
**MECHANISMS ENABLING COMPENSATION TO
THE LOSS OF ORIGIN OF REPLICATION IN
*ESCHERICHIA COLI***

A THESIS TO BE SUBMITTED TO
**THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH
SCIENCES AND TECHNOLOGY**



FOR THE AWARD OF THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY

RESHMA T VEETIL

UNDER THE GUIDANCE OF

DR. ASWIN SAI NARAIN SESHASAYEE
READER F, NCBS

NATIONAL CENTRE FOR BIOLOGICAL SCIENCES
BENGALURU

SEPTEMBER 2020

**THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND
TECHNOLOGY**

**Private University Established in Karnataka by ACT 35 of 2013
BENGALURU - 560064**

DECLARATION BY THE CANDIDATE

I declare that this thesis entitled “**Mechanisms enabling compensation to the loss of origin of replication in *Escherichia coli***” submitted for the award of Doctor of Philosophy to THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND TECHNOLOGY, Bengaluru, is my original work, conducted under the supervision of my guide Dr. Aswin Sai Narain Seshasayee (and co-guide, Dr. Sunil Laxman). I also wish to inform that no part of the research has been submitted for a degree or examination at any university. References, help and material obtained from other sources have been duly acknowledged.

I hereby confirm the originality of the work and that there is no plagiarism in any part of the dissertation.



Signature of the Candidate

Place: Bengaluru

Date: 14/09/2020

Name of candidate: Reshma T Veetil

Reg. No.: 20616030128

SEPTEMBER 2020

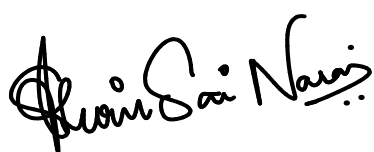
**THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES
AND TECHNOLOGY**

**Private University Established in Karnataka by ACT 35 of 2013
BENGALURU - 560064**

CERTIFICATE

This is to certify that the work incorporated in this thesis "**Mechanisms enabling compensation to the loss of origin of replication in *Escherichia coli***" submitted by **Reshma T Veetil** was carried out under my supervision. No part of this thesis has been submitted for a degree or examination at any university. References, help and material obtained from other sources have been duly acknowledged. I hereby confirm the originality of the work and that there is no plagiarism in any part of the dissertation.

Research Supervisor:



Aswin Sai Narain Seshasayee

Reader F

National Centre for Biological Sciences
Tata Institute of Fundamental Research
GKVK, Bellary Road, Bengaluru 560065

SEPTEMBER 2020

**THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES
AND TECHNOLOGY**

**Private University Established in Karnataka by ACT 35 of 2013
BENGALURU - 560064**

CERTIFICATE

This is to certify that the work incorporated in this thesis “**Mechanisms enabling compensation to the loss of origin of replication in *Escherichia coli***” submitted by **Reshma T Veetil** was carried out under my supervision. No part of this thesis has been submitted for a degree or examination at any university. References, help and material obtained from other sources have been duly acknowledged. I hereby confirm the originality of the work and that there is no plagiarism in any part of the dissertation.

Co-Supervisor:



Dr. Sunil Laxman

Assistant Investigator
Institute for Stem Cell Science
and Regenerative Medicine (inStem)
Bangalore: 560065
sunil@instem.res.in

SEPTEMBER 2020

Acknowledgement

This thesis would not have been possible without the help, encouragement and support of several people around me, who I wish to duly acknowledge in this section.

My sincere gratitude goes to my supervisor Dr. Aswin Sai Narain Seshasayee for his immense support and guidance. The discussions with him have always been motivational and cheerful. He gave me the freedom to work at my own pace and taught me the importance of doing what excites me to do in science. He has always been exceptionally tolerant to many mistakes I made while doing experiments and writing manuscript. I would like to thank his exceptional supportive efforts to improve my writing skills. I have always enjoyed his company starting from science to art, literature and sports.

I sincerely thank my thesis committee members Dr. Sunil Laxman, Dr. Shiva Prasad and Dr. Dasaradhi Palakodeti for their valuable suggestions and discussions. Sunil, my co-guide has been the person I have gone to for advice at times when I was confused about a particular problem, and his feedback during such occasions always helped me to take a decision wisely. I would like to thank prof. Jayaraman Gowrishankar and his lab at CDFD for discussions and materials. A Special thanks to Nalini Raghunathan and Sayantan Goswami for all their support with many of my experiments. Also, I am thankful to Dr. Anjana Badrinarayanan and Dr. Sandeep Krishna for scientific discussions. Thanks to all my collaborators for their help and discussions.

Thanks to members of our laboratory, both past and present for making the lab space most comfortable to do research. Especially, I would like to thank some of the senior students from our lab, Parul Singh, Supriya Khedkar, Aalap Mogre and Savita Chib for their help and support during my earlier days in the lab. Very special thanks to Mohak, Pabitra, Farhan, Shweta, Terence, Akshara and Nitish for being such a wonderful colleagues and good friends. I am also thankful to Harsha rani, for being a fantastic colleague and a supportive friend.

I would like to thank Dr Awadesh pandit and all the staff at Next Generation Genomics Facility, NCBS for providing sequencing services and the CIFF facility at NCBS for technical support. I would like to extend my gratitude to all the working staff at NCBS; the administrative section, instrumentation section; especially Alwyn, Purchase and store sections; laboratory support section; especially Ranjith and laboratory kitchen staff, for being so prompt and efficient in their duties so that our experiments can run smoothly. Also, I want to thank TDU administration especially Mr. Ravi Kumar for helping me with the university administrative procedures.

I want to thank Prof. Anil Kumar G and Prof. Sonali Barve for their care and support during the period. Also, I have enjoyed the friendship of many fantastic people in my life, and will remember them always with very warm memories, especially Amrita and Balu, Ashwin G, Binnu and their sweet little daughter Neelu, Dileep, Ismath, Calvin,

Sukanya, Vineeth, Shreya, Chitrang, Jagath, Arun have been great company to spend time away from the lab.

Also, I wish to thank late Pragati Dembla for being such a nice person to me and for teaching me the importance of care and love in a really hard way. You always remain in my heart and prayers.

Finally, I want to take this opportunity to thank my beloved parents (Suresh and Padmini) for their endless affection, support and love. Their confidence and faith in me has shaped me to the person I am today. A million thanks to mom and dad for everything. Also, I want to thank my first child, my sister vyshnavi for being an unconditional source of love and support in my life.

I am unable to find any words to express how fortune I feel to have the immeasurable love and constant support of my best friend, best half Sabari. I would like to thank him for always being there with me and lifting me up whenever I feel low. I am also thankful to his wonderful family (Mohandas and Prabhavathi) for their love, understanding, and constant support to complete my PhD.

I would like to dedicate this thesis to my beloved parents

Table of contents

Title	Pages
Chapter 1: Introduction: Replication oriented bacterial genome organization and alternative mechanisms	1
Chapter 2: Materials and Methods	16
Chapter 3: <i>oriC</i> independent DNA replication: Bacterial Growth and Gene expression	26
Chapter 4: <i>In vitro</i> Evolution Help Identify a Predominant Region of Constitutive Stable DNA Replication Initiation	53
Chapter 5: Conclusions and Discussion	89
Bibliography	100
Appendix Material	113

List of Tables

Title	Pages
Table 1.1: comparison of different replicative systems in <i>E.coli</i>	13
Table 2.1: M9 Minimal media composition (2X)	17
Table 3.1: identified peak positions and comparisons	46
Table 3.2: Functional classification of genes based on COG categories	47
Table 4.1: <i>Ori-to-ter</i> ratios for all strains calculated from MFA plots	79
Table 4.2: List of <i>E.coli</i> strains with chromosomal inversions around <i>oriC</i>	82
Table 4.3: Chromosomal Position of predicted <i>oriK</i> peaks from MFA plots	85
Table 4.4: probability of head on collisions in strains	88
Table A1: List of eubacterial strains which did not have a <i>dnaA</i> gene homologue	122
Table A2: Strains and plasmids used in this study	125
Table A3: Primers used in this study	126

