process of prokaryotes DNA replication, and it plays an important role in forming an optimal initiation complex for DNA strand opening at the origin (Ozaki and Katayama, 2009). Among bacteria, the initiation of replication is best understood in *E. coli*. All functions of bacterial DnaA protein depend on its ability to bind specifically to an asymmetric 9-bp recognition sequence, the typical DnaA box: 5′-TTATNCACA-3′. The interactions DnaA binding to 9-mer DnaA boxes of the *oriC* is a high-affinity interaction ($KD = 1$ nM) (Speck and Messer, 2001). The sequence of *oriC* usually consists of an array of several DnaA boxes and AT-rich regions. About 10–20 DnaA molecules form a homomultimeric initiation complex on the chromosomal replication origin, *oriC*. DnaA (52 kDa) consists of four functional domains, I, II, III, IV (Messer, 2002). The ssDNA-binding activity of DnaA domain I is weak (Abe et al., 2007), but

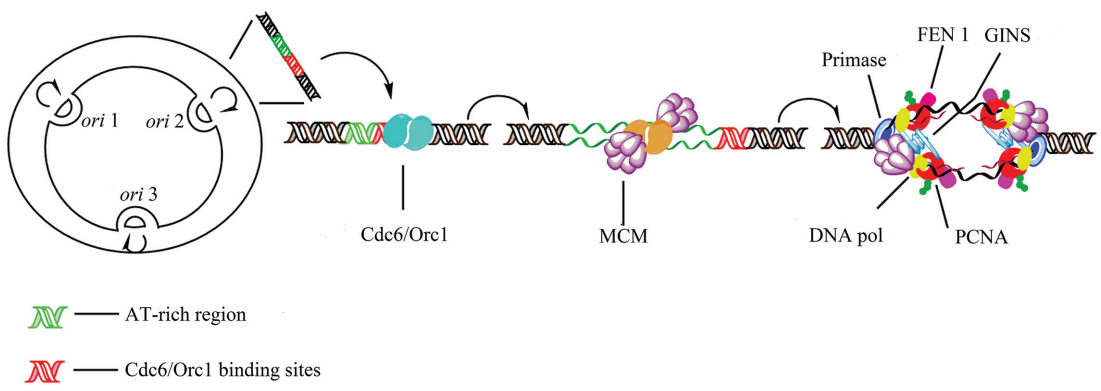


FIGURE 2 | Initiation of replication in archaea. Archaeal circular chromosome can contain a single or multiple origins (*oriC*). Archaea have the AAA+ Orc1/Cdc6 origin-binding protein, which shares extensive sequence homology with eukaryotic ORC proteins (Zhang et al., 2009). Minichromosome maintenance (MCM) proteins bind preferentially to the *oriC* region. ATP bound Cdc6/Orc1 associates with the Cdc6/Orc1-origin complex and with the MCM helicase. Following ATP hydrolysis the Cdc6/Orc1 protein releases the helicase, and the primase replaces the Cdc6/Orc1 protein binding to MCM. MCM interacts with the archaeal GINS (*go*, *i*chi, *ni*i, *san* [five, one, two, three in Japanese]) complex (Marinsek et al., 2006) which is additionally capable of binding primase. Each DNA Pol interacts with a trimer of PCNA (proliferating cell nuclear antigen). The flap endonuclease FEN1 and DNA ligase I are only assembled to PCNA clamp of similar structure to *E. coli* β (Michel and Bernander, 2014).

the interactions between domain I and several proteins, including domain I itself, DnaB helicase, and the initiation stimulator DiaA, are required for DnaB helicase loading onto *oriC* open complexes flexible linker (Weigel et al., 1999; Felczak and Kaguni, 2004; Abe et al., 2007; Keyamura et al., 2007; Nozaki and Ogawa, 2008). Domain III plays a major role in ATP and ADP binding, in ATP-dependent conformational changes of the DnaA multimer on *oriC*, in binding ssDNA of the *oriC* duplex unwinding element (DUE), and in ATP hydrolysis (Katayama, 2008; Ozaki et al., 2008). The C-terminal domain IV (∼10 kDa) has a typical helix–turn–helix fold that binds to DnaA box (Fujikawa et al., 2003). Domain IV Arg399 recognizes three more base pairs (5′ one-third of the DnaA box sequence: TTA) by base-specific hydrogen bonds in the minor groove of DNA (Fujikawa et al., 2003). Mostly, the C-terminal DnaA (IV) that was fused to a tag such as His_6 or GST in the C-terminus or N-terminus is necessary and sufficient for specific DNA binding (Richter and Messer, 1995; Roth and Messer, 1995; Sutton and Kaguni, 1997; Blaesing et al., 2000). DnaA binds to high- or low-affinity sites of origin and forms an oligomeric structure (Kawakami and Katayama, 2010) that involves two types of DnaA–DNA interactions, a double-stranded and a single-stranded DNA (Speck and Messer, 2001; Ozaki and Katayama, 2012). Furthermore, the DnaA protein is not only an initiator that binds to the specific site *oriC* but it is also a gene regulatory protein. There are about 300 high-affinity DnaA binding sites and a very large number of low-affinity sites around the chromosome (Kitagawa et al., 1996; Roth and Messer, 1998). Also, replication of microbial chromosome(s) occurs via the concerted action of many other origin binding proteins (oriBPs) which are cooperative with bacterial DnaA. The oriBPs includes factor for inversion stimulation (Fis), integration host factor (IHF), sequestration factor A (SeqA), aerobic respiration control (ArcA), inhibitor of chromosomal initiation (*lciA*) and that which binds specifically site(s) to *oriC* (Wolański et al., 2014). As reports have shown, only tens of origin regions of eubacteria and archaea have been confirmed experimentally (Myllykallio et al., 2000; Maisnier-Patin et al., 2002; Berquist and DasSarma, 2003; Matsunaga et al., 2003; Lundgren et al., 2004; Robinson et al., 2004; Norais et al., 2007; Coker et al., 2009).

A number of computational tools based on various skew types have been developed for predicting replication origins. Chromosome replication origins were mapped *in vivo* in the two hyperthermophilic archaea of *Sulfolobus acidocaldarius* (Duggin et al., 2008) and *Sulfolobus solfataricus* (Lundgren et al., 2004; Robinson et al., 2004), as well as in *Haloarcula hispanica* (Wu et al., 2013), *haloarchaeon Halobacterium* sp. NRC-1 model (Coker et al., 2009), *Pyrobaculum calidifontis* (Pelve et al., 2012), *Nitrosopumilus maritimus* (Pelve et al., 2013), and *Haloferax mediterranei* (Pelve et al., 2013), using high-throughput sequencing-based marker frequency (MF) analysis. MF analysis has been successfully used in combination with microarrays to study replication characteristics and to map chromosome replication origins in both bacteria (Khodursky et al., 2000) and eukaryotes (Raghuraman et al., 2001). Recently, the Web-based system Ori-Finder[1] and Ori-Finder 2[2] which utilize the Z-curve method and comparative genomics analysis were used to find *oriC*s in bacterial and archaeal genomes, respectively with high accuracy (Zhang and Zhang, 2005; Gao and Zhang, 2008; Gao, 2014; Luo et al., 2014). Ori-Finder 2 is also able to analyze the unannotated genome sequences by integrating them with gene prediction pipelines and BLAST software for gene identification and function annotation. The predicted *oriC* regions from Ori-Finder have been organized into an online database DoriC[3], which contains *oriC*s for >2000 bacterial genomes and 100 archaeal genomes, respectively (Gao and Zhang, 2007; Gao et al., 2012, 2013). Based on the predicted results, we can identify and confirm the *oriC* by its interaction with the initiator protein DnaA, and by its ability to form higher-order structures with DnaA that can be seen in the electron microscope.

Over the past several years, the rapid development of techniques used for confirming protein–DNA interaction *in vivo* and *in vitro,* such as gel retardation assay, surface plasmon resonance (SPR), electrophoretic mobility shift assays (EMSA), the DNase I footprinting technique, replication initiation point (RIP) mapping, isothermal titration calorimetry (ITC), chromatin immunoprecipitation (ChIP), and ChIP sequencing (ChIP-seq) have resulted in an increasingly refined picture of the biochemical rules governing protein–DNA interactions. Protein-DNA interactions can be explored by various *in vitro* and *in vivo* strategies, which present different advantages and disadvantages. This review begins with a discussion of the main existing experimental methods that are applied to verify protein–DNA interactions *in vivo* and *in vitro*, as well as explore some functional components of the complexes, especially applied in detecting transcription factor binding sites. Then, we outline the main advantages and limitations of these methods in **Table 1**. Through the listed methods, we could choose the most suitable experimental strategy for identifying replication origins.

Conventional Methods for Detecting Protein–DNA Interaction at Origins of Replication *In Vitro*

Electrophoretic Mobility Shift Assay

The EMSA, also known as the band shift, gel shift, or gel retardation assay (Lane et al., 1992), is one of the most sensitive and straightforward methods to determine the binding site-size of the DNA binding protein using a series of DNA polymers even when the protein is at a low concentration within the extract (Carey et al., 2012). It is based on the principle that DNA/RNA–protein complexes migrate more slowly when subjected to non-denaturing polyacrylamide or agarose gel electrophoresis as compared to unbound free probe (**Figure 3**). The DNA probes used may be radiolabeled or dyes specific to stain DNA and protein may be used to visualize the DNA/RNA–protein interaction. In general poly (dI-dC) is added to abolish any non-specific binding. Polyacrylamide gels offer better electrophoretic

[1]http://tubic.tju.edu.cn/Ori-Finder/

[2]http://tubic.tju.edu.cn/Ori-Finder2/

[3]http://tubic.tju.edu.cn/doric/

TABLE 1 | Summary of main experimental methods.

Methods	Benefits	Drawbacks	Extends
Electrophoretic mobility shift assay (EMSA)	(1) Qualitatively and quantitatively. (2) Effectively when protein at low concentration. (3) High sensitivity (capable to resolve complexes of different protein or DNA by stoichiometry, even to detect conformational changes). (4) Particularly useful for analyzing protein–DNA interactions on a small fragment (20–30 bp) and for determining cell types that contain each binding activity.	(1) Protein–DNA interaction must survive in gelelectrophoresis. (2) Disrupting specific interactions due to large amount of non-specific competitor. (3) Intact complexes to be highly stabled.	(1) MC-EMSA; (2) Supershift EMSA (SS-EMSA). (3) The two-color electrophoretic mobility shift assay. (4) The microfluidic mobility shift assays (MMSAs). (5) Reverse EMSA and the antibody supershift assay. (6) EMSA followed by SDS-PAGE with Western blot detection or followed by 2DE and MS were developed to identify the uncertain binding proteins. (7) Several classes of quantitative affinity-based microfluidic EMSAs including immunoassays (IAs), affinity EMSAs, and dragtag-based EMSAs. (8) EMSA linked with nanoparticle–aptamer conjugates (NP-EMSA)
Dnase I footprinting	(1) Quantitative. (2) Effective primarily when protein is at high concentration. (3) Identify DNA-binding proteins in crude extracts qualitatively. (4) Resolving the microscopic binding affinities containing multiple-binding sites with which DNA-binding proteins interact cooperatively.	(1) Binding must be stable only in solution. (2) Disrupting specific interactions by $MgCl_2$ and $CaCl_2$ which are required for DNase I activity.	(1) A method by using 96-well plates and capillary electrophoresis for high-throughput analysis of protein-binding sites in DNA. (2) GeF-seq (genome footprinting by high-throughput sequencing) combines *in vivo* DNase I digestion of genomic DNA with ChIP coupled with high-throughput sequencing.
Surface plasmon resonance (SPR)	(1) Label-free; (2) Versatility; (3) High sensitivity; (4) No purified sample. (5) No background interference. (6) Biological sample without tags. (7) Real-time monitoring dynamic response process. (8) Particularly useful for scanning a large DNA fragment (50–200 bp). (9) Flexibility (SPR sensors can perform continuous monitoring as well as one-time analyses)	(1) High cost. (2) Limitations in specificity of detection (biomolecular recognition elements may exhibit cross-sensitivity to structurally similar but non-target molecules) and sensitivity to interfering effects (includes adsorption of non-target molecules by the sensor surface and sample temperature and composition fluctuations).	(1) SPR-CELLIA system. (2) An automated system to analyze protein complexes by integrating a SPR biosensor with highly sensitive nanoflow liquid chromatography–tandem mass spectrometry (LC–MS/MS). (3) Localized-SPR (LSPR.) (4) A label-free method to immobilize basic proteins onto the C1 chip for SPR assay at physiological pH. (5) SPR effect of CD-Ag nanoparticles. (6) Spectroscopic SPR and imaging SPR.
Replication initiation point (RIP) mapping	(1) Faster technique and less laborious. (2) More reliable and sensible to compare between a wild-type protein and its mutants. (3) Observing and measuring the association and dissociation of the complex, and Gibbs free energy difference. (4) 1000-fold more sensitive and effective to separate the nascent DNA and nicked contaminating DNA by selective degradation of 5′ DNA by λ-exonuclease prior to primer extension.	(1) Only mapping the initiation point of eukaryotic and some archaea which have short eukaryotic or eukaryotic-like Okazaki fragments	

(Continued)

TABLE 1 | Continued

Methods	Benefits	Drawbacks	Extends
Isothermal titration calorimetry (ITC)	(1) Label-free; (2) Highly sensitive; (3) In solution; (4) The only method capable of determining the enthalpy, entropy, and the Gibbs free energy of a reaction in a single titration experiment.	(1) Modeling is limited to systems obeying simple Michaelis–Menton kinetics.	(1) In conjunction with complementary techniques such as X-ray crystallography, NMR spectroscopy, small angle X-ray scattering (SAXS), circular dichroism spectroscopy (CD), intrinsic fluorescence and immunoisolations; (2) the Omega ITC, MCS ITC, VP-ITC, Auto-ITC, Nano ITC-III, and ITC200 instruments
Chromatin immunoprecipitation (ChIP)	(1) Real-time monitoring dynamic response process; (2) The only way which allows one to determine the entire spectrum of DNA binding sites for any given protein *in vivo* with whole-genome DNA microarrays.	(1). Difficulty to find the specific antibody that the assay needs. (2) Necessarily to demonstrate the expression of target protein under the intracellular simulation environment.	(1) ChIP-Seq; (2) Re-ChIP; (3) ChIP-chip which combines ChIP and microarray technology. (4) ChIPon tiled arrays (ChIPOTIe). (5) Insertional chromatin immunoprecipitation (iChIP). (6) Engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP).

FIGURE 3 | The schematic illustration of electrophoretic mobility shift assays (EMSA). Typically one compound is labeled to follow its mobility during electrophoresis. In general, a single protein binds to a single site. Once the length of the nucleotides is sufficient for the binding of two or more proteins, the protein–DNA complexes migrate as distinct bands, usually referred to as a super shift. If the labeled nucleotides are bound by the proteins, then the mobility of the labeled nucleotides through the electrophoretic medium will be retarded.

resolution for protein–DNA and protein–RNA complexes of Mr $\leq$ 500,000 than agarose gel (Fried, 1989). Experimental procedures, announcements and guides for troubleshooting the most common problems that we have encountered were described detailedly by Hellman and Fried (2007) and Carey et al. (2013b).

The preponderances of EMSA account in large part for the application of a wide range of conditions and the continuing popularity of the assay. This assay can be applied to a wide range in size and structure of nucleic acids and proteins binding. Lengths from short oligonucleotides to several 1000 nt/bp of single-stranded, duplex, triplex, and quadruplex nucleic acids as well as small circular DNAs, and proteins size from small oligopeptides to transcription complexes with $M_r \geq 10^6$, all of these conditions are applicable in EMSA (Hellman and Fried, 2007; Alves and Cunha, 2012). EMSA also works well with both highly purified proteins and uncharacterized binding activities present in crude protein extracts (Memelink, 2013). Low concentrations (0.1 nM or less) and small sample volumes (20 μL or less) (Hellman and Fried, 2007) are performed by EMSA due to using radioisotopes to label nucleic acids and autoradiography.

Variants or the assay using fluorescence, chemiluminescence, and immunohistochemical detection are also available though less sensitive than radioisotopes.

Since its first publication in 1981, several improvements and variant techniques of EMSA have been developed. Reverse EMSA (rEMSA) and the antibody supershift assay were applied for identifying DNA–protein interactions (Tsai et al., 2012). EMSA followed by SDS-polyacrylamide gel electrophoresis with Western blot detection (Granger-Schnarr et al., 1988; Chen and Chang, 2001) or followed by two-dimensional electrophoresis (2DE) and mass spectrometry (MS) (Woo et al., 2002; Stead et al., 2006) were improved to identify the uncertain binding proteins. The supershift EMSA (SS-EMSA) can identify proteins that carrying a specific epitope in mobility-shifted complex(es) and validate previously identified proteins. Supershift EMSAs suggested the presence of transformation-specific DNA replication complexes in transformed human cells (Di Paola et al., 2010). MC-EMSA is a competition-based method developed by Smith and Humphries (2009) to identify unknown DNA binding proteins incubated with a pool of unlabeled DNA consensus competitors prior to adding the labeled DNA probe. A sensitive two-color EMSA was developed by Jing et al. (2003) for detecting both nucleic acids and protein that either free or bound conditions in gels. This assay is fast, simple, and needless the use of radioisotopes. The microfluidic mobility shift assays (MMSAs) as quantitative EMSA utilize affinity molecular probes (target) to induce a change in analyte molecule size and/or charge (Fourtounis et al., 2011; Karns et al., 2013). Several classes of quantitative affinity-based microfluidic EMSAs including immunoassays (IAs), affinity EMSAs, dragtag-based EMSAs, and other were elaborated by Pan et al. (2014). A separation technique for DNA–protein complex which called microchip electrophoretic mobility shift assay (μEMSA), based on EMSA by microchip electrophoresis was developed by Inoue et al. (2011). The performance of EMSA linked with nanoparticle–aptamer conjugates (NP-EMSA) was improved over the traditional EMSA (Wang and Reed, 2012). The most striking advantages of NP-EMSA as described in this research are real-time detection of protein–oligonucleotide interactions, the avoidance of harmful radioisotopes, and elimination of the need for expensive gel imagers.

Electrophoretic mobility shift assays is by far the most frequently used for detecting *oriC*-DnaA or -oriBPs complexes, ARS–ORC complexes, largely because it is technically the easiest and is often the most sensitive. The proteins which required in EMSA could be obtained from either purified proteins or crude extracts of cells. And the length of target DNA used in EMSA is best less than 300 bp. So, the electrophoresis separation effect of probe and protein–DNA complexes will be more obvious. Particularly, EMSA is useful for analyzing protein-DNA interactions on a small fragment (20–30 bp). So, EMSA could be used for identifying the interactions between oriBPs and *oriC*s, as well as the interactions between oriBPs and single or multiple DnaA boxes (Schaper et al., 2000; Zawilak et al., 2003; Robinson et al., 2004; Pei et al., 2007). For instance, by EMSA, the DnaA of *Thermoanaerobacter tengcongensis* was detected that it could achieve the efficient binding at a lower protein concentration (8 nM) when the DNA fragment containing two DnaA boxes with 3-bp spacing at 60°C, and the domain IV of DnaA is thermo-adaptive (Pei et al., 2007). All most the published papers for identifying origins of replication applied EMSA as the basic strategy as well as a standard to determine whether to do the following experiments.

DNase I Footprinting

The second most common assay is DNase I footprinting, although its use is rapidly declining. The limitation of this method is that it doesn't provide identity of the protein and requires higher concentration protein than EMSA (Leblanc and Moss, 2001). Even so, this method provides a myriad of applications both in determining the site of interaction of most sequence-specific DNA binding proteins and characterizing the binding interactions. The protein–DNA complexes are separated from free (unbound) DNA relies on a change that the protein prevents binding of DNase I in and around its binding site and thus generates a "footprint" in the cleavage ladder in denaturing acrylamide gel (shown in **Figure 4**). The distance from the end label to the edges of the footprint represents the position of the protein-binding site on the DNA fragment. In addition of DNase I, the enzymatic digestion methods also include the use of MNase (Fox and Waring, 1987), methidiumpropyl-EDTA_Fe(II) (MPE) (Van Dyke and Dervan, 1983), copper phenanthroline, uranyl photocleavage, hydroxyl radicals, DMS, and iron complexes (Dey et al., 2012). The classic experimental procedure, recipes, and consideration were detailed by Carey et al. (2013a). Recently, DNase I footprinting assay with fluorescent 6-carboxyfluorescein (FAM)-labeled probes was widely used for identifying the correct nucleotides regions that proteins protected (Zianni et al., 2006). The use of FAM-labeled primers eliminates the need for radioactively labeled nucleotides, slab gel electrophoresis, as well as commonly available automated fluorescent capillary electrophoresis instruments. The result of Thermo Sequenase outputted by Genemapper software was accurately aligned with DNase I digestion products, providing a ready means to assign correct nucleotides to each peak from the DNA footprint. Genome Footprinting by high-throughput sequencing (GeF-seq) was proved powerful to elucidate the molecular mechanism of target protein binding to its cognate DNA sequences (Chumsakul et al., 2013). In this research, GeF-seq combines *in vivo* DNase I digestion of genomic DNA with ChIP coupled with high-throughput sequencing.

Different with EMSA, DNase I footprinting is useful for scanning a large DNA fragment (50–200 bp) for DNA–protein interaction. Mostly, DNase I footprint assay was used for initially identifying the location and number of DnaA boxes from the whole region of *oriC* after EMSA. Through high-throughput analysis, the sequences of DnaA boxes could be confirmed and analyzed. DNase I footprinting widely applied in identification of *oriC*s in bacteria and archaea. The two *oriC*s of *S. solfataricus* have been identified before, DNase I footprinting assay has been fully used in the study (Robinson et al., 2004). Through DNase I footprinting, the precise sequences and locations of ORBs (origin recognition boxes) in *oriC*1 and *oriC*2 of *S. Solfataricus* which bind to three Orc1/Cdc6s have been directly

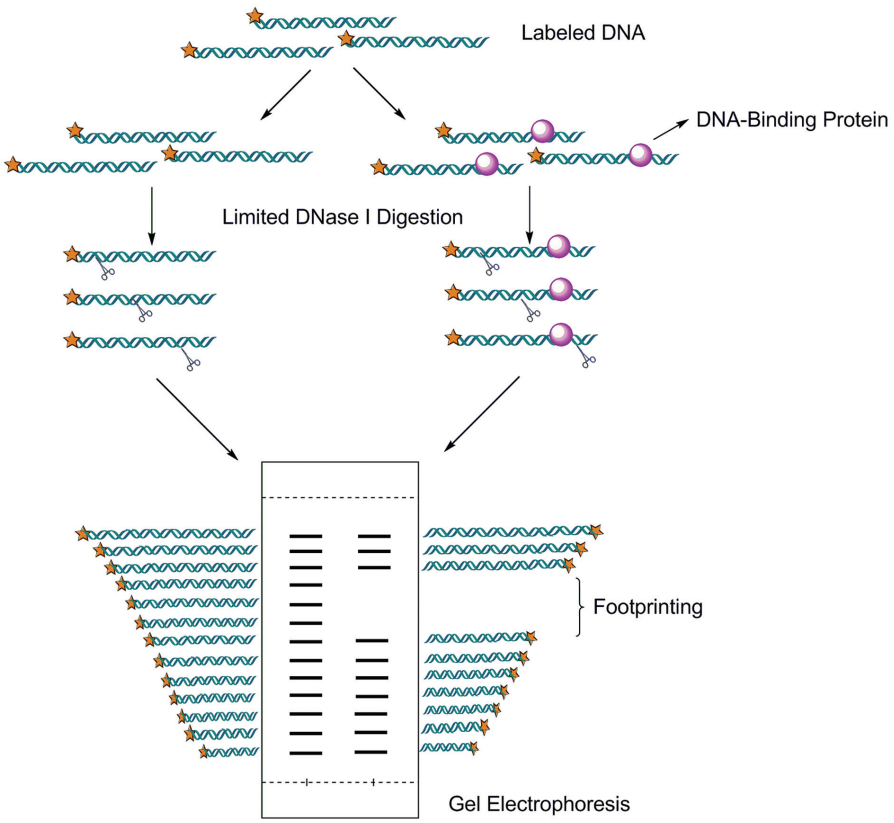


FIGURE 4 | Dnase I footprinting. This Analysis involves endonuclease treatment of an end labeled DNA fragment bound to a protein. This technique relies on the fact that fragments of DNA that have DNA-binding proteins bound will move more slowly through an acrlyamide gel. The enzyme DNaseI will only cut exposed DNA. Limited digestion yields fragments terminating everywhere except in the footprint region, which is protected from digestion.

identified, respectively. DNase I footprinting was also used for the identification *oriC*s of *E. coli* (Fuller et al., 1984), *Pyrococcus furiosus* (Robinson et al., 2004), and *Caulobacter crescentus* (Taylor et al., 2011). So, DNase I footprinting is one of the most useful method for identifying replication origins in microbial genomes.

Surface Plasmon Resonance

Since the SPR (surface plasmon resonance) technology was first used in chemical sensors, SPR sensors have gradually become an emerging alternative to the conventional *in vitro* techniques to study DNA–protein interactions, due to its label-free, high-sensitivity, real-time analysis, and flexible system design (Liedberg et al., 1995; Homola et al., 1999; Ladd et al., 2009). **Figure 5** depicts the basic principle and schematic illustration of SPR system. Compared to other methods studying protein interaction, such as direct protein interaction *in vitro* and co-immunoprecipitation, SPR is a more sensitive and quantitative biophysical approach that can measure binding affinity and kinetics simultaneously (Hoa et al., 2007). Furthermore, this technique is the basis of many lab-on-a-chip and biosensor applications. According to recent research, SPR technology can be particularly used to study the interactions between nucleic acids or protein-nucleic acids by real-time tracking of the nucleic acid reaction process. This application of SPR is unmatched by other techniques (Pattnaik, 2005; Sahai, 2011). The stoichiometry and kinetics of complex formation between DnaA protein and *oriC* could be analyzed using SPR experiments.

Surface plasmon resonance technique is an optical method for measuring the refractive index of very thin layers of material adsorbed on a metal. Its development will further extend the potential of SPR-sensing technology and allow SPR sensors to be used far more widely. Spectroscopic SPR and imaging SPR have been further adapted as affinity detection techniques in the proteomic and genomic fields, especially in a protein conformation study (Despeyroux et al., 2000), biomarker profiling, aptamer selections (Murphy et al., 2003), and antibody selections (Wilson and Howell, 2002). SPR-CELLIA system was configured for either whole cells or macromolecules in two parallel flow paths (Baird and Myszka, 2001). Applied Biosystems has also launched Affinity Sensor instrument based on SPR technology (Pattnaik, 2005). An automated system which developed for analyzing protein complexes by coupling a polymerization initiator to a biospecific interaction and inducing inline atom transfer radical polymerization (ATRP) was developed with highly sensitive nanoflow liquid chromatography-tandem mass spectrometry (LC–MS/MS) (Liu et al., 2010). Nanomaterials developed for localized surface

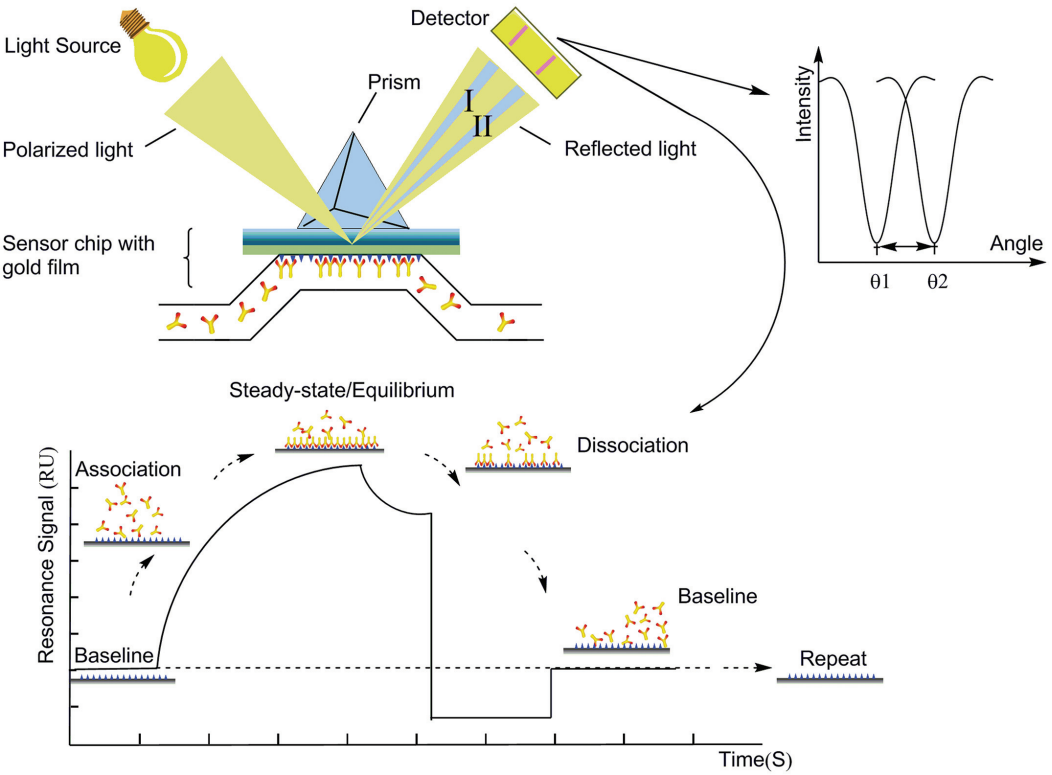


FIGURE 5 | The schematic illustration of surface plasmon resonance (SPR) system. SPR detects changes in the refractive index in the immediate vicinity of the surface layer of a sensor chip. The sensor surface is gold with antibodies attached to it. During the measurement, the chip is irradiated from the bottom with a beam of a wide angle range within that of total internal reflection. The SPR angle shifts (from I to II in the diagram) when biomolecules binding events cause changes in the refractive index at the surface layer. The detector will determine the angle of the intensity decrease. This change in resonant angle can be monitored non-invasively in real time as a plot of resonance signal (proportional to mass change) versus time (Sawhney and Singh, 2000; Cooper, 2002; Pattnaik, 2005; Sahai, 2011; Hou and Cronin, 2013).

plasmon resonance (LSPR) are increasingly integrated to classical prism-based SPR sensors, providing enhanced sensitivity and lower detection limits (Bolduc and Masson, 2011). Khan et al. developed a label-free method to immobilize basic proteins onto the C1 chip for SPR assay at physiological pH, which presents ligand with less conformational modification and thereby maintains the ligand at optimal biological activity (Khan et al., 2012). Besides, some materials have been proposed to improve the performance of SPR biosensors, such as gold nanoparticles, magnetic nanoparticles (MNP), carbon nanotubes, electropolymerized molecularly imprinted polythiophenes (Lyon et al., 1998; Wang, 2005; Parab et al., 2010; Pernites et al., 2010; Špringer et al., 2014).

Surface plasmon resonance-based biosensing is one of the most advanced label free, real time detection technologies. But, one of the main drawbacks that stem further development of SPR applications is the lack of sufficient sensitivity to reliably detect small changes in refractive index caused by compounds with low molecular weight or in low concentration at the sensing surface (Wang, 2005). So, several approaches have been reported to resolve such limitations. A modified SPR device achieved that the plasmonic detected single molecules in real time without the need for labeling or amplification by using a gold nanorod. And, the sensitivity of this device is ∼700 times higher than state-of-the-art plasmon sensors (Zijlstra et al., 2012). A new approach to SPR biosensors for rapid and highly sensitive detection of bacterial pathogens is based on the spectroscopy of grating-coupled long-range surface plasmons (LRSPs) combined with MNP assay (Wang et al., 2012). A highly efficient SPR immunosensor was effectively enhanced the sensitivity by using a non-covalently functionalized single graphene layer on a thin gold film (Singh et al., 2015).

DNA fragments were immobilized on a streptavidin matrix coated sensor chip by biotin covalent linkage. SPR analysis was performed by injecting solutions of replication origin protein from targeted bacteria or archaea followed by injection of replication origin protein from other bacteria or archaea for comparison (Jiang et al., 2003; Pei et al., 2007). Also, we can use SPR for analyzing the binding reactions of ATP- and ADP–DnaA protein to the *oriC* regions (Schaper et al., 2000; Pei et al., 2007). Based on the difference functions of ATP and ADP, the result revealed that DnaA proteins require ATP for site-specific unwinding at *T. tengcongensis oriC* region (Pei et al., 2007). Similar result was obtained in *Thermus thermophilus* (Schaper et al., 2000). This is similar to those in *E. coli* and *T. maritima*, further supporting that the ATP dependent

activation of DnaA in replication initiation is highly conserved in bacteria. The study of *S. solfataricus* eukaryote-like Orc1/Cdc6 initiators interacting with DNA polymerase B1 (Zhang et al., 2009) and *T. tengcongensis* DnaA initiators interacting with anti-terminator NusG (Liu et al., 2008) also profited from the widespread use of SPR. Messer et al. (2001) studied DnaA rules for DnaA binding and roles of DnaA in origin unwinding and helicase loading by SPR.

Replication Initiation Point mapping

Replication initiation point mapping method was developed by Gerbi and Bielinsky (1997) and Bielinsky and Gerbi (1998) to identify the RIPs by using the symmetry of a typical replication bubble that emerges once the bidirectionally moving forks have been established. This technique has been successfully used to detect the initiation sites of DNA replication (even locations of each DnaA box) at the nucleotide level in chromosomes (Matsunaga et al., 2003; Robinson et al., 2004; Pei et al., 2007) or plasmid (Sun et al., 2006) in many organisms. RIP mapping utilizes the shortest lengths of eukaryotic Okazaki fragments to map the transition point between leading and lagging strand synthesis by extending primers to various initiation points in an asynchronous population of replicating molecules schematic illustration (as shown in **Figure 6**). The extension products are fractionated on sequencing gels finally where maps that leading strand synthesis starts at a unique site, in both small and large origins.

Replication initiation point mapping is 1000-fold more sensitive and more effective to separate the nascent DNA and nicked contaminating DNA by selective degradation of 5′ DNA by λ-exonuclease prior to primer extension (Gerbi and Bielinsky, 1997) which ensures the integrity of RNA-primed DNA. Incipiently, this technology was used to identify the RIP of Eukaryote. Recently, works were demonstrated that archaea also have short eukaryotic-like Okazaki fragments allowing this technique to be used to map the initiation point of *P. abyssi* (Matsunaga et al., 2003). Robinson et al. (2004) performed RIP mapping to identify two origins of replication (*oriC*1 and the Cdc6-3 proximal origin-*oriC*2) in the single chromosome of the hyperthermophilic archaeon *S. solfataricus*. RIP mapping confirmed that the autonomously replicating sequence (ARS) elements corresponding to each replicon were functional in the chromosomal context of the halophilic archaeon *Haloferax volcanii* (Norais et al., 2007). But, because the exact size of the RNA primer synthesized by archaeon primase *in vivo* is not known, this technique does not allow the identification of the precise nucleotide at which replication initiates in archaeon that have multi-*oriC*s.

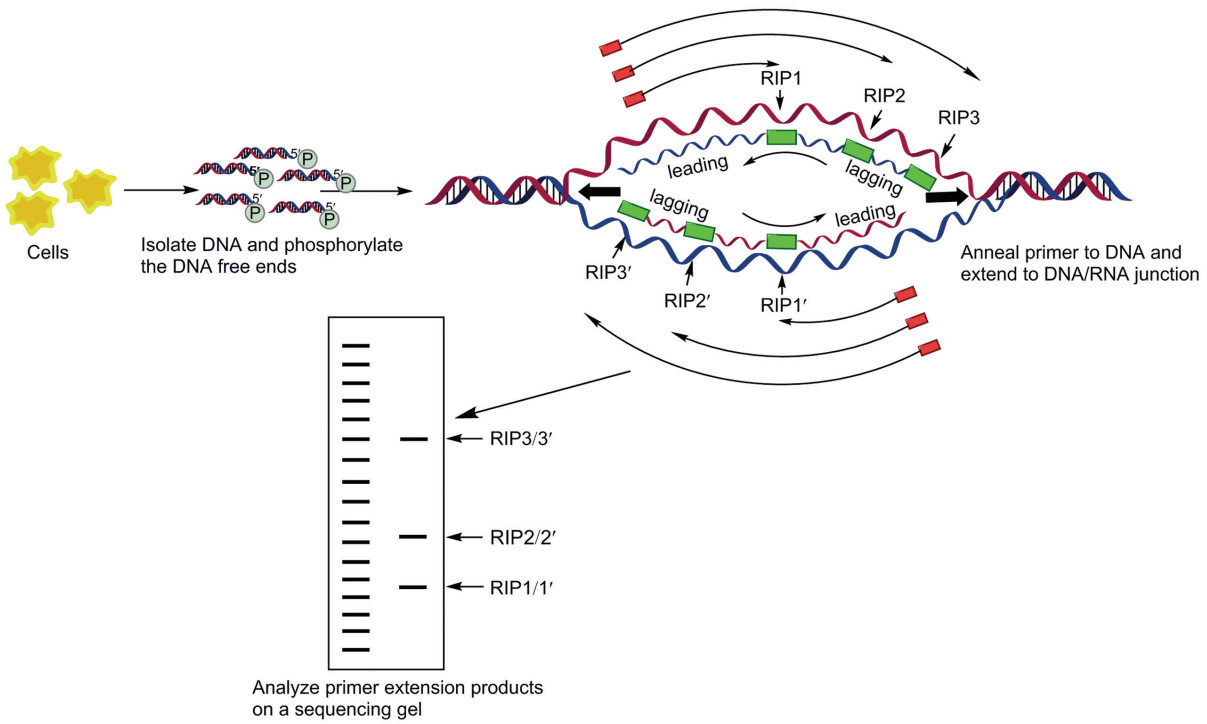


FIGURE 6 | Replication initiation point (RIP) mapping. The replication bubble of semi-discontinuous replication is diagrammed here. After phosphorylation of any 5′-OH ends with polynucleotide kinase, replication intermediates enriched on the BND-cellulose column are treated with λ-exonuclease to digest nicked DNA. Digestion is confirmed on the agarose gel before proceeding to the primer extension reaction. The primer extension products are showed as arrows outside the replication bubble, stopping at DNA/RNA junctions on the DNA. Extension stops at the points labeled RIP1, RIP2, RIP3. Green rectangles depict the RNA primers of nascent strands. PCR products are purified and analyzed on a denaturing polyacrylamide gel. Due to asynchrony, the replication bubble can be of various sizes, resulting in various length, is the transition point from discontinuous to continuous synthesis. Sequencing and RIP reactions were analyzed side by side on the same gel (Gerbi and Bielinsky, 1997; Matsunaga et al., 2003; Lee and Romero, 2012).