List of Figures

Title	Pages
Figure 1.1: The replication related organization of the bacterial chromosome.	9
Figure 3.1: Schematic representation of the steps in the construction of $\Delta rnhA$ - $\Delta dnaA$ strain of <i>E.coli</i> K12	30
Figure 3.2: Genetic characterization of $\Delta rnhA$ - $\Delta dnaA$ strain of <i>E.coli</i> K12	31
Figure 3.3: $\Delta rnhA$ - $\Delta dnaA$ strain shows reduced growth in LB medium	33
Figure 3.4: $\Delta rnhA$ - $\Delta dnaA$ shows reduced growth in M9 minimal media conditions.	34
Figure 3.5: 60X phase contrast microscopy images of strains	35
Figure 3.6: 60x phase contrast, DAPI staining, and the merged images of strains	36
Figure 3.7: Deep sequencing based MFA plots for $rnhA^+$ - $dnaA^+$, $\Delta rnhA$, $\Delta rnhA$ - $\Delta dnaA$ strains.	38
Figure 3.8: MFA plot of $\Delta rnhA$ - $\Delta dnaA$ strain at exponential phase of growth	40
Figure 3.9: Effect of DNA copy number on gene expression	42
Figure 4.1: Laboratory evolution of $\Delta rnhA$ - $\Delta dnaA$ mutant	57
Figure 4.2: Colony count during the evolution for independent lines	58
Figure 4.3: Growth characteristics of evolved mutants	59
Figure 4.4: Suppressor mutants show reduced cell length	61

Figure 4.5: Unique mutations in suppressor mutants	62
Figure 4.6: Deep sequencing based MFA plots for suppressor mutants.	64
Figure 4.7: chromosomal inversions around <i>oriC</i>	66
Figure 4.8: <i>oriK45</i> as a preferred initiation site for cSDR in suppressor mutants	68
Figure 4.9: effect of <i>oriK45</i> deletion	70
Figure 4.10: Clustering of strains based on logCPM values of RNA-seq data.	72
Figure 4.11: <i>oriC</i> independent DNA replication and global gene expression changes	74
Figure 4.12: Effect of DNA copy number on gene expression	76
Figure 5.1: Graphical Summary of the work	99
Figure A1: Growth characteristics of evolved populations	114
Figure A2: Clustering of differentially expressed genes on the chromosome	116
Figure A3: Clustering of differentially expressed genes on the chromosome	117
Figure A4: Deep Sequencing based MFA plots of suppressor mutants at the exponential phase of growth	118
Figure A5: Deep sequencing based MFA plots of suppressor mutants at the stationary phase of growth	121

Synopsis

Mechanisms enabling compensation to the loss of origin of replication in *Escherichia coli*

Replication of chromosome is among the most essential functions of a cell, and it influences many other cellular mechanisms including gene expression and cell division. In a cell, genome duplication is initiated and regulated by the recruitment of replisome proteins to specific initiation sites on the chromosome known as origin of replication (*ori*) (1). A unique feature of a bacterial cell is the presence of a single origin of replication per circular chromosome (2). Canonical chromosome replication in the bacterium *Escherichia coli* initiates at a single origin of replication (*oriC*) and proceeds by the bidirectional progression of the replication forks in opposite directions before terminating (*ter*) at a locus positioned diametrically opposite to *oriC* on the circular chromosome (3). This process is initiated and regulated by the specific recognition of the repetitive short sequence motifs within the origin of replication by the protein DnaA (4). Binding of DnaA is followed by the unwinding of DNA double helix and the synthesis of a RNA primer that can then be extended by the action of replicative DNA polymerase III (3).

Many fundamental genome structural features in *E.coli* and other fast growing bacteria have evolved around the concept of bidirectional replication from a single *oriC*. These features include the encoding of highly expressed, essential genes near *oriC* to leverage on higher copy numbers during the cell cycle, and on the leading strand to minimise detrimental head-on collisions between the DNA polymerase and RNA polymerases transcribing these genes (2). Maintenance of such genes closer to the origin proximal region is conserved, and more so in fast growing bacteria (5, 6). Positioning such genes away from *oriC* or placing them on lagging strand of the chromosome can lead to detrimental effects on fitness especially in rapid growth conditions (7–9).

Despite the central role played by *oriC* and its cognate recognition protein DnaA in DNA replication and growth, *E.coli* can survive independent of *oriC* using alternative modes of replication, referred to as Stable DNA Replication (SDR) (10). Constitutive Stable DNA Replication (cSDR) is a mode of SDR in which the DNA replication is primed by R-loops (RNA-DNA hybrids) or recombination intermediates. In general cSDR is activated by processes that stabilise RNA-DNA hybrids. Inactivation of several proteins directly or indirectly involved in the disruption of R-loops is known to activate cSDR in *E.coli* (10). These include genes such as (a) *rnhA* which encodes for the RNA-DNA hybrid nuclease RNaseHI (11); and *topA* which encodes for the enzyme topoisomerase I. RnaseHI is an endonuclease enzyme which is directly involved in the disruption of R-loops. Lack of Topoisomerase I results in hyper negative supercoiling of DNA and as a result of this there will be an elevated occurrence of R-loop structures (12). Inactivation of RecG, a protein with helicase activity for RNA-DNA hybrids along with its function in DNA recombination is also known to induce cSDR in *E.coli* (13–16). Most recently, it has been also reported that the deficiency of DNA methylase Dam also activate cSDR and this mechanism is known to be independent of RnaseHI activity (17).

In our study, we focus on *Arnha* induced cSDR in *AdnaA* mutants of *E. coli* K12. An important question in cSDR is: where does DNA replication initiate, and what consequence does this have on chromosome organisation? This has been an open question for over 30 years (18). Previous studies on identifying sites of cSDR initiation using high resolution Marker frequency analysis (MFA) have been complicated by the slow growth state of the mutant, which makes copy number gradients difficult to establish (19). Therefore, the problem of identifying predominant cSDR initiation sites as potential origin of replication remains unsolved.

This thesis broadly investigates the impact of chromosome organization on global gene expression and evolution of a bacterium *E.coli*. It does so using cSDR, a non-*oriC* non-DnaA dependent replication system as the model of interest. Two major questions addressed in this work include how *oriC* independent DNA replication effect growth and gene expression profiles in *E.coli* and whether and how the bacterium would adapt to the loss of canonical DNA replication system. Apart from

this, this thesis also discusses where on the chromosome does cSDR initiate and whether natural selection would favour initiation predominantly from one or a few specific sites, from loci distributed across the genome?

Chapter 1 reviews the existing knowledge on canonical replicative structure of the bacterial chromosome and the replication related genome organization in bacteria. This chapter also explains different classes of origin independent DNA replication mechanisms and the genetic requirements to initiate DNA replication process by these mechanisms. Among these, we majorly focus on cSDR, an *oriC-DnaA* independent DNA replication process in which initiations are primed by R-loops. This chapter describes various methods by which an *E.coli* cell can activate cSDR as well as what we know so far about the positions of these initiation sites on the chromosome.

Chapter 2 describes the materials and methods used in this study.

Chapter 3 focuses on how *oriC* independent DNA replication impact growth and global gene expression patterns of *E.coli*. This chapter describes the construction and characterization of a $\Delta rnhA-\Delta dnaA$ *E.coli* strain engineered to initiate replication from R-loops. Growth measurements of this strain showed a severe growth defect in different media condition, displaying an extended lag phase and a reduced maximal growth rate. Chapter 3 also talks about how *oriC* independent DNA replication disrupts the gene gradient established from *ori* to terminus in wild type cells. This chapter also asks where cSDR initiates. This includes identification of replication initiation sites from high resolution sequencing based MFA graphs of $\Delta rnhA-\Delta dnaA$ strain and comparing them with the previous literature.

Also, chapter 3 reports genome-wide expression studies to profile gene expression status in $\Delta rnhA-\Delta dnaA$ strain of *E.coli*. Altered gene expression profiles of this mutant indicate that the replicative structure of the chromosome largely plays a role in maintaining the gene expression homeostasis in *E.coli*.

Chapter 4 discusses how bacteria compensate for the loss of canonical origin of replication in *E.coli*. Performing an *in vitro* evolution experiment of $\Delta rnhA-\Delta dnaA$ strain of *E.coli* helped to generate suppressor mutants that grow fast and therefore display strong *ori-ter* gradients. In this chapter we show that selection favours cSDR

initiation predominantly at a broad region of the chromosome located around 0.4-0.7 Mb clockwise of *oriC*. Here we also report that this region of the chromosome includes many bisulfite-sensitive sites, as well as a site containing sequence motifs that favour R-loop formation.

However, replication initiation from this site would cause head on collisions between the DNA polymerase and the transcribing RNA polymerase particularly at rRNA loci. This chapter also focuses on how the bacterium adapts to this problem. Inverting chromosomal regions which encode for several rRNA loci is favoured by selection in order to reduce replication-transcription conflicts. Using transcriptome analysis of these suppressor mutants we show that these chromosomal inversions help the bacteria to partially restore the gene expression changes brought about by cSDR. In conclusion, this chapter summarizes different evolutionary strategies accessed by the bacterium *E.coli* to restore gene gradient and gene expression patterns in compensate for the loss of canonical replicative structure of the chromosome.

Finally, chapter 5 integrates these findings and presents a clearer picture of the impact of replicative structure of the chromosome on global gene expression and evolution of bacteria. This chapter includes a discussion on how the structure of the chromosome drives genome organization of bacteria. An elaborate review on different evolutionary strategies accessed by the bacterium to increase growth fitness in response to particular perturbations of the replicative structure of the chromosome is also included here. A major lacuna in this field has been the limited understanding of the presence of cSDR replication initiation sites on the chromosome due to the complicated slow growth phenotype of the strain. Our attempt to answer the question of the existence of preferred *oriK* sites using laboratory evolution experiments shows strong selection specific regions on the chromosome which shed some lights on this field. This chapter also points out problems that remain unanswered and suggests directions for the future study.

References

1. O'Donnell M, Langston L, Stillman B. 2013. Principles and Concepts of DNA Replication in Bacteria, Archaea, and Eukarya. *Cold Spring Harb Perspect Biol* 5.
2. Rocha EP. 2004. The replication-related organization of bacterial genomes. *Microbiology* 150:1609–1627.
3. Mott ML, Berger JM. 2007. DNA replication initiation: mechanisms and regulation in bacteria. *Nat Rev Microbiol* 5:343–354.
4. Katayama T. 2017. Initiation of DNA Replication at the Chromosomal Origin of *E. coli*, *oriC*. *Adv Exp Med Biol* 1042:79–98.
5. Couturier E, Rocha EPC. 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.
6. Khedkar S, Seshasayee ASN. 2016. Comparative Genomics of Interreplichore Translocations in Bacteria: A Measure of Chromosome Topology? *G3 (Bethesda)* 6:1597–1606.
7. Bryant JA, Sellars LE, Busby SJW, Lee DJ. 2014. Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res* 42:11383–11392.
8. Srivatsan A, Tehranchi A, MacAlpine DM, Wang JD. 2010. Co-orientation of replication and transcription preserves genome integrity. *PLoS Genet* 6:e1000810.
9. Wang JD, Berkmen MB, Grossman AD. 2007. Genome-wide coorientation of replication and transcription reduces adverse effects on replication in *Bacillus subtilis*. *Proc Natl Acad Sci USA* 104:5608–5613.

10. Kogoma T. 1997. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol Mol Biol Rev* 61:212–238.
11. Ogawa T, Pickett GG, Kogoma T, Kornberg A. 1984. RNase H confers specificity in the dnaA-dependent initiation of replication at the unique origin of the *Escherichia coli* chromosome in vivo and in vitro. *Proc Natl Acad Sci USA* 81:1040–1044.
12. Martel M, Balleydier A, Sauriol A, Drolet M. 2015. Constitutive stable DNA replication in *Escherichia coli* cells lacking type 1A topoisomerase activity. *DNA Repair (Amst)* 35:37–47.
13. Hong X, Cadwell GW, Kogoma T. 1995. *Escherichia coli* RecG and RecA proteins in R-loop formation. *The EMBO journal* 14:2385–2392.
14. Lloyd RG, Rudolph CJ. 2016. 25 years on and no end in sight: a perspective on the role of RecG protein. *Curr Genet* 62:827–840.
15. Rudolph CJ, Upton AL, Briggs GS, Lloyd RG. 2010. Is RecG a general guardian of the bacterial genome? *DNA Repair (Amst)* 9:210–223.
16. Midgley-Smith SL, Dimude JU, Taylor T, Forrester NM, Upton AL, Lloyd RG, Rudolph CJ. 2018. Chromosomal over-replication in *Escherichia coli* recG cells is triggered by replication fork fusion and amplified if replicore symmetry is disturbed. *Nucleic Acids Res* 46:7701–7715.
17. Raghunathan N, Goswami S, Leela JK, Pandiyan A, Gowrishankar J. 2019. A new role for *Escherichia coli* Dam DNA methylase in prevention of aberrant chromosomal replication. *Nucleic Acids Res* 47:5698–5711.
18. Gowrishankar J. 2015. End of the beginning: elongation and termination features of alternative modes of chromosomal replication initiation in bacteria. *PLoS genetics* 11:e1004909.

19. Maduiké NZ, Tehranchi AK, Wang JD, Kreuzer KN. 2014. Replication of the *Escherichia coli* chromosome in RNase HI-deficient cells: multiple initiation regions and fork dynamics. *Mol Microbiol* 91:39–56.

List of Publications

- 1) **Reshma T. Veetil**, Nitish Malhotra, Akshara Dubey, Aswin Sai Narain Seshasayee (2020) “Laboratory Evolution Experiments Help Identify a Predominant Region of Constitutive Stable DNA Replication Initiation”, *mSphere* 5(1) e00939-19; <https://doi.org/10.1128/mSphere.00939-19>.
- 2) Malikmohamed Yousuf, Ilaria Iuliani, **Reshma T Veetil**, Aswin Sai Narain Seshasayee, Bianca Sclavi, Marco Cosentino Lagomarsino (2020) “Early fate of exogenous promoters in *E.coli*”, *Nucleic Acids Research* 48(5) 2348-2356; <https://doi.org/10.1093/nar/gkz1196>.
- 3) Aalap Mogre, **Reshma T. Veetil** and Aswin Sai Narain Seshasayee (2017) “Modulation of Global Transcriptional Regulatory Networks as a Strategy for Increasing Kanamycin Resistance of the Translational Elongation Factor-G Mutants in *Escherichia coli*”, *G3: GENES, GENOMES, GENETICS* 7(12) 3955-3966; <https://doi.org/10.1534/g3.117.300284>
- 4) Aalap Mogre, Titas Sengupta, **Reshma T Veetil**, Preethi Ravi, Aswin Sai Narain Seshasayee (2014) “Genomic Analysis Reveals Distinct Concentration-Dependent Evolutionary Trajectories for Antibiotic Resistance in *Escherichia coli*” *DNA Research*, 21(6):711–726; doi:10.1093/dnares/dsu032.

Chapter 1

Introduction

Replication oriented bacterial genome organization and alternative mechanisms

Note:

Parts of the text in this chapter is included in a manuscript *Veetil. et. al, mSphere, 2020*

1.1 DNA replication: Mechanism and Regulation

DNA or Deoxyribonucleic acid encodes for the genetic information of most of the organisms including humans. Accurate duplication of genetic material is essential for the inheritance of traits that will determine the phenotype of a cell and an organism. Therefore, DNA replication is a highly ordered task and controlled by regulatory proteins throughout the process (1). During replication, DNA strands get separated and serve as templates for the production of its counterpart. This mode of replication process in which one strand of the chromosome is conserved from the parent is referred to as semiconservative mode of DNA replication. The core machinery of DNA replication is conserved in all three domains of life: bacteria, archaea, and eukaryotes. Each living cell invests a large amount of energy to make sure that the precise copying of its genome happens by a lesser error rate such as 1 mistake in about a 100 million bases (2).