Isothermal Titration Calorimetry

Isothermal titration calorimetry is a label-free, powerful, and highly sensitive technique for studying molecular interactions in solution. This method has been applied quite extensively to investigate the interaction of a macromolecule (in general, a protein) with small ligands (Sigurskjold, 2000; Velazquez-Campoy and Freire, 2006), other proteins (Pierce et al., 1999; Velazquez-Campoy et al., 2004), and nucleic acids (Matulis et al., 2000) as well as with drugs (Ward and Holdgate, 2001; Boonsongrit et al., 2008) and metal ions (Zhang et al., 2000), relies on the fact that such an interaction is accompanied by a heat effect. It does not rely on the presence of chromophores or fluorophores, nor does it require an enzymatic assay. A number of parameters such as enthalpy of binding (ΔH), entropy of binding (ΔS), association constant (Ka), binding stoichiometry (n), free energy of binding (ΔG), and potential site–site interactions (cooperativity) can be obtained from a single calorimetric titration, providing a full thermodynamic description of an interacting system (**Figure 7**).

Isothermal titration calorimetry has been one of the most common tools used for investigating interactions of protein association with nucleic acids. Recent advances in ITC instrumentation and data analysis software like the Omega ITC, MCS ITC, VP-ITC, Auto-ITC, Nano ITC-III, and ITC200 instruments have facilitated the development of experimental designs. It also can provide an informative thermodynamic when used in conjunction with complementary techniques such as X-ray crystallography, NMR spectroscopy, small angle x-ray scattering (SAXS), circular dichroism spectroscopy (CD), intrinsic fluorescence, and immunoisolations. Many particularly interesting reports employ ITC, with a focus on protein interactions with nucleic acids. Zhou et al. (2008) have utilized ITC in their study of the role of *E. coli* proline utilization A (PutA) flavoprotein, which acts as the transcriptional repressor of proline utilization genes putA and putP. ITC of PutA binding to the optimal oligonucleotide (O2) revealed a strongly endothermic interaction in Tris buffer but a weakly exothermic interaction in phosphate buffer. Kozlov and Lohman (2012) employed ITC to analyze the interaction about *E. coli* SSB and *D. radiodurans* SSB binding to ssDNA, respectively. Crane-Robinson et al. (2009) and Gilbert and Batey (2009) present an overview of ITC experiments on protein/DNA complexes, with detailed descriptions of the experimental methodologies. This review concentrates on the thermodynamics of interaction of protein DNA binding domains with DNA duplexes, and gives a thorough description of the joint implementation of ITC and differential scanning calorimetry (DSC) to provide a thorough description of the binding process. In spite of the widely using, there remain some important points to the use of ITC that should always be considered. Just as Falconer said in two reviews about ITC (Falconer et al., 2010; Falconer and Collins, 2011), several aspects of ITC data collection have been outlined in the reviews.

As more and more correlative analyses are performed and databases increased their informative capacity, ITC

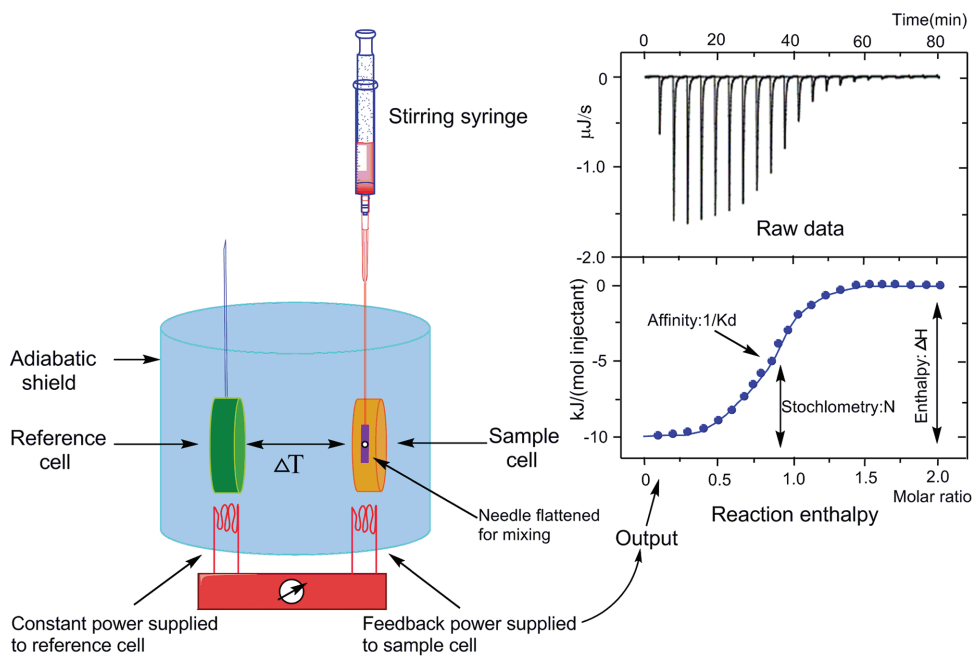


FIGURE 7 | Basic principle of isothermal titration calorimetry. Schematic representation of the isothermal titration calorimeter (left) and a characteristic titration experiment **(upper right)** with its evaluation **(lower right)**. In **(upper right)** picture, the titration thermogram is represented as heat per unit of time released after each injection of the ligand into the protein (black), as well as the dilution of ligand into buffer (blue). In **(lower right)** picture, the dependence of released heat in each injection versus the ratio between total ligand concentration and total protein concentration is represented. Circles represent experimental data and the line corresponds to the best fitting to a model considering *n* identical and independent sites. The syringe is inserted in the sample cell and a series of injections are made (Freyer and Lewis, 2008; Martinez et al., 2013).

should develop more accurate and powerful for estimating binding affinities from known structures and conversely to use thermodynamic data to make informed predictions regarding the properties of molecular interfaces. Although ITC is widely used in identification of protein–DNA interaction, the using in identification of replication origins is vacant.

Conventional Methods for Detecting Protein–DNA Interaction at Origins of Replication *In Vivo*

Chromatin Immunoprecipitation

Chromatin Immunoprecipitation is an excellent experimental method to determine the interactions of proteins with their binding sites *in vivo*. This technique is frequently used to detect the interactions between DnaA or oriBPs and replication origins due to the ability that ChIP assays allow one to determine the entire spectrum of DNA binding sites for any given protein *in vivo* with whole-genome DNA microarrays. ChIP also could be used for determining whether there were changes in the levels of binding *oriC*s and DnaA during different cell-cycle phase *in vivo* (Robinson et al., 2004; Duggin et al., 2008). As described in many papers, living cells should be handled with chemical cross-linkers to covalently bind proteins with each other and then with their DNA targets. Once cross-linked to associated proteins, sonication is used to extract and fragment chromatin, and specific antibodies against a target protein is employed to isolate protein–DNA complexes. The cross-links that is binding with proteins and DNA are then reversed, and the associated DNA was subjected to qPCR analysis to test for coprecipitation of specific DNA sequences (Orlando, 2000; Buck and Lieb, 2004). Using specific antibody or several antibodies together is one of key steps in ChIP assay. Antisera was obtained through recognizing one major chromatin associated band of approximately expected molecular weight in cell-free extracts of *Pyrococcus furiosus* (Komori and Ishino, 2001), *S. solfataricus* (Robinson et al., 2004), *C. crescentus* (Gorbatyuk and Marczynski, 2005; Taylor et al., 2011), and *Pyrococcus abyssi* (Matsunaga et al., 2001, 2007). The anti-DnaA antibody of *E. coli* (Sekimizu et al., 1988; Newman and Crooke, 2000) and *Bacillus subtilis* (Ogura et al., 2001; Gorbatyuk and Marczynski, 2005) was obtained by the same way. In a study of the identification of the the chromosomal *dif* site that binds Xer in *S. solfataricus in vivo* via ChIP and ChIP–chip, the antibodies required in ChIP assay were affinity purified from antisera that were raised against Xer-6H (His6-tagged Xer) in rabbits using Xer-6H immobilized on an NHS-activated agarose Hi-Trap column (GE Healthcare) (Duggin et al., 2011). The basic method is shown in **Figure 8**.

Despite the tremendous value of ChIP methods, it is important to be aware of their limitations. Carey et al. (2009) has listed three limitations of 'standard' ChIP experiment: (1) The ChIP assay often yields low signals in comparison to negative controls, which can lead to inconclusive results; (2) it is difficult to determine the precise binding site for a factor because of the limited resolution of the assay; and (3) ChIP is not a functional assay and cannot by itself demonstrate the functional significance of a protein or modified histone found to be located at a genomic

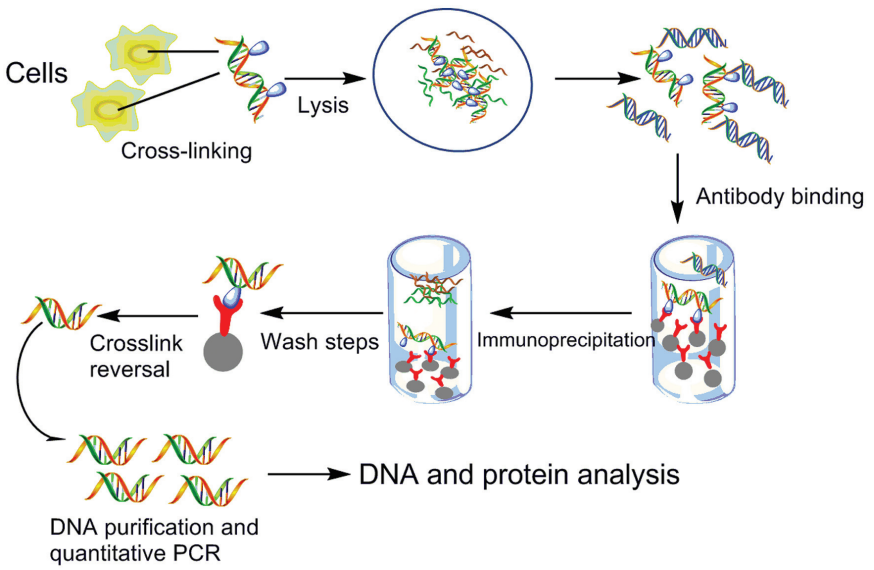


FIGURE 8 | Chromatin immunoprecipitation. Protein and associated chromatin in living cells or tissues are temporarily bonded, the DNA–protein complexes (chromatin-protein) are then sheared into ~500 bp DNA fragments using either enzymatic digestion or physical shearing by sonication. Cross-linked DNA fragments associated with the protein(s) of interest using formaldehyde are selectively immunoprecipitated from the cell debris using appropriate protein-specific antibody. After the cross-links are reversed, the associated DNA fragments are purified and their sequence is determined. These DNA sequences are supposed to be associated with the protein of interest *in vivo*. The DNA undergoes PCR amplification using primers targeting a particular genomic locus. These DNA sequences can be subjected to a number of downstream analysis techniques, including targeted approaches, like semiquantitative PCR and quantitative PCR, and genome-wide analyses using microarrays (ChIP–chip) and deep sequencing (ChIP-seq), ChIP-on-chip (Shah, 2009; Vinckevicius and Chakravarti, 2012).

region of interest. Recent advances in ChIP methodology have overcome some of the limitations, and the development of complementary assays, and analyses have expanded the number, types and resolution of protein–DNA interactions that have been discovered. Such as ChIP–chip (Horak and Snyder, 2002; Buck and Lieb, 2004), ChIP on tiled arrays (ChIPOTle) (Buck et al., 2005), ChIP-Seq (Robertson et al., 2007; Schmidt et al., 2009), ChIP-PaM (Wu et al., 2010), Re-ChIP (Truax and Greer, 2012) were developed for analyzing the more specific interactions between protein and DNA sequences. By means of ChIP coupled with hybridization on a whole genome microarray (ChIP–chip), researchers detected the binding of Cdc6/Orc1 to *oriC* of archaeon *P. abyssi in vivo*. And it was the first time that ChIP–chip method used for identifying the genome-wide distribution of the initiator of DNA replication in Archaea and Bacteria (Matsunaga et al., 2007). ChIP-on-chip was widely applied to genome-wide analysis, which combines the specificity of ChIP with the unbiased, high-throughput capabilities of microarrays (Testa et al., 2005; Huebert et al., 2006; Wyrick et al., 2009; Kim et al., 2014). Isolation of specific genomic regions retaining molecular interactions is necessary for their biochemical analysis. Insertional ChIP (iChIP) was a useful tool for dissecting chromatin structure of genomic region of interest. This technique can efficiently isolate of specific genomic domains (Hoshino and Fujii, 2009). In addition, a novel method called engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) was established, for purification of specific genomic regions retaining molecular interactions (Fujita et al., 2013). Here, we detailed analyze ChIP-seq.

ChIP Sequencing

Chromatin immunoprecipitation coupled with microarrays (ChIP–chip) or short-tag sequencing (ChIP-seq) has become the standard technique for identifying the locations and biochemical modifications of bound proteins genome-wide. ChIP-seq can be done without prior knowledge of the underlying sequence and relies only on the subsequent DNA sequence alignment to the reference genome of interest Compared to ChIP–chip. Furthermore, the nature of the microarray hybridization signal makes detection and rigorous quantification of low abundance signals problematic. Taken together, ChIP-seq can provide greater resolution, sensitivity, and specificity compared to ChIP–chip (Johnson et al., 2007; Robertson et al., 2007; Schmidt et al., 2009). Owing to the tremendous progress in next-generation sequencing technology including the Genome Analyzer (Illumina, formerly Solexa), SOLiD (Applied Biosystems), 454-FLX (Roche), and HeliScope (Helicos) (Morozova and Marra, 2008; Schmidt et al., 2009), ChIP-seq offers higher resolution, less noise, and greater coverage than its array-based predecessor ChIP–chip. With the decreasing cost of sequencing, ChIP-seq has become an indispensable tool for studying gene regulation and epigenetic mechanisms.

ChIP-seq experiments generate large quantities of data, and effective computational analysis will be crucial for uncovering biological mechanisms. An important consideration in experimental design is the minimum number of sequenced reads required to obtain statistically significant results. The standards and guidelines for carrying out ChIP-seq has been described based on the collective experience of laboratories involved in the Encyclopedia of DNA Elements (ENCODE) and model organism ENCODE (modENCODE) projects, including antibody validation, choosing appropriate sequencing depth, experimental replication, data quality assessment, data and metadata reporting (Landt et al., 2012). However, ChIP-seq has been proved to be a potential tool in the study of histone modifications, nucleosome positioning, and mapping of binding sites of various DNA binding proteins. Certainly, there are more and more researchers used ChIP coupled with high-throughput sequencing (ChIP-seq) to identify replication origins precisely, especially for the yeast genome or other eukaryotes (Eaton et al., 2010, 2011; Gilbert, 2010; Martin et al., 2011). Using ChIP or ChIP-seq, we can capture the change of DnaA protein level in the whole replication process of cells *in vivo*.

Other Methods and Applications

In addition to the methods described here, many methods were developed to identify the majority of origins found in the previous report. Complements and extends were achieved by direct, high resolution mapping of potential origins and proteins that could bind with the specific sites in the origins of replication, also something related to replication origins.

Owing to the pivotal role played by DNA-associating proteins in various cellular processes, many *in vitro*, *in vivo*, *in silico*, and biophysical techniques have been developed to study DNA–protein interactions. *In vitro* technique includes southwestern assay, yeast one-hybrid assay (Y1H), phage display and proximity ligation assay (PLA); scanning probe microscope (SPM) is a novel *in vivo* method on the interaction of protein–DNA; biophysical technique includes many methods, such as fluorescence-based techniques [time-resolved fluorescence depolarization, double labeled native gel electrophoresis and fluorescence-based imaging, fluorescence resonance energy transfer (FRET) techniques (Clegg, 1995)], capillary electrophoresis with laser-induced fluorescence (CE-LIF) (Riddick and Brumley, 2008), also some fluorescence-based protein or nucleic acids bioprobe like FRep (Shahravan et al., 2011), quantum dots (QDs) (Michalet et al., 2005), SPR, nuclear magnetic resonance, circular dichroism (CD), atomic force microscopy (AFM), and microcalorimetry (Dey et al., 2012).

ARS (autonomously replicating sequence) assays first utilized to prove that DNA sequences was important for replication by determining whether a given DNA fragment initiates replication when placed on a plasmid in *yeast* (Struhl et al., 1979). The plasmid-based ARS assay was used to identify numerous replication origins in budding and *fission yeasts* (Newlon, 1996; Huberman, 1999). PCR-based assay which is an alternative approach to the plasmid-based ARS assay was utilized to identify replicator at ectopic sites in the genome (Malott and Leffak, 1999; Vernis et al., 1999; Tao et al., 2000). In 1996, EMSA and DNase I footprint analysis were employed to detect the interaction of the IciA protein which is known to bind to the AT-rich repeat region in the *E. coli* origin of chromosome replication, with AT-rich regions in replication origins of plasmids F and R1 (Wei and Bernander, 1996). The direction of

replication fork movement is ascertained to pinpoint the origin located between the outwardly moving forks by neutral/alkaline gel electrophoresis (Nawotka and Huberman, 1988). Patrizia Contursi first described the functional cloning of a chromosomal *oriC* of the hyperthermophilic archaeon *S. solfataricus* from an archaeon and confirmed the proposed location by 2-D gel electrophoresis experiments. As described in the study, it represented an important step toward the reconstitution of an archaeal *in vitro* DNA replication system (Contursi et al., 2004). 2D neutral–neutral agarose gel analysis was used to test whether the loci associated with the cdc6 genes in the single chromosome of *S. solfataricus* might contain origins of replication (Robinson et al., 2004). Due to DNA isolated from asynchronously replicating cells and subjected the DNA to digestion with restriction enzymes, this technique can detect replication intermediates directly corresponding to the resolution of distinct arcs on the gel. Furthermore, RIP mapping was used to identify the RIPs at both origins in *S. solfataricus* and DNase I footprinting analysis, ChIP, EMSA were all utilized frequently to detect whether the Cdc6 could bind to the both origins in this study (Robinson et al., 2004). Zawilak et al. (2003) have presented the DNA recognition properties of the *H. pylori* DnaA protein. The interactions between the purified DnaA protein of *H. pylori* and its target were analyzed by gel retardation assay and SPR *in vitro*. A series of competition gel retardation assays were performed to elucidate the binding requirements and analyze the DNA–protein complexes (Zawilak et al., 2003). In the study of mechanism for the DnaA-*oriC* cooperative interaction at high temperature and duplex opening at an unusual AT-rich region in *T. tengcongensis*, many techniques for studying the interaction of protein–DNA complexes were utilized for different purposes. The GAL4-based yeast two-hybrid system, EMSA, RIP mapping, open-complex formation assay, SPR, nuclease P1 assay were used in this research for different interactions of protein–DNA complexes. It's proud that it's the first experimental demonstration of the chromosomal RIP in thermophilic bacteria at nucleotide level (Pei et al., 2007).

In the study of interactions of DnaA proteins from distantly related bacteria with the replication origin of the broad host range plasmid RK2, DNase I footprinting, gel mobility shift, and SPR analyses were utilized to compare the interactions of *oriV* with five different DnaA proteins from *E. coli*, *Pseudomonas putida*, *Pseudomonas aeruginosa, B. subtilis*, and *Streptomyces lividans* (Caspi et al., 2000). The results revealed that the DnaA proteins of a host bacterium were incapable to form a stable and functional complex with the DnaA boxes at *oriV* is a limiting step for plasmid host range (Caspi et al., 2000). Mode of initiator-*oriC* interactions with the loop formation between the subcomplexes of the discontinuous origin of *H. pylori* was revealed by the experimental analysis of RIP mapping, electron microscopy, and immunoprecipitation assay. *H. pylori oriC* exhibited bipartite structure and being the first such origin discovered in a Gram-negative bacterium (Donczew et al., 2012). Katarzyna et al. (2014) used SPR and EMSA methods to measure the sequence-specific interactions of Rep proteins with ssDNA within the DNA unwinding element (DUE) in the AT-rich region of the plasmid replication origin.

Conclusion

The relevant information of *oriC* could be found from the *oriC* predicting tool such as Ori-Finder as well as the online databases DoriC which include the locations of replication origins sites for thousands of bacterial chromosomes and archaeal genomes. Based on the predicted results, we can identify and confirm the interactions at origins of replication by experimental methods. Of course, purifying replication relevant protein is another pivotal step for the research. An ideal method would require minimal cell numbers or purified protein, could be able to detect rare interactions with high specificity and sensitivity, as well as it could be easily modified to quantify interactions and provide complete information on either of protein or DNA. *In vitro* techniques provide better quantitative characterization but require isolation of active, soluble protein, which can be challenging and impractical in high-throughput assays. Additionally, protein function may depend strongly on assay conditions; hence, a non-native *in vitro* environment can give rise to results contradictory to those performed in an *in vivo* assay. Alternatively, *in vivo* assays provide a native-like environment for studying the protein–DNA interaction. Due to the restriction of experimental conditions both *in vivo* and *in vitro*, as showed in the review, more than one method were applied in most of experiments to measure the multiple protein–DNA interactions that take place in and around replication origins. And outstanding results were received by them.

However, the sequence of replication origins must be known in methods of EMSA, SPR, ITC, and DNase I footprinting. ChIP and ChIP-seq detect replication origin interactions genome-wide under the condition of unknown or known binding sequences. Through ChIP-seq, the binding sequences could be confirmed precisely. And, the most important point is that we can visually observe the amount change of oriBPS during the cell cycle. Thus the results could help us to understand the mechanisms and regulations of microbial replication initiation clearly. As was showed in the research about how DnaA and essential response regulator CtrA compete to control *C. crescentus* chromosome replication, previous EMSA experiments was used for single DnaA binding site targets (G1 DnaA box), then DNase I footprinting assay was applied to identify replication origin (*Cori*) sites (G1, G2, W1, W2, W3, W4, W5) protected by DnaA and the position of CtrA binding site 'e' (Taylor et al., 2011). From the figure of autoradiograph of the sequencing gel, CtrA obscures some DnaA protected sites, and all others DnaA is displaced by CtrA binding. The result of DNase I footprinting assay showed the weaker binding ability of DnaA proteins of *C. crescentus* than CtrA. The followed ChIP assay *in vivo* and western blot showed that DnaA is continuously present during the cell cycle, and CtrA proteolysis coincides with DnaA binding to *Cori*. Therefore, series of assays proved that DnaA is regarded primarily as a chromosome replication regulator and secondarily

as a transcription regulator, CtrA is regarded primarily as a transcription regulator.

These methods have promoted the development in this field, however, a numerous of problems need to be solved timely. Many techniques were explored to detect the interaction of protein and nucleic acids, while how to improve these techniques to employ in the study of replication origins will be the further work that we do. Hence, we envisage that progress in these technologies will further improve detection abilities and allow sensitive, fast, and cost-effective biochemical analysis both in laboratories and in the field. This development will further extend the potential applications and allow them to be used far more widely. With the development of science and technology and strong cooperation between the various disciplines, research strategy with innovative thinking and novel methods will continue to emerge. It can be predicted that research on the regulation and mechanism of replication origins will make considerable progress in the near further.

Acknowledgments

The present work was partially supported by National High Technology Research and Development Program of China (Grant No. 2015AA020701) and National Natural Science Foundation of China (Grant No. 31470967).

References

Abe, Y., Jo, T., Matsuda, Y., Matsunaga, C., Katayama, T., and Ueda, T. (2007). Structure and function of DnaA N-terminal Domains: specific sites and mechanisms in inter-DnaA interaction and in DnaB helicase loading on oriC. *J. Biol. Chem.* 282, 17816–17827. doi: 10.1074/jbc.M701841200

Alves, C., and Cunha, C. (2012). Electrophoretic mobility shift assay: analyzing protein–nucleic acid interactions. *Nat. Protoc.* 2, 1849–1861. doi: 10.1038/nprot.2007.249

Arias-Palomo, E., O'shea, V. L., Hood, I. V., and Berger, J. M. (2013). The bacterial DnaC helicase loader is a DnaB ring breaker. *Cell* 153, 438–448. doi: 10.1016/j.cell.2013.03.006

Aves, S. J. (2009). "DNA replication initiation," in *DNA Replication,* eds S. Vengrova and J. Z. Dalgaard (New York, NY: Humana Press), 1–16. doi: 10.1093/jexbot/52.355.193

Baird, C. L., and Myszka, D. G. (2001). Current and emerging commercial optical biosensors. *J. Mol. Recognit.* 14, 261–268. doi: 10.1002/jmr.544

Baker, T. A., and Bell, S. P. (1998). Polymerases and the replisome: machines within machines. *Cell* 92, 295–305. doi: 10.1016/S0092-8674(00)80923-X

Barry, E. R., and Bell, S. D. (2006). DNA replication in the archaea. *Microbiol. Mol. Biol. Rev.* 70, 876–887. doi: 10.1128/MMBR.00029-06

Berquist, B. R., and DasSarma, S. (2003). An archaeal chromosomal autonomously replicating sequence element from an extreme halophile, *Halobacterium* sp. strain NRC-1. *J. Bacteriol.* 185, 5959–5966. doi: 10.1128/JB.185.20.5959-5966.2003

Bielinsky, A.-K., and Gerbi, S. A. (1998). Discrete start sites for DNA synthesis in the yeast ARS1 origin. *Science* 279, 95–98. doi: 10.1126/science.279.5347.95

Blaesing, F., Weigel, C., Welzeck, M., and Messer, W. (2000). Analysis of the DNA–binding domain of *Escherichia coli* DnaA protein. *Mol. Microbiol.* 36, 557–569. doi: 10.1046/j.1365-2958.2000.01881.x

Bolduc, O. R., and Masson, J.-F. (2011). Advances in surface plasmon resonance sensing with nanoparticles and thin films: nanomaterials, surface chemistry, and hybrid plasmonic techniques. *Anal. Chem.* 83, 8057–8062. doi: 10.1021/ac2012976

Boonsongrit, Y., Mueller, B. W., and Mitrevej, A. (2008). Characterization of drug–chitosan interaction by 1 H NMR, FTIR and isothermal titration calorimetry. *Eur. J. Pharm. Biopharm.* 69, 388–395. doi: 10.1016/j.ejpb.2007.11.008

Buck, M. J., and Lieb, J. D. (2004). ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* 83, 349–360. doi: 10.1016/j.ygeno.2003.11.004

Buck, M. J., Nobel, A. B., and Lieb, J. D. (2005). ChIPOTle: a user-friendly tool for the analysis of ChIP-chip data. *Genome Biol.* 6, R97. doi: 10.1186/gb-2005-6-11-r97

Carey, M. F., Peterson, C. L., and Smale, S. T. (2009). Chromatin immunoprecipitation (ChIP). *Cold Spring Harb. Protoc.* 2009, db-rot5279. doi: 10.1101/pdb.prot5279

Carey, M. F., Peterson, C. L., and Smale, S. T. (2012). Experimental strategies for the identification of DNA-binding proteins. *Cold Spring Harb. Protoc.* 2012, db. to067470. doi: 10.1101/pdb.top067470

Carey, M. F., Peterson, C. L., and Smale, S. T. (2013a). DNase I footprinting. *Cold Spring Harb. Protoc.* 2013, db. rot074328. doi: 10.1101/pdb.prot074328

Carey, M. F., Peterson, C. L., and Smale, S. T. (2013b). Electrophoretic mobility shift assays. *Cold Spring Harb. Protoc.* 2013, db. rot075861. doi: 10.1101/pdb.prot075861

Caspi, R., Helinski, D. R., Pacek, M., and Konieczny, I. (2000). Interactions of DnaA proteins from distantly related bacteria with the replication origin of the broad host range plasmid RK2. *J. Biol. Chem.* 275, 18454–18461. doi: 10.1074/jbc.M000552200

Chen, H., and Chang, G. D. (2001). Simultaneous immunoblotting analysis with activity gel electrophoresis in a single polyacrylamide gel. *Electrophoresis* 22, 1894–1899. doi: 10.1007/978-1-59745-542-8_4

Chumsakul, O., Nakamura, K., Kurata, T., Sakamoto, T., Hobman, J. L., Ogasawara, N., et al. (2013). High-resolution mapping of in vivo genomic transcription factor binding sites using in situ DNase I footprinting and ChIP-seq. *DNA Res.* 20, 325–338. doi: 10.1093/dnares/dst013

Clegg, R. M. (1995). Fluorescence resonance energy transfer. *Curr. Opin. Biotechnol.* 6, 103–110. doi: 10.1016/0958-1669(95)80016-6

Coker, J. A., Dassarma, P., Capes, M., Wallace, T., Mcgarrity, K., Gessler, R., et al. (2009). Multiple replication origins of *Halobacterium* sp. strain NRC-1: properties of the conserved orc7-dependent oriC1. *J. Bacteriol.* 191, 5253–5261. doi: 10.1128/JB.00210-09

Contursi, P., Pisani, F. M., Grigoriev, A., Cannio, R., Bartolucci, S., and Rossi, M. (2004). Identification and autonomous replication capability of a chromosomal replication origin from the archaeon *Sulfolobus solfataricus*. *Extremophiles* 8, 385–391. doi: 10.1007/s00792-004-0399-y

Cooper, M. A. (2002). Optical biosensors in drug discovery. *Nat. Rev. Drug Discov.* 1, 515–528. doi: 10.1038/nrd838

Crane-Robinson, C., Dragan, A. I., and Read, C. M. (2009). "Defining the thermodynamics of protein/DNA complexes and their components using micro-calorimetry," in *DNA-Protein Interactions,* Vol. 543, eds B. Leblanc and T. Moss (New York, NY: Humana Press), 625–651. doi: 10.1007/978-1-60327-015-1_37

Dame, R. T., Kalmykowa, O. J., and Grainger, D. C. (2011). Chromosomal macrodomains and associated proteins: implications for DNA organization and replication in gram negative bacteria. *PLoS Genet.* 7:e1002123. doi: 10.1371/journal.pgen.1002123

Despeyroux, D., Walker, N., Pearce, M., Fisher, M., Mcdonnell, M., Bailey, S., et al. (2000). Characterization of ricin heterogeneity by electrospray mass spectrometry, capillary electrophoresis, and resonant mirror. *Anal. Biochem.* 279, 23–36. doi: 10.1006/abio.1999.4423

Dey, B., Thukral, S., Krishnan, S., Chakrobarty, M., Gupta, S., Manghani, C., et al. (2012). DNA–protein interactions: methods for detection and analysis. *Mol. Cell. Biochem.* 365, 279–299. doi: 10.1007/s11010-012-1269-z

Di Paola, D., Rampakakis, E., Chan, M. K., Arvanitis, D. N., and Zannis-Hadjopoulos, M. (2010). Increased origin activity in transformed versus normal cells: identification of novel protein players involved in DNA replication and cellular transformation. *Nucleic Acids Res.* 38, 2314–2331. doi: 10.1093/nar/gkp1192

Donczew, R., Weigel, C., Lurz, R., Zakrzewska-Czerwińska, J., and Zawilak-Pawlik, A. (2012). *Helicobacter pylori* oriC—the first bipartite origin of chromosome replication in Gram-negative bacteria. *Nucleic Acids Res.* 40, 9647–9660. doi: 10.1093/nar/gks742

Duggin, I. G., Dubarry, N., and Bell, S. D. (2011). Replication termination and chromosome dimer resolution in the archaeon *Sulfolobus solfataricus*. *EMBO J.* 30, 145–153. doi: 10.1038/emboj.2010.301

Duggin, I. G., Mccallum, S. A., and Bell, S. D. (2008). Chromosome replication dynamics in the archaeon *Sulfolobus acidocaldarius*. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16737–16742. doi: 10.1073/pnas.0806414105

Eaton, M. L., Galani, K., Kang, S., Bell, S. P., and MacAlpine, D. M. (2010). Conserved nucleosome positioning defines replication origins. *Genes Dev.* 24, 748–753. doi: 10.1101/gad.1913210

Eaton, M. L., Prinz, J. A., MacAlpine, H. K., Tretyakov, G., Kharchenko, P. V., and MacAlpine, D. M. (2011). Chromatin signatures of the *Drosophila* replication program. *Genome Res.* 21, 164–174. doi: 10.1101/gr.116038.110

Falconer, R. J., and Collins, B. M. (2011). Survey of the year 2009: applications of isothermal titration calorimetry. *J. Mol. Recognit.* 24, 1–16. doi: 10.1002/jmr.1073

Falconer, R. J., Penkova, A., Jelesarov, I., and Collins, B. M. (2010). Survey of the year 2008: applications of isothermal titration calorimetry. *J. Mol. Recognit.* 23, 395–413. doi: 10.1002/jmr.1025

Felczak, M. M., and Kaguni, J. M. (2004). The box VII motif of *Escherichia coli* DnaA protein is required for DnaA oligomerization at the *E. coli* replication origin. *J. Biol. Chem.* 279, 51156–51162. doi: 10.1074/jbc.M409695200

Fourtounis, J., Falgueyret, J.-P., and Sayegh, C. E. (2011). Assessing protein–RNA interactions using microfluidic capillary mobility shift assays. *Anal. Biochem.* 411, 161–163. doi: 10.1016/j.ab.2010.11.042

Fox, K. R., and Waring, M. J. (1987). The use of micrococcal nuclease as a probe for drug-binding sites on DNA. *Biochim. Biophys. Acta* 909, 145–155. doi: 10.1016/0167-4781(87)90036-4

Freyer, M. W., and Lewis, E. A. (2008). Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods Cell Biol.* 84, 79–113. doi: 10.1016/S0091-679X(07)84004-0

Fried, M. G. (1989). Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay. *Electrophoresis* 10, 366–376. doi: 10.1002/elps.1150100515

Fujikawa, N., Kurumizaka, H., Nureki, O., Terada, T., Shirouzu, M., Katayama, T., et al. (2003). Structural basis of replication origin recognition by the DnaA protein. *Nucleic Acids Res.* 31, 2077–2086. doi: 10.1093/nar/gkg309

Fujita, T., Asano, Y., Ohtsuka, J., Takada, Y., Saito, K., Ohki, R., et al. (2013). Identification of telomere-associated molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP). *Sci. Rep.* 3:3171. doi: 10.1038/srep03171

Fuller, R. S., Funnell, B. E., and Kornberg, A. (1984). The dnaA protein complex with the *E. coli* chromosomal replication origin (oriC) and other DNA sites. *Cell* 38, 889–900. doi: 10.1016/0092-8674(84)90284-8

Gao, F. (2014). Recent advances in the identification of replication origins based on the Z-curve method. *Curr. Genomics* 15, 104–112. doi: 10.2174/1389202915999140328162938

Gao, F., Luo, H., and Zhang, C.-T. (2012). DeOri: a database of eukaryotic DNA replication origins. *Bioinformatics* 28, 1551–1552. doi: 10.1093/bioinformatics/bts151

Gao, F., Luo, H., and Zhang, C.-T. (2013). DoriC 5.0: an updated database of oriC regions in both bacterial and archaeal genomes. *Nucleic Acids Res.* 41, D90–D93. doi: 10.1093/nar/gks990

Gao, F., and Zhang, C.-T. (2007). DoriC: a database of oriC regions in bacterial genomes. *Bioinformatics* 23, 1866–1867. doi: 10.1093/bioinformatics/btm255

Gao, F., and Zhang, C.-T. (2008). Ori-Finder: a web-based system for finding oriCs in unannotated bacterial genomes. *BMC Bioinformatics* 9:79. doi: 10.1186/1471-2105-9-79

Garg, P., and Burgers, P. M. (2005). DNA polymerases that propagate the eukaryotic DNA replication fork. *Crit. Rev. Biochem. Mol. Biol.* 40, 115–128. doi: 10.1080/10409230590935433

Gerbi, S. A., and Bielinsky, A.-K. (1997). Replication initiation point mapping. *Methods* 13, 271–280. doi: 10.1006/meth.1997.0526

Gilbert, D. M. (2010). Evaluating genome-scale approaches to eukaryotic DNA replication. *Nat. Rev. Genet.* 11, 673–684. doi: 10.1038/nrg2830

Gilbert, S. D., and Batey, R. T. (2009). "Monitoring RNA–ligand interactions using isothermal titration calorimetry," in *Riboswitches*, ed. A. Serganov (New York, NY: Humana Press), 97–114. doi: 10.1007/978-1-59745-558-9_8

Gorbatyuk, B., and Marczynski, G. T. (2005). Regulated degradation of chromosome replication proteins DnaA and CtrA in *Caulobacter crescentus*. *Mol. Microbiol.* 55, 1233–1245. doi: 10.1111/j.1365-2958.2004.04459.x

Granger-Schnarr, M., Lloubes, R., De Murcia, G., and Schnarr, M. (1988). Specific protein-DNA complexes: immunodetection of the protein component after gel electrophoresis and Western blotting. *Anal. Biochem.* 174, 235–238. doi: 10.1016/0003-2697(88)90540-4

Hellman, L. M., and Fried, M. G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nat. Protoc.* 2, 1849–1861. doi: 10.1038/nprot.2007.249

Hoa, X., Kirk, A., and Tabrizian, M. (2007). Towards integrated and sensitive surface plasmon resonance biosensors: a review of recent progress. *Biosens. Bioelectron.* 23, 151–160. doi: 10.1016/j.bios.2007.07.001

Homola, J., Yee, S. S., and Gauglitz, G. (1999). Surface plasmon resonance sensors: review. *Sens. Actuators B Chem.* 54, 3–15. doi: 10.1016/S0925-4005(98)00321-9

Horak, C. E., and Snyder, M. (2002). ChIP-chip: a genomic approach for identifying transcription factor binding sites. *Methods Enzymol.* 350, 469–483. doi: 10.1016/S0076-6879(02)50979-4

Hoshino, A., and Fujii, H. (2009). Insertional chromatin immunoprecipitation: a method for isolating specific genomic regions. *J. Biosci. Bioeng.* 108, 446–449. doi: 10.1016/j.jbiosc.2009.05.005

Hou, W., and Cronin, S. B. (2013). A review of surface plasmon resonance-enhanced photocatalysis. *Adv. Funct. Mater.* 23, 1612–1619. doi: 10.1002/adfm.201202148

Huberman, J. A. (1999). Genetic methods for characterizing the cis-acting components of yeast dna replication origins. *Methods* 18, 356–367. doi: 10.1006/meth.1999.0792

Huebert, D. J., Kamal, M., O'donovan, A., and Bernstein, B. E. (2006). Genome-wide analysis of histone modifications by ChIP-on-chip. *Methods* 40, 365–369. doi: 10.1016/j.ymeth.2006.07.032

Inoue, S., Kaji, N., Kataoka, M., Shinohara, Y., Okamoto, Y., Tokeshi, M., et al. (2011). Rapid qualitative evaluation of DNA transcription factor NF-κB by microchip electrophoretic mobility shift assay in mammalian cells. *Electrophoresis* 32, 3241–3247. doi: 10.1002/elps.201100261

Jacob, F., Brenner, S., and Cuzin, F. (1963). *On the Regulation of DNA Replication in Bacteria. Cold Spring Harbor Symposia on Quantitative Biology*, Vol. 28. New York, NY: Cold Spring Harbor Laboratory Press, 329–348.