In a cell, genome duplication is initiated and regulated by the recruitment of replication machinery proteins to specific initiation sites on the chromosome known as the origin of replication. An origin of replication is a specific DNA sequence that possesses binding sites for the DNA replication initiator proteins. In eukaryotes, the DNA replication process is very complex and it initiates at multiple such sites whereas in bacteria the origin of replication exists in a single copy per circular chromosome (3). The process of replication initiates with the unwinding of DNA sequences at the origin of replication. This leads to the loading of replisome machinery and the bidirectional progression of replication forks from the origin. In general, replisome is composed of DNA polymerases, circular sliding clamps, a pentameric clamp loader, helicase, primase, and SSB (single-strand binding protein) even though the order in which they are arranged or connected varies among cell types (2, 4).

During DNA replication, the separation of two DNA strands by the ATP dependent *helicase* results in the synthesis of the complementary strand by a *DNA polymerase* enzyme. In general, *DNA polymerase* cannot initiate synthesis of new strands, but can only extend from an existing DNA or RNA strand paired with a template strand. In order

to accomplish this, a ~12 nt long fragment of RNA primer is produced and attached to the template strand by the action of an enzyme *primase* which helps the *DNA polymerase* to synthesise the complementary strand to complete the replication process. DNA polymerases have 5'-3' activity in which the enzyme adds a new base at the available 3'-OH of the nucleotide during the elongation process (2). This results in the replication of one strand continuously and the other in a discontinuous fashion. Nascent DNA which is synthesized continuously in the same direction as the growing replication fork is called as leading strand and the strand which synthesised in the opposite direction of the growing replication fork is called as lagging strand. Lagging strand synthesis happens with the formation of short Okazaki fragments which later get linked together (5). The termination of replication happens whenever replication forks in opposite direction meet. This occurs at many points in the eukaryotic chromosomes due to the presence of multiple origins. In bacteria, the circularity of the chromosome makes sure that the two replication forks meet each other on the opposite end of the parental chromosome known as the terminus region (6).

The expression of information contained in a section of a DNA (gene) is achieved by the process of transcription and translation and these processes are tightly regulated. Therefore, any change at the DNA level may lead to the loss of function or inactivation of a gene. Thus, the high accuracy of *DNA polymerase* is crucial to avoid any genetic changes such as mutations. However, most of the DNA damage which happens at the time of replication is primarily taken care of by the repair pathways present in cells (7). These repair machineries are capable to correct the mismatches to conserve the genetic information of a cell.

1.2 DNA replication in bacteria

In bacteria, DNA replication initiates from a single origin of replication and proceeds bidirectionally in opposite directions on the chromosome until the replication forks meet at the terminus (2, 8). The process of DNA replication is well understood in *Escherichia coli* (*E.coli*) which is a rod-shaped, gram negative bacterium found in the gut of warm-

blooded animals. The specific recognition of the single origin sequence is important for initiating DNA replication process in bacteria. *E.coli* contains a ~4.7Mbp long single chromosome with the presence of a single origin of replication known as *oriC*. *oriC* is a 245-bp AT-rich DNA segment which is recognized by an initiator protein **DnaA**. Specific binding of ATP dependent **DnaA** filament at the 9 mer repeats of *oriC* starts unwinding of 13 mer repeats in the same locus which creates a loop structure to facilitate *helicase* based unwinding process (9–11).

The binding of **DnaA** at the *oriC* is controlled by a protein called **SeqA** which sequesters the origin and prevents access to **DnaA**. This process is highly dependent on the methylation status of *oriC* where **SeqA** binds only to the newly synthesised hemimethylated DNA (2, 12). **DnaA** binding followed by the unwinding of *oriC* sequence leads to the loading of **DnaB** *helicase* by the help of a clamp protein **DnaC**. Following this, **DnaG**, a single subunit primase enzyme binds to **DnaB** and stimulates the release of the regulatory protein **DnaC** from **DnaB** (13). The synthesis of RNA primer by **DnaG** is important to initiate the synthesis of the new copy of DNA strand by DNA polymerase III in order to complete the replication process (2). These cascade of events indicate that the initiation and elongation of replication process is tightly coordinated in bacteria.

The termination of replication in *E.coli* happens at a defined zone of the chromosome known as terminus where the replication forks get trapped. *E.coli ter* region contains a series of termination or *ter* sites (*terA-terJ*), that block the replication forks moving in one direction but not the other (2). This allows the forks to enter but not to leave the terminus region which results in “Replication fork trap”. This phenomenon is achieved by the specific binding of **Tus** protein at the *ter* sites of the chromosome to infer the replication fork speed (14). The bidirectional movement of replication forks around the circular chromosome happens at a rate of 1,000 nucleotides per second to complete a full circle of the *E.coli* chromosome in an about 40 min after the initiation process (15).

1.3 Replication related bacterial genome organization

E.coli, a fast growing bacteria maintains a chromosomal symmetry by positioning *oriC* and terminus region approximately diametrically opposite to each other on the chromosome (16). This arrangement of the genome makes sure that the replichores have similar sizes so that the distance travelled by both replication forks will be similar in length. In fast growing conditions, the generation time of *E.coli* is less than that of DNA replication time. To achieve this, *E. coli* reinitiates DNA synthesis before completing the previous round of replication. This yields multiple copies of the chromosome in a cell at a time which will eventually get segregated into individual cells. The single origin-based DNA replicative structure of the *E.coli* chromosome plays an important role in shaping the genome organization. Many fundamental genome structural features have evolved around the single origin concept in bacteria which includes polarity in gene content and in gene expression between the *ori* and the *ter* (16). As follows, some of these concepts are described in detail in this section (Figure 1.1).

1.31 Gene expression gradients

In fast growing *E.coli*, multiple replication initiations at *oriC* results in a higher gene dosage near to the origin compared to the *ter* region. The gradient in gene dosage is expected to result in a certain polarity in gene organization on the chromosome (16). Thereby, rapid growth related, highly expressed genes tend to be encoded near the *oriC*. A classic example of this is the presence of six of the seven ribosomal RNA operons in *ori*-proximal regions of the chromosome. This characteristic might strongly depend on the growth rate of the bacteria (17). In the past, a bioinformatic analysis of numerous complete bacterial genomes showed the presence of significant maintenance of those genes involved in translation and transcription near the origin in fast-growing than in slow-growing bacterial species. Also, there is an enrichment of highly expressed essential genes near *ori* which strongly suggests that there might be a selection pressure to maintain the positioning of these genes at the *ori*-proximal region of the chromosome (17, 18). Recently, transcriptome data from a growth phase-dependent study showed that

genes expressed during phases of rapid growth are encoded closer to the *ori*, whereas those activated during the stationary phase tend to be encoded closer to *ter* (19). Apart from this, horizontally acquired genes that are typically maintained in a transcriptionally silenced state including the defective prophages are also located around the *ter* half of the genome (20). More direct evidence for the impact of gene positioning on the expression levels of specific genes also exists. For example, in an early study, authors showed that the expression of the *his* gene can be modulated by translocating it across the chromosome, and that differences in expression level between *ori*-proximal and *ter*-proximal loci were greater during rapid growth and presumably consistent with differences in gene dosage (21). Apart from this, Bryant. *et. al* showed that the expression level of a *lac-gfp* fusion varies with gene position but, in a way several orders of magnitude more than what would be expected from gene dosage effects (22). This study indicates an important role for local chromatin structure in gene expression states.

1.32 Strand biases in gene encoding

It has been observed that a majority of genes in any bacterial genome are encoded on the leading strand of the chromosome (23). Specifically, all ribosomal RNA operons, which together account for most cellular transcription are encoded on the leading strand. This strand bias in gene encoding is particularly severe in *Firmicutes* where ~75% of genes are encoded on the leading strand (24). It is understood that head-on collisions between the replicating DNA polymerase and the transcribing RNA polymerase are detrimental to the host organism, by causing lesions on the DNA. Encoding the most highly expressed genes on the leading strand minimises the extent of head-on collisions between the two polymerases (25). In *E.coli*, ~80% of the essential genes are encoded on the leading strand of the chromosome. This is consistent with the idea that head-on collision between the replication fork and the transcription bubble can be mutagenic which is unaffordable by those genes whose inactivation is lethal to the cell (26).