Jiang, Y., Pacek, M., Helinski, D. R., Konieczny, I., and Toukdarian, A. (2003). A multifunctional plasmid-encoded replication initiation protein both recruits and positions an active helicase at the replication origin. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8692–8697. doi: 10.1073/pnas.1532393100

Jing, D., Agnew, J., Patton, W. F., Hendrickson, J., and Beechem, J. M. (2003). A sensitive two-color electrophoretic mobility shift assay for detecting both nucleic acids and protein in gels. *Proteomics* 3, 1172–1180. doi: 10.1002/pmic.200300438

Johnson, A., and O'Donnell, M. (2005). Cellular DNA replicases: components and dynamics at the replication fork. *Annu. Rev. Biochem.* 74, 283–315. doi: 10.1146/annurev.biochem.74.011303.073859

Johnson, D. S., Mortazavi, A., Myers, R. M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316, 1497–1502. doi: 10.1126/science.1141319

Karns, K., Vogan, J. M., Qin, Q., Hickey, S. F., Wilson, S. C., Hammond, M. C., et al. (2013). Microfluidic screening of electrophoretic mobility shifts elucidates riboswitch binding function. *J. Am. Chem. Soc.* 135, 3136–3143. doi: 10.1021/ja310742m

Katarzyna, W., Eugenia, F.-P. M., Katarzyna, B., Magdalena, R., Fernando, M.-H., and Igor, K. (2014). Sequence-specific interactions of Rep proteins with ssDNA in the AT-rich region of the plasmid replication origin. *Nucleic Acids Res.* 42, gku453. doi: 10.1093/nar/gku453

Katayama, T. (2008). Roles for the AAA+ motifs of DnaA in the initiation of DNA replication. *Biochem. Soc. Trans.* 36(Pt 1), 78–82. doi: 10.1042/BST0360078

Kawakami, H., and Katayama, T. (2010). DnaA, ORC, and Cdc6: similarity beyond the domains of life and diversity. *Biochem. Cell Biol.* 88, 49–62. doi: 10.1139/o09-154

Kelman, Z., and O'Donnell, M. (1995). DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. *Annu. Rev. Biochem.* 64, 171–200. doi: 10.1146/annurev.bi.64.070195.001131

Keyamura, K., Fujikawa, N., Ishida, T., Ozaki, S., Su'etsugu, M., Fujimitsu, K., et al. (2007). The interaction of DiaA and DnaA regulates the replication cycle in *E. coli* by directly promoting ATP DnaA-specific initiation complexes. *Genes Dev.* 21, 2083–2099. doi: 10.1101/gad.1561207

Khan, S. H., Farkas, K., Kumar, R., and Ling, J. (2012). A versatile method to measure the binding to basic proteins by surface plasmon resonance. *Anal. Biochem.* 421, 385–390. doi: 10.1016/j.ab.2011.12.006

Khodursky, A. B., Peter, B. J., Schmid, M. B., Derisi, J., Botstein, D., Brown, P. O., et al. (2000). Analysis of topoisomerase function in bacterial replication fork movement: use of DNA microarrays. *Proc. Natl. Acad. Sci. U.S.A.* 97, 9419–9424. doi: 10.1073/pnas.97.17.9419

Kim, J.-M., To, T. K., Tanaka, M., Endo, T. A., Matsui, A., Ishida, J., et al. (2014). "Highly reproducible ChIP-on-chip analysis to identify genome-wide protein binding and chromatin status in *Arabidopsis thaliana*," in *Arabidopsis Protocols*, eds J. J. Sanchez-Serrano and J. Salinas (Berlin: Springer), 405–426. doi: 10.1007/978-1-62703-580-4_22

Kitagawa, R., Mitsuki, H., Okazaki, T., and Ogawa, T. (1996). A novel DnaA protein-binding site at 94.7 min on the *Escherichia coli* chromosome. *Mol. Microbiol.* 19, 1137–1147. doi: 10.1046/j.1365-2958.1996.453983.x

Komori, K., and Ishino, Y. (2001). Replication protein A in *Pyrococcus furiosus* is involved in homologous DNA recombination. *J. Biol. Chem.* 276, 25654–25660. doi: 10.1074/jbc.M102423200

Kozlov, A. G., and Lohman, T. M. (2012). "SSB binding to ssDNA using isothermal titration calorimetry," in *Single-Stranded DNA Binding Proteins*, ed. J. L. Keck (New York, NY: Springer), 37–54. doi: 10.1007/978-1-62703-032-8_3

Ladd, J., Lu, H., Taylor, A. D., Goodell, V., Disis, M. L., and Jiang, S. (2009). Direct detection of carcinoembryonic antigen autoantibodies in clinical human serum samples using a surface plasmon resonance sensor. *Colloids Surf. B Biointerfaces* 70, 1–6. doi: 10.1016/j.colsurfb.2008.11.032

Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 22, 1813–1831. doi: 10.1101/gr.136184.111

Lane, D., Prentki, P., and Chandler, M. (1992). Use of gel retardation to analyze protein-nucleic acid interactions. *Microbiol. Rev.* 56, 509–528.

Leblanc, B., and Moss, T. (2001). "DNase I footprinting," in *DNA-Protein Interactions,* ed. T. Moss (New York, NY: Humana Press), 31–38. doi: 10.1385/1-59259-208-2:031

Lee, H., and Romero, J. (2012). Origin of DNA replication at the human lamin B2 locus: OBR or ABR? *Cell Cycle* 11, 4281–4283. doi: 10.4161/cc.21992

Liedberg, B., Nylander, C., and Lundström, I. (1995). Biosensing with surface plasmon resonance—how it all started. *Biosens. Bioelectron.* 10, i–ix. doi: 10.1016/0956-5663(95)96965-2

Liu, J., Pei, H., Mei, S., Li, J., Zhou, L., and Xiang, H. (2008). Replication initiator DnaA interacts with an anti-terminator NusG in *T. tengcongensis*. *Biochem. Biophys. Commun.* 371, 573–577. doi: 10.1016/j.bbrc.2008.04.131

Liu, Y., Dong, Y., Jauw, J., Linman, M. J., and Cheng, Q. (2010). Highly sensitive detection of protein toxins by surface plasmon resonance with biotinylation-based inline atom transfer radical polymerization amplification. *Anal. Chem.* 82, 3679–3685. doi: 10.1021/ac1000114

Lundgren, M., Andersson, A., Chen, L., Nilsson, P., and Bernander, R. (2004). Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. *Proc. Natl. Acad. Sci. U.S.A.* 101, 7046–7051. doi: 10.1073/pnas.0400656101

Luo, H., Zhang, C.-T., and Gao, F. (2014). Ori-Finder 2, an integrated tool to predict replication origins in the archaeal genomes. *Front. Microbiol.* 5:482. doi: 10.3389/fmicb.2014.00482

Lyon, L. A., Musick, M. D., and Natan, M. J. (1998). Colloidal Au-enhanced surface plasmon resonance immunosensing. *Anal. Chem.* 70, 5177–5183. doi: 10.1021/ac9809940

Maisnier-Patin, S., Malandrin, L., Birkeland, N. K., and Bernander, R. (2002). Chromosome replication patterns in the hyperthermophilic euryarchaea *Archaeoglobus fulgidus* and *Methanocaldococcus* (Methanococcus) jannaschii. *Mol. Microbiol.* 45, 1443–1450. doi: 10.1046/j.1365-2958.2002.03111.x

Malott, M., and Leffak, M. (1999). Activity of the c-myc replicator at an ectopic chromosomal location. *Mol. Cell. Biol.* 19, 5685–5695.

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.embor.7400649

Martin, M. M., Ryan, M., Kim, R., Zakas, A. L., Fu, H., Lin, C. M., et al. (2011). Genome-wide depletion of replication initiation events in highly transcribed regions. *Genome Res.* 21, 1822–1832. doi: 10.1101/gr.124644.111

Martinez, J. C., Murciano-Calles, J., Cobos, E. S., Iglesias-Bexiga, M., Luque, I., and Ruiz-Sanz, J. (2013). "Isothermal titration calorimetry: thermodynamic analysis of the binding thermograms of molecular recognition events by using equilibrium models," in *Applications of Calorimetry in a Wide Context – Differential Scanning Calorimetry, Isothermal Titration Calorimetry and Microcalorimetry,* Chap. 4, ed. A. A. Elkordy (Rijeka: INTECH). doi: 10.5772/53311

Matsunaga, F., Forterre, P., Ishino, Y., and Myllykallio, H. (2001). In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11152–11157. doi: 10.1073/pnas.191387498

Matsunaga, F., Glatigny, A., Mucchielli-Giorgi, M.-H., Agier, N., Delacroix, H., Marisa, L., et al. (2007). Genomewide and biochemical analyses of DNA-binding activity of Cdc6/Orc1 and Mcm proteins in *Pyrococcus* sp. *Nucleic Acids Res.* 35, 3214–3222. doi: 10.1093/nar/gkm212

Matsunaga, F., Norais, C., Forterre, P., and Myllykallio, H. (2003). Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin. *EMBO Rep.* 4, 154–158. doi: 10.1038/sj.embor.embor732

Matulis, D., Rouzina, I., and Bloomfield, V. A. (2000). Thermodynamics of DNA binding and condensation: isothermal titration calorimetry and electrostatic mechanism. *J. Mol. Biol.* 296, 1053–1063. doi: 10.1006/jmbi.1999.3470

Memelink, J. (2013). "Electrophoretic mobility shift assay for the analysis of interactions of jasmonic acid-responsive transcription factors with dna," in *Jasmonate Signaling*, eds A. Goossens and L. Pauwels (Berlin: Springer), 209–225. doi: 10.1007/978-1-62703-414-2_17

Messer, W. (2002). The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiol. Rev.* 26, 355–374. doi: 10.1111/j.1574-6976.2002.tb00620.x

Messer, W., Blaesing, F., Jakimowicz, D., Krause, M., Majka, J., Nardmann, J., et al. (2001). Bacterial replication initiator DnaA. Rules for DnaA binding and roles of DnaA in origin unwinding and helicase loading. *Biochimie* 83, 5–12. doi: 10.1016/S0300-9084(00)01216-5

Michalet, X., Pinaud, F. F., Bentolila, L. A., Tsay, J. M., Doose, S., Li, J. J., et al. (2005). Quantum dots for live cells, in vivo imaging, and diagnostics. *Science* 307, 538–544. doi: 10.1126/science.1104274

Michel, B., and Bernander, R. (2014). Chromosome replication origins: do we really need them? *Bioessays* 36, 585–590. doi: 10.1002/bies.201400003

Morozova, O., and Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92, 255–264. doi: 10.1016/j.ygeno.2008.07.001

Murphy, M. B., Fuller, S. T., Richardson, P. M., and Doyle, S. A. (2003). An improved method for the in vitro evolution of aptamers and applications in protein detection and purification. *Nucleic Acids Res.* 31, e110–e110. doi: 10.1093/nar/gng110

Myllykallio, H., Lopez, P., López-García, P., Heilig, R., Saurin, W., Zivanovic, Y., et al. (2000). Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* 288, 2212–2215. doi: 10.1126/science.288.5474.2212

Nawotka, K. A., and Huberman, J. A. (1988). Two-dimensional gel electrophoretic method for mapping DNA replicons. *Mol. Cell. Biol.* 8, 1408–1413. doi: 10.1128/MCB.8.4.1408

Newlon, C. S. (1996). *DNA Replication in Eukaryotic Cells*. Plainview, NY: Cold Spring Harbor Laboratory Press, 873–914.

Newman, G., and Crooke, E. (2000). DnaA, the initiator of *Escherichia coli* chromosomal replication, is located at the cell membrane. *J. Bacteriol.* 182, 2604–2610. doi: 10.1128/JB.182.9.2604-2610.2000

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 *Haloferax volcanii*. *PLoS Genet.* 3:e77. doi: 10.1371/journal.pgen.0030077

Nozaki, S., and Ogawa, T. (2008). Determination of the minimum domain II size of *Escherichia coli* DnaA protein essential for cell viability. *Microbiology* 154(Pt 11), 3379–3384. doi: 10.1099/mic.0.2008/019745-0

O'Donnell, M., Langston, L., and Stillman, B. (2013). Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb. Perspect. Biol.* 5:a010108. doi: 10.1101/cshperspect.a010108

Ogura, Y., Imai, Y., Ogasawara, N., and Moriya, S. (2001). Autoregulation of the dnaA-dnaN operon and effects of DnaA protein levels on replication initiation in *Bacillus subtilis*. *J. Bacteriol.* 183, 3833–3841. doi: 10.1128/JB.183.13.3833-3841.2001

Orlando, V. (2000). Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem. Sci.* 25, 99–104. doi: 10.1016/S0968-0004(99)01535-2

Ozaki, S., and Katayama, T. (2009). DnaA structure, function, and dynamics in the initiation at the chromosomal origin. *Plasmid* 62, 71–82. doi: 10.1016/j.plasmid.2009.06.003

Ozaki, S., and Katayama, T. (2012). Highly organized DnaA–oriC complexes recruit the single-stranded DNA for replication initiation. *Nucleic Acids Res.* 40, 1648–1665. doi: 10.1093/nar/gkr832

Ozaki, S., Kawakami, H., Nakamura, K., Fujikawa, N., Kagawa, W., Park, S. Y., et al. (2008). A common mechanism for the ATP–DnaA-dependent formation of open complexes at the replication origin. *J. Biol. Chem.* 283, 8351–8362. doi: 10.1074/jbc.M708684200

Pan, Y., Karns, K., and Herr, A. E. (2014). Microfluidic electrophoretic mobility shift assays for quantitative biochemical analysis. *Electrophoresis* 35, 2078–2090. doi: 10.1002/elps.201300500

Parab, H. J., Jung, C., Lee, J.-H., and Park, H. G. (2010). A gold nanorod-based optical DNA biosensor for the diagnosis of pathogens. *Biosens. Bioelectron.* 26, 667–673. doi: 10.1016/j.bios.2010.06.067

Pattnaik, P. (2005). Surface plasmon resonance. *Appl. Biochem. Biotechnol.* 126, 79–92. doi: 10.1385/ABAB:126:2:079

Pei, H., Liu, J., Li, J., Guo, A., Zhou, J., and Xiang, H. (2007). Mechanism for the TtDnaA–Tt-oriC cooperative interaction at high temperature and duplex opening at an unusual AT-rich region in *Thermoanaerobacter tengcongensis*. *Nucleic Acids Res.* 35, 3087–3099. doi: 10.1093/nar/gkm137

Pelve, E. A., Lindås, A. C., Knöppel, A., Mira, A., and Bernander, R. (2012). Four chromosome replication origins in the archaeon *Pyrobaculum calidifontis*. *Mol. Microbiol.* 85, 986–995. doi: 10.1111/j.1365-2958.2012.08155.x

Pelve, E. A., Martens-Habbena, W., Stahl, D. A., and Bernander, R. (2013). Mapping of active replication origins in vivo in thaum- and euryarchaeal replicons. *Mol. Microbiol.* 90, 538–550. doi: 10.1111/mmi.12382

Pernites, R. B., Ponnapati, R. R., and Advincula, R. C. (2010). Surface plasmon resonance (SPR) detection of theophylline via electropolymerized molecularly imprinted polythiophenes. *Macromolecules* 43, 9724–9735. doi: 10.1021/ma101868y

Pierce, M. M., Raman, C., and Nall, B. T. (1999). Isothermal titration calorimetry of protein–protein interactions. *Methods* 19, 213–221. doi: 10.1006/meth.1999.0852

Raghuraman, M., Winzeler, E. A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., et al. (2001). Replication dynamics of the yeast genome. *Science* 294, 115–121. doi: 10.1126/science.294.5540.115

Richter, S., and Messer, W. (1995). Genetic structure of the dnaA region of the cyanobacterium *Synechocystis* sp. strain PCC6803. *J. Bacteriol.* 177, 4245–4251.

Riddick, L., and Brumley, W. C. (2008). "Capillary electrophoresis with laser-induced fluorescence," in *Capillary Electrophoresis*, ed. P. Schmitt-Kopplin (New York, NY: Humana Press), 119–134. doi: 10.1007/978-1-59745-376-9_6

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., et al. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* 4, 651–657. doi: 10.1038/nmeth1068

Robinson, N. P., and Bell, S. D. (2007). Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5806–5811. doi: 10.1073/pnas.0700206104

Robinson, N. P., Dionne, I., Lundgren, M., Marsh, V. L., Bernander, R., and Bell, S. D. (2004). Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell* 116, 25–38. doi: 10.1016/S0092-8674(03)01034-1

Roth, A., and Messer, W. (1995). The DNA binding domain of the initiator protein DnaA. *EMBO J.* 14, 2106–2111.

Roth, A., and Messer, W. (1998). High-affinity binding sites for the initiator protein DnaA on the chromosome of *Escherichia coli*. *Mol. Microbiol.* 28, 395–401. doi: 10.1046/j.1365-2958.1998.00813.x

Sahai, N. (2011). Surface plasmon resonance. *Encycl. Astrobiol.* 1642–1642. doi: 10.1007/978-3-642-11274-4_1551

Sawhney, S., and Singh, R. (2000). *Introductory Practical Biochemistry*. Oxford: Alpha Science International Limited.

Schaper, S., Nardmann, J., Lüder, G., Lurz, R., Speck, C., and Messer, W. (2000). Identification of the chromosomal replication origin from *Thermus thermophilus* and its interaction with the replication initiator DnaA. *J. Mol. Biol.* 299, 655–665. doi: 10.1006/jmbi.2000.3764

Schmidt, D., Wilson, M. D., Spyrou, C., Brown, G. D., Hadfield, J., and Odom, D. T. (2009). ChIP-seq: using high-throughput sequencing to discover protein–DNA interactions. *Methods* 48, 240–248. doi: 10.1016/j.ymeth.2009.03.001

Sekimizu, K., Yung, B. Y.-M., and Kornberg, A. (1988). The dnaA protein of *Escherichia coli*. Abundance, improved purification, and membrane binding. *J. Biol. Chem.* 263, 7136–7140.

Shah, A. (2009). Chromatin immunoprecipitation sequencing (ChIP-Seq) on the SOLiD™ system. *Nat. Methods* 6.

Shahravan, S. H., Li, I. T., Truong, K., and Shin, J. A. (2011). FRep: a fluorescent protein-based bioprobe for in vivo detection of protein–dna interactions. *Anal. Chem.* 83, 9643–9650. doi: 10.1021/ac2024602

Sigurskjold, B. W. (2000). Exact analysis of competition ligand binding by displacement isothermal titration calorimetry. *Anal. Biochem.* 277, 260–266. doi: 10.1006/abio.1999.4402

Singh, M., Holzinger, M., Tabrizian, M., Winters, S., Berner, N. C., Cosnier, S., et al. (2015). Non-covalently functionalized monolayer graphene for sensitivity enhancement of surface plasmon resonance immunosensors. *J. Am. Chem. Soc.* 137, 2800–2803. doi: 10.1021/ja511512m

Skarstad, K., and Katayama, T. (2013). Regulating DNA replication in bacteria. *Cold Spring Harb. Perspect. Biol.* 5:a012922. doi: 10.1101/cshperspect.a012922

Smith, A. J., and Humphries, S. E. (2009). Characterization of DNA-binding proteins using multiplexed competitor EMSA. *J. Mol. Biol.* 385, 714–717. doi: 10.1016/j.jmb.2008.11.035

Speck, C., and Messer, W. (2001). Mechanism of origin unwinding: sequential binding of DnaA to double– and single-stranded DNA. *EMBO J.* 20, 1469–1476. doi: 10.1093/emboj/20.6.1469

Stead, J. A., Keen, J. N., and Mcdowall, K. J. (2006). The identification of nucleic acid-interacting proteins using a simple proteomics-based approach that directly incorporates the electrophoretic mobility shift assay. *Mol. Cell. Proteom.* 5, 1697–1702. doi: 10.1074/mcp.T600027-MCP200

Struhl, K., Stinchcomb, D. T., Scherer, S., and Davis, R. W. (1979). High-frequency transformation of yeast: autonomous replication of hybrid DNA molecules. *Proc. Natl. Acad. Sci. U.S.A.* 76, 1035–1039. doi: 10.1073/pnas.76.3.1035

Sun, C., Zhou, M., Li, Y., and Xiang, H. (2006). Molecular characterization of the minimal replicon and the unidirectional theta replication of pSCM201 in extremely halophilic archaea. *J. Bacteriol.* 188, 8136–8144. doi: 10.1128/JB.00988-06

Sutton, M. D., and Kaguni, J. M. (1997). Novel alleles of the *Escherichia coli* dnaA gene. *J. Mol. Biol.* 271, 693–703. doi: 10.1006/jmbi.1997.1209

Špringer, T. S., Ermini, M. L., ŠPačKová, B., JabloňKů, J., and Homola, J. (2014). Enhancing sensitivity of surface plasmon resonance biosensors by functionalized gold nanoparticles: size matters. *Anal. Chem.* 86, 10350–10356. doi: 10.1021/ac502637u

Tao, L., Dong, Z., Leffak, M., Zannis-Hadjopoulos, M., and Price, G. (2000). Major DNA replication initiation sites in the c-myc locus in human cells. *J. Cell. Biochem.* 78, 442–457. doi: 10.1002/1097-4644(20000901)78:3<442::AID-JCB9>3.0.CO;2-1

Taylor, J. A., Ouimet, M. C., Wargachuk, R., and Marczynski, G. T. (2011). The *Caulobacter crescentus* chromosome replication origin evolved two classes of weak DnaA binding sites. *Mol. Microbiol.* 82, 312–326. doi: 10.1111/j.1365-2958.2011.07785.x

Testa, A., Donati, G., Yan, P., Romani, F., Huang, T. H.-M., Viganò, M. A., et al. (2005). Chromatin immunoprecipitation (ChIP) on chip experiments

uncover a widespread distribution of NF-Y binding CCAAT sites outside of core promoters. *J. Biol. Chem.* 280, 13606–13615. doi: 10.1074/jbc.M414039200

Truax, A. D., and Greer, S. F. (2012). "ChIP and Re-ChIP assays: investigating interactions between regulatory proteins, histone modifications, and the DNA sequences to which they bind," in *Transcriptional Regulation*, ed. A. Vancura (New York, NY: Springer), 175–188. doi: 10.1007/978-1-61779-376-9_12

Tsai, C., Smider, V., Hwang, B. J., and Chu, G. (2012). "Electrophoretic mobility shift assays for protein–DNA complexes involved in DNA repair," in *DNA Repair Protocols*, ed. L. Bjergbæk (New York, NY: Humana Press), 53–78. doi: 10.1007/978-1-61779-998-3_5

Van Dyke, M. W., and Dervan, P. B. (1983). Methidiumpropyl-EDTA-Fe (II) and DNase I footprinting report different small molecule binding site sizes on DNA. *Nucleic Acids Res.* 11, 5555–5567. doi: 10.1093/nar/11.16.5555

Velazquez-Campoy, A., and Freire, E. (2006). Isothermal titration calorimetry to determine association constants for high-affinity ligands. *Nat. Protoc.* 1, 186–191. doi: 10.1038/nprot.2006.28

Velazquez-Campoy, A., Leavitt, S. A., and Freire, E. (2004). "Characterization of protein-protein interactions by isothermal titration calorimetry," in *Protein-Protein Interactions*, ed. H. Fu (New York, NY: Humana Press), 35–54. doi: 10.1385/1-59259-762-9:035

Vernis, L., Chasles, M., Pasero, P., Lepingle, A., Gaillardin, C., and Fournier, P. (1999). Short DNA fragments without sequence similarity are initiation sites for replication in the chromosome of the yeast *Yarrowia lipolytica*. *Mol. Biol. Cell* 10, 757–769. doi: 10.1091/mbc.10.3.757

Vinckevicius, A., and Chakravarti, D. (2012). Chromatin immunoprecipitation: advancing analysis of nuclear hormone signaling. *J. Mol. Endocrinol.* 49, R113–R123. doi: 10.1530/JME-12-0016

Waga, S., and Stillman, B. (1998). The DNA replication fork in eukaryotic cells. *Annu. Rev. Biochem.* 67, 721–751. doi: 10.1146/annurev.biochem.67.1.721

Wahle, E., Lasken, R., and Kornberg, A. (1989). The dnaB-dnaC replication protein complex of *Escherichia coli*. I. Formation and properties. *J. Biol. Chem.* 264, 2463–2468.

Wang, J. (2005). Nanomaterial-based amplified transduction of biomolecular interactions. *Small* 1, 1036–1043. doi: 10.1002/smll.200500214

Wang, M. S., and Reed, S. M. (2012). Direct visualization of electrophoretic mobility shift assays using nanoparticle–aptamer conjugates. *Electrophoresis* 33, 348–351. doi: 10.1002/elps.201100308

Wang, Y., Knoll, W., and Dostalek, J. (2012). Bacterial pathogen surface plasmon resonance biosensor advanced by long range surface plasmons and magnetic nanoparticle assays. *Anal. Chem.* 84, 8345–8350. doi: 10.1021/ac301904x

Ward, W. H., and Holdgate, G. A. (2001). 7 isothermal titration calorimetry in drug discovery. *Prog. Med. Chem.* 38, 309–376. doi: 10.1016/S0079-6468(08)70097-3

Wei, T., and Bernander, R. (1996). Interaction of the IciA protein with AT-rich regions in plasmid replication origins. *Nucleic Acids Res.* 24, 1865–1872. doi: 10.1093/nar/24.10.1865

Weigel, C., Schmidt, A., Seitz, H., Tüngler, D., Welzeck, M., and Messer, W. (1999). The N-terminus promotes oligomerization of the *Escherichia coli* initiator protein DnaA. *Mol. Microbiol.* 34:53–66. doi: 10.1046/j.1365-2958.1999.01568.x

Wilson, S., and Howell, S. (2002). High-throughput screening in the diagnostics industry. *Biochem. Soc. Trans.* 30, 794–796. doi: 10.1042/bst0300794

Wolański, M., Donczew, R., Zawilak-Pawlik, A., and Zakrzewska-Czerwińska, J. (2014). oriC-encoded instructions for the initiation of bacterial chromosome replication. *Front. Microbiol.* 5:735. doi: 10.3389/fmicb.2014.00735

Woo, A. J., Dods, J. S., Susanto, E., Ulgiati, D., and Abraham, L. J. (2002). A proteomics approach for the identification of DNA binding activities observed in the electrophoretic mobility shift assay. *Mol. Cell. Proteomics* 1, 472–478. doi: 10.1074/mcp.T200003-MCP200

Wu, S., Wang, J., Zhao, W., Pounds, S., and Cheng, C. (2010). ChIP-PaM: an algorithm to identify protein-DNA interaction using ChIP-Seq data. *Theor. Biol. Med. Model.* 7, 18. doi: 10.1186/1742-4682-7-18

Wu, Z., Liu, J., Yang, H., Liu, H., and Xiang, H. (2013). Multiple replication origins with diverse control mechanisms in *Haloarcula hispanica*. *Nucleic Acids Res.* 42, gkt1214. doi: 10.1093/nar/gkt1214

Wyrick, J., Young, R. A., Ren, B., Francois, R., and Simon, I. (2009). Genome wide location and function of DNA binding proteins. *Science* 290, 2306–2309. doi: 10.1126/science.290.5500.2306

Zawilak, A., Durrant, M. C., Jakimowicz, P., Backert, S., and Zakrzewska-Czerwińska, J. (2003). DNA binding specificity of the replication initiator protein, DnaA from *Helicobacter pylori*. *J. Mol. Biol.* 334, 933–947. doi: 10.1016/j.jmb.2003.10.018

Zhang, L., Zhang, L., Liu, Y., Yang, S., Gao, C., Gong, H., et al. (2009). Archaeal eukaryote-like Orc1/Cdc6 initiators physically interact with DNA polymerase B1 and regulate its functions. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7792–7797. doi: 10.1073/pnas.0813056106

Zhang, R., and Zhang, C.-T. (2003). Multiple replication origins of the archaeon *Halobacterium* species NRC-1. *Biochem. Biophys. Res. Commun.* 302, 728–734. doi: 10.1016/S0006-291X(03)00252-3

Zhang, R., and Zhang, C.-T. (2005). Identification of replication origins in archaeal genomes based on the Z-curve method. *Archaea* 1, 335–346. doi: 10.1155/2005/509646

Zhang, Y., Akilesh, S., and Wilcox, D. E. (2000). Isothermal titration calorimetry measurements of Ni (II) and Cu (II) binding to His, GlyGlyHis, HisGlyHis, and bovine serum albumin: a critical evaluation. *Inorg. Chem.* 39, 3057–3064. doi: 10.1021/ic000036s

Zhou, Y., Larson, J. D., Bottoms, C. A., Arturo, E. C., Henzl, M. T., Jenkins, J. L., et al. (2008). Structural basis of the transcriptional regulation of the proline utilization regulon by multifunctional PutA. *J. Mol. Biol.* 381, 174–188. doi: 10.1016/j.jmb.2008.05.084

Zianni, M., Tessanne, K., Merighi, M., Laguna, R., and Tabita, F. (2006). Identification of the DNA bases of a DNase I footprint by the use of dye primer sequencing on an automated capillary DNA analysis instrument. *J. Biomol. Tech.* 17, 103–113.

Zijlstra, P., Paulo, P. M., and Orrit, M. (2012). Optical detection of single non-absorbing molecules using the surface plasmon resonance of a gold nanorod. *Nat. Nanotechnol.* 7, 379–382. doi: 10.1038/nnano.2012.51

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.

Copyright © 2015 Song, Zhang and Huang. 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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 10 December 2015
doi: 10.3389/fmicb.2015.01370

Identification of the Replication Origins from *Cyanothece* ATCC 51142 and Their Interactions with the DnaA Protein: From *In Silico* to *In Vitro* Studies

*He Huang[1,2,3], Cheng-Cheng Song[1,2,3], Zhi-Liang Yang[1,2,3], Yan Dong[1,2,3], Yao-Zhong Hu[1,2,3] and Feng Gao[2,3,4]**

[1] Department of Biochemical Engineering, School of Chemical Engineering and Technology, Tianjin University, Tianjin, China, [2] Key Laboratory of Systems Bioengineering, Ministry of Education, Tianjin University, Tianjin, China, [3] Collaborative Innovation Center of Chemical Science and Engineering, Tianjin, China, [4] Department of Physics, Tianjin University, Tianjin, China

OPEN ACCESS

Edited by:
Frank T. Robb, University of Maryland, Baltimore, USA

Reviewed by:
Gregory Marczynski, McGill University, Canada
Justine Collier, University of Lausanne, Switzerland

***Correspondence:**
Feng Gao
fgao@tju.edu.cn

Specialty section:
This article was submitted to Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *22 July 2015*
Accepted: *17 November 2015*
Published: *10 December 2015*

Citation:
Huang H, Song C-C, Yang Z-L, Dong Y, Hu Y-Z and Gao F (2015) Identification of the Replication Origins from Cyanothece ATCC 51142 and Their Interactions with the DnaA Protein: From In Silico to In Vitro Studies. Front. Microbiol. 6:1370. doi: 10.3389/fmicb.2015.01370

Based on the complete genome of *Cyanothece* ATCC 51142, the *oriC*s of both the circular and linear chromosomes in *Cyanothece* ATCC 51142 have been predicted by utilizing a web-based system Ori-Finder. Here, we provide experimental support for the results of Ori-Finder to identify the replication origins of *Cyanothece* ATCC 51142 and their interactions with the initiator protein, DnaA. The two replication origins are composed of three characteristically arranged DnaA boxes and an AT-rich stretch, and the *oriC* in the circular chromosome is followed by the *dnaN* gene. The *dnaA* gene is located downstream of the origin of the circular chromosome and it expresses a typical DnaA protein that is divided into four domains (I, II, III, IV), as with other members of the DnaA protein family. We purify DnaA (IV) and characterize the interaction of the purified protein with the replication origins, so as to offer experimental support for the prediction. The results of the electrophoretic mobility shift assay and DNase I footprint assay demonstrate that the C-terminal domain of the DnaA protein from *Cyanothece* ATCC 51142 specifically binds the *oriC*s of both the circular and linear chromosomes, and the DNase I footprint assay demonstrates that DnaA (IV) exhibits hypersensitive affinity with DnaA boxes in both *oriC*s.

Keywords: DnaA, DnaA (IV), DNA binding, origin of chromosomal replication (*oriC*), initiation complex, *Cyanothece* ATCC 51142

INTRODUCTION

The replication initiation of bacteria requires two basic elements: the discrete origin of replication (*oriC*) for positively acting replication factors and the initiator protein (DnaA) to which other replication proteins bind, promoting origin unwinding and the subsequent initiation of DNA polymerization (Duderstadt and Berger, 2008). The initiation of bacterial chromosome replication

is mediated by the initiator protein DnaA, which recognizes and specifically binds to non-palindromic, repetitive, 9-mer consensus sequences termed "DnaA boxes". DnaA boxes are present in most replication origins in bacterial chromosomes, as well as in the replication origins of some plasmids (Messer, 2002; Mott and Berger, 2007). The replication origins vary in terms of both the size and number of DnaA boxes across different species.

The initiator protein, DnaA, is a key protein in the initiation of chromosome replication. DnaA is highly conserved among different bacteria (Fujita et al., 1990; Kaguni, 1997). The bacterial consensus DnaA box sequence, 5′-TT$^{A}/_{T}$TNCACA-3′, is highly conserved with only one or two nt differences. As for *Synechococcus* sp. Strain PCC 7942, the consensus sequence of the DnaA box is TTTTCCACA, as it was found in seven of the eleven repeats (Liu and Tsinoremas, 1996). Six DnaA boxes of *Anabaena* sp. strain PCC 7120, were predicted by Ori-Finder; their sequence is TTTTCCACA, and the assays confirmed the predicted result (Zhou et al., 2011). DnaA was assigned to four functional domains, I, II, III, and IV based on the degree of sequence similarity (Roth and Messer, 1995; Erzberger et al., 2002). The ssDNA-binding activity of DnaA domain I is weak (Abe et al., 2007); however, the interactions between domain I and several proteins, including domain I itself, as well as DnaB helicase, and the initiation stimulator DiaA, are required in order for DnaB helicase to load onto *oriC* open complex flexible linker (Weigel et al., 1999; Felczak and Kaguni, 2004; Abe et al., 2007; Keyamura et al., 2007; Nozaki and Ogawa, 2008). Domain II is highly variable in sequence and length in different bacteria (Messer, 2002). Domain III plays a major role in adenosine triphosphate (ATP) and adenosine diphosphate (ADP) binding, as well as in ATP-dependent conformational changes of the DnaA multimer on *oriC*, in binding ssDNA of the *oriC* duplex unwinding element (DUE), and in ATP hydrolysis (Katayama, 2008; Ozaki et al., 2008). Domain IV, the DNA-binding region [denoted as DnaA (IV) herein], comprises three potential α-helices that feature a highly conserved basic loop and a long connector helix (α12) linked to a helix-turn-helix (HTH) motif (α15 and α16), which is buttressed by two additional helices (α14 and α17) (Erzberger et al., 2002). In *Escherichia coli*, the DnaA protein is also a transcriptional regulator; it is autoregulated and interferes with the activity of other genes through DnaA binding at their promoter region (Messer and Weigel, 1997). DnaA may either bind to a DnaA box downstream of the promoter region to block transcription, such as the *mioC* promoter in *E. coli* (Theisen et al., 1993; Blaesing et al., 2000), or it may activate the expression of a gene such as *fliC* (Mizushima et al., 1994). Furthermore, with the respect to sporulation in *B. subtilis*, the DnaA protein not only initiates DNA replication, but it also regulates other aspects of cell activities (Veening et al., 2009).

Cyanobacteria, also called blue–green bacteria, blue–green algae, cyanophyceae, or cyanophytes, represent a large and widespread group of photoautotrophic microorganisms (Stanier and Bazine, 1977; Whitton and Potts, 2012). Cyanobacteria are the only diazotrophs that produce molecular oxygen as a byproduct of photosynthesis; they have evolved a variety of mechanisms to accommodate the activity of an oxygen-sensitive enzyme (Berman-Frank et al., 2003). As we know, nitrogen fixation has played a crucial role in marine environments where the bioavailability of nitrogen determines the level of primary productivity (Montoya et al., 2004). *Cyanothece* sp. ATCC 51142 (hereafter referred to as *Cyanothece* 51142), a marine unicellular diazotrophic strain, features a robust diurnal cycle in which the processes of oxygenic photosynthesis and nitrogen fixation are performed and separated temporally within the same cell during the diurnal cycle (Sherman et al., 1998; Welsh et al., 2008a). The complete genome of *Cyanothece* 51142 was reported in 2008, and it was the first strain of the genus to be sequenced (Welsh et al., 2008a). In addition, *Cyanothece* 51142 was the first to report that there is a linear element in the genome of a photosynthetic bacterium (Welsh et al., 2008a). However, the origin of replication could not be determined for either the circular or linear chromosome using standard GC skew and DnaA box analysis at that time (Welsh et al., 2008a). By utilizing a web-based system called Ori-Finder[1], we have identified the locations of *oriC*s for both the circular and linear chromosomes in *Cyanothece* 51142 (Gao and Zhang, 2008b), which may provide clarity on the replication origins in *Cyanothece* 51142, as well as other cyanobacteria (Welsh et al., 2008b). Meanwhile, only experimental work will finally answer this question.

In the present study, we have amplified the putative *oriC*s of the circular and linear chromosomes of *Cyanothece* 51142 via polymerase chain reaction (PCR). We have also cloned the *dnaA* gene of *Cyanothece* 51142. Domain IV of the *dnaA* gene has been cloned into an overexpressing vector from *E. coli* for purification. According to the prediction of Ori-Finder, both of *oriC*s contain three putative DnaA boxes. Here, we characterize the interaction of the purified *Cyanothece* 51142 DnaA protein, domain IV [DnaA (IV)] with *oriC*s via gel retardation assays and a DNase I footprinting assay to provide experimental support for the prediction by Ori-Finder.

MATERIALS AND METHODS

Bacterial Strains, Media, and Culture Conditions

Escherichia coli DH5a (F$^-$,Φ80d*lac*ZΔM15, *rec*A1, *end*A1, *gyr*A96, *thi*-1, *hsd*R17, [$r_k{}^-$, mk^+], *sup*E44, *rel*A1, *deo*R, Δ[*lac*ZYA-*arg*F]U169) (Sambrook et al., 1989) served as a host for plasmids, while *E. coli* BL21 (DE3) acted as the host for the overproduction of recombinant proteins DnaA (IV) (**Table 1**). The *E. coli* strains were grown in Luria-Bertani (LB) medium at 37°C. The plasmids and oligonucleotides used in the present study were described in **Tables 1** and **2,** respectively. Antibiotics were used in the following concentrations: ampicillin (100 μg/mL) for *E. coli*, and kanamycin (50 μg/mL) for plasmids.

[1]http://tubic.tju.edu.cn/Ori-Finder/

TABLE 1 | Bacterial strains and plasmids used in this study.

Strain/plasmid	Genotype/relevant characteristics	Source
Escherichia coli DH5α	*supE*44, *hsdR*17, *recA*1, *endA*1, *gyrA*96, *thi*-1, *relA*1	Novagen
E. coli BL21 (DE3)	Overexpression of recombinant protein His_6-DnaA (IV)	Novagen
pET-28a(+)	Kan^r, expression vector, His-tag coding sequence	Novagen
pEASY-T1	Amp^r, Kan^r, T-vector for cloning PCR-amplified fragments	TransGen
pET28a-DnaA (IV)	Expression plasmid derived from pET-28a for the Domain IV of *Cyanothece* 51142 His_6-DnaA (IV)	This study
pEASY-*oriC*-C	Cloning plasmid derived from pEASY-T1 of *oriC* region of circular chromosome	This study
pEASY-*oriC*-L	Cloning plasmid derived from pEASY-T1 of *oriC* region of circular chromosome	This study

TABLE 2 | Oligonucleotides used in this study.

Names	Sequence(5′ to 3′)[a]
dnaA-F	ATGACTATTTCCCCTCAATATATTTGGAACC
dnaA-R	TTAATTTTGATTCCGACTAGCAATATTAATG
dnaA(IV)-F	CG**GGATCC**ATTTCCATTTCAGGACTATCGATGA
dnaA(IV)-R	CC**AAGCTT**TTAATTTTGATTCCGACTAGCAATA
oriC-C-F	GCTCTAACTTTCGCTTCTCTTGGAAGCCAT
oriC-C-R	ACGGCCCGACTGACTAAAGATAAATTGCT
oriC-L-F	CCGCCGTAGATATTTTAACGAAGGCCCCTC
oriC-L-R	TTGGTAAGCAACTTGTTGAGCCGTCAACTC
Cbox1F	GTCCTGTGGGGGTC TGTGGAAAA CATTAAAATTC
Cbox1R	GAATTTTAATGTTTTCCACAGACCCCCACAGGAC
Cbox23F[b]	CCTTTAAA TTTTTCCACA GT TTTTCCACA GCATTA
Cbox23R[b]	TAATGCTGTGGAAAAACTGTGGAAAAATTTAAAGG
Lbox23F[b]	ATCTTGAA TGTTGATAA ATTTAAGTAT TTATCCTCA ACATTG
Lbox23R[b]	CAATGTTGAGGATAAATACTTAAATTTATCAACATTCAAGAT
EboxF[b]	GATCCTTG TTATCCACA GGGCAGTGC
EboxR[b]	GCACTGCCCTGTGGATAACAAGGATC

[a] *Restriction sites are underlined; DnaA boxes are boxed.*
[b] *Prior to use in the EMSA assay, the oligonucleotides were annealed with their complementary oligonucleotides. The boxed bases represent the DnaA box(es). Ebox represents the E. coli DnaA box.*

DNA Manipulations

Plasmids and DNA fragments were purified using purification kits according to the manufacturer's protocols (TransGen, Beijing, China). The genomic DNA of *Cyanothece* 51142 was bought from the American Type Culture Collection (ATCC; Manassas, VA, USA). The *Cyanothece* 51142 *dnaA* gene was amplified using *dnaA*-F/*dnaA*-R as the primers and the genomic DNA of *Cyanothece* 51142 as the template (**Table 2**). The *dnaA* gene of domain IV of *Cyanothece* 51142 was amplified by PCR using the *dnaA* gene as the template and *dnaA* (IV)-F/*dnaA* (IV)-R as the primers (**Table 2**; **Figure 3**). To achieve high-level expression of the DnaA (IV) protein, the *dnaA* (IV) genes were subcloned into an expression vector of the pET series (TransGen) to produce $(His)_6$-tagged fusion proteins in conventional *E. coli* BL21 (DE3) cells. The coding region of the *dnaA* (IV) gene was cut out by *BamH* I and *Hind* III, and cloned in frame into the T7 promotor-driven expression vector pET-28a(+) (TransGen) using the same

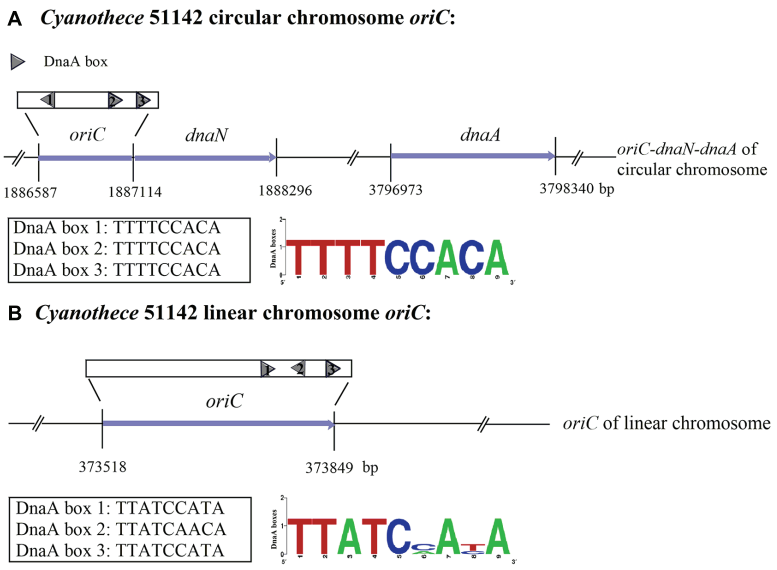


FIGURE 1 | The sequences and structure of the *Cyanothece* 51142 *oriC* regions (not to scale). The predicted DnaA boxes 1–3 of the circular and linear chromosomes and their orientation are represented by gray pentagons. The DnaA box sequence for the circular chromosome is all TTTTCCACA, as shown in the DnaA boxes logo. The sequences of the two DnaA boxes from the linear chromosome are TTATCCATA and the other one is TTATCAACA. **(A)** The *oriC* of the circular chromosome of *Cyanothece* 51142 was predicted to be next to the start of the *dnaN* gene, and the *dnaA* gene is located downstream of the origin of the circular chromosome. The sequence of DnaA boxes is TTTTCCACA. **(B)** The *oriC* of the linear chromosome has been found in the region from 373518 to 373849 nt. The three DnaA boxes of both circular and linear chromosomes have been tagged.

FIGURE 2 | The supplemental phylogenetic tree of cyanobacteria and information about the predicted *oriC* regions. The figure was drawn in a similar style to that of Zhou et al. (2011), in which 38 species of cyanobacteria were shown. The phylogenetic tree was constructed with *dnaN* genes from 68 species of cyanobacteria. The sequences of DnaA boxes in most of the cyanobacteria are TTTTCCACA. Fifty-five species (indicated by an asterisk) have *dnaN* on one side, and the genes located on the other side of *oriC* are indicated on the right. Note that the *oriC* regions of the *Synechocystis* sp. PCC 6803 substr. GT-I, substr. PCC-N, and substr. PCC-P overlap the membrane protein coding sequences.

restriction sites. The pET-28a(+) vector contains an N-terminal His·Tag and an optional C-terminal His·Tag sequence, which can be used as an affinity ligand for purification purposes. The authenticity of the pET28a-DnaA (IV) construction was verified by sequencing both strands. Enzymes were supplied by TransGen Biotech (Beijing, China) and TaKara (Dalian, China). The oligonuucleotides used for PCR were from TransGen (Beijing, China).

To clone the chromosomal replication origin regions of *Cyanothece* 51142, the putative *oriC* region of the circular chromosome located between ORF cce_1862 and the *dnaN* gene (cce_1864) (from 1,886,587 to 1,887,114 nt) (Gao and Zhang, 2008a) was amplified by PCR with primers *oriC*-C-F and *oriC*-C-R (**Table 2**). This *oriC* fragment was inserted into the pEASY-T1 (TransGen Biotech, China; **Table 1**), resulting in the plasmid pEASY-*oriC*-C; the putative *oriC* region of the

linear chromosome located between 2 ORFs, cce_5168, and cce_5169 (from 373,518 to 373,849 nt) (Gao and Zhang, 2008a) was amplified by PCR with primer *oriC*-L-F and *oriC*-L-R (**Table 2**). The fragment then also was inserted into pEASY-T1, resulting in the plasmid pEASY-*oriC*-L. The two recombinant vectors were used for subcloning or as templates for *oriC* sequencing.

DnaA (IV) Protein Purification

The *E. coli* BL21 (DE3) cells were transformed with the pET28a-DnaA (IV) plasmid. The pI value (5.67) and molecular weight (14,069.0 Da) of DnaA (IV) were predicted with ProtParam. A 100 mL culture, in LB broth supplemented with 50 μg/mL of kanamycin, was induced by the addition of 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG) at $A_{600\ nm} = 0.6$–0.8. Incubation then continued at 20°C for 12 h. The cells were harvested by centrifugation (10,000 *g*, 5 min, 4°C). The pellet was washed twice with phosphate buffered saline (PBS) (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) and subsequently centrifuged (10,000 *g*, 5 min, 4°C). Then, the pellet was frozen at −80°C and retained until required for the further purification steps. The bacterial pellet was thawed and suspended in His-tag binding buffer (10 mM Na_2HPO_4, 10 mM KH_2PO_4, 0.5 M NaCl, 20 mM imidazole, pH 7.8) (20 mL/g of wet biomass, ice bath). The lysozyme was added to a final concentration of 1 mg/mL and the cell suspension was incubated on ice for 30 min. The cells were lysed by sonication (ice bath) for 1 h and centrifuged at 3,000 *g* for 10 min, 4°C. The supernatant was purified using a 6× His trap column according to the standard protocol of AKTA prime plus. The sample was loaded onto a Ni-NTA (Ni^{2+}-nitrilotriacetate)-agarose column, previously equilibrated with His-tag binding buffer. The His-tag DnaA (IV) protein was eluted after washing with the His-tag elution buffer (10 mM Na_2HPO_4, 10 mM KH_2PO_4, 0.5 M NaCl, 0.5 M imidazole, pH 7.8). The purified DnaA (IV) protein was analyzed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE). Protein concentrations were determined to be about 1300 μg/mL by using the BCATM protein assay kit (PIERCE). The purified DnaA (IV) protein was checked by SDS/PAGE, and the protein purity was >98%.

SDS-PAGE

SDS-PAGE was performed according to the method established by Laemmli (1970). The purified protein was separated by PAGE (5% stacking gel, 15% separating gel). Gels were analyzed by a Vilber Lourmat Fusion 3500 Molecular Imager and using the ImageQuant software program.

Electrophoretic Mobility Shift Assay

Unless otherwise indicated, electrophoretic mobility shift assay (EMSA) was carried out, as described by Schaper and Messer (1995). Here, we used PCR fragments encompassing the *oriC* regions, as well as the double-stranded oligonucleotides containing various numbers and combinations of DnaA boxes from the origins. For the EMSA analysis, the predicted *oriC* regions of the circular and linear chromosomes were PCR-amplified using a pair of primers: *oriC*-C-F/R and *oriC*-L-F/R, respectively (**Table 2**). In each binding reaction, PCR fragments or nucleotides were incubated with various amounts of purified His_6-DnaA (IV) (2–20 μg) in the presence of a nonspecific competitor (poly [dI/dC]) at 37°C for 30 min in a binding buffer (20 mM Hepes/KOH [pH 8.0], 5 mM magnesium acetate, 1 mM Na_2EDTA, 4 mM dithiothreitol, 0.2% [v/v] Triton X-100, 3 mM ATP, and 5 mg/mL of bovine serum albumin [BSA]). The bound complexes were separated by electrophoresis in 8% polyacrylamide gels (0.25× TBE, at 4 V/cm, 4°C). Then, the gels were stained by 1× SYBR Green I solution. Gels were washed and analyzed via a Vilber Lourmat Fusion 3500 Molecular Imager and using the ImageQuant software program.

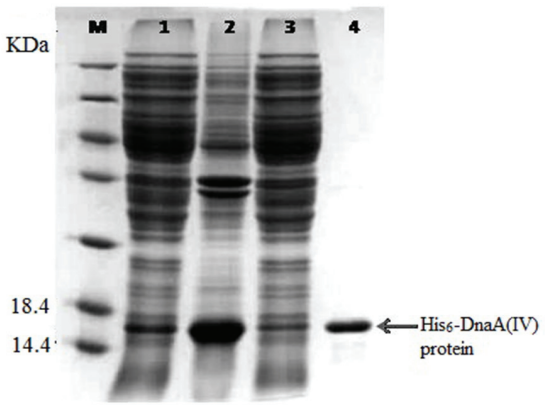


FIGURE 3 | Expression of the His_6-DnaA (IV) in *Escherichia coli* cells. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of the total *Cyanothece* 51142 cell proteins. Molecular mass markers (M) are indicated on the left and the positions of the protein are indicated on the right. Lane 1, supernatant of cell lysate; Lane 2, precipitant of cell lysate; Lane 3, breakthrough peak in the affinity chromatogram; Lane 4, targeted protein in the affinity chromatogram.

DNase I Footprinting Assay with FAM-labeled Primers

For the preparation of fluorescent FAM-labeled probes, the *oriC*s of circular and linear chromosomes were PCR amplified with Dpx DNA polymerase (TOLO Biotech, Shanghai, China) from the plasmids, pEASY-*oriC*-C and pEASY-*oriC*-L, using M13F (fluorescent 6-carboxyfluorescein [FAM]-labeled) and M13R primers. The FAM-labeled probes were purified by the Wizard® SV Gel and PCR Clean-Up System (Promega Corporation, Madison, WI, USA) and they were quantified with NanoDrop 2000C (Thermo Fisher Scientific, Waltham, MA, USA).

DNase I footprinting assays were performed similar to Wang et al. (2012). For each assay, 400 ng probes were incubated with different amounts of the His_6-DnaA (IV) protein in a total volume of 40 μL. Following incubation for 30 min at 25°C, a 10 μL solution containing about 0.015 units of DNase I (Promega Corporation) and 100 nmol of freshly prepared $CaCl_2$ was added and further incubated for 1 min at 25°C.

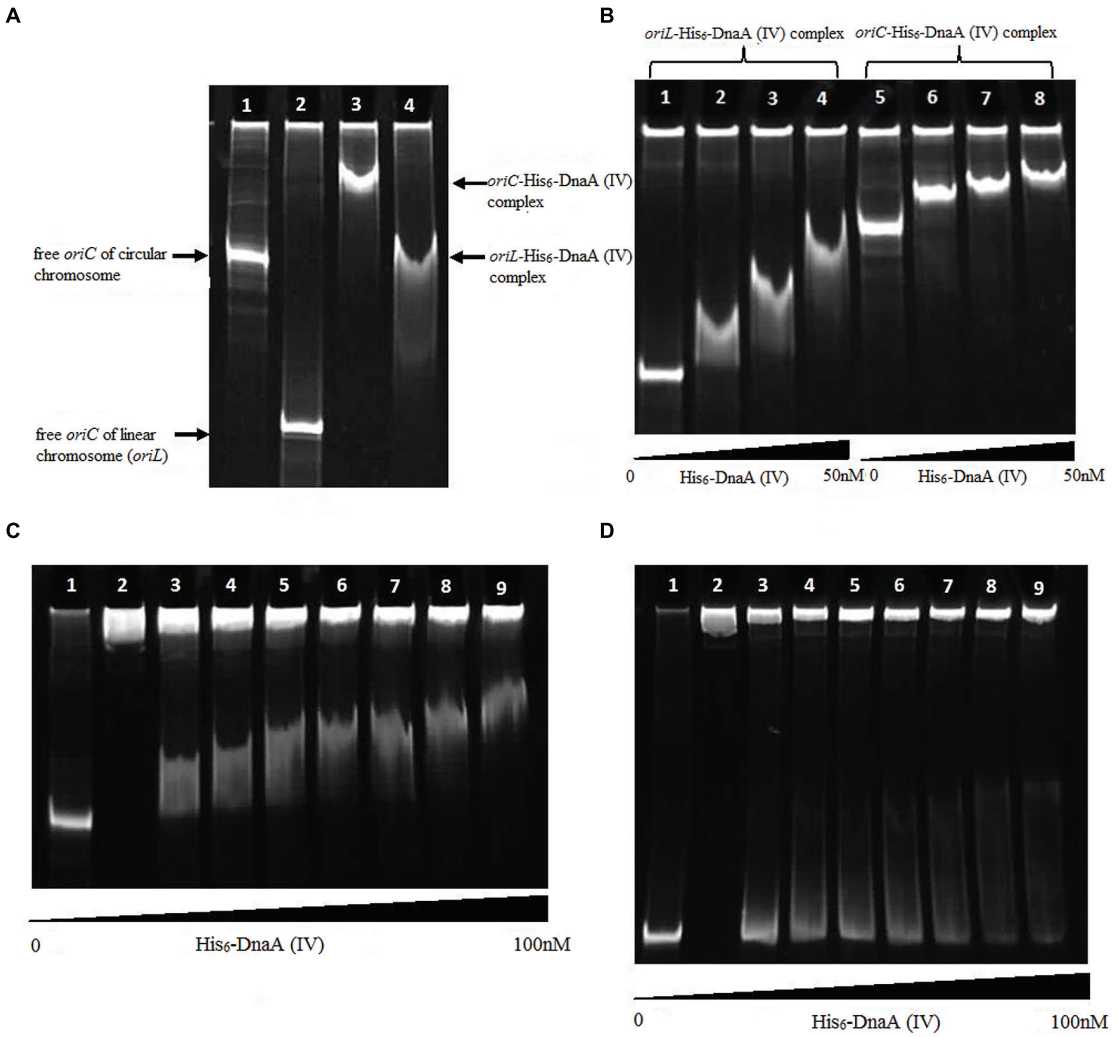


FIGURE 4 | Interaction of the DnaA (IV) protein with the *Cyanothece* 51142 *oriC*. (A) Interaction of the DnaA (IV) protein with the *Cyanothece* 51142 *oriC* (5% gel). (1) Free *oriC* fragment of the circular chromosome. (2) Free *oriC* fragment of of the linear chromosome. (3) His_6-DnaA (IV) binding to the circular chromosome *oriC*. (4) His_6-DnaA (IV) binding to the linear chromosome *oriC*. **(B)** Interaction of the DnaA (IV) protein with the *Cyanothece* 51142 *oriC* (5% gel). (1) Free *oriC* fragment of the linear chromosome. (2–4) His_6-DnaA (IV) binding to the linear chromosome *oriC* (10, 25, 50 nM). (5) Free *oriC* fragment of the circular chromosome. (6–8) His_6-DnaA (IV) binding to the circular chromosome *oriC* (10, 25, 50 nM). **(C)** Interaction of the His_6-DnaA (IV) protein with the *oriC* of the circular chromosome (5% gel). (1) *oriC* fragment of the circular chromosome [without DnaA (IV)]. (2) His_6-DnaA (IV) protein (without the *oriC* fragment). (3–9) His_6-DnaA (IV) binding to the circular chromosome *oriC* (3: 2.5; 4: 5.0; 5: 10.0; 6: 25.0; 7: 50.0; 8: 75.0; 9: 100.0 nM). **(D)** Interaction of the His_6-DnaA (IV) protein with the *oriC* of linear chromosome (5% gel). (1) *oriC* fragment of the linear chromosome [without His_6-DnaA (IV)]. (2) His_6-DnaA (IV) protein (without the *oriC* fragment). (3–9) His_6-DnaA (IV) binding to the linear chromosome *oriC* (3: 2.5; 4: 5.0; 5: 10.0; 6: 25.0; 7: 50.0; 8: 75.0; 9: 100.0 nM).

The reaction was stopped by adding 140 μL of DNase I stop solution (200 mM unbuffered sodium acetate, 30 mM of EDTA, and 0.15% SDS). Samples were first extracted with phenol/chloroform; they were then precipitated with ethanol and the pellets were dissolved in 30 μL of MiniQ water. The preparation of the DNA ladder, electrophoresis, and data analysis were the same as the methods described previously (Wang et al., 2012), with the exception that the GeneScan-LIZ500 size standard (Applied Biosystems; Thermo Fisher Scientific) was used. The protected sites correspond to the locations of the DnaA (IV)-*oriC*s complexes visualized using the Peak Scanner software v1.0 (Applied Biosystems; Thermo Fisher Scientific).

RESULTS

Prediction of Cyanothece 51142 Replication Origins and Comparison of the oriC Regions in Different Cyanobacteria

Based on the *Z*-curve method, employing the means of comparative genomics, a web-based system Ori-Finder has been developed to identify *oriC*s in bacterial and archaeal genomes with high accuracy and reliability (Gao and Zhang, 2008b; Luo et al., 2014). By utilizing Ori-Finder, the locations of *oriC*s for

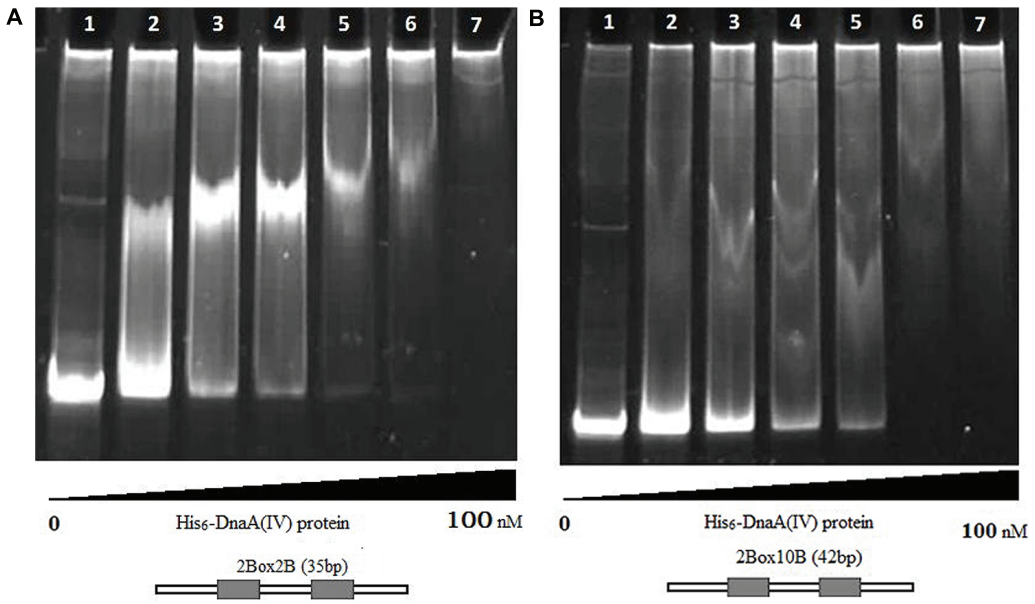


FIGURE 5 | Interaction of the His_6-DnaA (IV) protein with the circular and linear chromosomes DnaA boxes (8% gel). (A) Interaction of the His_6-DnaA (IV) protein with the circular chromosome DnaA boxes. (1) Double-stranded oligonucleotides [Cbox23F:Cbox23R, without DnaA (IV)]. (2–7) His_6-DnaA (IV) binding to double-stranded oligonucleotides (2: 5.0; 3: 10.0; 4: 25.0; 5: 50.0; 6: 75.0; 7: 100.0 nM). 2Box2B (35 nt): the 35-nt double-stranded oligonucleotides containing two DnaA boxes with 2 nt space. **(B)** Interaction of the His_6-DnaA (IV) protein with the linear chromosome DnaA boxes. (1) double-stranded oligonucleotides [Lbox23F:Lbox23R, without DnaA (IV)]. (2–7) His_6-DnaA (IV) binding to double-stranded oligonucleotides (2: 5.0; 3: 10.0; 4: 25.0; 5: 50.0; 6: 75.0; 7: 100.0 nM). 2Box10B (42 nt): the 42-nt double-stranded oligonucleotides containing two DnaA boxes with 10 nt space.

both the circular and linear chromosomes in *Cyanothece* 51142 have been identified (Gao and Zhang, 2008a; **Figure 1**). For the circular chromosome, the *oriC* is predicted to be within the intergenic region ranging from 1,886,587 to 1,887,114 nt between the ORF cce_1862 and the *dnaN* gene (cce_1864). For the linear chromosome, the *oriC* is predicted to be within the intergenic region, ranging from 373,518 to 373,849 nt, between the ORFs cce_5168 and cce_5169. Both of *oriC*s contain three DnaA boxes, which differ by one position from the most stringent consensus sequence of the *E. coli* DnaA box TTATCCACA. With respect to the *oriC* of the circular chromosome, the DnaA boxes match perfectly to TTTTCCACA, the "species-specific" DnaA box motif for cyanobacteria, and these DnaA boxes have been found next to the start of the *dnaN* gene (**Figure 1**). A comparison of DnaA boxes from the identified *Cyanothece* ATCC51142 *oriC* regions allows us to propose the consensus (**Figure 1**). The analysis of replication origins for bacteria in DoriC[2], a database of *oriC* regions in bacterial and archaeal genomes (Gao and Zhang, 2007; Gao et al., 2013), has also shown the conserved features associated with the *oriC* regions in the phylum cyanobacteria, such as the adjacent gene, *dnaN*, and the consensus sequence TTTTCCACA of the DnaA boxes (Gao, 2014). However, the results obtained by Ori-Finder were not sufficient to unequivocally determine the origins of replication. As such, we also performed experimental validation to confirm the results predicted by Ori-Finder.

The identified *oriC* region of *Anabaena* PCC 7120 was also predicted by Ori-Finder (Zhou et al., 2011). Previously, Zhou et al. (2011) compared the *oriC* regions predicted by the Ori-Finder in 32 species of cyanobacteria. We added the other 36 species of cyanobacteria which were also predicted by Ori-Finder, and listed them in DoriC 5.0 database. Similarly, most of the 68 species of cyanobacteria have a *dnaN*-coding region nearby (**Figure 2**). Therefore, we also constructed a phylogenetic tree featuring *dnaN* gene sequences with phylogeny.fr[3] to compare the changes of the *oriC* regions during cyanobacterial evolution (**Figure 2**). The putative *oriC* region of *Thermosynechococcus elongatus* BP-1 is bordered by genes that encode proteins of unknown functions on both sides. The range of the number of DnaA boxes at the *oriC* regions have changed, ranging from one in *Cyanothece* sp. PCC 8801 to 12 in *T. elongatus* BP-1. Furthermore, the sequences of DnaA boxes in most of the cyanobacteria are TTTTCCACA, which is the same as *Cyanothece* 51142, *Anabaena* sp. PCC 7120 (Zhou et al., 2011), and *Synechococcus elongatus* PCC 7942 (Liu and Tsinoremas, 1996). However, the *oriC* regions of *Synechocystis* sp. PCC 6803 substr. GT-I, substr. PCC-N, and substr. PCC-P overlap with the membrane protein-coding sequences; as such, it may be interesting to verify these predicted results in further experiments.

[2]http://tubic.tju.edu.cn/doric/

[3]http://www.phylogeny.fr/simple_phylogeny.cgi

Characterization of *Cyanothece* 51142 Replication Origins, the *dnaA* Gene, and Its Product, the Binding Domain of Initiation Protein-DnaA (IV)

The predicted replication origin (*oriC*) region of the circular chromosome from *Cyanothece* 51142 is located upstream of the replication initiator gene (*dnaA*). It is composed of three putative DnaA boxes, each with a perfect match to TTTTCCACA, the "species-specific" DnaA box motif for cyanobacteria (**Figure 1**). For the linear chromosome, *oriC* is located within the intergenic region between two hypothetical open reading frames (ORFs). This identified *oriC* region is located around the minimum point of the GC disparity curve and contains three predicted DnaA boxes, each of which has no more than one mismatch from the DnaA box motif for *E. coli*: TTATCCACA. It is also observed that the *oriC* for the linear chromosome contains a reverse repeat.

The *dnaA* gene is located downstream of the *oriC* region of the circular chromosome, and it encodes a DnaA protein of 455 amino acid residues (∼52.4 kDa). Based on the structural and functional analysis of DnaA homologs, four domains of DnaA were deduced and the *dnaA* IV was found to express the DnaA IV protein of 125 amino acid residues (∼14.1 kDa). Previous studies have shown that the larger part of domain IV of the DnaA protein (the C-terminal 94 amino acid residues) was necessary and sufficient for DNA binding. To determine whether the C-terminus of the *Cyanothece* 51142 DnaA protein is responsible for DNA binding, its interaction with *oriC*s was analyzed by gel retardation assay. The PCR-amplified DNA fragment of the *dnaA* (IV) gene fused to the His·Tag sequence (see Materials and Methods for details) was overexpressed in *E. coli* BL21 (DE3). The fusion protein, His_6-DnaA (IV) (∼14.1 kDa) was purified by affinity chromatography on the Ni-NTA-agarose column as described in the Section "Materials and Methods".

Identification of the Replication Origins from *Cyanothece* 51142

Electrophoretic mobility shift assay was performed to determine whether the His_6-DnaA (IV) protein interacted with the DnaA boxes of the *oriC*s of *Cyanothece* 51142. The C-terminus of DnaA domain IV – namely DnaA (IV) – was used for all binding experiments given that domain IV of DnaA has been shown to be essential and sufficient for binding.

The protein–DNA complexes were analyzed by 5 or 8% native PAGE. When the PCR fragments containing each *oriC* (**Figure 4**), or oligonucleotides containing the DnaA boxes, were used to bind DnaA (IV), protein–DNA complexes were formed. As shown by the EMSA, increasing nucleoprotein complexes were observed as the DnaA (IV) concentration increased. When nucleotides with two DnaA boxes were used for binding, protein-DNA complexes were also observed (**Figures 5** and **6**). DnaA (IV) showed better affinity for the circular DnaA box motif (with two perfect DnaA boxes separated by 2 nt) than the linear DnaA box motif (with two imperfect DnaA boxes spaced by 10 nt) (**Figure 5**). Subsequently, to identify the exact DNA sequences that His_6-DnaA (IV) protected in the *oriC* regions of the circular and linear chromosomes, a DNase I footprinting assay, combined with FAM-labeled primers using purified His_6-DnaA (IV), was performed. Two DNA fragments (representing the entire *Cyanothece* 51142 *oriC* region of the circular and linear chromosomes) with FAM-labeled probes at the 5-end (upper strands) were incubated with different amounts of the His_6-DnaA (IV) protein. The precipitated DNA sequences, which were protected by His_6-DnaA (IV), were sequenced (**Figure 7**). According to the merged figure [with and without DnaA (IV)], a clearly protected region (39 nt) relative to the second and third DnaA boxes sites was found in the *oriC* region of the circular chromosome (**Figure 7A**), although the first DnaA box was not bound to DnaA(IV). Within the protected regions of the linear chromosome *oriC*, the results from the DNase I footprinting assay revealed that His_6-DnaA (IV) protected two specific regions: an AT-rich region, as well as the region containing two DnaA boxes and an incomplete DnaA box (TATG) (**Figure 7B**). Moreover, the hypersensitive sites, which are consistent with the locations of the DnaA boxes of the circular and linear chromosomes, corroborated the results obtained with Ori-Finder and EMSA (**Figure 7**).

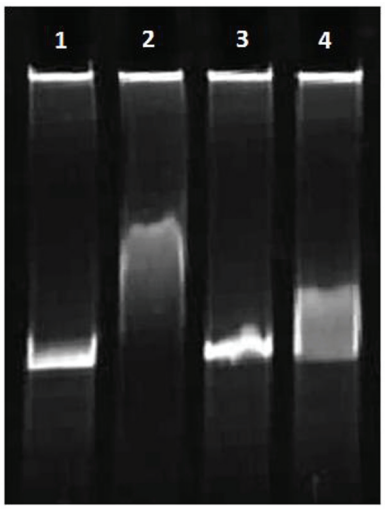


FIGURE 6 | Interaction of the His_6-DnaA (IV) protein with linear chromosome DnaA box and *E. coli* DnaA box (8% gel). (1,3) double-stranded oligonucleotides (1: Cbox1F:Cbox1R; 3: EboxF:EboxR). (2,4) 50.0 nM His_6-DnaA (IV) binding to double-stranded oligonucleotides (2: Cbox1F:Cbox1R; 4: EboxF:EboxR).

DISCUSSION

The origins of replication in *Cyanothece* 51142 have been difficult to determine using classic algorithms due to a lack of distinct patterns in strand asymmetry, although the complete genome sequence has already been determined. By utilizing the web-based system Ori-Finder, the locations of *oriC*s for both the circular and linear chromosomes in *Cyanothece* 51142 have been identified. Subsequently, we confirmed the predicted results *in vitro*. As demonstrated by EMSA and the DNase I footprinting assay,

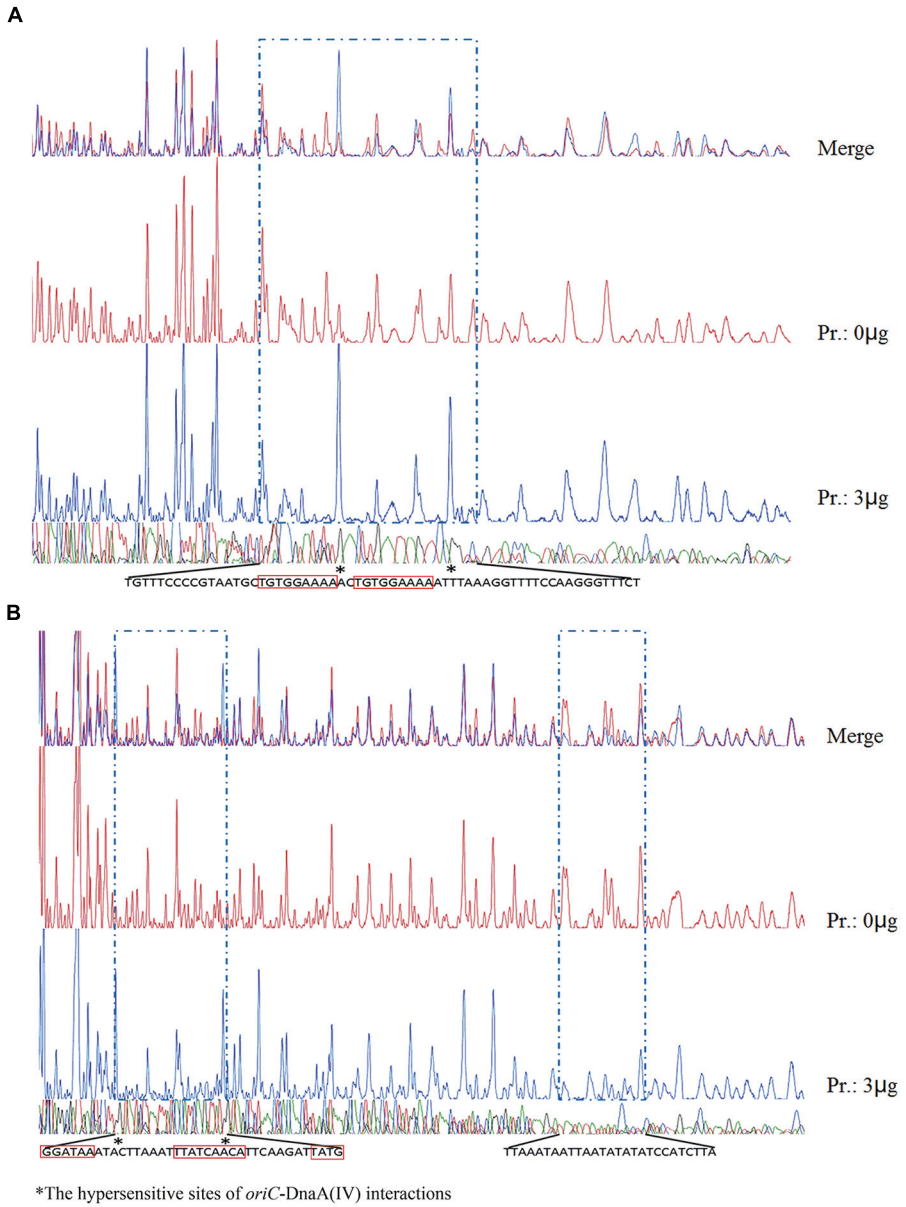


FIGURE 7 | DNase I footprint of protein His_6-DnaA (IV) to the *oriC*s of the circular and linear chromosomes. (A) Along with the increase of His_6-DnaA (IV), a clearly protected region of 39 nt relative to the second and third DnaA boxes sites in the *oriC* region of the circular chromosome was found, although the first DnaA box was not bound to the protein [the sequences in **(A)** are reverse complementary]. The asterisks mark the hypersensitive sites, which are consistent with the locations of two DnaA boxes of the circular chromosome. **(B)** Within the protected regions of the linear chromosome *oriC*, the results from the DNase I footprinting assay revealed that His_6-DnaA (IV) protected two specific regions- an AT-rich region, as well as a region containing two DnaA boxes and an incomplete DnaA box (TATG) [the sequences in **(B)** are reverse complementary]. The asterisks mark the hypersensitive sites, which are consistent with the locations of two DnaA boxes of the linear chromosome.

His_6-DnaA (IV) does not clearly bind a single DnaA box; rather, it binds two DnaA boxes (the second and third DnaA boxes) from the *oriC*s of the circular chromosome, as well as three DnaA boxes (the first DnaA box is incomplete). Our results suggest that interactions of the *Cyanothece* 51142 DnaA (IV) with several DnaA boxes exhibit cooperativity.

In most bacteria, DnaA is essential for initiating chromosomal replication and recognizing the DnaA boxes near *oriC* (Skarstad and Boye, 1994). However, most of the cyanobacteria have an exceptionally low strand bias, which suggests that the process of DNA replication in these species is somehow different from that of other bacteria (Worning et al., 2006). Indeed, *Synechocystis* sp. PCC 6803 and *Prochlorococcus* lack a DnaA-binding box near *dnaA*, and they display an unusual gene arrangement in this region (Richter and Messer, 1995; Partensky et al., 1999). In addition, the *dnaA* gene of *Synechocystis* sp. PCC

6803 could be deleted without phenotypic effect (Richter et al., 1998). Furthermore, the *dnaA* genes of *Nostoc azollae* 0708, *Cyanobacterium aponinum* PCC 10605 and *C. stanieri* PCC 7202 were not detected from the genomes (data from National Center for Biotechnology Information). However, DnaA has been conserved during evolution, and its transcription is not autoregulated in the same manner as *E. coli* but light dependent instead, and follows the circadian rhythm of DNA synthesis (Richter et al., 1998). Studying the mechanism and regulation of DNA replication should reveal clues about the evolution of the DNA replication mechanism, and it will allow us to better understand the relationship between the various timing circuits involved in the circadian clock and cell division cycles.