To address the importance of gene orientation, Srivatsan *et al.* constructed bacterial strains containing chromosomal inversions around the *oriC* and measured the fitness of the strain. These inversions reverted the orientation of several ribosomal RNA operons and these bacterial strains show a reduction in fitness especially in rich media conditions (27). Despite this, a two origin strain of *E. coli*, constructed in (28), did not show any apparent growth defects, even though the second origin being positioned in such a way that it would cause replication-transcription conflicts at rRNA operons. However, a later attempt by Ivanova and colleagues to create a similar strain revealed a strong growth defect and the genome analysis of the previously reported two origin strain by this group indicated the presence of a large chromosomal inversion, affecting several rRNA operons (29). Thus, Ivanova *et al.* could conclude that replication-transcription conflicts are one of the key determinants of fitness in *E. coli*.

1.33 Chromosome topology and replication-dependent gene organisation

Order and orientation affects the expression of a particular gene. Apart from this, the local chromosome topology and the three-dimensional structure of the genome might also have an impact on gene expression (30). Recombination-based contact maps of the bacterial chromosomes revealed the presence of *macrodomains* which is defined as a linear stretch of the chromosome such that recombination frequencies between loci were significantly higher within a macrodomain than across macrodomains. The *E. coli* chromosome contains such four macrodomains, one around the *ori*, another around the *ter*, and one on each replicore. Two non-structured elements, which were promiscuous in recombining with more distal elements were defined on either side of the *ori* macrodomain (31). Recently, high-resolution genome-wide chromosome conformation capture (Hi-C) maps generated for *E. coli* chromosome were also able to distinguish these similar localised structural domains (32). Further exploration of this structure identified molecules that define certain macrodomains, most notably the MatP protein which specifically binds to the *ter* macrodomain (33). Apart from this, Esnault *et al.* revealed that the effect of

chromosomal inversions on growth is also dependent on the local chromosome topology; i.e. inversions that span multiple topological macrodomains of the chromosome are more detrimental than those that are limited to a single domain (34). Taken together, it emerges that the functional value of chromosome organisation, in terms of the interplay between replication and transcription, also incorporates a topological element to it.

How does local chromosome topology affect gene expression? The folded bacterial genome is known to be regulated by Nucleoid-associated proteins (NAPs) and topoisomerases. DNA supercoiling is a dynamic topological property of the chromosome which impacts gene expression (35). *In vivo* binding profiles of DNA gyrase, an enzyme that creates negative supercoils in *E.coli*, shows a gradient decreasing from the *oriC* towards the *ter*, similar to a gradient in gene expression (19). The binding of NAPs that are global regulators of gene expression, also known to affect chromosome topology. For example, over-expression of H-NS (Histone-like nucleoid structuring protein), a global silencer of horizontally-acquired genes in *E.coli* results in compaction of the chromosome (36). On the other hand, the loss of HNS will lead to de-silencing of horizontally acquired genes that are positioned near terminus. A recent report suggests that *E.coli* can adapt to this loss of fitness due to H-NS deletion by duplicating ~40% of the chromosome centered around *oriC* region of the chromosome (37). It has also been shown that the binding sites of certain nucleoid-associated transcription regulators are clustered in definite chromosomal regions related to macrodomain boundaries (38). As an example, in *E.coli*, H-NS binding sites are enriched around the *ter* domain of the chromosome (20). Thus, nucleoid-associated proteins could act as a link between chromosome topology and gene expression.

Recently, an analysis of genome-wide detection of supercoiling patterns showed that there is an existence of a supercoiling gradient from origin to terminus only at stationary phase of growth in *E.coli*. In this case, *ter* region of the chromosome is more negatively supercoiled than the *ori* region and this pattern is correlated with the binding of the nucleoid-associated protein HU (39). A more recent exploration of *E.coli* chromosome

architecture using Hi-C analysis shows that the *E.coli* chromosome is partitioned into two structurally distinct entities; the terminus region and the rest of the chromosome. MatP protein activity is restricting the structure of *ter* domain whereas binding of HU and MukBEF proteins maintains the DNA contacts outside *ter* (32). Together, there is evidence to suggest that the replicative nature of the bacterial chromosome impacts gene organisation and probably chromosome topology.

Figure 1.1

Strand bias in gene encoding to avoid replication-transcription conflicts

Highly expressed genes cluster around *oriC* for replication associated gene dosage effects

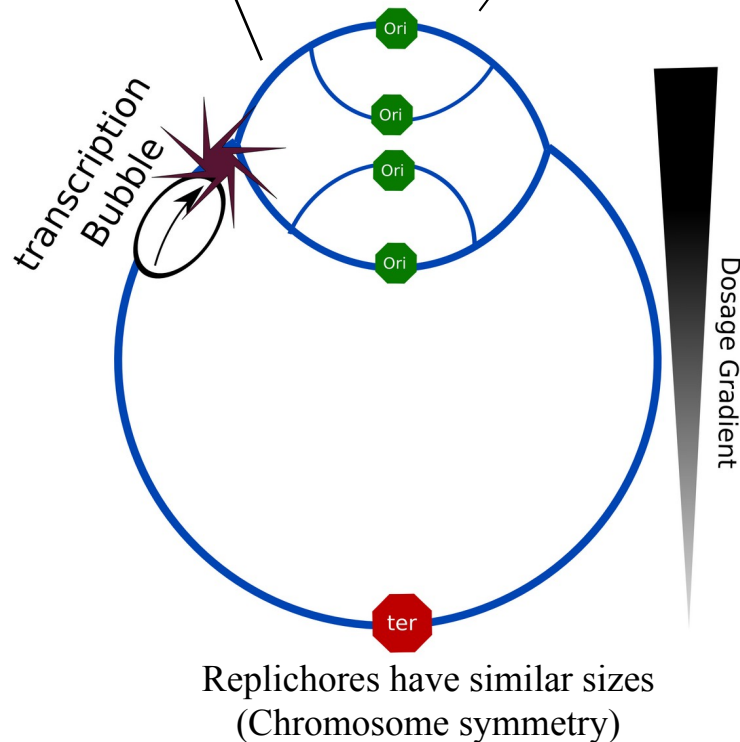


Figure 1.1: The replication related organization of bacterial chromosome. (16)

1.4 *oriC* independent DNA Replication in *E.coli*

Can the *oriC*-DnaA dependent mechanism of replication initiation in bacteria be dispensed with? Though DnaA is highly conserved across bacteria, it cannot be detected by sequence homology in a few (Figure 1.21). Mitochondria are not known to use *oriC*-DnaA-based DNA replication initiation. There are many models proposed to explain the process of mammalian mitochondrial replication so far including the “Strand displacement mechanism model”. However, mitochondrial replication systems are far more complex and the complete picture of the replication process is not yet clear. (40, 41) In *E.coli*, it is known that DnaA dependent initiation of replication is sensitive to chloramphenicol, an inhibitor of translation (42). This is probably because one or more protein factors involved in replication initiation are “unstable”. The realization that replication initiation by DnaA is sensitive to inhibition of translation resulted in the discovery of non-*oriC*, non-DnaA dependent alternative mode of replication process called “Stable DNA Replication” (SDR) (43).

There are multiple SDR mechanisms defined in the literature where they only differ from each other in the manner in which they initiate the replication process (43). Multiple broad types of SDR - each with its own set of genetic requirements - have been described here in detail (Table 1.1).

1.41 inducible Stable DNA replication (iSDR)

inducible SDR (iSDR) is known to be a recombination dependent replication process which gets activated upon induction of SOS DNA damage response in *E.coli* (44). iSDR is repressed under normal conditions and dependent on a number of homologous recombination proteins which includes RecA and RecBC (45). RecA is a protein that mediates the process of strand exchange between single stranded DNA and double stranded DNA during homologous recombination. Due to the dependency of iSDR on recombination proteins it has been hypothesised that the mode of duplex opening during iSDR occurs at a D-loop (DNA-DNA hybrid) which is a by-product of double strand

break and RecA activity. These replication initiation sites are called as *oriM* sites in iSDR conditions. The conditions which activates iSDR includes stress conditions such as Thiamine starvation, UV irradiation and Genotoxic agents (43).

According to Kogoma's D-loop model, introducing a double strand break at or near *oriM* site is sufficient to activate iSDR (43). This has been tested using a study by generating an artificial double strand break on a plasmid and monitoring the replication of the plasmid in SOS-induced cells grown in the presence of chloramphenicol and rifampicin antibiotics (46). These antibiotics are known to inhibit plasmid replication from the canonical origin of replication so the result of this experiment in which they showed that the plasmid replicates in presence of these antibiotics provide a strong support for the D-loop model.