The freshwater cyanobacteria *S. elongatus* PCC 7942 and *Synechocystis* sp. PCC 6803, have been used as model organisms for phototrophs because their transformation efficiency and growth rate are superior to those of marine cyanobacteria and their complete genome sequences have been published. In cyanobacteria, the cell division cycle is strongly light dependent, and light is the most important factor that affects the circadian clock. Most publications pertaining to the replication origins of cyanobacteria also focus on light-dependent DNA replication processes (Liu and Tsinoremas, 1996; Richter et al., 1998; Ohbayashi et al., 2013). It has been reported that for *S. elongatus* PCC 7942 and *Synechocystis* sp. PCC 6803, DNA replication depends on photosynthetic electron transport (Hihara et al., 2003; Ohbayashi et al., 2013). We successfully identified the *oriC* regions of *Cyanothece* 51142; however, it is regrettable that we do not have live cells to further assess whether or not the DNA replication of *Cyanothece* 51142 is light dependent. Knowledge about the interactions between the *Cyanothece* 51142 DnaA protein and *oriC*s may provide fresh insights into the function of this protein, as well as into the regulation of the initiation of cyanobacterial chromosome replication. Further studies are required to understand the exact mechanism underlying the phenotypes of *oriC*s of both circular and linear chromosomes. The exact mechanism underlying DnaA protein regulation in the replication, and other functions, of linear chromosome also need to be investigated.

ACKNOWLEDGMENTS

We would like to thank Prof. Chun-Ting Zhang for providing invaluable assistance and inspiring discussions. The present work was supported in part by the National Natural Science Foundation of China (Grant Nos. 31571358, 31470967, 31171238, and 30800642), the Tianjin Municipal Natural Science Foundation of China (Grant No. 09JCZDJC17100), the Program for New Century Excellent Talents in University (No. NCET-12-0396), and the China National 863 High-Tech Program (2015AA020101).

REFERENCES

Abe, Y., Jo, T., Matsuda, Y., Matsunaga, C., Katayama, T., and Ueda, T. (2007). Structure and function of DnaA N-terminal domains: specific sites and mechanisms in inter-DnaA interaction and in DnaB helicase loading on oriC. *J. Biol. Chem.* 282, 17816–17827. doi: 10.1074/jbc.M701841200

Berman-Frank, I., Lundgren, P., and Falkowski, P. (2003). Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. *Res. Microbiol.* 154, 157–164. doi: 10.1016/S0923-2508(03)00029-9

Blaesing, F., Weigel, C., Welzeck, M., and Messer, W. (2000). Analysis of the DNA-binding domain of *Escherichia coli* DnaA protein. *Mol. Microbiol.* 36, 557–569. doi: 10.1046/j.1365-2958.2000.01881.x

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

Erzberger, J. P., Pirruccello, M. M., and Berger, J. M. (2002). The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation. *EMBO J.* 21, 4763–4773. doi: 10.1093/emboj/cdf496

Felczak, M. M., and Kaguni, J. M. (2004). The box VII motif of *Escherichia coli* DnaA protein is required for DnaA oligomerization at the *E. coli* replication origin. *J. Biol. Chem.* 279, 51156–51162. doi: 10.1074/jbc.M409695200

Fujita, M. Q., Yoshikawa, H., and Ogasawara, N. (1990). Structure of the dnaA region of *Micrococcus luteus*: conservation and variations among eubacteria. *Gene* 93, 73–78. doi: 10.1016/0378-1119(90)90138-H

Gao, F. (2014). Recent advances in the identification of replication origins based on the Z-curve method. *Curr. Genomics* 15:104. doi: 10.2174/1389202915999140328162938

Gao, F., Luo, H., and Zhang, C.-T. (2013). DoriC 5.0: an updated database of oriC regions in both bacterial and archaeal genomes. *Nucleic Acids Res.* 41, D90–D93. doi: 10.1093/nar/gks990

Gao, F., and Zhang, C.-T. (2007). DoriC: a database of oriC regions in bacterial genomes. *Bioinformatics* 23, 1866–1867. doi: 10.1093/bioinformatics/btm255

Gao, F., and Zhang, C.-T. (2008a). Origins of replication in Cyanothece 51142. *Proc. Natl. Acad. Sci. U.S.A.* 105:E125. doi: 10.1073/pnas.0809987106

Gao, F., and Zhang, C. T. (2008b). Ori-Finder: a web-based system for finding oriCs in unannotated bacterial genomes. *BMC Bioinformatics* 9:79. doi: 10.1186/1471-2105-9-79

Hihara, Y., Sonoike, K., Kanehisa, M., and Ikeuchi, M. (2003). DNA microarray analysis of redox-responsive genes in the genome of the cyanobacterium *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* 185, 1719–1725. doi: 10.1128/JB.185.5.1719-1725.2003

Kaguni, J. M. (1997). *Escherichia coli* DnaA protein: the replication initiator. *Mol. Cells* 7:145.

Katayama, T. (2008). Roles for the AAA+ motifs of DnaA in the initiation of DNA replication. *Biochem. Soc. Trans.* 36(Pt 1), 78–82. doi: 10.1042/BST0360078

Keyamura, K., Fujikawa, N., Ishida, T., Ozaki, S., Su'etsugu, M., Fujimitsu, K., et al. (2007). The interaction of DiaA and DnaA regulates the replication cycle in *E. coli* by directly promoting ATP DnaA-specific initiation complexes. *Genes Dev.* 21, 2083–2099. doi: 10.1101/gad.1561207

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

Liu, Y., and Tsinoremas, N. F. (1996). An unusual gene arrangement for the putative chromosome replication origin and circadian expression of dnaN in *Synechococcus* sp. strain PCC 7942. *Gene* 172, 105–109. doi: 10.1016/0378-1119(96)00160-6

Luo, H., Zhang, C.-T., and Gao, F. (2014). Ori-Finder 2, an integrated tool to predict replication origins in the archaeal genomes. *Front. Microbiol.* 5:482. doi: 10.3389/fmicb.2014.00482

Messer, W. (2002). The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiol. Rev.* 26, 355–374. doi: 10.1111/j.1574-6976.2002.tb00620.x

Messer, W., and Weigel, C. (1997). DnaA initiator-also a transcription factor. *Mol. Microbiol.* 24, 1–6. doi: 10.1046/j.1365-2958.1997.3171678.x

Mizushima, T., Tomura, A., Shinpuku, T., Miki, T., and Sekimizu, K. (1994). Loss of flagellation in dnaA mutants of *Escherichia coli*. *J. Bacteriol.* 176, 5544–5546.

Montoya, J. P., Holl, C. M., Zehr, J. P., Hansen, A., Villareal, T. A., and Capone, D. G. (2004). High rates of N2 fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature* 430, 1027–1032. doi: 10.1038/nature02824

Mott, M. L., and Berger, J. M. (2007). DNA replication initiation: mechanisms and regulation in bacteria. *Nat. Rev. Microbiol.* 5, 343–354. doi: 10.1038/nrmicro1640

Nozaki, S., and Ogawa, T. (2008). Determination of the minimum domain II size of *Escherichia coli* DnaA protein essential for cell viability. *Microbiology* 154(Pt 11), 3379–3384. doi: 10.1099/mic.0.2008/019745-0

Ohbayashi, R., Watanabe, S., Kanesaki, Y., Narikawa, R., Chibazakura, T., Ikeuchi, M., et al. (2013). DNA replication depends on photosynthetic electron transport in cyanobacteria. *FEMS Microbiol. Lett.* 344, 138–144. doi: 10.1111/1574-6968.12166

Ozaki, S., Kawakami, H., Nakamura, K., Fujikawa, N., Kagawa, W., Park, S. Y., et al. (2008). A common mechanism for the ATP–DnaA-dependent formation of open complexes at the replication origin. *J. Biol. Chem.* 283, 8351–8362. doi: 10.1074/jbc.M708684200

Partensky, F., Hess, W., and Vaulot, D. (1999). Prochlorococcus, a marine photosynthetic prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* 63, 106–127.

Richter, S., Hagemann, M., and Messer, W. (1998). Transcriptional analysis and mutation of adnaa-like gene in *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* 180, 4946–4949.

Richter, S., and Messer, W. (1995). Genetic structure of the dnaA region of the cyanobacterium *Synechocystis* sp. strain PCC6803. *J. Bacteriol.* 177, 4245–4251.

Roth, A., and Messer, W. (1995). The DNA binding domain of the initiator protein DnaA. *EMBO J.* 14:2106.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning*. New York, NY: Cold Spring Harbor Laboratory Press.

Schaper, S., and Messer, W. (1995). Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target. *J. Biol. Chem.* 270, 17622–17626. doi: 10.1074/jbc.270.29.17622

Sherman, L. A., Meunier, P., and Colon-Lopez, M. S. (1998). Diurnal rhythms in metabolism: a day in the life of a unicellular, diazotrophic cyanobacterium. *Photosynth. Res.* 58, 25–42. doi: 10.1023/A:1006137605802

Skarstad, K., and Boye, E. (1994). The initiator protein DnaA: evolution, properties and function. *Biochim. Biophys. Acta* 1217, 111–130. doi: 10.1016/0167-4781(94)90025-6

Stanier, R., and Bazine, G. (1977). Phototrophic prokaryotes: the cyanobacteria. *Annu. Rev. Microbiol.* 31, 225–274. doi: 10.1146/annurev.mi.31.100177.001301

Theisen, P. W., Grimwade, J. E., Leonard, A. C., Bogan, J. A., and Helmstetter, C. E. (1993). Correlation of gene transcription with the time of initiation of chromosome replication in *Escherichia coli*. *Mol. Microbiol.* 10, 575–584. doi: 10.1111/j.1365-2958.1993.tb00929.x

Veening, J.-W., Murray, H., and Errington, J. (2009). A mechanism for cell cycle regulation of sporulation initiation in *Bacillus subtilis*. *Genes Dev.* 23, 1959–1970. doi: 10.1101/gad.528209

Wang, Y., Cen, X.-F., Zhao, G.-P., and Wang, J. (2012). Characterization of a new GlnR binding box in the promoter of amtB in *Streptomyces coelicolor* inferred a PhoP/GlnR competitive binding mechanism for transcriptional regulation of amtB. *J. Bacteriol.* 194, 5237–5244. doi: 10.1128/JB.00989-12

Weigel, C., Schmidt, A., Seitz, H., Tüngler, D., Welzeck, M., and Messer, W. (1999). The N-terminus promotes oligomerization of the *Escherichia coli* initiator protein DnaA. *Mol. Microbiol.* 34, 53–66. doi: 10.1046/j.1365-2958.1999.01568.x

Welsh, E. A., Liberton, M., Stöckel, J., Loh, T., Elvitigala, T., Wang, C., et al. (2008a). The genome of Cyanothece 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle. *Proc. Natl. Acad. Sci. U.S.A.* 105, 15094–15099. doi: 10.1073/pnas.0805418105

Welsh, E. A., Liberton, M., Stöckel, J., and Pakrasi, H. B. (2008b). Reply to Zhang et al.: Identification of origins of replication in the Cyanothece 51142 genome. *Proc. Natl. Acad. Sci. U.S.A.* 105:E126. doi: 10.1073/pnas.0810681106

Whitton, B. A., and Potts, M. (2012). "Introduction to the cyanobacteria," in *Ecology of Cyanobacteria II,* ed. B. A. Whitton (Berlin: Springer), 1–13.

Worning, P., Jensen, L. J., Hallin, P. F., Stærfeldt, H. H., and Ussery, D. W. (2006). Origin of replication in circular prokaryotic chromosomes. *Environ. Microbiol.* 8, 353–361. doi: 10.1111/j.1462-2920.2005.00917.x

Zhou, Y., Chen, W.-L., Wang, L., and Zhang, C.-C. (2011). Identification of the oriC region and its influence on heterocyst development in the filamentous cyanobacterium *Anabaena* sp. strain PCC 7120. *Microbiology* 157, 1910–1919. doi: 10.1099/mic.0.047241-0

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.

Copyright © 2015 Huang, Song, Yang, Dong, Hu and Gao. 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.

frontiers in
MICROBIOLOGY

ORIGINAL RESEARCH ARTICLE
published: 18 November 2014
doi: 10.3389/fmicb.2014.00574

Sequence analysis of origins of replication in the *Saccharomyces cerevisiae* genomes

Wen-Chao Li[1], Zhe-Jin Zhong[1], Pan-Pan Zhu[1], En-Ze Deng[1], Hui Ding[1]*, Wei Chen[2]* and Hao Lin[1]*

[1] Key Laboratory for Neuro-Information of Ministry of Education, Center of Bioinformatics, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China
[2] Department of Physics, School of Sciences and Center for Genomics and Computational Biology, Hebei United University, Tangshan, China

Edited by:
Feng Gao, Tianjin University, China

Reviewed by:
Dong Wang, Harbin Medical University, China
Lu Cai, Inner Mongolia University of Science and Technology, China

***Correspondence:**
Hui Ding and Hao Lin, Key Laboratory for Neuro-Information of Ministry of Education, Center of Bioinformatics, School of Life Science and Technology, University of Electronic Science and Technology of China, No. 4 Section 2 North Jianshe Road, Chengdu 610054, Sichuan, China
e-mail: hding@uestc.edu.cn; hlin@uestc.edu.cn;
Wei Chen, Department of Physics, School of Sciences and Center for Genomics and Computational Biology, Hebei United University, Tangshan 063000, Hebei, China
e-mail: greatchen@heuu.edu.cn

DNA replication is a highly precise process that is initiated from origins of replication (ORIs) and is regulated by a set of regulatory proteins. The mining of DNA sequence information will be not only beneficial for understanding the regulatory mechanism of replication initiation but also for accurately identifying ORIs. In this study, the GC profile and GC skew were calculated to analyze the compositional bias in the *Saccharomyces cerevisiae* genome. We found that the GC profile in the region of ORIs is significantly lower than that in the flanking regions. By calculating the information redundancy, an estimation of the correlation of nucleotides, we found that the intensity of adjoining correlation in ORIs is dramatically higher than that in flanking regions. Furthermore, the relationships between ORIs and nucleosomes as well as transcription start sites were investigated. Results showed that ORIs are usually not occupied by nucleosomes. Finally, we calculated the distribution of ORIs in yeast chromosomes and found that most ORIs are in transcription terminal regions. We hope that these results will contribute to the identification of ORIs and the study of DNA replication mechanisms.

Keywords: ***Saccharomyces cerevisiae*, origin of replication, GC profile, GC skew, information redundancy, distribution of ORIs**

INTRODUCTION

The well-known replication theory was proposed in 1963 based on a large number of experiments using the sexual system of *Escherichia coli* (Jacob et al., 1963). DNA replication is an orchestrated process. When a cell enters the S phase of replication, the DNA double helix of this cell is unwound. Then, replication forks are generated to allow the DNA synthesis machinery to copy each DNA strand in a bidirectional manner. In the process of replication, the specific regions that are responsible for the initiation of the replication of the genome are called origin of replication (ORI) regions. These regions are recognized by the origin recognition complex (ORC). The DNA replication process is usually activated only once per cell cycle to avoid amplification and maintain genome integrity (Cayrou et al., 2012).

Although most of bacterial genomes have only a single ORI region (Gao and Zhang, 2007) and some archaea use more than one ORI region to initiate DNA replication (Luo et al., 2014), the fungus, *Saccharomyces cerevisiae* (*S. cerevisiae*) has multiple ORIs on its chromosomes to perform complete replication in a reasonable period of time because of the large size of its genomes and the limitation of nucleotide incorporation during DNA synthesis. Therefore, predicting ORIs is more difficult in the *S. cerevisiae* genome than that in bacterial genomes. Several experiments have revealed that the activity of ORIs in yeast depends on a *cis*-acting replicator sequence termed autonomous replication sequence (ARS). These regulatory sequences are generally found in AT-rich regions in yeast genome. The ARS generally contains three domains: A, B, and C. An essential ARS consensus sequence (ACS) (T/A)TTTAT(A/G)TTT(T/A) usually appears in the A domain (Wu et al., 2014). The B domain contains a number of short sequence motifs that contribute to origin activity (Dhar et al., 2012). The motifs in the C domain are responsible for the interaction between DNA and regulatory proteins (Crampton et al., 2008). However, these motif sequences are not conserved enough to be used to identify ORIs (Nieduszynski et al., 2006). Thus, the discovery of the hidden intrinsic characteristics at the sequence level is helpful not only for understanding the regulatory mechanism but also for accurately identifying ORIs.

With the accumulation of experimental data (Levitsky et al., 2005; Yamashita et al., 2011; Gao et al., 2012), some researchers have analyzed features of replication. Recently, by analyzing four highly active origins, Chang et al. (2011) revealed that sequences adjacent to the ACS contributed substantially to origin activity and ORC binding. Yin et al. (2009) found that the nucleosome depletion regions are preferentially permissive for replication and proposed that the ORI organization imposed by nucleosome positioning is phylogenetically widespread in eukaryotes. DNA structure may also influence the distribution of ORIs. Chen et al. (2012) found that the DNA bendability and cleavage intensity in

ORIs are dramatically lower than those in both upstream and downstream regions of ORIs.

Although some characteristics of ORIs have been described, the available information about ORIs is still far from satisfactory. Therefore, to clarify replication mechanisms, it is still necessary to discover the intrinsic characteristics of ORIs. With this in mind, we performed a series of analyses to investigate the composition bias and correlation of nucleotides in ORIs, the distribution of ORIs in genomes, and the relationships between ORIs and regulatory elements.

MATERIALS AND METHODS

DATASETS

The *S. cerevisiae* ORIs were collected from OriDB (Siow et al., 2012; http://www.oridb.org/). The confidence of the ORI data has three levels: confirmed, likely, and dubious. To provide a reliable and high-quality dataset, only the 410 experimentally confirmed ORIs were selected and used in the following analysis.

The complete *S. cerevisiae* genome was downloaded from GenBank (Benson et al., 2013). The 5015 transcription start sites (TSSs) of *S. cerevisiae* were previously published (Lee et al., 2007). The *in vitro* nucleosome data and nucleosome data from three growth conditions [ethanol, yeast extract, peptone, and dextrose (YPD) medium, and galactose] were previously reported (Yuan et al., 2005; Lee et al., 2007; Kaplan et al., 2009)

SEQUENCE COMPOSITION ANALYSIS

The GC profile represents the variation in GC content along the genomic sequence (Gao and Zhang, 2006), which can be defined by the following equation (Zhang et al., 2005; Xing et al., 2014):

$$GC\ \ profile[\mathrm{i}] = \frac{f_i(G) + f_i(C)}{f_i(A) + f_i(C) + f_i(G) + f_i(T)} \quad (1)$$

where f_i(A), f_i(C), f_i(G), and f_i(T) are the frequencies of adenine(A), cytosine(C), guanine (G), and thymine(T), respectively, in the i-th sliding window along the sequence. The range of values for the GC profile is between 0 and +1. Values ranging from 0 to 0.5 indicate that the GC content in the i-th sliding window is lower than the AT content, while values ranging from 0.5 to 1 indicate that the GC content in the i-th sliding window is higher than the AT content.

GC skew was the first proposed computational method to identify ORIs in bacterial genomes (Lobry, 1996a,b). For a given sequence, the GC skew is defined by the following equation (McLean et al., 1998):

$$GC\ \ skew[\mathrm{i}] = \frac{f_i(G) - f_i(C)}{f_i(G) + f_i(C)} \quad (2)$$

where f_i(C) and f_i(G) are the frequencies of cytosine(C), and guanine (G) in the i-th sliding window along a sequence, respectively. The range of values for GC skew is between −1 and +1. Values ranging from −1 to 0 indicate that $f_i(G) < f_i(C)$, and values ranging from 0 to +1 indicate that $f_i(G) > f_i(C)$.

INFORMATION REDUNDANCY

As a genetic language, the nucleic acid sequence can be investigated through an information-theoretic method (Luo et al., 1998). In recent years, informational entropy was widely applied in the recognition and evolution research of DNA sequences (Grosse et al., 2000; Yu and Jiang, 2001; Otu and Sayood, 2003; Xing et al., 2013). The average mutual information profile is an excellent candidate for a species signature (Bauer et al., 2008). Based on these studies, we introduced the k-order information redundancy, which can be defined as follows (Luo et al., 1998):

$$D_{k+2} = 2H + \sum_{i,j} p_{i(k)j} \log_2 p_{i(k)j} \qquad k = 0, 1, 2, \ldots \quad (3)$$

where $p_{i(k)j}$ is the joint probability of base j occurring after base i at a distance k along the sequence. The term $k = 0$ indicates the adjacent correlation between two bases. D_{k+2} describes the divergence of the sequence from independence and the correlation between nucleotides with the gap of k nucleotides. In general, the larger the D_{k+2} value is, the stronger the divergence degree of the sequence from independence is. The H value is the informational entropy and is defined by the following equation

$$H = -\sum_{a} p_a \log_2 p_a \quad (4)$$

where p_a is the probability of base a (a = A, G, C, or T) occurring in the sequence.

RESULTS AND DISCUSSION

GC CONTENT SURROUNDING ORIs

DNA sequence information is the most basic but important genetic information. It also plays an important role in the determination of ORIs in the *S. cerevisiae* genome. However, the extent to which ORIs are determined *in vivo* by *cis*-acting sequence is poorly understood. To investigate the compositional bias of ORIs, we calculated the GC content of 300 bp of each ORI. As a comparison, the GC content of the genome sequence was also calculated by using a window of 300 bp with a step of 300 bp. The mean GC content of ORIs is 0.3168 (SD = 0.23×10^{-2}), which is significantly lower ($P < 2.3 \times e^{-133}$, Mann–Whitney U-test) than the genome-wide GC content (0.3796; SD = 0.24×10^{-2}). In other words, ORIs are AT-rich. The high AT content of ORI sequences contributes to the opening of the DNA double helix structure for the initiation of DNA replication.

GC PROFILE AND GC-SKEW SURROUNDING ORI

To investigate the compositional bias, the GC profile and GC skew surrounding ORIs was calculated using a 50 bp sliding window with a step of 1 bp. The average scores of the GC profile and GC-skew are plotted in **Figure 1**. As illustrated in **Figure 1A**, the score of the GC profile in the ORI regions was statistically lower than that in the surrounding regions ($P < 2.0 \times e^{-86}$, Mann-Whitney U-test).

To further investigate the sequence mode of ORI sequences, MEME (Multiple Em for Motif Elicitation; Bailey and Elkan, 1994) was used to discover the consensus motifs in ORI sequences. We found that the consensus sequences are all AT-rich motifs. It has been reported that ORIs contain some AT-rich elements for interactions with regulatory proteins (Reeves and Beckerbauer,

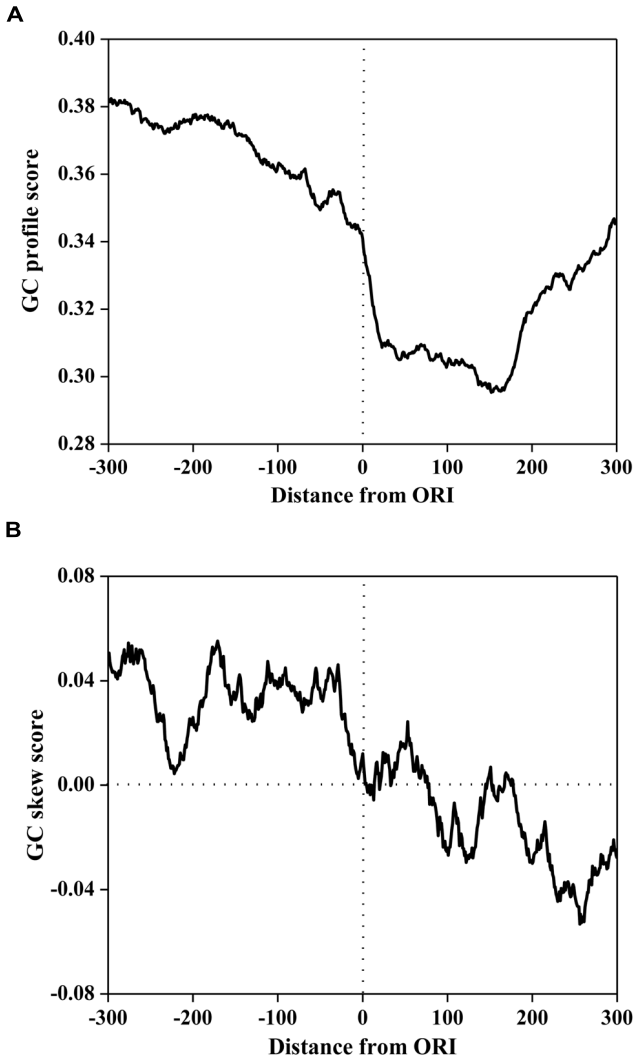


FIGURE 1 | The GC profile (A) and the GC skew (B) of origins of replication (in ORIs) and surrounding regions. The profiles are plotted using a 50-bp sliding window with a 1-bp step. The horizontal axis represents the nucleotide position, which ranges from −300 to +300 bp relative to ORIs (denoted as 0). The vertical axis represents the GC content score **(A)** and GC skew score **(B)**.

2001; Takahashi et al., 2003). Previous research demonstrated that the information encoded in the high AT content can be recognized by the Orc4 subunit of ORC (Mojardin et al., 2013). This can be attributed to the enrichment of the ACS around ORIs in *S. cerevisiae*, which is an AT-rich motif that contains the binding site for ORC. Recent research also revealed that a conspicuous feature of a replication regulatory protein was the presence of nine AT-hook domains in its amino terminus (Chuang and Kelly, 1999) that were essential for the binding of ORC to ORIs.

However, the GC skew in **Figure 1B** displays a different trend. The GC skew score in the core ORI regions was statistically lower than that in the upstream regions ($P < 2.3 \times e^{-80}$, Mann-Whitney U-test), but higher than that in the downstream regions ($P < 5.0 \times e^{-40}$, Mann-Whitney U-test). We noticed that the GC skew score conversed from positive to negative at the 0^{th} site corresponding to the DNA replication initiation site. In bacterial genomes, GC skew changes sign at the boundaries of the two replichores, which correspond to the DNA replication origin or terminus (Lobry, 1996a; Necsulea and Lobry, 2007). Thus, our finding implies that the *S. cerevisiae* genome may have a replication mechanism that is similar to that of bacterial genomes.

CORRELATION OF NUCLEOTIDES SURROUNDING ORIs

Based on Eq. 3, we calculated information redundancies D_{k+2} of ORI sequences. The average values are illustrated in **Figure 2A**. The main maxima for most ORI sequences are located on D_2. This result demonstrates that D_2 is the maximum among all considered D_{k+2} ($k = 0, 1, \ldots, 48$), indicating that ORI sequences have a short-range dominance of base correlations. Subsequently, we calculated D_2 in a 150 bp window with a step of 1 bp for ORI

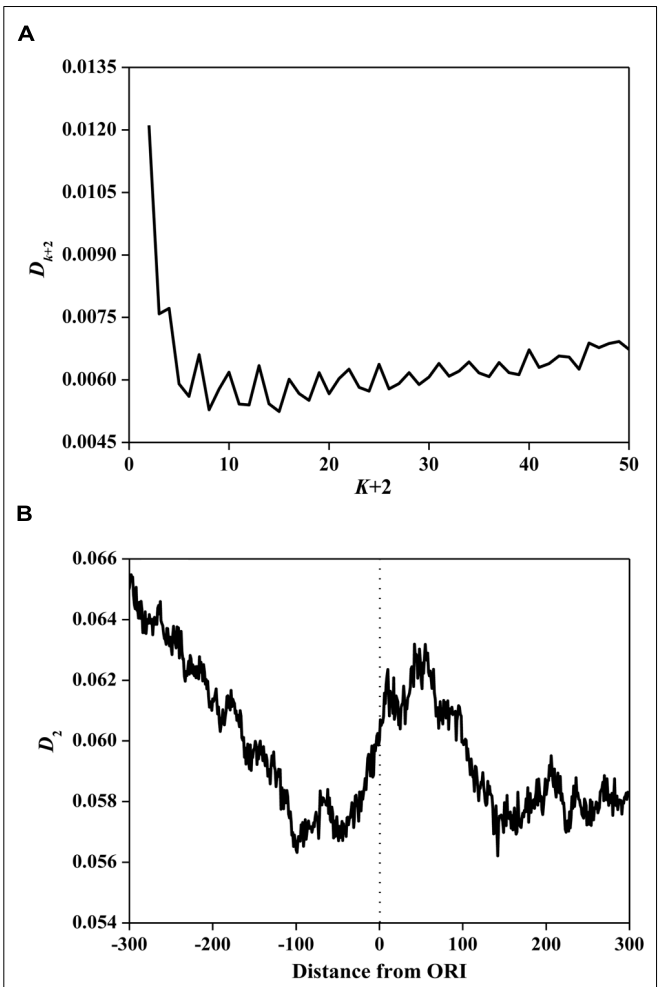


FIGURE 2 | (A) Average D_{k+2} vs. $k+2$ for the ORI sequences. The horizontal axis represents the gap of $k+2$. The vertical axis represents the value of D_{k+2}. **(B)** The distribution of D_2 surrounding ORIs. The horizontal axis represents the nucleotide position, which ranges from −300 bp to +300 bp relative to ORIs (denoted as 0). The vertical axis represents the value of D_2.

sequences. As shown in **Figure 2B**, a peak near the ORIs and two valleys flanking the ORIs were observed, suggesting that the ORI sequences have very strong short-range correlations. It has been reported that D_2 is correlated with the evolutionary active region (Du et al., 2006). As a special region in the replication process, ORIs have a high probability of deletion, insertion, and mismatch (Umar and Kunkel, 1996). Thus, the evolutionary force reflected by the D_2 constraint indicates the diversity of ORI sequences. However, the evolutionary mechanism of fungi ORIs needs further investigation.

DISTRIBUTION OF ORIs IN THE GENOME

It is widely accepted that functional regions are not randomly distributed in the genome (Zhang et al., 2007). Based on this hypothesis, we statistically analyzed the distribution of ORIs in the yeast genome.

First, we investigated the position relationship between ORIs and nucleosomes. Nucleosomes are the elementary units of chromatin organization and are composed of a ∼147 bp stretch of DNA that is tightly wrapped around a histone core (Richmond and Davey, 2003; Segal et al., 2006). Nucleosome positioning affects nearly every cellular process that requires protein access to genomic DNA (Lee et al., 2007; Kaplan et al., 2009). Thus, it is worth studying the nucleosome occupancy around ORIs. To examine the distribution of nucleosomes around ORIs, we selected regions from −1000 to 1000 bp flanking ORIs and then mapped the nucleosomes in these regions. The average nucleosome occupancy scores surrounding ORIs *in vitro* and *in vivo* (ethanol, YPD, and galactose) are shown in **Figure 3**. The nucleosome occupancies around ORIs both *in vitro* and *in vivo* display a similar tendency: i.e., the nucleosome occupancy scores in ORIs are significantly lower than those in flanking regions, indicating that ORIs always appear in the nucleosome-free regions. This result can be explained as follows: once wrapped around the histone core, it is difficult for regulatory proteins to access the regions, which makes it difficult to open the DNA double helix (Kass and Wolffe, 1998).

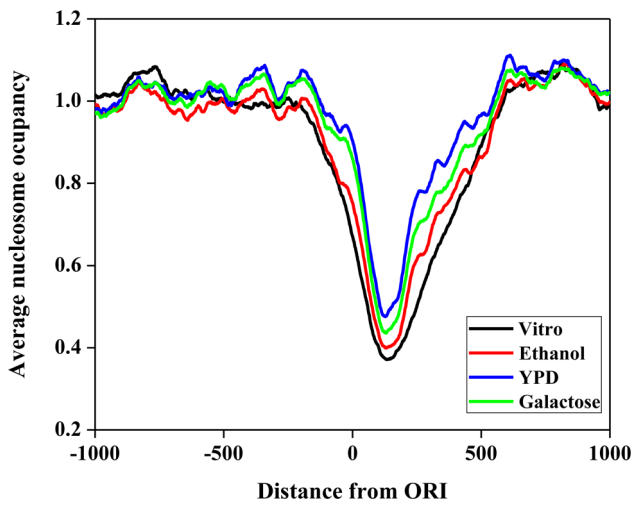


FIGURE 3 | Nucleosome occupancy around ORIs. The black curve represents the *in vitro* data. The red, blue, and green curves represent *in vivo* experimental maps for three growth conditions (ethanol, yeast extract, peptone, and dextrose medium [YPD] and galactose).

Gene transcription also requires the opening of the DNA double helix. Thus, there are coupling effects between ORIs and promoters. In fact, several studies focused on replication–transcription interactions (Rocha, 2004; Sequeira-Mendes and Gomez, 2012; Helmrich et al., 2013; Lubelsky et al., 2014). Here, the distance between ORIs and TSSs in the yeast genome was calculated. For over 31.46% of cases, the distance between ORI and TSS was less than 500 bp. These promoters are also AT-rich sequences (Lee et al., 2001). Thus, these promoters might share elements with ORIs.

Origins of replications are associated with bias in gene density (Necsulea et al., 2009). To further investigate the relationship between replication and transcription, we analyzed the distribution of ORIs in three kinds of intergenic regions. We obtained 2770 tandem, 1514 divergent, and 1497 convergent intergenic regions based on the orientations of the adjacent gene pair from the GenBank database. The tandem and divergent intergenic regions usually contain promoters; especially, each divergent intergenic region has two promoters for the transcription of two genes, whereas no promoter appears in convergent intergenic regions. By mapping ORIs in these regions, we found that 12.9% of ORIs are located in convergent regions, 25.1% are located in tandem regions, and 12.9% are located in divergent regions. The remaining ORIs (about 46.8%) overlap with coding regions, including 16.3% that are found in the tail of coding regions and 6.6% that are in the head of genes. These results suggest that most ORIs are not biased to transcription start regions, which may guarantee the coordination of replication and transcription.

Table 1 | Predicted results of different parameters using a support vector machine[a].

Method	Performance evaluation[b]		
	Sn	***Sp***	***Acc***
GC profile	0.7605	0.7728	0.7667
GC skew	0.6247	0.5778	0.6012
D_2	0.5309	0.5704	0.5506
Nucleosome (*in vitro*)	0.7448	0.7575	0.7511
Nucleosome (ethanol)	0.7071	0.7840	0.7456
Nucleosome (YPD)	0.7567	0.7811	0.7689
Nucleosome (galactose)	0.7485	0.7910	0.7697

[a] *The software package LIBSVM (version 3.17) was used to implement the support vector machine. The best separating hyperplane was constructed using the basis of radial basis kernel function. The regularization parameter C and the kernel width parameter γ were optimized using the grid-search approach.*

[b] *The three metrics, sensitivity (Sn), specificity (Sp), and overall accuracy (Acc), we re defined as Sn = TP/(TP+FN), Sp = TN/(TN+FP), and Acc = (TP+TN)/(TP+TN+FP+FN), respectively, where TP denotes the number of correctly predicted ORIs, FN denotes the number of ORIs that were predicted as non-ORIs, FP denotes the number of non-ORIs that were predicted as ORIs, and TN denotes the number of correctly predicted non-ORIs.*

PREDICTION OF ORIs

The aim of the above statistical analysis was to gain intrinsic observations to understand the replication initiation mechanism and to provide enough information for ORI prediction. Thus, we evaluated the predicted accuracies of the GC profile, GC skew, information redundancy D_2, and nucleosome occupancy to discriminate the ORIs from non-ORIs using a support vector machine. Here, 300 bp of each ORI was selected as the positive set, while the 300 bp upstream of ORIs was extracted as the negative set. The 10-fold cross-validated results are recorded in **Table 1**. It is obvious that the nucleosome occupancy feature can more accurately predict ORIs than GC skew and D_2. The comparative accuracy was also obtained with the GC profile. However, these results are still far from satisfactory. The features of GC profile, GC skew, and D_2 are based on the nucleotide sequence content, in which little sequence-order effect was considered. In the future, we will consider the sequence-order effect to improve the prediction quality.

CONCLUSION

Despite several studies focusing on DNA replication, the mechanism of replication initiation remains elusive. This study focused on the ORIs of *S. cerevisiae* and systematically analyzed the sequences surrounding ORIs. We found that the sequence around ORIs had a lower GC profile score and a higher nucleotide correlation than the sequence in flanking regions. DNA replication is a highly regulated process that relies on interactions between regulatory proteins and DNA sequences. The AT-rich motif is easily recognized by ORC. By studying the distribution of ORIs in genomes, we found that DNA replication initiation usually occurs in nucleosome-free regions. The short distance between ORIs and TSSs suggested that the expression of genes may be influenced by DNA replication. We expect that the observed properties of ORIs in this work will influence research related to ORIs and provide novel insights into regulatory mechanisms of DNA replication.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their constructive comments. This work was supported by the National Nature Scientific Foundation of China (No. 61202256, 61301260, and 61100092), the Nature Scientific Foundation of Hebei Province (No. C2013209105), and the Fundamental Research Funds for the Central Universities (No. ZYGX2012J113, ZYGX2013J102).

REFERENCES

Bailey, T. L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28–36.

Bauer, M., Schuster, S. M., and Sayood, K. (2008). The average mutual information profile as a genomic signature. *BMC Bioinformatics* 9:48. doi: 10.1186/1471-2105-9-48

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., et al. (2013). GenBank. *Nucleic Acids Res.* 41, D36–D42. doi: 10.1093/nar/gks1195

Cayrou, C., Coulombe, P., Puy, A., Rialle, S., Kaplan, N., Segal, E., et al. (2012). New insights into replication origin characteristics in metazoans. *Cell Cycle* 11, 658–667. doi: 10.4161/cc.11.4.19097

Chang, F. J., May, C. D., Hoggard, T., Miller, J., Fox, C. A., and Weinreich, M. (2011). High-resolution analysis of four efficient yeast replication origins reveals new insights into the ORC and putative MCM binding elements. *Nucleic Acids Res.* 39, 6523–6535. doi: 10.1093/nar/gkr301

Chen, W., Feng, P., and Lin, H. (2012). Prediction of replication origins by calculating DNA structural properties. *FEBS Lett.* 586, 934–938. doi: 10.1016/j.febslet.2012.02.034

Chuang, R. Y., and Kelly, T. J. (1999). The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. *Proc. Natl. Acad. Sci. U.S.A.* 96, 2656–2661. doi: 10.1073/pnas.96.6.2656

Crampton, A., Chang, F., Pappas, D. L. Jr., Frisch, R. L., and Weinreich, M. (2008). An ARS element inhibits DNA replication through a SIR2-dependent mechanism. *Mol. Cell* 30, 156–166. doi: 10.1016/j.molcel.2008.02.019

Dhar, M. K., Sehgal, S., and Kaul, S. (2012). Structure, replication efficiency and fragility of yeast ARS elements. *Res. Microbiol.* 163, 243–253. doi: 10.1016/j.resmic.2012.03.003

Du, D., Xie, T. U., and Li, H. (2006). Evolutional heterogeneity of DNA sequence of the *C. elegans* along the chromosome. *Acta Scientiarum Naturalium Universitatis NeiMongol* 37, 516–523.