1.42 constitutive Stable DNA replication(cSDR)

The efforts to find the genetic basis of SDR by Kogoma group resulted in generating mutants that express SDR in a constitutive manner (42). Consequently, this phenomenon is termed as constitutive Stable DNA replication (cSDR). Similar to iSDR, cSDR also shows a dependency on RecA function however, no other recombination proteins are known to be involved in cSDR (43). Although cSDR can initiate in the absence of protein synthesis it is observed that this process is sensitive to Rifampicin, a transcription inhibitor molecule (43). Later, the viability of cSDR mutants in a $\Delta oriC$ strain indicated that cSDR function is independent of *oriC* (47).

cSDR in *E.coli* is primed by RNA-DNA hybrids (R-loops) or recombination intermediates and these replication initiation sites are known as *oriK* sites in order to distinguish them from *oriC* (43). R-loops refers to a nucleic acid structure that is comprised of single stranded RNA base-paired with one strand of duplex DNA. Nascent untranslated transcription in bacteria is prone to generate R-loops (48).

In *E.coli*, stable R-loop formation is inhibited by the action of several proteins. This results in the repression of cSDR under normal conditions. cSDR is activated by processes that stabilise R-loops (49), such as the inactivation of (a) RnhA, the RNA-DNA hybrid nuclease RNaseHI; and that of (b) the topoisomerase I TopA, which results in

hyper negative supercoiling and elevated occurrence of RNA-DNA hybrids (50). Excessive R-loops have also been proposed to occur in strains defective for Rho-dependent transcription termination (48, 51–53), though to our knowledge Rho-dependent transcription termination has not been explicitly associated with cSDR. Very recently Raghunathan *et al.* demonstrated the role of the DNA *methylase* Dam in suppressing aberrant *oriC*-independent chromosomal replication and showed that the deficiency of this protein conferred SDR by a mechanism which is resistant to RNaseHI over-expression (54).

The first specific gene mutation that is known to activate cSDR involved inactivation of *rnhA* gene which encodes for RnaseH1 enzyme (42). RnaseH1 is an endonuclease that specifically cleaves RNA strands from RNA-DNA hybrids (55). The presence of this protein specifically ensures the exclusivity of initiation at *oriC*. Inactivation of RecG, a helicase for RNA-DNA hybrids with roles in DNA recombination is also known to activate cSDR (56–60). RecG function is hypothesized to compensate for the loss of RnaseH1 to some extent in *E.coli* cells since the deletion of *rnhA* and *recG* together in a cell is lethal (56).

Apart from this, at least one report has described “nutritional shift-up-activatable SDR (nSDR), as a non-*oriC*, non-DnaA dependent mechanism of chromosome replication employed by *E. coli* cells transiently during the stationary phase (61). As similar to cSDR, nSDR is dependent on RecA function and independent of RecB and DnaA. The similarities between nSDR and cSDR suggests that nSDR may be a manifestation of cSDR (43). However, DNA replication by SDR is sub-optimal under normal conditions relative to canonical DNA replication. Thus, it has been argued that this could be a potential primordial mechanism of DNA replication initiation (43).

Table 1.1

Mode of Replication	Type of cells	Mode of initiation	Origin used	Other requirements
DnaA/ <i>oriC</i>	Normal cells	DnaA- <i>oriC</i> interaction	<i>oriC</i>	DnaA, protein synthesis, RNA synthesis
iSDR	SOS induced	D-loop	<i>oriM</i>	RecA, RecBCD, PriA,
cSDR	<i>recG, rnhA, topA, dam</i> mutants	R-loop/recombination intermediates	<i>oriK</i>	RecA, PriA, RNA synthesis
nSDR	Nutrient upshifted	R-loop?	<i>oriK</i> ?	RecA, RNA synthesis? PriA?

Table 1.1: comparison of different replicative systems in *E.coli*. (43, 54, 60, 62)

1.5 cSDR and Replication initiation sites

An important question in cSDR is: where does DNA replication initiate and what consequence does this have on chromosome organisation? The Kogoma group, employing traditional *marker frequency analysis* (MFA), had identified five *oriK* loci at which replication might initiate (63). MFA analysis uses the argument that origin-proximal loci have a higher copy number than the rest of the chromosome in growing cells, even if they are not synchronised, to identify potential origins. Recently, Maduiké *et al.* used deep sequencing-based high resolution version of MFA to identify potential *oriK* sites, which were proximal to those identified by Kogoma's group. The strongest signal in the Maduiké *et al.* study mapped within the terminus of replication (64).

Nishitani and colleagues cloned and screened for fragments of the *E. coli* chromosome with potential for autonomous self-replication, and thereby identified a cluster of fragments again from within the terminus (65). However, both Maduiké *et al.* and Nishitani *et al.* appear to agree that the terminus sites identified in their studies are not bonafide *oriK* sites (64, 65). In the Maduiké *et al.* study, these terminus signals disappeared in a *Δtus* background in which replication forks trapped within the terminus are released. The authors conclude that the terminus signal may represent trapping of

forks originating from initiation sites elsewhere on the chromosome (64). Some of the *ter* sites identified by the Horiuchi group lost their activities in *Δtus*, but others did not. The Horiuchi group argued that an increased copy number of fragments from the terminus can be attributed to homologous recombination-based events and not autonomous replication (65). Gowrishankar has synthesised these arguments (66), and in conjunction with his lab's finding that RNA-DNA hybrids can occur throughout the chromosome (67), presented the case that cSDR can initiate anywhere on the chromosome; individual cells can initiate replication at different sites thus generating population-level heterogeneity, and; these can well explain the prominent MFA signal within the terminus. In a recent paper, Brochu *et al.* argue that *ΔtopA-topB* (more so than *ΔtopA-rnhA*) cSDR cells show a strong copy number peak within the terminus suggesting an *oriK* site here, but do not evaluate it in a *Δtus* background (62). These authors however observe that the *ter* peak is maintained in a strain with a large inversion around the *ter*, arguing against this peak being merely a consequence of replication fork trapping events.

Peak identification in high resolution MFA studies of cSDR is complicated by the slow growth phenotype of the mutant strain, which results in weak origin to terminus copy number gradients. Thus, the problem of identifying a predominant cSDR initiation site remains unsolved.

Objectives of this study

The replicative structure of the chromosome based on a single origin of replication is a central feature of bacterial physiology. A number of fundamental genome structural characteristics are evolved around the single origin concept in bacteria (16). This led us to ask how does bacteria compensate for the loss of a highly evolved replicative structure? In this thesis, we address the following objectives using a sub-optimal replication condition as our model of interest.

1) Investigate the impact of the replicative structure of the bacterial chromosome on growth and genome-wide gene expression patterns.

- Growth measurements, copy number analysis and RNA-seq experiments for a cSDR strain of *E.coli* in which *oriC* dependent DNA replication is inactivated.

2) Investigate the evolutionary strategies accessed by bacteria in the face of sub-optimal chromosome replication conditions to redress the loss of fitness?

- Laboratory evolution experiments to select for genetic variants with enhanced fitness.
- Whole-genome analysis and transcriptome analysis to understand the nature of mutations and other adaptive strategies evolved to compensate for the loss of canonical replication.

Chapter 2

Materials and Methods

Note:

Parts of the text in this chapter is included in a manuscript *Veetil. et. al, mSphere, 2020*

2.1 Strains, plasmids and Media conditions

Wild type (*rnhA⁺dnaA⁺*) strain mentioned in this study is a derivative of non pathogenic *E.coli* K12 MG1655 strain denoted as GJ13519 in (67). Gene deletions were performed using the one-step inactivation method described by Datsenko and Wanner (68) or by P1 phage mediated transduction (69). pHYD2388, a plasmid mentioned in this study is a single copy, plasmid derivative of the vector pMU575. This plasmid encodes for a region of *S. Enterica* genome from 4045412 bp to 4043222 bp which encompasses *dnaA* gene including its upstream regulatory region and a part of the downstream gene *dnaN*. pHYD2388 plasmid also encodes for a *lacZ* gene in the same orientation of cloned *dnaA* gene sequence. (54).