Gao, F., Luo, H., and Zhang, C. T. (2012). DeOri: a database of eukaryotic DNA replication origins. *Bioinformatics* 28, 1551–1552. doi: 10.1093/bioinformatics/bts151

Gao, F., and Zhang, C. T. (2006). GC-Profile: a web-based tool for visualizing and analyzing the variation of GC content in genomic sequences. *Nucleic Acids Res.* 34, W686–W691. doi: 10.1093/nar/gkl040

Gao, F., and Zhang, C. T. (2007). DoriC: a database of oriC regions in bacterial genomes. *Bioinformatics* 23, 1866–1867. doi: 10.1093/bioinformatics/btm255

Grosse, I., Herzel, H., Buldyrev, S. V., and Stanley, H. E. (2000). Species independence of mutual information in coding and noncoding DNA. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* 61, 5624–5629. doi: 10.1103/PhysRevE.61.5624

Helmrich, A., Ballarino, M., Nudler, E., and Tora, L. (2013). Transcription-replication encounters, consequences and genomic instability. *Nat. Struct. Mol. Biol.* 20, 412–418. doi: 10.1038/nsmb.2543

Jacob, F., Brenner, S., and Cuzin, F. (1963). On the regulation of DNA replication in bacteria. *Cold Spring Harb. Symp. Quant. Biol.* 28, 329–348. doi: 10.1101/sqb.1963.028.01.048

Kaplan, N., Moore, I. K., Fondufe-Mittendorf, Y., Gossett, A. J., Tillo, D., Field, Y., et al. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* 458, U362–U129. doi: 10.1038/nature07667

Kass, S. U., and Wolffe, A. P. (1998). DNA methylation, nucleosomes and the inheritance of chromatin structure and function. *Novartis Found. Symp.* 214, 22–35.

Lee, M., Yu, S. H., Lee, Y., and Park, J. S. (2001). Characterization of nuclear factors binding to AT-rich element in the rat p53 promoter. *J. Cell. Biochem.* 80, 580–588. doi: 10.1002/1097-4644(20010315)80:4<580::aid-jcb1012>3.0.co;2-c

Lee, W., Tillo, D., Bray, N., Morse, R. H., Davis, R. W., Hughes, T. R., et al. (2007). A high- resolution atlas of nucleosome occupancy in yeast. *Nat. Genet.* 39, 1235–1244. doi: 10.1038/ng2117

Levitsky, V. G., Katokhin, A. V., Podkolodnaya, O. A., Furman, D. P., and Kolchanov, N. A. (2005). NPRD: nucleosome positioning region database. *Nucleic Acids Res.* 33, D67–D70. doi: 10.1093/nar/gki049

Lobry, J. R. (1996a). Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* 13, 660–665. doi: 10.1093/oxfordjournals.molbev.a025626

Lobry, J. R. (1996b). Origin of replication of *Mycoplasma genitalium*. *Science* 272, 745–746. doi: 10.1126/science.272.5262.745

Lubelsky, Y., Prinz, J. A., Denapoli, L., Li, Y. L., Belsky, J. A., and Macalpine, D. M. (2014). DNA replication and transcription programs respond to the same chromatin cues. *Genome Res.* 24, 1102–1114. doi: 10.1101/gr.160010.113

Luo, H., Zhang, C. T., and Gao, F. (2014). Ori-Finder 2, an integrated tool to predict replication origins in the archaeal genomes. *Front. Microbiol.* 5:482. doi: 10.3389/fmicb.2014.00482

Luo, L., Lee, W., Jia, L., Ji, F., and Tsai, L. (1998). Statistical correlation of nucleotides in a DNA sequence. *Phys. Rev. E* 58, 861–871. doi: 10.1103/PhysRevE.58.861

McLean, M. J., Wolfe, K. H., and Devine, K. M. (1998). Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. *J. Mol. Evol.* 47, 691–696. doi: 10.1007/PL00006428

Mojardin, L., Vazquez, E., and Antequera, F. (2013). Specification of DNA replication origins and genomic base composition in fission yeasts. *J. Mol. Biol.* 425, 4706–4713. doi: 10.1016/j.jmb.2013.09.023

Necsulea, A., Guillet, C., Cadoret, J. C., Prioleau, M. N., and Duret, L. (2009). The relationship between DNA replication and human genome organization. *Mol. Biol. Evol.* 26, 729–741. doi: 10.1093/molbev/msn303

Necsulea, A., and Lobry, J. R. (2007). A new method for assessing the effect of replication on DNA base composition asymmetry. *Mol. Biol. Evol.* 24, 2169–2179. doi: 10.1093/molbev/msm148

Nieduszynski, C. A., Knox, Y., and Donaldson, A. D. (2006). Genome-wide identification of replication origins in yeast by comparative genomics. *Genes Dev.* 20, 1874–1879. doi: 10.1101/gad.385306

Otu, H. H., and Sayood, K. (2003). A divide-and-conquer approach to fragment assembly. *Bioinformatics* 19, 22–29. doi: 10.1093/bioinformatics/19.1.22

Reeves, R., and Beckerbauer, L. (2001). HMGI/Y proteins: flexible regulators of transcription and chromatin structure. *Biochim. Biophys. Acta* 1519, 13–29. doi: 10.1016/s0167-4781(01)00215-9

Richmond, T. J., and Davey, C. A. (2003). The structure of DNA in the nucleosome core. *Nature* 423, 145–150. doi: 10.1038/nature01595

Rocha, E. P. C. (2004). The replication-related organization of bacterial genomes. *Microbiology* 150, 1609–1627. doi: 10.1099/mic.0.26974-0

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I. K., et al. (2006). A genomic code for nucleosome positioning. *Nature* 442, 772–778. doi: 10.1038/nature04979

Sequeira-Mendes, J., and Gomez, M. (2012). On the opportunistic nature of transcription and replication initiation in the metazoan genome. *Bioessays* 34, 119–125. doi: 10.1002/bies.201100126

Siow, C. C., Nieduszynska, S. R., Muller, C. A., and Nieduszynski, C. A. (2012). OriDB, the DNA replication origin database updated and extended. *Nucleic Acids Res.* 40, D682–D686. doi: 10.1093/nar/gkr1091

Takahashi, T., Ohara, E., Nishitani, H., and Masukata, H. (2003). Multiple ORC-binding sites are required for efficient MCM loading and origin firing in fission yeast. *EMBO J.* 22, 964–974. doi: 10.1093/emboj/cdg079

Umar, A., and Kunkel, T. A. (1996). DNA-replication fidelity, mismatch repair and genome instability in cancer cells. *Eur. J. Biochem.* 238, 297–307. doi: 10.1111/j.1432-1033.1996.0297z.x

Wu, L., Liu, Y., and Kong, D. (2014). Mechanism of chromosomal DNA replication initiation and replication fork stabilization in eukaryotes. *Sci. China Life Sci.* 57, 482–487. doi: 10.1007/s11427-014-4631-4

Xing, Y. Q., Liu, G. Q., Zhao, X. J., and Cai, L. (2013). An analysis and prediction of nucleosome positioning based on information content. *Chromosome Res.* 21, 63–74. doi: 10.1007/s10577-013-9338-z

Xing, Y. Q., Liu, G. Q., Zhao, X. J., Zhao, H. Y., and Cai, L. (2014). Genome-wide characterization and prediction of *Arabidopsis thaliana* replication origins. *Biosystems* 124C, 1–6. doi: 10.1016/j.biosystems.2014.07.001

Yamashita, R., Sathira, N. P., Kanai, A., Tanimoto, K., Arauchi, T., Tanaka, Y., et al. (2011). Genome-wide characterization of transcriptional start sites in humans by integrative transcriptome analysis. *Genome Res.* 21, 775–789. doi: 10.1101/gr.110254.110

Yin, S. Y., Deng, W. J., Hu, L. D., and Kong, X. Y. (2009). The impact of nucleosome positioning on the organization of replication origins in eukaryotes. *Biochem. Biophys. Res. Commun.* 385, 363–368. doi: 10.1016/j.bbrc.2009.05.072

Yu, Z. G., and Jiang, P. (2001). Distance, correlation and mutual information among portraits of organisms based on complete genomes. *Phys. Lett. A* 286, 34–46. doi: 10.1016/S0375-9601(01)00336-X

Yuan, G. C., Liu, Y. J., Dion, M. F., Slack, M. D., Wu, L. F., Altschuler, S. J., et al. (2005). Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* 309, 626–630. doi: 10.1126/science.1112178

Zhang, C. T., Gao, F., and Zhang, R. (2005). Segmentation algorithm for DNA sequences. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 72, 041917. doi: 10.1103/PhysRevE.72.041917

Zhang, Z. D., Paccanaro, A., Fu, Y., Weissman, S., Weng, Z., Chang, J., et al. (2007). Statistical analysis of the genomic distribution and correlation of regulatory elements in the ENCODE regions. *Genome Res.* 17, 787–797. doi: 10.1101/gr.5573107

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: 05 September 2014; accepted: 11 October 2014; published online: 18 November 2014.

*Citation: Li W-C, Zhong Z-J, Zhu P-P, Deng E-Z, Ding H, Chen W and Lin H (2014) Sequence analysis of origins of replication in the Saccharomyces cerevisiae genomes. Front. Microbiol. **5**:574. doi: 10.3389/fmicb.2014.00574*

This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology.

Copyright © 2014 Li, Zhong, Zhu, Deng, Ding, Chen and Lin. 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.

frontiers in
MICROBIOLOGY

REVIEW ARTICLE
published: 29 October 2014
doi: 10.3389/fmicb.2014.00573

An integrated overview of spatiotemporal organization and regulation in mitosis in terms of the proteins in the functional supercomplexes

Yueyuan Zheng[1†], Junjie Guo[1†], Xu Li[2†], Yubin Xie[1], Mingming Hou[1], Xuyang Fu[1], Shengkun Dai[1], Rucheng Diao[1], Yanyan Miao[1]* and Jian Ren[1]*

[1] *Cancer Center, School of Life Sciences, School of Advanced Computing, Cooperative Innovation Center for High Performance Computing, Sun Yat-sen University, Guangzhou, China*
[2] *Orthopaedic Department of Anhui Medical University Affiliated Provincial Hospital, Hefei, China*

Edited by:
Feng Gao, Tianjin University, China

Reviewed by:
Zhixiang Zuo, University of Chicago, USA
Ziding Zhang, China Agricultural University, China
Penghui Zhou, Dana Farber Cancer Institute, USA

****Correspondence:***
Yanyan Miao and Jian Ren, Cancer Center, School of Life Sciences, School of Advanced Computing, Cooperative Innovation Center for High Performance Computing, Sun Yat-sen University, 135 West Xingang Road, Guangzhou 510275, China
e-mail: myy@mail.ustc.edu.cn; renjian.sysu@gmail.com

† The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

Eukaryotic cells may divide via the critical cellular process of cell division/mitosis, resulting in two daughter cells with the same genetic information. A large number of dedicated proteins are involved in this process and spatiotemporally assembled into three distinct super-complex structures/organelles, including the centrosome/spindle pole body, kinetochore/centromere and cleavage furrow/midbody/bud neck, so as to precisely modulate the cell division/mitosis events of chromosome alignment, chromosome segregation and cytokinesis in an orderly fashion. In recent years, many efforts have been made to identify the protein components and architecture of these subcellular organelles, aiming to uncover the organelle assembly pathways, determine the molecular mechanisms underlying the organelle functions, and thereby provide new therapeutic strategies for a variety of diseases. However, the organelles are highly dynamic structures, making it difficult to identify the entire components. Here, we review the current knowledge of the identified protein components governing the organization and functioning of organelles, especially in human and yeast cells, and discuss the multi-localized protein components mediating the communication between organelles during cell division.

Keywords: super-complex structures, cell division/mitosis, protein components, centrosome, kinetochore, midbody

INTRODUCTION

Cell division/mitosis is a precisely modulated process of chromosome segregation and nuclear division in which one eukaryotic cell divides into two daughter cells with identical chromosomes in order to produce more cells for growth and replace any damaged, dying or senescent cells (Sancar et al., 2004). Mitosis is always accompanied by a separation of the cell cytoplasm, known as cytokinesis, in which the daughter cells become completely separated (Wheatley et al., 2001; Straight et al., 2003). Mitosis (nuclear division) and cytokinesis (cytoplasmic division), which define the mitotic (M) phase, are the most crucial and fundamental activities of the eukaryotic cell cycle. Before entering the M phase of the cell cycle, the cell undergoes a period of growth and maturation during the interphase, duplicating genetic materials and organelles for the performance of cell division (Heun et al., 2001). The interphase and M phase of the cell cycle are complex and highly regulated by numerous proteins which are spatially and temporally organized as protein super-complexes. The super-complexes carry out chromosome replication and alignment, sister chromatid separation and cytoplasm division (Straight et al., 2003).

Chromosome must be precisely replicated once per cell cycle to maintain genome integrity. Eukaryotic cells may use multiple proteins, many of which are also involved in super-complex formation to regulate chromosome alignment, separation and cytoplasm division, to control the origins of chromosome replication. During the interphase, the origin recognition complex (ORC), a six-subunit complex comprised of ORC1-6, binds to chromosomes at the replication origin sites and acts as a central component for eukaryotic chromosome replication initiation (Bell and Dutta, 2002). As the initiation of replication is a central event in cell cycle, the identification of replication origin sites and its binding proteins is essential to the understanding of DNA replication. Benefit from recent genome-wide approaches, a huge number of replication origins and ORC proteins were identified. Also, several specialized databases, such as DeOri (Gao et al., 2012), have been developed to assist the comprehensive study on eukaryotic DNA replication. Over the years, new roles for many ORC proteins were revealed in cells. Unlike their regular function that controls the initiation of DNA replication, a fair amount of ORC proteins also binds to other cell cycle-related organelles, including centrosome, kinetochore and midbody. Evidences have shown that ORC1 and

ORC2 can regulate centrosome duplication and a depletion of them resulted in abnormal centrosome copy number (Prasanth et al., 2004; Hemerly et al., 2009). Coincidentally, researches also demonstrated that ORC6 and ORC2 can localize to kinetochore. The absence of ORC proteins may lead to kinetochore dysfunction (Shimada and Gasser, 2007). Furthermore, in anaphase, the ORC6 may target to the midbody in controlling of chromosome segregation (Prasanth et al., 2002). In the process of chromosome replication, the enzymes that catalyze DNA duplication are unable to reach the very end of the chromosome. Chromosome has a special DNA structure named telomere at the end. Thus, in the course of each replication, the length of the telomeres is shortened (Von Zglinicki, 2002). Once the telomeres shrink to a critical minimum size, the cells no longer divide and ultimately become senescent or die (Hahn et al., 1999; Henson et al., 2002). However, telomerase, a unique protein-RNA complex that is activated in certain cells (Rudolph et al., 1999; Hanahan and Weinberg, 2000), such as yeast cells, stem cells, reproductive cells and cancer cells, is responsible for elongating telomeres (Herbert et al., 1999; Dunham et al., 2000). It thus prevents the chromosome degradation, maintains the stability of the genome and assists cells to escape the fate of being unable to continue division (Hoeijmakers, 2001).

In the M phase of the cell cycle, multiple proteins assemble in the three distinct regions of the centrosome/spindle pole body, kinetochore/centromere and cleavage furrow/midbody/bud neck, directing the process of cell division. The centrosomes in animal cells, spindle pole bodies (SPB) in budding yeast and related/homologous structures in other organisms have been characterized as the microtubule organizing centers (MTOCs) (Veith et al., 2005), which participate in the organization and orientation of the mitotic spindle apparatus, and thus direct the chromosome alignment and sister chromatids segregation during cell division. In addition, the kinetochore, a specialized protein complex which is dynamically assembled around the centromere of chromosomes (Ditchfield et al., 2003), acts as the "handle" of the chromosome and specifies the attachments between the chromosomes and spindle to ensure accurate chromosome segregation (Hauf et al., 2003). Dysfunction of the centrosome/spindle pole body and kinetochore/centromere is catastrophic for cells and contributes to aberrant division and chromosome instability (Fodde et al., 2001), both of which are hallmarks of cancer cells (Schuyler et al., 2012). The chromosome separation in animal cells is always accompanied by cytokinesis, which begins with ingression of the cleavage furrow mediated by the actomyosin ring (Somers and Saint, 2003), followed by the formation dense structure of the midbody (Gromley et al., 2005) which is also known as the phragmoplast in plants (Van Damme et al., 2004) and the bud neck in budding yeast (Vallen et al., 2000; Caviston et al., 2003). Numerous proteins are recruited to the midbody and form a super-complex which mediates the midbody abscission in order to perform cytokinesis, with complete separation of the two daughter cells (Adams et al., 2001; Wheatley et al., 2001; Mollinari et al., 2002).

Although the importance of organelles to cell biology has been repeatedly demonstrated by multiple reports over the past decades, many aspects of their function, structure and composition are still largely unknown. In this regard, comprehensive identification of the protein components of the super-complex structures will be one of the keys to understanding the mechanisms of chromosome segregation and cytokinesis, and may provide important clues for the discovery and validation of new therapeutic targets. Recently, many protein components have been identified, but according to a combination of proteomic analysis, biochemical studies and genetic screening, there still remain a large number of proteins that are predicted to be associated with these organelles (**Table 1**). In this review, we will present a general overview of the identified components of the super-complexes involved in the mitosis and cytokinesis with the aim of integrating the relevant information of organelles and thus broadening the knowledge of cell division. Remarkably, the process of the cell division is highly conserved in eukaryotic cells, we therefore briefly review the two commonly studied systems, human and yeast cells.

THE CENTROSOME

As a complex and dynamic organelle, the MTOC contributes to both microtubule organization and nucleation, which are important for chromosomes separation during mitosis (Brinkley, 1985; Luders and Stearns, 2007). Multiple proteins must be involved in manipulating MTOC functions, controlling its duplication and driving maturation. To further clarify the functional processes of the organization and regulation of the MTOC, the protein components must be identified. Recently, evidence from

Table 1 | The number of proteins located in centrosome, kinetochore, midbody with experimental verification and predicted in 7 different species from MiCroKiTS (Updated June 27, 2014).

Organism	Centrosome		Kinetochore		Midbody	
	Experimental verification	Predicted from orthologs	Experimental verification	Predicted from orthologs	Experimental verification	Predicted from orthologs
H. sapiens	516	112	203	82	229	92
M. musculus	131	477	33	249	22	289
X. laevis	36	0	29	0	6	0
C. elegans	35	227	59	129	20	148
D. melanogaster	67	240	58	146	29	152
S. cerevisiae	89	107	102	56	133	50
S. pombe	48	145	91	66	38	108

a combination of genetic and biochemical studies has revealed many important MTOC-associated proteins in a variety of species (Masuda et al., 2013). However, according to the MiCroKiTS database (Ren et al., 2010) (http://microkit.biocuckoo.org/), an integrated database of the midbody, centrosome and kinetochore most recently updated in June 27, 2014, a large number of proteins that are predicted to be located on the MTOC are still not well validated (**Table 1**). Confirmation of the functions of these predicted proteins has broad implications for the understanding of the MTOC.

The centrosome is the primary MTOC, which contains two orthogonally arranged centrioles and the surrounding pericentriolar material (PCM) (Nigg and Raff, 2009). The centriole, composed mainly of tublin, is a typically cylindrical organelle made up of nine triplets of microtubules in most animal eukaryotic cells (Kitagawa et al., 2011), although absent in most fungi and high plant cells (Bell and Dutta, 2002; Gao et al., 2012). In the G1 phase of the cell cycle, the paired centrioles, termed the mother and the daughter centrioles, are connected via interconnecting fibers. Morphologically distinct from the daughter centriole, the mother centriole has both distal and subdistal appendages that serve to anchor the centrioles to the plasma membrane (Bettencourt-Dias and Glover, 2009). Recently, several components of the centriole appendages have been described, such as the distal appendage proteins CEP164 and CEP89, as well as three novel components of CEP83, the Sodium channel and clathrin linker 1 (SCLT1) and the Fas-binding factor 1 (FBF1) (Tanos et al., 2013; Kloc et al., 2014). The subdistal appendage proteins include Outer dense fiber 2 (ODF2; also known as cenexin) (Chang et al., 2003), ninein (Graser et al., 2007), epsilon-tubulin (Chang et al., 2003), Centriolin (Gromley et al., 2003), and CC2D2A (Veleri et al., 2014). However, the molecular composition and the exact functions of the appendages remain largely unclear. The mechanisms underlying the assembly of the centriole are still poorly understood. In recent years, identification of the proteins that are responsible for centriole formation has advanced the understanding of the assembly mechanisms. In human cells, spindle assembly abnormal 6 (HsSAS-6), Polo-like kinase 4 (PLK4), SCL/TAL1 interrupting locus (STIL), centrosomal P4.1-associated protein (CPAP) (Brownlee and Rogers, 2013), located at the centriole, have been identified as the core components required for centriole assembly. Using proteomic and biochemical analysis as well as genetic screening, a list of the proteins associated with centriole, such as centrosomal protein of 135 kDa (CEP135), CEP152, CEP63, spindle and centriole-associated protein (SPICE), CP110, centrobin, CEP120, and CEP192 (Gonczy, 2012), are considered to govern the centriole assembly. The maintenance of a constant centriole number is critical for the progression of the cell cycle, and precisely controlled by numerous proteins which are involved in regulating centrosome duplication in the G1 and S phases, centrosome maturation in the G2/M phase and separation in the mitotic phase (Brownlee and Rogers, 2013). In human cells, PLK4, hsSAS-6 and STIL are three regulators necessary for centrosome duplication (Vulprecht et al., 2012). Following the activation of PLK4 and accumulation of STIL around the mother centriole, F-box protein FBXW5 stabilizes HsSAS-6 (Puklowski et al., 2011). In addition, several other proteins, such as CEP135, CPAP (Tang et al., 2009), γ-tubulin, CEP192, BRCA2, CP110 and its interaction protein USP33 (Li et al., 2013) that are essential for centrosome duplication, are recruited to the centriole, thus orchestrating centrosome duplication (Brownlee and Rogers, 2013). Additionally, cell cycle kinase CDK2 as well as potential partners of cyclin A and cyclin E are required for the two centrioles to split during centrosome duplication (Stearns, 2001). At the onset of mitosis, NEK2 and centrin are required for the sister centrosome disjunction as well as the formation of the two spindle poles during mitosis (Hinchcliffe and Sluder, 2001). Centrosome maturation is accompanied by the recruitment of many proteins to the centrioles and a dramatic expansion of the pericentriolar matrix (PCM) (Mennella et al., 2014). Phosphorylation is considered to be a key mechanism underlying centrosome maturation. The Polo-like kinases1 (PLK1) (Barr et al., 2004; Conduit et al., 2014) and Aurora kinases (Carmena and Earnshaw, 2003) have been identified as two important regulators of centrosome maturation. The specific phosphorylation of pericentrin (PCNT) by PLK1 results in the recruitment of many centrosomal proteins, such as γ-tubulin, Aurora A, PLK1, CEP192, and GCP-WD (γ-complex protein with WD repeats), to the centrosome during mitosis (Lee and Rhee, 2011). PCM is a matrix of proteins involved in centrosomal organization, microtubule nucleation and anchoring. The main components of PCM exist in the form of two proteins layers, one comprising a large number of coiled-coil proteins, such as pericentrin/pericentrin-like protein (PLP) and CEP152, with the other one including CEP215, γ-tubulin and CEP192 (Mennella et al., 2014). In PCM, γ-tubulin and other proteins such as γ tubulin complex protein (GCP) family can be assembled as γ-tubulin ring complexes (γ-TuRCs) for microtubule nucleation. The GCP family is also involved in the γ-TuRCs function, regulation and localization of γ-TuRCs (Kollman et al., 2011). The centrosomal protein pericentrin and the ninein-like protein (NLP) have been shown to anchor γ-TuRCs at the spindle poles (Zimmerman et al., 2004). Meanwhile, the precise components and regulators of γ-TuRCs remains incompletely understood.

Collectively, the identification of the structural and functional proteins of centrosome is clearly crucial for elucidating the structure of the centrosome and uncovering the underlying mechanisms in centrosome organization and regulation. Up to now, only a portion of the centrosome components have been detected (Tables S1, S2), and more efforts are required for the experimental validation of the remaining components. The centrosome in yeast cells is termed the spindle pole body (SPB), which is composed of a half-bridge for new SPB assembly, and three plaques, including an inner plaque for nuclear microtubules forms as the mitotic spindles originate, a central plaque spanning the nuclear membrane, and an outer plaque for cytoplasmic microtubules that used for karyogamy, nuclear positioning and spindle orientation (Seybold and Schiebel, 2013). The identified SPB proteins involved in the organization and regulation of SPB are listed in Table S2. However, there is still a large number of proteins located on the SPB that need to be further validated (**Table 1**).

The centrosome is a complex and precisely regulated organelle for bipolar spindle assembly, primary cilia formation, cell division and certain other cellular processes, including cell migration,

protein degradation and axonal growth in human cells. More recent studies have shown that aberrant organization of centrosome resulting from defects in structural and functional proteins of the centrosome (Ganem et al., 2009; Nigg and Raff, 2009) is linked to neurodegenerative, Bardet–Biedl syndrome (Swaminathan, 2004), microcephaly (Marthiens et al., 2013), cystic kidney disease (Ong and Wheatley, 2003) and tumorigenesis (Marina and Saavedra, 2014). Thus, identification of the centrosomal proteins and clarification of the mechanisms underlying the centrosome assembly and regulation may lead to new drug targets, diagnostics or therapeutic approaches.

THE KINETOCHORE

During mitosis in eukaryotic cells, a large number of proteins are assembled as a unique protein complex called the kinetochore, at the surface of the centromeric chromatin/centromere. The kinetochore functions as the binding site of the spindle microtubules to chromatin and directs sister chromatid segregation (Cheeseman and Desai, 2008). The protein components of the kinetochore modulate the connection between the centromeric chromatin and microtubules from the mitotic spindle to facilitate the proper segregation of the chromosomes during cell division (Gonen et al., 2012). According to the MiCroKiTS database and a comprehensive literature review (Cheeseman and Desai, 2008; Gonen et al., 2012), many kinetochore proteins have been identified in different species (**Table 1**). However, there are still a number of proteins localized at the kinetochore without any functional validation, as shown in the MiCroKiTS database.

The kinetochore is a complex and dynamic structure of variable size and shape. It is difficult to obtain the structural information on the complete kinetochore, so the structure is still not entirely clear. Previous studies have revealed that the overall positioning, main components and architecture of kinetochore are highly conserved from yeast to human (Quarmby and Parker, 2005). Many copies of centromeric proteins are assembled as a trilaminar kinetochore structure with the inner layer, a platform for kinetochore assembly that is located on the centromeric chromatin, the outer layer, responsible for the interaction with spindle microtubules, and the central layer, a region that links the inner and outer layers. In vertebrate cells, the inner layer consists of at least 18 centromeric proteins (Santaguida and Musacchio, 2009). Histone H3 variant centromeric protein A (CENP-A), also known as Cse4 in budding yeast, is one inner layer protein that may function as an early epigenetic marker for centromere localization and formation by making the centromeres distinct from the rest of the chromosome (Barnhart et al., 2011; Guse et al., 2011; Henikoff et al., 2014). CENP-A, together with CENP-B and CENP-C, are three main auto-antigens recognized by anticentromeric antibodies (Masumoto et al., 1989). Many other CENPs are also included in the inner layer, such as CENP-H, CENP-I, and CENP-K–W, all of which along with CENP-C colocalize with CENP-A and constitute the constitutive centromere-associated network (CCAN) (Cheeseman and Desai, 2008) (Table S3). Most of the components of the inner layer are evolutionarily conserved. They are responsible for keeping the kinetochore tethered to the centromere throughout the cell cycle and are essential for outer layer assembly (Carroll and Straight, 2006; Okada et al., 2006; Tanaka et al., 2009). The outer layer of the kinetochore is composed of several super-complexes, including Mis12, Ndc80 and Ska. The Mis12 complex provides the main platform for outer layer assembly, and consists of MIS12, NSL1, NNF1, and DSN1 (Screpanti et al., 2011). The Knl1 complex, which consists of KNL-1 and ZWINT, has been shown to recruit other outer layer proteins, such as spindle assembly checkpoint (SAC) proteins, CENP-F and the Rod–ZW10–Zwilch (RZZ) complex. The Ndc80 complex (NDC80, NUF2, SPC24, and SPC25) is one of the core binding sites of kinetochore-microtubules (kMTs) (Malvezzi et al., 2013). The Ska complex, composed of SKA1, SKA2, and SKA3, is essential for stabilizing kMT attachement (Welburn et al., 2009). In addition, the Knl1 complex, together with the Mis12 and Ndc80 complexes, forms the core of a highly conserved KMN network. This network is required for effective kMT attachment and force generation, and regulated by the Ska complex (the Dam/Dash complex in yeast) (Varma and Salmon, 2012). During mitosis, the components of the SAC, a mechanism that acts in response to unattached kinetochores, are recruited to the kinetochore monitor the correct kMT attachment by inhibiting the polyubiquitylation activities of the anaphase promoting complex (APC) (Peters, 2006). Several SAC components have been identified to date, including the non-kinase components Mad1, Mad2 and Bub3, the kinase components BubR1 (Mad3 in budding yeast), Bub1 and Mps1, the RZZ complex and other proteins (Lara-Gonzalez et al., 2012), as shown in Table S4. Among these components, Mad2 can interact with the APC activator of CDC20 to negatively regulate its function for the purpose of APC inhibition (Yu, 2002). In recent studies, several other mitotic protein kinases, including Aurora B (Chan et al., 2012) and PLK1 (Kang et al., 2006), PP2A phosphatase (Schmitz et al., 2010) and a number of nuclear pore proteins, including the Nup107–160 complex and SEH1 (D'angelo and Hetzer, 2008), have also been shown to transiently localize to the kinetochore during mitosis. They are involved in accurate segregation of chromosomes and controlling kinetochore function (Cheeseman and Desai, 2008), possibly by modulating checkpoint signaling. As the main structural features of kinetochores are conserved from yeast to human, the kinetochore also consists of the inner and outer layers in yeast, and the kMT attachment is regulated by numerous SAC proteins (Tables S3, S4). Among the components of the kinetochore found in yeast and human, the Ndc80 complex and some of the SAC proteins are highly conserved and exist in both species, indicating the importance of these proteins for correct chromosome segregation during cell division.

A combination of biochemical, fluorescence-microscopy and electron microscopy (EM) studies has led to the proposal of several structural models of the kinetochore with only weak supporting evidence. However, in recent studies, the first three-dimensional images of the kinetochore core structure have been obtained from budding yeast. These images show that the size of the kinetochore is approximately 126 nm, with a large central hub surrounded by multiple outer globular domains that form a ring-like structure around the microtubules (Gonen et al., 2012). This finding is important and extends the knowledge of the kinetochore. To further the understanding of the assembly process of the kinetochore and the mechanisms underlying

chromosome segregation, additional kinetochore components and higher resolution images of kinetochore are needed to assist the elucidation of the structure and regulatory network. These are key elements in advancing our understanding of the mechanisms of the kinetochore-associated diseases, such as cancer, and may contribute to the development of early-stage clinical treatments (Gonen et al., 2012).

THE MIDBODY

During cytokinesis, many proteins promote furrow ingression, dividing one cell into two daughter cells still connected by midbody, a cellular substructure contains many transient protein complexes formed at the narrow intracellular bridge (Steigemann and Gerlich, 2009). The midbody is generally considered to be an important structure for directing the abscission and completely separating the two daughter cells at the final stage of cytokinesis (Pohl and Jentsch, 2008). However, more functions of the midbody are still unclear. According to recent studies, the midbody may also be involved in cell-fate determination. Morphologically, the midbody is a dense structure formed by a tightly packed anti-parallel microtubule array, and many proteins are recruited to this site to assist in the cytokinesis process (Mullins and Biesele, 1977; Steigemann and Gerlich, 2009). However, the current knowledge of the midbody components and the way the midbody proteins are organized is limited. To further clarify the functions and the processes of assembly and regulation of the midbody, the primary task is to identify its protein components. Although there are approximately 229 proteins identified as being associated with the midbody in human cells, and 133 proteins in yeast cells (**Table 1**), there are still many remaining components that urgently need to be uncovered and validated.

Previous studies have shown that the midbody proteins are organized in three parts, the bulge, the dark zone and the flanking zone (Mullins and Biesele, 1977; Steigemann and Gerlich, 2009). The bulge is at the center of the midbody, containing few bundled anti-parallel microtubules and various proteins. In human cells, centralspindlin, a key component of the bulge, is a complex of the human GTPase-activating protein MgcRacGAP and Mitotic kinesin-like protein 1 (MKLP1). Centralspindlin is essential for the midbody formation and links the midbody to the plasma membrane. Many of the identified bulge proteins are associated with centralspindlin. The ADP-ribosylation factor 6 (ARF6) GTPase can interact with centralspindlin and may be respectively responsible for midbody stabilization (Joseph et al., 2012). The Rho guanine nucleotide exchange factor (RhoGEF) Ect2 is also a centralspindlin-interacting protein and localizes at the bulge to facilitate midbody abscission (Yuce et al., 2005). The coiled-coil protein centriolin, recruited to the midbody by centralspindlin, is important for integrating the process of membrane-vesicle fusion with abscission by interacting with the exocyst components and SNARE complexes (Gromley et al., 2005). Another centralspindlin binding protein is a centrosomal protein of 55 kDa (CEP55) that is persistently localized at the midbody bulge during cytokinesis. The tumor-susceptibility gene 101 (TSG101) has been observed at the bulge. TSG101 and another midbody protein called Alg2-interacting protein X (ALIX) are associated with CEP55, and are proposed to be responsible for recruiting ESCRT-III components to the dark zone and thus assisting with the midbody abscission (Morita et al., 2007; Lee et al., 2008; Elia et al., 2011). The dark zone is a narrow region in the center of the midbody where antiparallel microtubules overlap. The microtubule-associated protein regulator of cytokinesis 1 (PRC1), in association with a microtubule-based motor protein of kinesin superfamily protein member 4 (KIF4), colocalizes at the midbody dark zone and together they are essential for cytokinesis (Kurasawa et al., 2004). Wnt5a signaling is important for stabilization. In recent studies, Wnt receptor Frizzled 2 (FZD2), which has been observed in the dark zone and has a similar localization pattern as the ESCRT-III subunit of CHMP4B, may regulate ESCRT-III localization via a Wnt5a-mediated β-catenin-independent signaling pathway (Fumoto et al., 2012). The midbody flanking zone resides outside of the dark zone, containing multi-proteins (Hu et al., 2012), such as the negative cytokinesis-regulator of centromere protein E (CENPE) (Liu et al., 2006), mitotic kinesin-like protein 2 (MKLP2) that regulates the localization of the chromosomal passenger complex (CPC) during cytokinesis (Gruneberg et al., 2004), and a CPC subunit of the Aurora B kinase-mediated abscission checkpoint (Steigemann et al., 2009). In yeast cells, the bud neck, which is analogous to the midbody, is responsible for cytokinesis and abscission (Guertin et al., 2002). And the main components of the organism are evolutionarily conserved from yeast to vertebrate (Otegui et al., 2005).

Actually, many components of each substructure of the midbody and bud neck subregions listed in Tables S5, S6 display a dynamic localization pattern, but the detailed composition of midbody and bud neck is still not known. In human cells, the midbody contains secretory or membrane-trafficking proteins, actin-associated proteins, microtubule-associated proteins, kinases proteins, and other uncharacterized or other function proteins, involved in many processes, such as the cytoskeleton, lipid rafts and vesicle trafficking (Skop et al., 2004). In addition, recent studies indicate that the functions of the midbody are not only related to abscission, but also involved in patterning, morphogenesis and development during embryogenesis (Chai et al., 2012). The accumulation of midbodies has been shown to correlate with the pluripotency of stem cells and to increase the tumorigenicity of cancer cells, while in differentiated cells, the midbody is degraded through an autophagy pathway (Ettinger et al., 2011; Kuo et al., 2011; Schink and Stenmark, 2011). Thus, identification of the midbody components is essential for advancing our knowledge of midbody and cell-fate determination, and also for exploring new therapeutic strategies for midbody related diseases treatment, such as cancer.

DISCUSSION

A large number of proteins have been shown to participate in the process of cell division and spatiotemporally assemble as super-complexes at defined subcellular localizations, such as kinetochores at the centromeric chromatin, the centrosome near the nucleus, and the midbody between two daughter cells. According to the MiCroKiTS database search, during cell division, there are a total of approximately 754 identified proteins localized at the organelles of the centrosome, kinetochore and midbody in

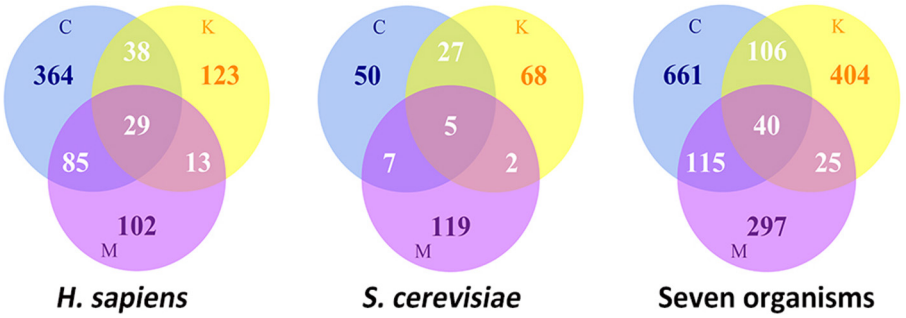


FIGURE 1 | The statistics of Location distributions of MicroKiTS proteins. C refers to Centrosome. K refers to Kinetochore. M refers to Midbody. Seven organisms include of *H. sapiens*, *S. cerevisiae*, *C. elegans*, *D. melanogaster*, *X. laevis*, *M. musculus*, and *S. prombe*.

Homo sapiens, and 278 in *Saccharomyces cerevisiae* (**Figure 1**). Despite the fact that the protein components of each organelle are recruited to a specific subcellular localization, some proteins exhibit multi-localization in various species. Collectively, there are approximately 165 proteins which have more than one subcellular localizations in *Homo sapiens*, while there are 41 proteins in *S. cerevisiae*.