All experiments were conducted in Lysogeny Broth (LB; Hi-Media, India, M575-500) or freshly prepared M9 minimal media supplemented with Glucose and/or casaminoacids. When required the strains were plated on agar plates with the same media conditions in the presence or absence of X-gal and antibiotics. Antibiotics Kanamycin, Ampicillin and Trimethoprim were used at a final concentration of 50, 50 and 10 µg/ml respectively.

Table 2.1

Minimal salts	(2X)
Di-potassium phosphate (K ₂ HPO ₄)	21g
Potassium dihydrogen phosphate(KH ₂ PO ₄)	9g
Ammonium Sulfate (NH ₄) ₂ SO ₄	2g
Sodium acetate CH ₃ COONa.2H ₂ O	1g
H ₂ O	1000ml
MgSO ₄ (1M)	1ml
Glucose(20%)	10ml

Table 2.1: M9 Minimal media composition (2X)

2.2 Growth measurements

Growth was measured in 250ml flasks or 24-well plates in Lysogeny broth (LB; Hi-Media, India, M575-500) media at 37°C with shaking at 200 rpm. Optical density (OD) measurements were carried out at 600 nm (OD 600) using a UV-visible spectrophotometer (SP-8001) or multi well plate reader (Infinite F200pro, Tecan). Growth rates were calculated using Growthcurver (<https://CRAN.R-project.org/package=growthcurver>) and all plots were generated using in-house R scripts.

2.3 Spotting Assay

Spotting assay was performed for all strains at μ_{max} , which corresponds to the maximum growth at the exponential phase of growth. Overnight grown bacterial cultures were diluted in LB media to achieve 0.03 OD and incubated at 37 °C, 200 rpm until μ_{max} . Serial 10- fold dilutions of cultures were spotted (as 3 μ l spots) on LB agar plates. The plates were imaged after 30 hours of incubation at 37 °C.

2.4 Cell imaging and Cell length estimation

Overnight grown bacterial cultures were diluted in LB media and incubated at 37 °C with constant shaking at 200 rpm. Cells were harvested at μ_{max} and pelleted by centrifugation at 4000 rpm for 3 minutes. Pellets were resuspended in twice the volume of phosphate buffered saline (1X PBS, pH 7). 5 μ l of cells were then embedded under a 1% agarose pad prepared by dissolving UltraPure™ agarose (1%) (Invitrogen 16500-100) in PBS. The embedded cells were imaged using a Nikon Eclipse Ti-2 widefield microscope. Phase contrast images were taken using a 60X lens with oil immersion. The image analysis was performed using a software package called *Oufti* using *E.coli_LB_subpixel.set* parameters (70). The cell length in the number of pixels was converted to microns by multiplying with a pixel to micron conversion factor (0.064). Around 600 cells of each strain were imaged from multiple experiments to obtain a distribution of cell length.

2.5 DAPI Staining

Cells were harvested at μ_{max} as explained in the previous section (2.4) and pellets were resuspended in twice the volume of phosphate buffered saline (1X PBS, pH 7). 1 μ l of DAPI (4',6-diamidino-2-phenylindole) stain was added in to 1ml of PBS containing cells and mixed. 5 μ l of cells were then embedded under a 1% agarose pad prepared by dissolving UltraPureTM agarose (1%) (Invitrogen 16500-100) in PBS. The embedded cells were imaged using a Nikon Eclipse Ti-2 widefield microscope. Phase contrast images were taken using a 60X lens with oil immersion and DAPI fluorescence was imaged using a blue fluorescence channel (358nm excitation and 461nm emission). The images were analysed using Fiji-ImageJ software (71).

2.6 Genomic DNA isolation and Whole-genome sequencing

For genomic DNA extraction, overnight cultures were inoculated in 50 ml of fresh LB media to bring the initial Optical Density (OD) of the culture to 0.03 and the flasks were incubated at 37°C with shaking at 200 rpm. Cells were harvested at μ_{max} and genomic DNA was isolated using *GenEluteTM Bacterial Genomic DNA Kit* (NA2120-1KT, Sigma-Aldrich) using the manufacturer's protocol. For stationary phase whole-genome sequencing, the cultures were harvested after 16 hours of growth. Library preparation was carried out using *Truseq Nano DNA low throughput Library preparation kit* (15041757) following random shearing of genomic DNA into 350bp fragments using sonication (Covaris S220). Pooled libraries were sequenced using a paired end (2X100) sequencing approach on the Illumina Hiseq 2500 platform. The samples were multiplexed to obtain a minimum coverage of 10 million reads per sample.

2.7 DNA copy number analysis

The sequencing reads were aligned and mapped to the reference genome (NC_000913.3) using Burrows Wheeler Aligner (BWA) (72). BWA is a reference based mapping method that performs gapped alignment for Illumina reads and generates default output alignment

format file SAM. We chose alignment quality and mapping quality thresholds as 20 for this analysis. Read coverage across the genome was calculated for non-overlapping windows of 200nt each using custom PERL scripts and the values were normalized by the mode of the distribution across these bins. The normalized values in the logarithmic scale (\log_2) were plotted against chromosome coordinates to get measures of DNA copy number from *ori-to-ter*. The coordinates were repositioned in such a way that the numbering starts from *oriC* position in either direction. LOESS polynomial regression analysis was used for curve fitting.

LOESS polynomial regression is a non-parametric method used to fit a smooth curve between two variables. The fit is done locally which means for the fit at point x, the model considers only neighbourhood points of x weighted by their distance. This method uses an estimate considering the whole curve rather than a single estimate so that the overall uncertainty is measured.

2.8 *oriK* peak prediction

oriK positions were predicted from the loess fitted copy number plots using custom R scripts. Outliers were removed by visualization from the copy number data before fitting the curve. The loess fit was derived after removing known deletions and reversing the copy number curve around inversions. A position was called as an *oriK* peak if it has a negative slope, measured relative to the peak position, up to 100 kbp in both directions in the loess predicted values. Peak range is defined from the minimum to maximum position predicted for each peak site across strains.

2.9 RNA extraction, mRNA enrichment, and sequencing

Overnight cultures were inoculated in 100 ml of fresh LB media to bring the initial Optical Density (OD) of the culture to 0.03 and the flasks were incubated at 37°C with shaking at 200 rpm. Samples were collected at the maximum growth rate and two biological replicates were performed for each sample. The samples were immediately processed for total RNA isolation using Trizol method (15596018; Invitrogen). DNase

treated RNA was depleted of ribosomal RNA using the Ambion MicrobeExpress™ Kit (AM1905). Libraries were prepared for RNA-sequencing using RNA Sample preparation Kit without poly-A selection (NEB #E7645S/L) and single end sequencing for 50 cycles was done using the Illumina HiSeq 2500 platform.

2.10 Transcriptome analysis

The sequencing reads were aligned and mapped to the reference genome (NC_000913.3) using Burrows Wheeler Aligner (BWA) (72). The reference genome sequence (.fna) and annotation (.gff) files for the same strain were downloaded from the NCBI FTP website(<ftp://ftp.ncbi.nlm.nih.gov>). The raw read quality was checked using the FastQC software (version v0.11.5). SAMTOOLS (version 1.2) and BEDTOOLS (version 2.25.0) were used to calculate the read count per gene using the annotation file (.bed). The format of the annotation file (.gff) was changed to .bed using an in-house python script. The normalization and differential gene expression analysis for the two conditions were carried out using the edgeR pipeline (73). EdgeR uses a negative binomial distribution which helps the model to separate biological variation from technical variation. Log fold change expression values in comparison to *ArnhA-ΔdnaA* were plotted using In-house R scripts and the Pearson correlation values were predicted for the same. The genes that are differentially expressed by a log (base 2) fold change of 1.5 or above with a FDR value of 0.01 were considered for further analysis.