Proteins with multiple localizations are the key factors for mediating the communication between the organelles. In human cells, Ndc80 complex dynamically localizes at centrosome, and then concentrates at centromere and becomes a stable component of kinetochore until completion of the mitosis (Hori et al., 2003). Ndc80 complex is required for the stable kinetochore-spindle microtubule attachments, which controls the chromosome alignment and segregation in mitosis (Wei et al., 2005). The kinetochore protein components of INCENP (Cooke et al., 1987), CENP-A(Liu et al., 2013) and Aurora B (Kimura and Okano, 2005) for the chromosome biorientation, and the centrosome proteins of BARD1 (Ryser et al., 2009), BRCA2 (Daniels et al., 2004) and CEP55 (Fabbro et al., 2005), can be recruited to the midbody for the progression of cytokinesis. PLK1 (Cdc5 in yeast), a key mitotic regulator that phosphorylates substrate proteins on several different mitotic structures in human cells, first localizes at the centrosome before associating with kinetochore, and then is recruited to the midbody (Petronczki et al., 2008). The dynamic localization of PLK1, mediated by the polo-box domain (PBD) and kinase activity, is critical for chromosome alignment, spindle assembly and cytokinesis (Petronczki et al., 2008; Liu et al., 2012). A ubiquitin-ligase complex of APC and the HECT E3 ligase Smurf2, both of which control the progression of mitosis and cytokinesis through ubiquitin modification of substrate proteins and thus altering the protein localization and stability, have also been found to be dynamically localized at the centrosome, kinetochore and midbody (Kurasawa and Todokoro, 1999; Osmundson et al., 2008). In yeast cells, 5 proteins, including Cdc5 (Snead et al., 2007), protein phosphatase 2A regulatory subunit RTS1 (Gentry and Hallberg, 2002) and TPD3 (Gentry and Hallberg, 2002), Casein kinase I homolog HRR25 (Lusk et al., 2007) and protein phosphatase PP1-2 (Bloecher and Tatchell, 2000), are spatiotemporally recruited to the SPB, kinetochore and bud neck, and precisely regulate the cell division progression by altering the phosphorylation state of the substrates proteins. The subcellular localization determines the biological activities of multi-localized proteins through controlling the access of these proteins to different interaction partners, and is critical for the formation of the dynamic protein-protein interaction network to govern the process of the cell division/mitosis. Meanwhile, the posttranslational modifications (PTMs), including phosphorylation and ubiquitylation, as well as altering of the subcellular localizations, are essential mechanisms used by multi-localized proteins to diversify function and regulate cell division. A latest analysis of the dynamics of proteome and phosphoproteome during the cell division of the fission yeast revealed that changes of proteome level are weak, whereas changes of protein phosphorylation states are the predominant events occurred in mitosis, indicating that phosphorylation is probably associated with the functions and localizations of the proteins, which are involved in regulating mitotic progression and completion(Carpy et al., 2014). Additionally, the progresses in proteome-wide analysis of ubiquitination modifications in cell division demonstrated that ubiquitination, which affect protein stability, activity, and localization, plays an important role in regulating the mitotic progression (Chuang et al., 2010; Merbl et al., 2013). Certainly, the current understanding of the mechanisms used by multi-localized proteins to dynamically control the formation and functions of subcellular structures is still limited. Future studies are needed to identify the components of the subcellular structures as well as the multi-localized proteins, and also to characterize their functions, on–off mechanisms and crosstalk.

AUTHOR'S CONTRIBUTIONS

Yueyuan zheng, Junjie Guo, and Xu Li contribute to the literature collection and help to draft the manuscript. Mingming Hou, Yubin Xie, Xuyang Fu, Shengkun Dai, and Rucheng Diao participate in drafting the manuscript and revising it critically for important intellectual content. Yanyan Miao writes the manscript and interprets the data. Jian Ren contributes to the conception and design of the work.

ACKNOWLEDGMENTS

This work was supported by grants from the National Basic Research Program (973 project) [2013CB933900,

2012CB911201]; National Natural Science Foundation of China [31471252, 81201383]; Guangdong Natural Science Funds for Distinguished Young Scholar [S20120011335]; Zhujiang Nova Program of Guangzhou [2011J2200042]; Program of International S&T Cooperation [2014DFB30020]; Fundamental Research Funds for the Central Universities [14lgjc14, 14lgpy02]; Program for New Century Excellent Talents in University [NCET-13-0610]; China Postdoctoral Science Foundation [2014M562238].

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fmicb.2014.00573/abstract

REFERENCES

Adams, R. R., Maiato, H., Earnshaw, W. C., and Carmena, M. (2001). Essential roles of Drosophila inner centromere protein (INCENP) and aurora B in histone H3 phosphorylation, metaphase chromosome alignment, kinetochore disjunction, and chromosome segregation. *J. Cell Biol.* 153, 865–879. doi: 10.1083/jcb.153.4.865

Barnhart, M. C., Kuich, P. H. J. L., Stellfox, M. E., Ward, J. A., Bassett, E. A., Black, B. E., et al. (2011). HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. *J. Cell Biol.* 194, 229–243. doi: 10.1083/jcb.201012017

Barr, F. A., Sillje, H. H., and Nigg, E. A. (2004). Polo-like kinases and the orchestration of cell division. *Nat. Rev. Mol. Cell Biol.* 5, 429–440. doi: 10.1038/nrm1401

Bell, S. P., and Dutta, A. (2002). DNA replication in eukaryotic cells. *Annu. Rev. Biochem.* 71, 333–374. doi: 10.1146/annurev.biochem.71.110601.135425

Bettencourt-Dias, M., and Glover, D. M. (2009). SnapShot: centriole biogenesis. *Cell* 136, 188–188.e1. doi: 10.1016/j.cell.2008.12.035

Bloecher, A., and Tatchell, K. (2000). Dynamic localization of protein phosphatase type 1 in the mitotic cell cycle of *Saccharomyces cerevisiae*. *J. Cell Biol.* 149, 125–140. doi: 10.1083/jcb.149.1.125

Brinkley, B. R. (1985). Microtubule organizing centers. *Annu. Rev. Cell Biol.* 1, 145–172. doi: 10.1146/annurev.cb.01.110185.001045

Brownlee, C. W., and Rogers, G. C. (2013). Show me your license, please: deregulation of centriole duplication mechanisms that promote amplification. *Cell. Mol. Life Sci.* 70, 1021–1034. doi: 10.1007/s00018-012-1102-6

Carmena, M., and Earnshaw, W. C. (2003). The cellular geography of aurora kinases. *Nat. Rev. Mol. Cell Biol.* 4, 842–854. doi: 10.1038/nrm1245

Carpy, A., Krug, K., Graf, S., Koch, A., Popic, S., Hauf, S., et al. (2014). Absolute proteome and phosphoproteome dynamics during the cell cycle of *Schizosaccharomyces pombe* (Fission Yeast). *Mol. Cell. Proteomics* 13, 1925–1936. doi: 10.1074/mcp.M113.035824

Carroll, C. W., and Straight, A. F. (2006). Centromere formation: from epigenetics to self-assembly. *Trends Cell Biol.* 16, 70–78. doi: 10.1016/j.tcb.2005.12.008

Caviston, J. P., Longtine, M., Pringle, J. R., and Bi, E. (2003). The role of Cdc42p GTPase-activating proteins in assembly of the septin ring in yeast. *Mol. Biol. Cell* 14, 4051–4066. doi: 10.1091/mbc.E03-04-0247

Chai, Y., Tian, D., Yang, Y., Feng, G., Cheng, Z., Li, W., et al. (2012). Apoptotic regulators promote cytokinetic midbody degradation in *C. elegans*. *J. Cell Biol.* 199, 1047–1055. doi: 10.1083/jcb.201209050

Chan, Y. W., Jeyaprakash, A. A., Nigg, E. A., and Santamaria, A. (2012). Aurora B controls kinetochore–microtubule attachments by inhibiting Ska complex–KMN network interaction. *J. Cell Biol.* 196, 563–571. doi: 10.1083/jcb.201109001

Chang, P., Giddings, T. H. Jr., Winey, M., and Stearns, T. (2003). Epsilon-tubulin is required for centriole duplication and microtubule organization. *Nat. Cell Biol.* 5, 71–76. doi: 10.1038/ncb900

Cheeseman, I. M., and Desai, A. (2008). Molecular architecture of the kinetochore-microtubule interface. *Nat. Rev. Mol. Cell Biol.* 9, 33–46. doi: 10.1038/nrm2310

Chuang, C., Lin, S. H., Huang, F., Pan, J., Josic, D., and Yu-Lee, L. Y. (2010). Acetylation of RNA processing proteins and cell cycle proteins in mitosis. *J. Proteome Res.* 9, 4554–4564. doi: 10.1021/pr100281h

Conduit, P. T., Feng, Z., Richens, J. H., Baumbach, J., Wainman, A., Bakshi, S. D., et al. (2014). The centrosome-specific phosphorylation of Cnn by Polo/Plk1 drives Cnn scaffold assembly and centrosome maturation. *Dev. Cell* 28, 659–669. doi: 10.1016/j.devcel.2014.02.013

Cooke, C. A., Heck, M. M., and Earnshaw, W. C. (1987). The inner centromere protein (INCENP) antigens: movement from inner centromere to midbody during mitosis. *J. Cell Biol.* 105, 2053–2067. doi: 10.1083/jcb.105.5.2053

D'angelo, M. A., and Hetzer, M. W. (2008). Structure, dynamics and function of nuclear pore complexes. *Trends Cell Biol.* 18, 456–466. doi: 10.1016/j.tcb.2008.07.009

Daniels, M. J., Wang, Y., Lee, M., and Venkitaraman, A. R. (2004). Abnormal cytokinesis in cells deficient in the breast cancer susceptibility protein BRCA2. *Science* 306, 876–879. doi: 10.1126/science.1102574

Ditchfield, C., Johnson, V. L., Tighe, A., Ellston, R., Haworth, C., Johnson, T., et al. (2003). Aurora B couples chromosome alignment with anaphase by targeting BubR1, Mad2, and Cenp-E to kinetochores. *J. Cell Biol.* 161, 267–280. doi: 10.1083/jcb.200208091

Dunham, M. A., Neumann, A. A., Fasching, C. L., and Reddel, R. R. (2000). Telomere maintenance by recombination in human cells. *Nat. Genet.* 26, 447–450. doi: 10.1038/82586

Elia, N., Sougrat, R., Spurlin, T. A., Hurley, J. H., and Lippincott-Schwartz, J. (2011). Dynamics of endosomal sorting complex required for transport (ESCRT) machinery during cytokinesis and its role in abscission. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4846–4851. doi: 10.1073/pnas.1102714108

Ettinger, A. W., Wilsch-Brauninger, M., Marzesco, A. M., Bickle, M., Lohmann, A., Maliga, Z., et al. (2011). Proliferating versus differentiating stem and cancer cells exhibit distinct midbody-release behaviour. *Nat. Commun.* 2, 503. doi: 10.1038/ncomms1511

Fabbro, M., Zhou, B. B., Takahashi, M., Sarcevic, B., Lal, P., Graham, M. E., et al. (2005). Cdk1/Erk2- and Plk1-dependent phosphorylation of a centrosome protein, Cep55, is required for its recruitment to midbody and cytokinesis. *Dev. Cell* 9, 477–488. doi: 10.1016/j.devcel.2005.09.003

Fodde, R., Kuipers, J., Rosenberg, C., Smits, R., Kielman, M., Gaspar, C., et al. (2001). Mutations in the APC tumour suppressor gene cause chromosomal instability. *Nat. Cell Biol.* 3, 433–438. doi: 10.1038/35070129

Fumoto, K., Kikuchi, K., Gon, H., and Kikuchi, A. (2012). Wnt5a signaling controls cytokinesis by correctly positioning ESCRT-III at the midbody. *J. Cell Sci.* 125, 4822–4832. doi: 10.1242/jcs.108142

Ganem, N. J., Godinho, S. A., and Pellman, D. (2009). A mechanism linking extra centrosomes to chromosomal instability. *Nature* 460, 278–282. doi: 10.1038/nature08136

Gao, F., Luo, H., and Zhang, C. T. (2012). DeOri: a database of eukaryotic DNA replication origins. *Bioinformatics* 28, 1551–1552. doi: 10.1093/bioinformatics/bts151

Gentry, M. S., and Hallberg, R. L. (2002). Localization of *Saccharomyces cerevisiae* protein phosphatase 2A subunits throughout mitotic cell cycle. *Mol. Biol. Cell* 13, 3477–3492. doi: 10.1091/mbc.02-05-0065

Gonczy, P. (2012). Towards a molecular architecture of centriole assembly. *Nat. Rev. Mol. Cell Biol.* 13, 425–435. doi: 10.1038/nrm3373

Gonen, S., Akiyoshi, B., Iadanza, M. G., Shi, D., Duggan, N., Biggins, S., et al. (2012). The structure of purified kinetochores reveals multiple microtubule-attachment sites. *Nat. Struct. Mol. Biol.* 19, 925–929. doi: 10.1038/nsmb.2358

Graser, S., Stierhof, Y. D., Lavoie, S. B., Gassner, O. S., Lamla, S., Le Clech, M., et al. (2007). Cep164, a novel centriole appendage protein required for primary cilium formation. *J. Cell Biol.* 179, 321–330. doi: 10.1083/jcb.200707181

Gromley, A., Jurczyk, A., Sillibourne, J., Halilovic, E., Mogensen, M., Groisman, I., et al. (2003). A novel human protein of the maternal centriole is required for the final stages of cytokinesis and entry into S phase. *J. Cell Biol.* 161, 535–545. doi: 10.1083/jcb.200301105

Gromley, A., Yeaman, C., Rosa, J., Redick, S., Chen, C. T., Mirabelle, S., et al. (2005). Centriolin anchoring of exocyst and SNARE complexes at the midbody is required for secretory-vesicle-mediated abscission. *Cell* 123, 75–87. doi: 10.1016/j.cell.2005.07.027

Gruneberg, U., Neef, R., Honda, R., Nigg, E. A., and Barr, F. A. (2004). Relocation of Aurora B from centromeres to the central spindle at the metaphase to anaphase transition requires MKlp2. *J. Cell Biol.* 166, 167–172. doi: 10.1083/jcb.200403084

Guertin, D. A., Trautmann, S., and McCollum, D. (2002). Cytokinesis in eukaryotes. *Microbiol. Mol. Biol. Rev.* 66, 155–178. doi: 10.1128/MMBR.66.2.155-178.2002

Guse, A., Carroll, C. W., Moree, B., Fuller, C. J., and Straight, A. F. (2011). *In vitro* centromere and kinetochore assembly on defined chromatin templates. *Nature* 477, 354–U136. doi: 10.1038/nature10379

Hahn, W. C., Stewart, S. A., Brooks, M. W., York, S. G., Eaton, E., Kurachi, A., et al. (1999). Inhibition of telomerase limits the growth of human cancer cells. *Nat. Med.* 5, 1164–1170. doi: 10.1038/13495

Hanahan, D., and Weinberg, R. A. (2000). The hallmarks of cancer. *Cell* 100, 57–70. doi: 10.1016/S0092-8674(00)81683-9

Hauf, S., Cole, R. W., Laterra, S., Zimmer, C., Schnapp, G., Walter, R., et al. (2003). The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint. *J. Cell Biol.* 161, 281–294. doi: 10.1083/jcb.200208092

Hemerly, A. S., Prasanth, S. G., Siddiqui, K., and Stillman, B. (2009). Orc1 controls centriole and centrosome copy number in human cells. *Science* 323, 789–793. doi: 10.1126/science.1166745

Henikoff, S., Ramachandran, S., Krassovsky, K., Bryson, T. D., Codomo, C. A., Brogaard, K., et al. (2014). The budding yeast Centromere DNA Element II wraps a stable Cse4 hemisome in either orientation *in vivo*. *Elife* 3:e01861. doi: 10.7554/eLife.01861

Henson, J. D., Neumann, A. A., Yeager, T. R., and Reddel, R. R. (2002). Alternative lengthening of telomeres in mammalian cells. *Oncogene* 21, 598–610. doi: 10.1038/sj.onc.1205058

Herbert, B. S., Pitts, A. E., Baker, S. I., Hamilton, S. E., Wright, W. E., Shay, J. W., et al. (1999). Inhibition of human telomerase in immortal human cells leads to progressive telomere shortening and cell death. *Proc. Natl. Acad. Sci. U.S.A.* 96, 14276–14281. doi: 10.1073/pnas.96.25.14276

Heun, P., Laroche, T., Shimada, K., Furrer, P., and Gasser, S. M. (2001). Chromosome dynamics in the yeast interphase nucleus. *Science* 294, 2181–2186. doi: 10.1126/science.1065366

Hinchcliffe, E. H., and Sluder, G. (2001). "It takes two to tango": understanding how centrosome duplication is regulated throughout the cell cycle. *Genes Dev.* 15, 1167–1181. doi: 10.1101/gad.894001

Hoeijmakers, J. H. J. (2001). Genome maintenance mechanisms for preventing cancer. *Nature* 411, 366–374. doi: 10.1038/35077232

Hori, T., Haraguchi, T., Hiraoka, Y., Kimura, H., and Fukagawa, T. (2003). Dynamic behavior of Nuf2-Hec1 complex that localizes to the centrosome and centromere and is essential for mitotic progression in vertebrate cells. *J. Cell Sci.* 116, 3347–3362. doi: 10.1242/jcs.00645

Hu, C. K., Coughlin, M., and Mitchison, T. J. (2012). Midbody assembly and its regulation during cytokinesis. *Mol. Biol. Cell* 23, 1024–1034. doi: 10.1091/mbc.E11-08-0721

Joseph, N., Hutterer, A., Poser, I., and Mishima, M. (2012). ARF6 GTPase protects the post-mitotic midbody from 14-3-3-mediated disintegration. *EMBO J.* 31, 2604–2614. doi: 10.1038/emboj.2012.139

Kang, Y. H., Park, J.-E., Yu, L.-R., Soung, N.-K., Yun, S.-M., Bang, J. K., et al. (2006). Self-regulated Plk1 recruitment to kinetochores by the Plk1-PBIP1 interaction is critical for proper chromosome segregation. *Mol. Cell* 24, 409–422. doi: 10.1016/j.molcel.2006.10.016

Kimura, M., and Okano, Y. (2005). [Aurora kinases and cancer]. *Gan To Kagaku Ryoho* 32, 1–5.

Kitagawa, D., Vakonakis, I., Olieric, N., Hilbert, M., Keller, D., Olieric, V., et al. (2011). Structural basis of the 9-fold symmetry of centrioles. *Cell* 144, 364–375. doi: 10.1016/j.cell.2011.01.008

Kloc, M., Kubiak, J. Z., Li, X. C., and Ghobrial, R. M. (2014). The newly found functions of MTOC in immunological response. *J. Leukoc. Biol.* 95, 417–430. doi: 10.1189/jlb.0813468

Kollman, J. M., Merdes, A., Mourey, L., and Agard, D. A. (2011). Microtubule nucleation by gamma-tubulin complexes. *Nat. Rev. Mol. Cell Biol.* 12, 709–721. doi: 10.1038/nrm3209

Kuo, T. C., Chen, C. T., Baron, D., Onder, T. T., Loewer, S., Almeida, S., et al. (2011). Midbody accumulation through evasion of autophagy contributes to cellular reprogramming and tumorigenicity. *Nat. Cell Biol.* 13, 1214–1223. doi: 10.1038/ncb2332

Kurasawa, Y., Earnshaw, W. C., Mochizuki, Y., Dohmae, N., and Todokoro, K. (2004). Essential roles of KIF4 and its binding partner PRC1 in organized central spindle midzone formation. *EMBO J.* 23, 3237–3248. doi: 10.1038/sj.emboj.7600347

Kurasawa, Y., and Todokoro, K. (1999). Identification of human APC10/Doc1 as a subunit of anaphase promoting complex. *Oncogene* 18, 5131–5137. doi: 10.1038/sj.onc.1203133

Lara-Gonzalez, P., Westhorpe, F. G., and Taylor, S. S. (2012). The spindle assembly checkpoint. *Curr. Biol.* 22, R966–R980. doi: 10.1016/j.cub.2012.10.006

Lee, H. H., Elia, N., Ghirlando, R., Lippincott-Schwartz, J., and Hurley, J. H. (2008). Midbody targeting of the ESCRT machinery by a noncanonical coiled coil in CEP55. *Science* 322, 576–580. doi: 10.1126/science.1162042

Lee, K., and Rhee, K. (2011). PLK1 phosphorylation of pericentrin initiates centrosome maturation at the onset of mitosis. *J. Cell Biol.* 195, 1093–1101. doi: 10.1083/jcb.201106093

Li, J., D'angiolella, V., Seeley, E. S., Kim, S., Kobayashi, T., Fu, W., et al. (2013). USP33 regulates centrosome biogenesis via deubiquitination of the centriolar protein CP110. *Nature* 495, 255–259. doi: 10.1038/nature11941

Liu, D., Davydenko, O., and Lampson, M. A. (2012). Polo-like kinase-1 regulates kinetochore-microtubule dynamics and spindle checkpoint silencing. *J. Cell Biol.* 198, 491–499. doi: 10.1083/jcb.201205090

Liu, D., Zhang, N., Du, J., Cai, X., Zhu, M., Jin, C., et al. (2006). Interaction of Skp1 with CENP-E at the midbody is essential for cytokinesis. *Biochem. Biophys. Res. Commun.* 345, 394–402. doi: 10.1016/j.bbrc.2006.04.062

Liu, R. M., Tian, X. Y., Huang, X. T., and Zhou, H. (2013). Transfer of Ser7 phosphorylated CENP-A from centromere to midbody during mitosis in MCF-7 cells. *Folia Biol. (Praha.)* 59, 105–109.

Luders, J., and Stearns, T. (2007). Microtubule-organizing centres: a re-evaluation. *Nat. Rev. Mol. Cell Biol.* 8, 161–167. doi: 10.1038/nrm2100

Lusk, C. P., Waller, D. D., Makhnevych, T., Dienemann, A., Whiteway, M., Thomas, D. Y., et al. (2007). Nup53p is a target of two mitotic kinases, Cdk1p and Hrr25p. *Traffic* 8, 647–660. doi: 10.1111/j.1600-0854.2007.00559.x

Malvezzi, F., Litos, G., Schleiffer, A., Heuck, A., Mechtler, K., Clausen, T., et al. (2013). A structural basis for kinetochore recruitment of the Ndc80 complex via two distinct centromere receptors. *EMBO J.* 32, 409–423. doi: 10.1038/emboj.2012.356

Marina, M., and Saavedra, H. I. (2014). Nek2 and Plk4: prognostic markers, drivers of breast tumorigenesis and drug resistance. *Front. Biosci. (Landmark Ed.)* 19, 352–365. doi: 10.2741/4212

Marthiens, V., Rujano, M. A., Pennetier, C., Tessier, S., Paul-Gilloteaux, P., and Basto, R. (2013). Centrosome amplification causes microcephaly. *Nat. Cell Biol.* 15, 731–740. doi: 10.1038/ncb2746

Masuda, H., Mori, R., Yukawa, M., and Toda, T. (2013). Fission yeast MOZART1/Mzt1 is an essential gamma-tubulin complex component required for complex recruitment to the microtubule organizing center, but not its assembly. *Mol. Biol. Cell* 24, 2894–2906. doi: 10.1091/mbc.E13-05-0235

Masumoto, H., Masukata, H., Muro, Y., Nozaki, N., and Okazaki, T. (1989). A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J. Cell Biol.* 109, 1963–1973. doi: 10.1083/jcb.109.5.1963

Mennella, V., Agard, D. A., Huang, B., and Pelletier, L. (2014). Amorphous no more: subdiffraction view of the pericentriolar material architecture. *Trends Cell Biol.* 24, 188–197. doi: 10.1016/j.tcb.2013.10.001

Merbl, Y., Refour, P., Patel, H., Springer, M., and Kirschner, M. W. (2013). Profiling of ubiquitin-like modifications reveals features of mitotic control. *Cell* 152, 1160–1172. doi: 10.1016/j.cell.2013.02.007

Mollinari, C., Kleman, J. P., Jiang, W., Schoehn, G., Hunter, T., and Margolis, R. L. (2002). PRC1 is a microtubule binding and bundling protein essential to maintain the mitotic spindle midzone. *J. Cell Biol.* 157, 1175–1186. doi: 10.1083/jcb.200111052

Morita, E., Sandrin, V., Chung, H. Y., Morham, S. G., Gygi, S. P., Rodesch, C. K., et al. (2007). Human ESCRT and ALIX proteins interact with proteins of the midbody and function in cytokinesis. *EMBO J.* 26, 4215–4227. doi: 10.1038/sj.emboj.7601850

Mullins, J. M., and Biesele, J. J. (1977). Terminal phase of cytokinesis in D-98s cells. *J. Cell Biol.* 73, 672–684. doi: 10.1083/jcb.73.3.672

Nigg, E. A., and Raff, J. W. (2009). Centrioles, centrosomes, and cilia in health and disease. *Cell* 139, 663–678. doi: 10.1016/j.cell.2009.10.036

Okada, M., Cheeseman, I. M., Hori, T., Okawa, K., McLeod, I. X., Yates, J. R. 3rd., et al. (2006). The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. *Nat. Cell Biol.* 8, 446–457. doi: 10.1038/ncb1396

Ong, A. C., and Wheatley, D. N. (2003). Polycystic kidney disease–the ciliary connection. *Lancet* 361, 774–776. doi: 10.1016/S0140-6736(03)12662-1

Osmundson, E. C., Ray, D., Moore, F. E., Gao, Q., Thomsen, G. H., and Kiyokawa, H. (2008). The HECT E3 ligase Smurf2 is required for Mad2-dependent spindle assembly checkpoint. *J. Cell Biol.* 183, 267–277. doi: 10.1083/jcb.200801049

Otegui, M. S., Verbrugghe, K. J., and Skop, A. R. (2005). Midbodies and phragmoplasts: analogous structures involved in cytokinesis. *Trends Cell Biol.* 15, 404–413. doi: 10.1016/j.tcb.2005.06.003

Peters, J. M. (2006). The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nat. Rev. Mol. Cell Biol.* 7, 644–656. doi: 10.1038/nrm1988

Petronczki, M., Lenart, P., and Peters, J. M. (2008). Polo on the Rise-from Mitotic Entry to Cytokinesis with Plk1. *Dev. Cell* 14, 646–659. doi: 10.1016/j.devcel.2008.04.014

Pohl, C., and Jentsch, S. (2008). Final stages of cytokinesis and midbody ring formation are controlled by BRUCE. *Cell* 132, 832–845. doi: 10.1016/j.cell.2008.01.012

Prasanth, S. G., Prasanth, K. V., Siddiqui, K., Spector, D. L., and Stillman, B. (2004). Human Orc2 localizes to centrosomes, centromeres and heterochromatin during chromosome inheritance. *EMBO J.* 23, 2651–2663. doi: 10.1038/sj.emboj.7600255

Prasanth, S. G., Prasanth, K. V., and Stillman, B. (2002). Orc6 involved in DNA replication, chromosome segregation, and cytokinesis. *Science* 297, 1026–1031. doi: 10.1126/science.1072802

Puklowski, A., Homsi, Y., Keller, D., May, M., Chauhan, S., Kossatz, U., et al. (2011). The SCF-FBXW5 E3-ubiquitin ligase is regulated by PLK4 and targets HsSAS-6 to control centrosome duplication. *Nat. Cell Biol.* 13, 1004–1009. doi: 10.1038/ncb2282

Quarmby, L. M., and Parker, J. D. (2005). Cilia and the cell cycle? *J. Cell Biol.* 169, 707–710. doi: 10.1083/jcb.200503053

Ren, J., Liu, Z., Gao, X., Jin, C., Ye, M., Zou, H., et al. (2010). MiCroKit 3.0: an integrated database of midbody, centrosome and kinetochore. *Nucleic Acids Res.* 38, D155–D160. doi: 10.1093/nar/gkp784

Rudolph, K. L., Chang, S., Lee, H. W., Blasco, M., Gottlieb, G. J., Greider, C., et al. (1999). Longevity, stress response, and cancer in aging telomerase-deficient mice. *Cell* 96, 701–712. doi: 10.1016/S0092-8674(00)80580-2

Ryser, S., Dizin, E., Jefford, C. E., Delaval, B., Gagos, S., Christodoulidou, A., et al. (2009). Distinct roles of BARD1 isoforms in mitosis: full-length BARD1 mediates Aurora B degradation, cancer-associated BARD1beta scaffolds Aurora B and BRCA2. *Cancer Res.* 69, 1125–1134. doi: 10.1158/0008-5472.CAN-08-2134

Sancar, A., Lindsey-Boltz, L. A., Unsal-Kacmaz, K., and Linn, S. (2004). Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.* 73, 39–85. doi: 10.1146/annurev.biochem.73.011303.073723

Santaguida, S., and Musacchio, A. (2009). The life and miracles of kinetochores. *EMBO J.* 28, 2511–2531. doi: 10.1038/emboj.2009.173

Schink, K. O., and Stenmark, H. (2011). Cell differentiation: midbody remnants - junk or fate factors? *Curr. Biol.* 21, R958–960. doi: 10.1016/j.cub.2011.10.035

Schmitz, M. H., Held, M., Janssens, V., Hutchins, J. R., Hudecz, O., Ivanova, E., et al. (2010). Live-cell imaging RNAi screen identifies PP2A-B55 [alpha] and importin-[beta] 1 as key mitotic exit regulators in human cells. *Nat. Cell Biol.* 12, 886–893. doi: 10.1038/ncb2092

Schuyler, S. C., Wu, Y. F., and Kuan, V. J. W. (2012). The Mad1-Mad2 balancing act - a damaged spindle checkpoint in chromosome instability and cancer. *J. Cell Sci.* 125, 4197–4206. doi: 10.1242/jcs.107037

Screpanti, E., De Antoni, A., Alushin, G. M., Petrovic, A., Melis, T., Nogales, E., et al. (2011). Direct binding of Cenp-C to the Mis12 complex joins the inner and outer kinetochore. *Curr. Biol.* 21, 391–398. doi: 10.1016/j.cub.2010.12.039

Seybold, C., and Schiebel, E. (2013). Spindle pole bodies. *Curr. Biol.* 23, R858–R860. doi: 10.1016/j.cub.2013.07.024

Shimada, K., and Gasser, S. M. (2007). The origin recognition complex functions in sister-chromatid cohesion in *Saccharomyces cerevisiae*. *Cell* 128, 85–99. doi: 10.1016/j.cell.2006.11.045

Skop, A. R., Liu, H., Yates, J. 3rd., Meyer, B. J., and Heald, R. (2004). Dissection of the mammalian midbody proteome reveals conserved cytokinesis mechanisms. *Science* 305, 61–66. doi: 10.1126/science.1097931

Snead, J. L., Sullivan, M., Lowery, D. M., Cohen, M. S., Zhang, C., Randle, D. H., et al. (2007). A coupled chemical-genetic and bioinformatic approach to Polo-like kinase pathway exploration. *Chem. Biol.* 14, 1261–1272. doi: 10.1016/j.chembiol.2007.09.011

Somers, W. G., and Saint, R. (2003). A RhoGEF and Rho family GTPase-activating protein complex links the contractile ring to cortical microtubles at the onset of cytokinesis. *Dev. Cell* 4, 29–39. doi: 10.1016/S1534-5807(02)00402-1

Stearns, T. (2001). Centrosome duplication. a centriolar pas de deux. *Cell* 105, 417–420. doi: 10.1016/S0092-8674(01)00366-X

Steigemann, P., and Gerlich, D. W. (2009). Cytokinetic abscission: cellular dynamics at the midbody. *Trends Cell Biol.* 19, 606–616. doi: 10.1016/j.tcb.2009.07.008

Steigemann, P., Wurzenberger, C., Schmitz, M. H., Held, M., Guizetti, J., Maar, S., et al. (2009). Aurora B-mediated abscission checkpoint protects against tetraploidization. *Cell* 136, 473–484. doi: 10.1016/j.cell.2008.12.020

Straight, A. F., Cheung, A., Limouze, J., Chen, I., Westwood, N. J., Sellers, J. R., et al. (2003). Dissecting temporal and spatial control of cytokinesis with a myosin II inhibitor. *Science* 299, 1743–1747. doi: 10.1126/science.1081412

Swaminathan, S. (2004). Human disease: the centrosome connection. *Nat. Cell Biol.* 6, 383. doi: 10.1038/ncb0504-383

Tanaka, K., Li Chang, H., Kagami, A., and Watanabe, Y. (2009). CENP-C functions as a scaffold for effectors with essential kinetochore functions in mitosis and meiosis. *Dev. Cell* 17, 334–343. doi: 10.1016/j.devcel.2009.08.004

Tang, C. J. C., Fu, R. H., Wu, K. S., Hsu, W. B., and Tang, T. K. (2009). CPAP is a cell-cycle regulated protein that controls centriole length. *Nat. Cell Biol.* 11, 825–U103. doi: 10.1038/ncb1889

Tanos, B. E., Yang, H. J., Soni, R., Wang, W. J., Macaluso, F. P., Asara, J. M., et al. (2013). Centriole distal appendages promote membrane docking, leading to cilia initiation. *Genes Dev.* 27, 163–168. doi: 10.1101/gad.207043.112

Vallen, E. A., Caviston, J., and Bi, E. (2000). Roles of Hof1p, Bni1p, Bnr1p, and Myo1p in cytokinesis in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 11, 593–611. doi: 10.1091/mbc.11.2.593

Van Damme, D., Bouget, F. Y., Van Poucke, K., Inze, D., and Geelen, D. (2004). Molecular dissection of plant cytokinesis and phragmoplast structure: a survey of GFP-tagged proteins. *Plant J.* 40, 386–398. doi: 10.1111/j.1365-313X.2004.02222.x

Varma, D., and Salmon, E. D. (2012). The KMN protein network–chief conductors of the kinetochore orchestra. *J. Cell Sci.* 125, 5927–5936. doi: 10.1242/jcs.093724

Veith, D., Scherr, N., Efimov, V. P., and Fischer, R. (2005). Role of the spindle-pole-body protein ApsB and the cortex protein ApsA in microtubule organization and nuclear migration in Aspergillus nidulans. *J. Cell Sci.* 118, 3705–3716. doi: 10.1242/jcs.02501

Veleri, S., Manjunath, S. H., Fariss, R. N., May-Simera, H., Brooks, M., Foskett, T. A., et al. (2014). Ciliopathy-associated gene Cc2d2a promotes assembly of subdistal appendages on the mother centriole during cilia biogenesis. *Nat. Commun.* 5, 4207. doi: 10.1038/ncomms5207

Von Zglinicki, T. (2002). Oxidative stress shortens telomeres. *Trends Biochem. Sci.* 27, 339–344. doi: 10.1016/S0968-0004(02)02110-2

Vulprecht, J., David, A., Tibelius, A., Castiel, A., Konotop, G., Liu, F., et al. (2012). STIL is required for centriole duplication in human cells. *J. Cell Sci.* 125, 1353–1362. doi: 10.1242/jcs.104109

Wei, R. R., Sorger, P. K., and Harrison, S. C. (2005). Molecular organization of the Ndc80 complex, an essential kinetochore component. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5363–5367. doi: 10.1073/pnas.0501168102

Welburn, J. P., Grishchuk, E. L., Backer, C. B., Wilson-Kubalek, E. M., Yates, J. R. 3rd., and Cheeseman, I. M. (2009). The human kinetochore Ska1 complex facilitates microtubule depolymerization-coupled motility. *Dev. Cell* 16, 374–385. doi: 10.1016/j.devcel.2009.01.011

Wheatley, S. P., Carvalho, A., Vagnarelli, P., and Earnshaw, W. C. (2001). INCENP is required for proper targeting of Survivin to the centromeres and the anaphase spindle during mitosis. *Curr. Biol.* 11, 886–890. doi: 10.1016/S0960-9822(01)00238-X

Yu, H. (2002). Regulation of APC-Cdc20 by the spindle checkpoint. *Curr. Opin. Cell Biol.* 14, 706–714. doi: 10.1016/S0955-0674(02)00382-4

Yuce, O., Piekny, A., and Glotzer, M. (2005). An ECT2-centralspindlin complex regulates the localization and function of RhoA. *J. Cell Biol.* 170, 571–582. doi: 10.1083/jcb.200501097

Zimmerman, W. C., Sillibourne, J., Rosa, J., and Doxsey, S. J. (2004). Mitosis-specific anchoring of gamma tubulin complexes by pericentrin controls spindle organization and mitotic entry. *Mol. Biol. Cell* 15, 3642–3657. doi: 10.1091/mbc.E03-11-0796

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: 05 September 2014; accepted: 11 October 2014; published online: 29 October 2014.

Citation: Zheng Y, Guo J, Li X, Xie Y, Hou M, Fu X, Dai S, Diao R, Miao Y and Ren J (2014) An integrated overview of spatiotemporal organization and regulation in mitosis in terms of the proteins in the functional supercomplexes. Front. Microbiol. ***5****:573. doi: 10.3389/fmicb.2014.00573*

This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology.

Copyright © 2014 Zheng, Guo, Li, Xie, Hou, Fu, Dai, Diao, Miao and Ren. 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.

frontiers in
MICROBIOLOGY

MINI REVIEW ARTICLE
published: 19 February 2015
doi: 10.3389/fmicb.2015.00117

Recent advances in the genome-wide study of DNA replication origins in yeast

Chong Peng[1], Hao Luo[1], Xi Zhang[1] and Feng Gao[1,2,3]*

[1] *Department of Physics, Tianjin University, Tianjin, China*
[2] *Key Laboratory of Systems Bioengineering (Ministry of Education), Tianjin University, Tianjin, China*
[3] *SynBio Research Platform, Collaborative Innovation Center of Chemical Science and Engineering, Tianjin, China*

Edited by:
John R. Battista, Louisiana State University and Agricultural and Mechanical College, USA

Reviewed by:
Suleyman Yildirim, Istanbul Medipol University, Turkey
Abd El-Latif Hesham, Assiut University, Egypt

***Correspondence:**
Feng Gao, Department of Physics, Tianjin University, No. 92 Weijin Road, Nankai District, Tianjin 300072, China
e-mail: fgao@tju.edu.cn

DNA replication, one of the central events in the cell cycle, is the basis of biological inheritance. In order to be duplicated, a DNA double helix must be opened at defined sites, which are called DNA replication origins (ORIs). Unlike in bacteria, where replication initiates from a single replication origin, multiple origins are utilized in the eukaryotic genomes. Among them, the ORIs in budding yeast *Saccharomyces cerevisiae* and the fission yeast *Schizosaccharomyces pombe* have been best characterized. In recent years, advances in DNA microarray and next-generation sequencing technologies have increased the number of yeast species involved in ORIs research dramatically. The ORIs in some non-conventional yeast species such as *Kluyveromyces lactis* and *Pichia pastoris* have also been genome-widely identified. Relevant databases of replication origins in yeast were constructed, then the comparative genomic analysis can be carried out. Here, we review several experimental approaches that have been used to map replication origins in yeast and some of the available web resources related to yeast ORIs. We also discuss the sequence characteristics and chromosome structures of ORIs in the four yeast species, which can be utilized to improve yeast replication origins prediction.

Keywords: DNA replication, replication origin, *Saccharomyces cerevisiae, Schizosaccharomyces pombe, Kluyveromyces lactis, Pichia pastoris*

INTRODUCTION

DNA replication is one of the crucial steps for cell cycle. During cell division, accurate and complete duplication of the genome is required to ensure the faithful inheritance of genetic information from one cell generation to the next. To be duplicated, a DNA double helix must be opened at defined sites, termed DNA replication origins (MacAlpine and Bell, 2005; Mechali, 2010; Schepers and Papior, 2010). In general terms, the number of origins (ORIs) in a genome is bound up with the size of the chromosome. Bacterial genomes frequently have a single replication origin, because they usually consist of a small circular chromosome (Gao and Zhang, 2007; Gao et al., 2013; Leonard and Mechali, 2013). In contrast, eukaryotic DNA replication initiates at multiple origins due to their enormous genomic information and the complexity of their chromosome structures (Mechali, 2010). Budding yeast *Saccharomyces cerevisiae* and the fission yeast *Schizosaccharomyces pombe* have the best characterized ORIs in eukaryotes. In *S. cerevisiae*, origin selection is mediated by the formation of a multi-protein complex termed the pre-replicative complex (pre-RC), whose activation leads to DNA unwinding and the assembly of replisomes to carry out DNA synthesis (Bell and Dutta, 2002). Proteins required for pre-RC formation include the origin recognition complex (ORC), the pre-RC assembly factors Cdc6 and Cdt1 and the putative replicative DNA helicase, the MCM2-7 complex (Bell, 2002; Bowers et al., 2004).

Recent advances in DNA microarray technology and next-generation sequencing technologies have brought a dramatic increase in the number of ORIs identified in eukaryotic genomes, such as human (Cadoret et al., 2008; Karnani et al., 2010), mouse (Sequeira-Mendes et al., 2009; Cayrou et al., 2011), *Arabidopsis thaliana* (Costas et al., 2011), and *Drosophila melanogaster* (Cayrou et al., 2011; Gao et al., 2012). The ORIs in some non-conventional yeast species such as *Kluyveromyces lactis* (Liachko et al., 2010) and *Pichia pastoris* (Liachko et al., 2014) have also been genome-widely identified. Because of the increasing data of eukaryotic ORIs, some secondary databases with comprehensive and intuitive ORIs' information have been constructed. In this review, we summarize several experimental approaches that have been used to identify replication origins in yeast and list some available web resources relevant to yeast ORIs. In addition, we also discuss the characteristics of ORIs in the four yeast species based on the sequence data in the Database of Eukaryotic ORIs (DeOri), including the significant motifs found by the MEME-ChIP web service, the chromosome structures of ORIs, and the origin replication timing and efficiency features.

EXPERIMENTAL METHODS TO IDENTIFY YEAST REPLICATION ORIGINS

Primal efforts to identify origins across an entire chromosome were two-dimensional gel agarose electrophoresis, which utilized

the fact that non-linear DNA molecule does not migrate in gels at the same rate as a linear molecule of equal mass (Bell and Byers, 1983; Brewer and Fangman, 1987). Partially unwound DNA are likely to form only in the vicinity of replication origins, and such structures can be mapped by virtue of being branched. For the relatively low throughput of two-dimensional gel agarose electrophoresis, just a small set of activity origins in the smallest chromosomes in *S. cerevisiae* were located by this method (Reynolds et al., 1989; Newlon et al., 1993; Friedman et al., 1997; Besnard et al., 2014).

To comprehensively identify the location of origins and characterize the ORIs, microarray-based approaches were developed. The combination of fluorescently labeled DNA and microarray representing all the yeast open reading frames (ORFs) can reveal the replicating details of the DNA sequence. Even though they are time consuming and the resolution may not be ideal, these studies make it possible to locate ORIs genome-widely.

There are three widely used microarray-based techniques. (a) By generating a replication timing profile and taking advantage of the fact that ORIs replicate earlier than its neighboring sequences. Methods to differentiate replicated from non-replicated DNA in the progression of DNA replication are diversiform. Both density transfer approach by isotopically labeling of DNA (heavy : light study) and copy number approach by monitoring the change of copy number (Raghuraman et al., 2001; Yabuki et al., 2002; Heichinger et al., 2006) were involved. (b) By identifying pre-replicative complexes at origins of replication using chromatin immunoprecipitation (ChIP). The genome-wide identification of ORC- and MCM-bound sites can reveal the locations of DNA replication origins (Wyrick et al., 2001; Nieduszynski et al., 2006; Xu et al., 2006; Hayashi et al., 2007). (c) By measuring the accumulation of single-stranded DNA (ssDNA) in the presence of a replication-impeding drug, hydroxyurea (HU). This technique makes use of the observation that ssDNA formation is restricted to origins of replication in the checkpoint-deficient mutant *rad*53 (Feng et al., 2006; Masai et al., 2010).

In recent years, the next-generation sequencing technology has also been combined into replication origins identifying methods. Sequencing of replication intermediates or direct sequencing of short, newly replicated DNA strands can help locate replication origins. Compared with microarray-based approaches, deep-sequencing-based approach is characterized by high efficiency, low cost and high resolution. Some methods can even define replication origin sequences throughout the genome with single-nucleotide resolution. On the other hand, next-generation sequencing technologies exhibit coverage biases, which should be avoided to ensure the accuracy of whole-genome origin maps (Besnard et al., 2014).

ChIP-seq, ChIP followed by direct high-throughput sequencing, is the most representative application (Kharchenko et al., 2008). Xu et al. (2012) identified ORIs in three distantly related fission yeasts, *Schizosaccharomyces pombe*, *Schizosaccharomyces octosporus*, and *Schizosaccharomyces japonicas* at high resolution with a generally applicable deep-sequencing-based approach. They counted the frequency of each region of the genome in S-phase arrested cells by deep sequencing, then produced replication timing profiles by mapping all the sites with increased DNA copy number (Xu et al., 2012). Autonomously replicating sequences ARS-seq followed with miniARS-seq is another sequencing-based method. The most recently updated ORIs in *S. cerevisiae* and the firstly reported ORIs in *P. pastoris* are identified with this method (Liachko et al., 2013, 2014). We take *P. pastoris* for instance here to represent the operation steps of this technique. Liachko et al. (2014) firstly constructed a ∼15 × library of genomic DNA in a non-replicating *URA*3 shuttle vector, then screened for ARS activity. ARS inserts were amplified by vector-specific Illumina primers and sequenced by paired-end deep sequencing. Short subfragments of ARSs isolated from the initial ARS-seq screen were then constructed as an input library for a follow-up ARS screen. The subsequent usage of miniARS-seq generated a high-resolution map of ARS sites in the *P. pastoris* genome (Liachko et al., 2014).

In **Figure 1**, we present DNA replication data from different experimental approaches of chromosome 1 in *S. cerevisiae*. The data of microarray-based techniques including heavy : light study, copy number study, ORC-ChIP, and MCM-ChIP, as well as ssDNA in HU study were downloaded from the DNA replication origin database OriDB (Nieduszynski et al., 2007). We also mark the ORIs identified by ARS-seq method on the figure (Liachko et al., 2013). Obvious overlaps exist among the different groups of data.

DATABASES RELEVANT TO THE STUDY OF YEAST REPLICATION ORIGINS

Due to the increasing data of eukaryotic ORIs, developing repositories of these information became feasible and necessary. We list some of the available web resources relevant to DNA replication in yeast, and discuss their contents in this section.

OriDB[1] is the most widely used database of DNA replication origins, which is limited to budding yeast (*S. cerevisiae*) and fission yeast (*S. pombe*) by present. The data of *S. cerevisiae* replication origins in OriDB was collated from four microarray-based studies, each of which separately mapped the approximate location of ORIs throughout the yeast genome, and the fifth study that used analysis of phylogenetic conservation and provided another list of origin sites. After amalgamating the data of each study, OriDB produced an integrated list of origin sites. Each proposed origin site is assigned a status (confirmed, likely, or dubious) that indicates the assurance of the site genuinely corresponding to an origin. In 2012, origin sites from *S. pombe* were collected. OriDB provides lots of assistance to researchers working in the DNA replication field because it brings together comprehensive information which was difficult to access and compare (Nieduszynski et al., 2007; Siow et al., 2012).

DeOri[2] was constructed in the year of 2012 and has been updated constantly. When the original version was constructed, DeOri contained replication origins from six eukaryotic organisms. Now the entries have been increased to 173,988 ORIs from eight eukaryotic organisms, including human, mouse, *A. thaliana*, *D. melanogaster*, *K. lactis*, *S. pombe*, *P. pastoris*, and *S. cerevisiae*.

[1]http://cerevisiae.oridb.org/
[2]http://tubic.tju.edu.cn/deori/

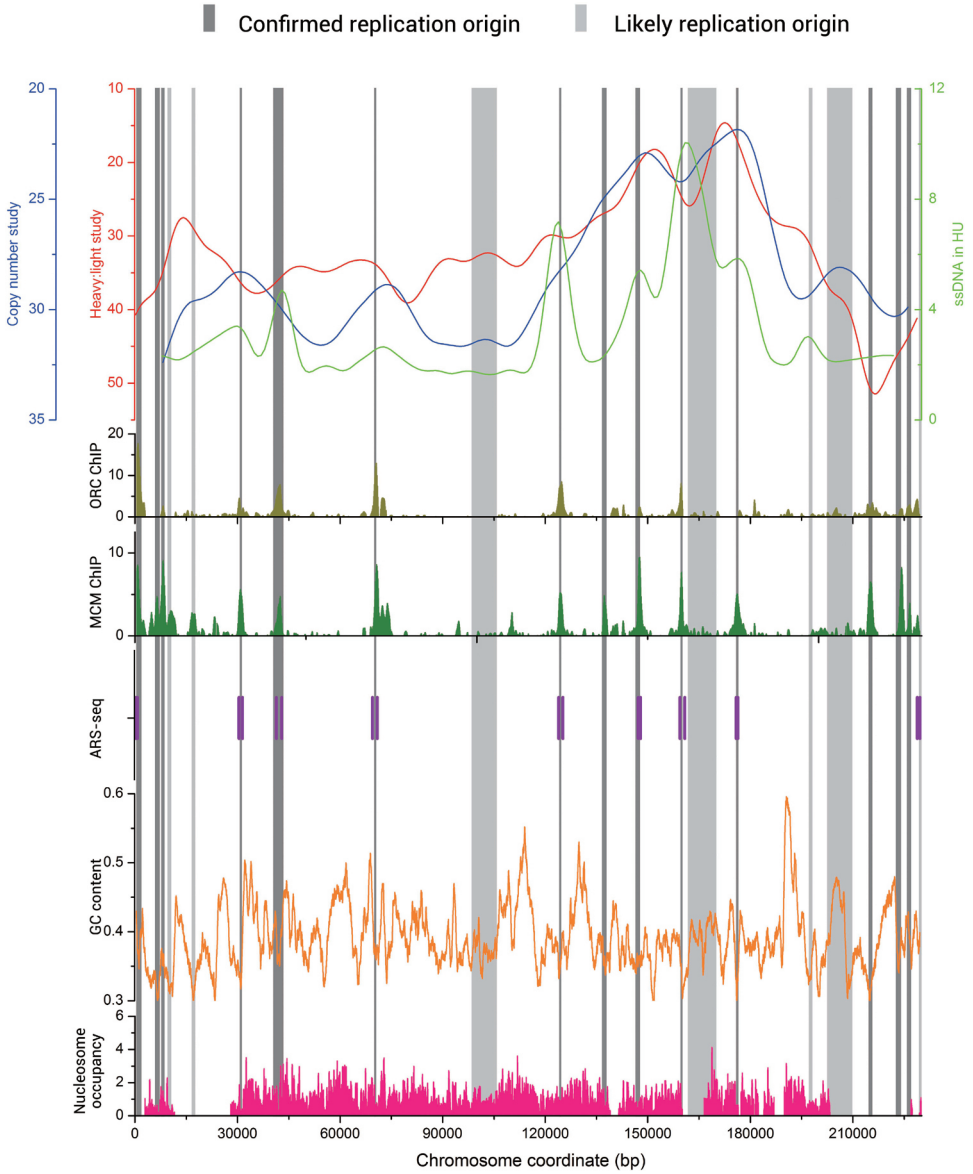


FIGURE 1 | Graph view of genome-wide data relevant to the replication origins in *Saccharomyces cerevisiae* chromosome 1. The genome-wide data including heavy: light study (red line), copy number study (blue line), ssDNA in HU study (light green line) are visualized at the top of Figure. The bottom five plots show the genome-wide data of ORC-ChIP (olive bars), MCM-ChIP (green bars), ARS-seq (vertical purple bars), GC content (orange line), and nucleosome occupancy (pink bars), respectively. The replication origin sites are indicated by vertical bars (dark gray for confirmed and light gray for likely).

We have filtered the replication origin data in the four yeasts for the following sequence analyzing. This database aims to contribute in the comparative genomic analysis of replication origins, and provides some insights into the nature of replication origins on a genome scale (Gao et al., 2012).

DNAReplication[3] is a database aimed to provide information and resources for the eukaryotic DNA replication community. Organism-sorted data on replication proteins are presented in this database, and are summarized in the categories of nomenclature, biochemical properties, motifs, interactions, modifications, structure, cell localization and expression, and general comments. Users are also provided with links to recent replication papers, other useful replication websites, and homepages of replication labs. All these functions make this database a valuable tool for the study of eukaryotic DNA replication (Cotterill and Kearsey, 2009).

ReplicationDomain[4] is a comparative web-based database for storing, sharing and visualizing DNA replication timing data.

[3]http://www.dnareplication.net/

[4]http://www.replicationdomain.org

Other genome-wide chromatin features as well as comparative information of transcriptional expression are also provided in this database. Replication Domain is also a valuable resource for the scientific community because users not only can download the publicly available microarray data, but also are allowed to upload their own data sets and share them with colleagues prior to providing public access (Weddington et al., 2008).

SGD (*Saccharomyces* Genome Database, available at http://www.yeastgenome.org/) is a genomic resource of the budding yeast *S. cerevisiae*. The highest-quality comprehensive information, including the complete *S. cerevisiae* reference genome DNA sequence, its genes and their products, the phenotypes of its mutants, and the literatures supporting these data, are provided in the SGD project (Cherry et al., 2012). ARSs mentioned in peer-reviewed literatures are also integrated in this database. For each ARS, the details about its sequence, location, relative literatures, and history can be obtained. Users can also use the analysis tools such as BLAST provided in SGD to explore these data.

SEQUENCE CHARACTERISTICS OF YEAST REPLICATION ORIGINS

In budding yeast *S. cerevisiae*, replication origins are defined as ARS because they can support the maintenance of a plasmid in growing yeast cells (Stinchcomb et al., 1979). Every replication origin contains a conserved 11-bp motif (sometimes assigned as 17 bp in length) called the ARS consensus sequence (ACS) that is essential for the binding of the initiator protein ORC (Rao and Stillman, 1995; Rowley et al., 1995; Theis and Newlon, 1997). A match to the ACS is essential but not sufficient for origin function. Even though, some bioinformatic algorithms for predicting the location of yeast replication origins have been developed based on ACS. For example, to predict the location of ORIs in the *S. cerevisiae* genome, Breier et al. (2004) developed an algorithm called Oriscan. This method utilized 268 bp of sequence, including the T-rich ACS and a 3′ A-rich region to identify ORI candidates. It then ranked potential origins by their likelihood of activity. A large proportion of origins in the genome were recognized by Oriscan with near-perfect specificity (Breier et al., 2004). Another computational study made use of the discovery that most replication origin sequences are phylogenetically conserved among closely related *Saccharomyces* species. It combined motif searches, phylogenetic conservation, and microarray data together to identify replication origin sequences throughout the *S. cerevisiae* genome (Nieduszynski et al., 2006). Analogously, the ORIs in *K. lactis* also contain a 50-bp ACS. The difference is that ACS in *K. lactis* ARSs is both necessary and largely sufficient for ARS activity (Liachko et al., 2010).

Abundant research was also conducted on the replication origins in fission yeast *S. pombe*, where replication sequences also function as autonomous replicators. However, ORIs in *S. pombe* do not have recognizable consensus elements but have a 500–1000 bp extended AT-rich structure (Dubey et al., 1994; Clyne and Kelly, 1995). Segurado et al. (2003) identified 384 potential origins by this feature. It was previously believed that replication origins in plant and metazoan are G/C-rich while in yeasts are A/T-rich. However, an industrially important methylotrophic budding yeast, *P. pastoris*, owed different characteristics in its ORIs compared with other studied yeasts. In this kind of yeast, two different types of ORIs exist simultaneously. In addition to an A/T-rich type more reminiscent of typical budding and fission yeast origins, there is also a G/C-rich type of replication origins associated with transcription start sites (Liachko et al., 2014). We calculate the GC content along *S. cerevisiae* chromosome 1 with sliding window algorithm (window size: 1000, shift: 20) and present it in **Figure 1** by the orange line. This line indicates that GC contents of the ORIs sequences are significantly lower than those of the entire genome sequences. In fact, this status exists in all the four kinds of yeasts, even in *P. pastoris*, the one includes G/C-rich type of ORIs.

To gain a comprehensive view of the conserved motifs in the origin sequences, we use the MEME-ChIP web service to discovery enriched motifs in the ORI sequences in the four kinds of yeasts. MEME-ChIP web service is designed especially for discovering motifs in the large sets of short DNA sequences (Bailey et al., 2009; Machanick and Bailey, 2011). The motifs we found are displayed in **Figure 2A**. ORIs in *S. cerevisiae*, *K. lactis*, and *S. pombe* contain AT-rich motifs, whereas GC-rich motifs are found in *P. pastoris* ORIs. We also construct the phylogenetic tree (**Figure 2A**) of the four organisms based on the cytochrome c downloaded from NCBI. The tree was constructed using the MEGA6 program (Statistical Method: Maximum Likelihood, Test of Phylogeny: Bootstrap method, No. of Bootstrap Replications: 1000; Tamura et al., 2013). Conserved motifs found in the four yeasts ORIs show no significant correlation with their phylogenetic relationships.

In addition, regions of local similarity in sequences between each pair of organisms are searched by the BLAST program (Altschul et al., 1997). **Figure 2B** is created by circos (Krzywinski et al., 2009), and shows the ORIs that share similar sequences. Each number around the circle is the ORI's serial number in DeOri. When two ORIs share similar local regions, a line will be drawn between them. For example, eori001300188, eori001300214, and eori001300331 have local regions similar with eori000800141, eori000800068, and eori000800010, respectively, hence the three pairs of ORIs are connected. No significant similarity is found between sequences in *S. pombe* ORIs and any other three groups of sequences. This may be caused by the large phylogenetic distance of *S. pombe*.

A new study suggests that in budding yeast, specific origin sequences are not strictly required for DNA replication *in vitro*, although they are essential for plasmid replication *in vivo*. The observation supports the notion that DNA replication specification in budding yeast is not completely dependent on DNA sequences, and epigenetic mechanisms are also important for determining replication origin sites (Gros et al., 2014).

DISTRIBUTION AND ORGANIZATION OF YEAST REPLICATION ORIGINS

Despite the lack of uniform feature of replication origin sequences, ORIs do not randomly locate on chromosome. Indeed, in all the four kinds of yeasts, origins have a significant preference for intergenic regions (Hayashi et al., 2007; Liachko et al., 2010, 2014; Renard-Guillet et al., 2014). We find that the correlation

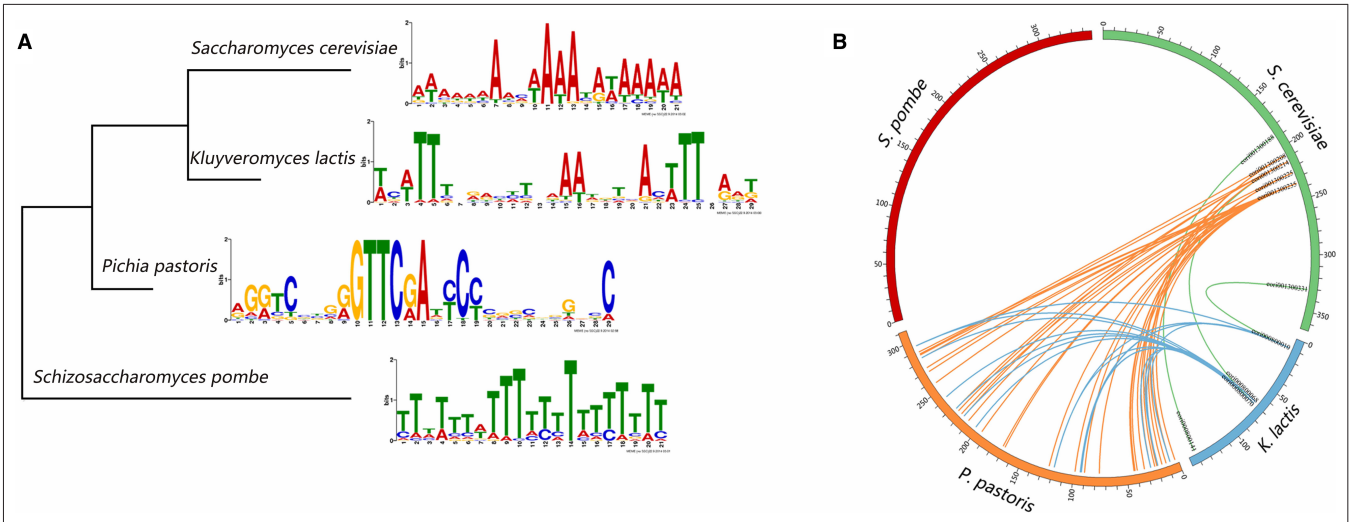


FIGURE 2 | Sequence characteristics of yeast replication origins. (A) The significant motifs found in the replication origin sequences and the phylogenetic tree of the four yeasts. **(B)** The circos plot of replication origins that share similar sequences. Each number around the circle is the ORI's serial number in DeOri.

coefficient values (R values) between the chromosome length and replication origins number are 0.956, 0.999, 0.966, and 0.854 for *S. cerevisiae*, *S. pombe*, *K. lactis*, and *P. pastoris*, respectively, which indicates that longer chromosomes tend to have more ORIs. In addition, ORIs always appear in the nucleosome-free regions (Li et al., 2014; Sherstyuk et al., 2014). We collect the nucleosome occupancy data in *S. cerevisiae* chromosome 1 (Kaplan et al., 2009) and map it in **Figure 1** by pink bars. The nucleosome occupancy scores in ORIs are significantly lower, which agrees well with the above conclusions. An asymmetric pattern of positioned nucleosomes has been verified at origins in both *S. cerevisiae* and *K. lactis* (Eaton et al., 2010; Tsai et al., 2014). These nucleosome occupancy information has been successfully used to train a machine learning algorithm to predict the position of active arm origins in the *Candida albicans* genome (Tsai et al., 2014).

Two other important features of ORIs are origin replication timing and efficiency. Origins are fired at various time throughout the S phase. *S. cerevisiae* ORIs can be separated into early and late origins. They present different nucleosomal architectures, which are already established in G1 phase. A higher occupancy of nucleosomes and broader nucleosome-depleted region (NDR) features appear in early origins, while late origins display a lower occupancy and tighter NDR (Soriano et al., 2014). In *S. pombe*, early and late origins tend to distribute separately in large chromosome regions (Hayashi et al., 2007). The dynamics of replication in *P. pastoris* shows an unexpected difference in replication timing between GC-ARSs and AT-ARSs. GC-rich ORIs replicate remarkably earlier and/or more efficiently than AT-rich ORIs (Liachko et al., 2014). In regard to origin replication efficiency, not all origins are used at each cell cycle. The overall efficiency of origin firing is less than 50% in *S. cerevisiae* and *S. pombe* (Friedman et al., 1997; Heichinger et al., 2006). It appears to be that the replication stress presented by different growth conditions affects the number of sites being activated (Tuduri et al., 2010). The flexibility of replication origins may be an obstacle in the thorough genome-wide understanding of ORIs in yeast.

ACKNOWLEDGMENTS

The authors would like to thank Prof. Chun-Ting Zhang for the invaluable assistance and inspiring discussions. The present work was supported in part by National Natural Science Foundation of China (Grant Nos. 31171238 and 30800642), Program for New Century Excellent Talents in University (No. NCET-12-0396), and the China National 863 High-Tech Program (2015AA020101).

REFERENCES

Altschul, S. F., Madden, T. L., Schaffer, 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

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335

Bell, L., and Byers, B. (1983). Separation of branched from linear DNA by two-dimensional gel electrophoresis. *Anal. Biochem.* 130, 527–535. doi: 10.1016/0003-2697(83)90628-0

Bell, S. P. (2002). The origin recognition complex: from simple origins to complex functions. *Genes Dev.* 16, 659–672. doi: 10.1101/gad.969602

Bell, S. P., and Dutta, A. (2002). DNA replication in eukaryotic cells. *Annu. Rev. Biochem.* 71, 333–374. doi: 10.1146/annurev.biochem.71.110601.135425

Besnard, E., Desprat, R., Ryan, M., Kahli, M., Aladjem, M. I., and Lemaitre, J.-M. (2014). Best practices for mapping replication origins in eukaryotic chromosomes. *Curr. Protoc. Cell Biol.* 64, 22.18.21–22.18.13. doi: 10.1002/0471143030.cb2218s64

Bowers, J. L., Randell, J. C. W., Chen, S. Y., and Bell, S. P. (2004). ATP hydrolysis by ORC catalyzes reiterative Mcm2-7 assembly at a defined origin of replication. *Mol. Cell* 16, 967–978. doi: 10.1016/j.molcel.2004.11.038

Breier, A. M., Chatterji, S., and Cozzarelli, N. R. (2004). Prediction of *Saccharomyces cerevisiae* replication origins. *Genome Biol.* 5:R22. doi: 10.1186/gb-2004-5-4-r22

Brewer, B. J., and Fangman, W. L. (1987). The localization of replication origins on ARS plasmids in *S. cerevisiae*. *Cell* 51, 463–471. doi: 10.1016/0092-8674(87)90642-8

Cadoret, J. C., Meisch, F., Hassan-Zadeh, V., Luyten, I., Guillet, C., Duret, L., et al. (2008). Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 15837–15842. doi: 10.1073/pnas.0805208105

Cayrou, C., Coulombe, P., Vigneron, A., Stanojcic, S., Ganier, O., Peiffer, I., et al. (2011). Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res.* 21, 1438–1449. doi: 10.1101/gr.121830.111

Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., et al. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* 40, D700–D705. doi: 10.1093/nar/gkr1029

Clyne, R. K., and Kelly, T. J. (1995). Genetic analysis of an ARS element from the fission yeast *Schizosaccharomyces pombe*. *EMBO J.* 14, 6348–6357.

Costas, C., Sanchez, M. D., Stroud, H., Yu, Y., Oliveros, J. C., Feng, S., et al. (2011). Genome-wide mapping of *Arabidopsis thaliana* origins of DNA replication and their associated epigenetic marks. *Nat. Struct. Mol. Biol.* 18, 395–400. doi: 10.1038/nsmb.1988

Cotterill, S., and Kearsey, S. E. (2009). DNAReplication: a database of information and resources for the eukaryotic DNA replication community. *Nucleic Acids Res.* 37, D837–D839. doi: 10.1093/nar/gkn726

Dubey, D. D., Zhu, J., Carlson, D. L., Sharma, K., and Huberman, J. A. (1994). Three ARS elements contribute to the ura4 replication origin region in the fission yeast, *Schizosaccharomyces pombe*. *EMBO J.* 13, 3638–3647.

Eaton, M. L., Galani, K., Kang, S., Bell, S. P., and Macalpine, D. M. (2010). Conserved nucleosome positioning defines replication origins. *Genes Dev.* 24, 748–753. doi: 10.1101/gad.1913210

Feng, W. Y., Collingwood, D., Boeck, M. E., Fox, L. A., Alvino, G. M., Fangman, W. L., et al. (2006). Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. *Nat. Cell Biol.* 8, 148–155. doi: 10.1038/ncb1358

Friedman, K. L., Brewer, B. J., and Fangman, W. L. (1997). Replication profile of *Saccharomyces cerevisiae* chromosome VI. *Genes Cells* 2, 667–678. doi: 10.1046/j.1365-2443.1997.1520350.x

Gao, F., Luo, H., and Zhang, C.-T. (2012). DeOri: a database of eukaryotic DNA replication origins. *Bioinformatics* 28, 1551–1552. doi: 10.1093/bioinformatics/bts151

Gao, F., Luo, H., and Zhang, C.-T. (2013). DoriC 5.0: an updated database of oriC regions in both bacterial and archaeal genomes. *Nucleic Acids Res.* 41, D90–D93. doi: 10.1093/nar/gks990

Gao, F., and Zhang, C. T. (2007). DoriC: a database of oriC regions in bacterial genomes. *Bioinformatics* 23, 1866–1867. doi: 10.1093/bioinformatics/btm255

Gros, J., Devbhandari, S., and Remus, D. (2014). Origin plasticity during budding yeast DNA replication in vitro. *EMBO J.* 33, 621–636. doi: 10.1002/embj.201387278

Hayashi, M., Katou, Y., Itoh, T., Tazumi, M., Yamada, Y., Takahashi, T., et al. (2007). Genome-wide localization of pre-RC sites and identification of replication origins in fission yeast. *EMBO J.* 26, 1327–1339. doi: 10.1038/sj.emboj.7601585

Heichinger, C., Penkett, C. J., Bähler, J., and Nurse, P. (2006). Genome-wide characterization of fission yeast DNA replication origins. *EMBO J.* 25, 5171–5179. doi: 10.1038/sj.emboj.7601390

Kaplan, N., Moore, I. K., Fondufe-Mittendorf, Y., Gossett, A. J., Tillo, D., Field, Y., et al. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* 458, 362–366. doi: 10.1038/nature07667

Karnani, N., Taylor, C. M., Malhotra, A., and Dutta, A. (2010). Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. *Mol. Biol. Cell* 21, 393–404. doi: 10.1091/mbc.E09-08-0707

Kharchenko, P. V., Tolstorukov, M. Y., and Park, P. J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat. Biotechnol.* 26, 1351–1359. doi: 10.1038/nbt.1508

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

Leonard, A. C., and Mechali, M. (2013). DNA replication origins. *Cold Spring Harb. Perspect. Biol.* 5:a010116. doi: 10.1101/cshperspect.a010116

Li, W.-C., Zhong, Z.-J., Zhu, P.-P., Deng, E.-Z., Ding, H., Chen, W., et al. (2014). Sequence analysis of origins of replication in the *Saccharomyces cerevisiae* genomes. *Front. Microbiology* 5:574. doi: 10.3389/fmicb.2014.00574

Liachko, I., Bhaskar, A., Lee, C., Chung, S. C. C., Tye, B.-K., and Keich, U. (2010). A comprehensive genome-wide map of autonomously replicating sequences in a naive genome. *PLoS Genet.* 6:e1000946. doi: 10.1371/journal.pgen.1000946

Liachko, I., Youngblood, R. A., Keich, U., and Dunham, M. J. (2013). High-resolution mapping, characterization, and optimization of autonomously replicating sequences in yeast. *Genome Res.* 23, 698–704. doi: 10.1101/gr.144659.112

Liachko, I., Youngblood, R. A., Tsui, K., Bubb, K. L., Queitsch, C., Raghuraman, M. K., et al. (2014). GC-rich DNA elements enable replication origin activity in the methylotrophic yeast *Pichia pastoris*. *PLoS Genet.* 10:e1004169. doi: 10.1371/journal.pgen.1004169

MacAlpine, D., and Bell, S. (2005). A genomic view of eukaryotic DNA replication. *Chromosome Res.* 13, 309–326. doi: 10.1007/s10577-005-1508-1

Machanick, P., and Bailey, T. L. (2011). MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* 27, 1696–1697. doi: 10.1093/bioinformatics/btr189

Masai, H., Matsumoto, S., You, Z., Yoshizawa-Sugata, N., and Oda, M. (2010). Eukaryotic chromosome DNA replication: where, when, and how? *Annu. Rev. Biochem.* 79, 89–130. doi: 10.1146/annurev.biochem.052308.103205

Mechali, M. (2010). Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat. Rev. Mol. Cell Biol.* 11, 728–738. doi: 10.1038/nrm2976

Newlon, C. S., Collins, I., Dershowitz, A., Deshpande, A. M., Greenfeder, S. A., Ong, L. Y., et al. (1993). Analysis of replication origin function on chromosome III of *Saccharomyces cerevisiae*. *Cold. Spring Harb. Symp. Quant. Biol.* 58, 415–423.

Nieduszynski, C. A., Hiraga, S., Ak, P., Benham, C. J., and Donaldson, A. D. (2007). OriDB: a DNA replication origin database. *Nucleic Acids Res.* 35, D40–D46. doi: 10.1093/nar/gkl758

Nieduszynski, C. A., Knox, Y., and Donaldson, A. D. (2006). Genome-wide identification of replication origins in yeast by comparative genomics. *Genes Dev.* 20, 1874–1879. doi: 10.1101/gad.385306

Raghuraman, M. K., Winzeler, E. A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., et al. (2001). Replication dynamics of the yeast genome. *Science* 294, 115–121. doi: 10.1126/science.294.5540.115

Rao, H., and Stillman, B. (1995). The origin recognition complex interacts with a bipartite DNA binding site within yeast replicators. *Proc. Natl. Acad. Sci. U.S.A.* 92, 2224–2228. doi: 10.1073/pnas.92.6.2224

Renard-Guillet, C., Kanoh, Y., Shirahige, K., and Masai, H. (2014). Temporal and spatial regulation of eukaryotic DNA replication: from regulated initiation to genome-scale timing program. *Semin. Cell Dev. Biol.* 30, 110–120. doi: 10.1016/j.semcdb.2014.04.014

Reynolds, A. E., Mccarroll, R. M., Newlon, C. S., and Fangman, W. L. (1989). Time of replication of ARS elements along yeast chromosome III. *Mol. Cell. Biol.* 9, 4488–4494.

Rowley, A., Cocker, J. H., Harwood, J., and Diffley, J. F. (1995). Initiation complex assembly at budding yeast replication origins begins with the recognition of a bipartite sequence by limiting amounts of the initiator, ORC. *EMBO J.* 14, 2631–2641.

Schepers, A., and Papior, P. (2010). Why are we where we are? Understanding replication origins and initiation sites in eukaryotes using ChIP-approaches. *Chromosome Res.* 18, 63–77. doi: 10.1007/s10577-009-9087-1

Segurado, M., De Luis, A., and Antequera, F. (2003). Genome-wide distribution of DNA replication origins at A^+ T-rich islands in *Schizosaccharomyces pombe*. *EMBO Rep.* 4, 1048–1053. doi: 10.1038/sj.embor.7400008

Sequeira-Mendes, J., Diaz-Uriarte, R., Apedaile, A., Huntley, D., Brockdorff, N., and Gomez, M. (2009). Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet.* 5:e1000446. doi: 10.1371/journal.pgen.1000446

Sherstyuk, V. V., Shevchenko, A. I., and Zakian, S. M. (2014). Epigenetic landscape for initiation of DNA replication. *Chromosoma* 123, 183–199. doi: 10.1007/s00412-013-0448-3

Siow, C. C., Nieduszynska, S. R., Muller, C. A., and Nieduszynski, C. A. (2012). OriDB, the DNA replication origin database updated and extended. *Nucleic Acids Res.* 40, D682–D686. doi: 10.1093/nar/gkr1091

Soriano, I., Morafraile, E. C., Vazquez, E., Antequera, F., and Segurado, M. (2014). Different nucleosomal architectures at early and late replicating origins in *Saccharomyces cerevisiae*. *BMC Genomics* 15:791. doi: 10.1186/1471-2164-15-791

Stinchcomb, D., Struhl, K., and Davis, R. (1979). Isolation and characterisation of a yeast chromosomal replicator. *Nature* 282, 39–43. doi: 10.1038/282039a0

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

Theis, J. F., and Newlon, C. S. (1997). The ARS309 chromosomal replicator of *Saccharomyces cerevisiae* depends on an exceptional ARS consensus sequence. *Proc. Natl. Acad. Sci. U.S.A.* 94, 10786–10791. doi: 10.1073/pnas.94.20.10786

Tsai, H.-J., Baller, J. A., Liachko, I., Koren, A., Burrack, L. S., Hickman, M. A., et al. (2014). Origin replication complex binding, nucleosome depletion patterns, and a primary sequence motif can predict origins of replication in a genome with epigenetic centromeres. *mBio* 5, e01703–e01714. doi: 10.1128/mBio.01703-14

Tuduri, S., Tourriere, H., and Pasero, P. (2010). Defining replication origin efficiency using DNA fiber assays. *Chromosome Res.* 18, 91–102. doi: 10.1007/s10577-009-9098-y

Weddington, N., Stuy, A., Hiratani, I., Ryba, T., Yokochi, T., and Gilbert, D. M. (2008). ReplicationDomain: a visualization tool and comparative database for genome-wide replication timing data. *BMC Bioinformatics* 9:530. doi: 10.1186/1471-2105-9-530

Wyrick, J. J., Aparicio, J. G., Chen, T., Barnett, J. D., Jennings, E. G., Young, R. A., et al. (2001). Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. *Science* 294, 2357–2360. doi: 10.1126/science.1066101

Xu, J., Yanagisawa, Y., Tsankov, A. M., Hart, C., Aoki, K., Kommajosyula, N., et al. (2012). Genome-wide identification and characterization of replication origins by deep sequencing. *Genome Biol.* 13:R27. doi: 10.1186/gb-2012-13-4-r27

Xu, W., Aparicio, J. G., Aparicio, O. M., and Tavaré, S. (2006). Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in *S. cerevisiae*. *BMC Genomics* 7:276. doi: 10.1186/1471-2164-7-276

Yabuki, N., Terashima, H., and Kitada, K. (2002). Mapping of early firing origins on a replication profile of budding yeast. *Genes Cells* 7, 781–789. doi: 10.1046/j.1365-2443.2002.00559.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: 08 December 2014; accepted: 29 January 2015; published online: 19 February 2015.

Citation: Peng C, Luo H, Zhang X and Gao F (2015) Recent advances in the genome-wide study of DNA replication origins in yeast. Front. Microbiol. ***6****:117. doi: 10.3389/fmicb.2015.00117*

This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology.

Copyright © 2015 Peng, Luo, Zhang and Gao. 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

frontiers

EPFL Innovation Park · Building I · 1015 Lausanne · Switzerland
T +41 21 510 17 00 · info@frontiersin.org · frontiersin.org