2.11 Laboratory Evolution of cSDR mutant

Laboratory evolution experiment was carried out for overnight grown cultures of eight independent isolates of *ArnhA-ΔdnaA*. Cells were grown in 24-well plates at 37°C, shaking at 200 rpm, until late exponential phase and diluted by a factor of 1:100 into fresh LB broth. Bacterial populations were stored as 50% glycerol stocks at -80°C before the next sub-culturing. Contamination check was done for each population using PCR amplification of *rnhA* and *dnaA* genes from isolated genomic DNA samples. Alternative passages were plated on Luria agar plates (10^{-6} and 10^{-7} dilution) and counted CFU/ml for

each sample during evolution. Number of generations of evolution (N) was calculated using the minimum and maximum OD values per passage. The growth characteristics of evolved populations were monitored in 96-well plates at 37°C, 200 rpm using a Plate reader (Tecan, infinite® F200 PRO). Randomly chosen colonies from different passages were selected for whole-genome sequencing.

2.12 Mutation analysis and *ori-to-ter* ratio calculation

SNPs and indels were identified from the genome sequencing data using the BRESEQ (version 0.33.1) pipeline (74) which uses Bowtie (a short read aligner similar to BWA) for sequence alignment. A mutational matrix representing the presence and absence of mutations were generated from BRESEQ output file using custom R scripts and heat maps were generated using Matrix2png. Copy number plots for each sample at the maximum growth rate were used to determine *ori-to-ter* ratios. The ratio of maximum loess fit value (excluding *ter*) to the loess fit value of *dif* site (1588800) for each evolved strain was calculated using custom R scripts.

2.13 R-loop predictions using QmRLFs Finder

To predict RNA-DNA hybrids on the chromosome, we used QmRLFs model (75, 76) on *Escherichia coli* K12 MG1655 (NC_000913.3) genome with default parameters. QmRLFs model predicts R-loop structures in nucleic acid sequences based on experimentally supported structural models. This model is mainly based on three parameters which are a short G-cluster-rich region, a linker and a high G-density, downstream R-loop elongation zone. From the output file we considered starting position of a predicted R-loop and plotted a line plot for these positions using custom R scripts for both the models (m1 and m2) separately.

2.14 Nanopore sequencing and assembly of genomes

Genome assembly using short-read sequencing technologies are challenging due to the difficulty in accurately analysing repetitive regions and large structural variations. This problem can be greatly simplified using nanopore methods which generate extremely

long reads. Thus, Nanopore sequencing followed by a complete genome assembly is highly recommended to detect chromosome rearrangements such as inversions (77).

Genomic DNA isolated using GenElute™ Bacterial Genomic DNA Kit (NA2120-1KT, Sigma Aldrich) were subjected to Nanopore sequencing. Sequencing library preparation was carried out with Nanopore Genomic Sequencing Kit SQK-108 and a PCR-free ‘native barcoding’ kit following manufacturer's protocol. Barcoded samples were pooled and loaded onto a MinION™ Flow Cell MIN106 controlled by MinKNOW version V1.2.8 software (ONT). Base calling was performed using albacore Basecall Barcoding workflow (version 1.11.5) (ONT). The Fasta files of reads obtained from sequencing were subjected to a Denovo assembler Canu (<https://github.com/marbl/canu>) (78) using default parameters. Assembled contigs were analysed using sequence aligner Mauve (<http://darlinglab.org/mauve/mauve.html>) to find chromosomal rearrangements.

2.15 *oriC* inversion prediction in *E.coli* genomes

675 complete *E.coli* genomes downloaded from ncbi ftp site were used for this analysis. For finding *E.coli* strains that possess a chromosomal inversion of *oriC* region, BLASTn was performed on genomes with *E.coli* K12 MG1655 (NC_000913.3) genome as query and reference for inversion. The inverted regions from BLAST output of complete genomes were stitched and added together to calculate the total inverted region, thus an inference was made on the status of inversion of the region involving *oriC*. *oriC* positions in these genomes were predicted for all *E.coli* strains by performing blastn using *E.coli* K12 MG1655 (NC_000913.3) *oriC* region as query.

2.16 *hotH* Deletion and Blue-white screening

Appropriate dilution (10^{-6}) of overnight cultures of $\Delta rnhA-\Delta dnaA$ -pHYD2388 (*dnaA*⁺*lacZ*⁺) and $\Delta rnhA-\Delta dnaA-\Delta hotH$ -pHYD2388 (*dnaA*⁺*lacZ*⁺) $\Delta rnhA-\Delta dnaA-\Delta fimD$ -pHYD2388 (*dnaA*⁺*lacZ*⁺), $\Delta rnhA-\Delta dnaA-\Delta m1Rloop$ -pHYD2388 (*dnaA*⁺*lacZ*⁺) were plated on M9 minimal agar plates containing X-gal. Plates were incubated at 37°C for 30 hours and the number of blue and white colonies appeared on these plates were counted

separately. Using the colony count data, we compared the percentage of white colonies across strains.

2.17 Probability of head-on collision prediction

The probability of head-on collisions in evolved and parental strains from RNA sequencing data was calculated for the chromosome region 3.3Mb to 4.6Mb, which includes the inversion. The rate of head-on collisions in the presence or absence of the inversion was calculated by assuming the activation of a single predominant origin of replication in evolved and parental clones (either *oriC* or *oriK45*). The fractional score of head-on replication-transcription conflicts was defined as the ratio of the number of reads mapping to genes encoded on the lagging strand to the total number of reads mapping to the region for each strain. The strand information for genes were adapted from NC_000913 (version 3) .ptt or .rnt files.

2.18 DNA Copy number and transcriptome comparison

The sequencing reads of both DNA and RNA isolated at the exponential phase of growth were analysed as similar to transcriptome analysis described in materials and methods. The normalization and differential gene expression analysis for the two conditions were carried out using the edgeR (73) pipeline. Smoothed Log₂FC expression values in comparison to *ΔrnhA-ΔdnaA* were plotted against chromosome coordinates using in-house R scripts.

2.19 Promoter activity assay

The promoter activity of the mutant and wildtype *rrnD* promoter region (*rrsD-yrdA* Intergenic region (IGR)) was monitored by transforming the pUA139 plasmid containing cloned construct of the IGR region in to wild type *E.coli*. M9 medium with 0.2% glucose was used to culture the strains. Overnight cultures containing the plasmid strains were inoculated at a ratio of 1:100 in 100 ml media and the samples were isolated at various intervals to measure GFP fluorescence using FACS caliber instrument.

Around 25,000 cells were acquired for each sample using a 488-nm excitation laser, and the emission was recorded from the FL1 channel that uses a 530/30 band-pass (BP) filter to collect the GFP intensity. An empty vector of pUA139 plasmid was used to set the background fluorescence, and GFP intensity above this background was marked as positive. The FL1 histograms of samples across the same time point were compared by thresholding according to the empty vector FL1 curve using Flowing software (Cell Imaging Core of the Turku Centre for Biotechnology [<http://www.flowingsoftware.com/>]).

Chapter 3

***oriC* independent DNA replication: Bacterial Growth and Gene expression**

Note:

Parts of the text in this chapter is included in a manuscript *Veetil. et. al, mSphere, 2020*

3.1 Introduction

A unique feature of a bacterial cell is the presence of a single origin of replication (*ori*) per circular chromosome. Canonical chromosome replication in the bacterium *Escherichia coli* is initiated by the specific recognition of repetitive short sequence motifs within the origin of replication *oriC* by the protein DnaA. This is followed by DNA unwinding and the synthesis of an RNA primer that can then be extended by the replicative DNA polymerase III (8). Replication proceeds bidirectionally outwards of *oriC* before terminating at a locus positioned diametrically opposite to *oriC* on the circular chromosome (79).

Bidirectional replication from a single *oriC* might have been the selective force behind the evolution of several organisational features of the genomes of Bacteria, especially of those capable of rapid growth. These features include the encoding of highly expressed essential genes close to *oriC* to take advantage of the higher copy number of these loci while replication is in progress, and on the leading strand of replication to minimise the detrimental effects of head-on collisions between the DNA polymerase and RNA polymerases transcribing these genes (16). The positioning of such genes close to *oriC* is conserved, and more so in fast growing bacteria (17, 18). Repositioning of such genes away from *oriC* or on the lagging strand can be detrimental to fitness, especially in nutrient rich conditions (22, 27).

Although single origin centric DNA replication and its impact on chromosome organization is well established in bacteria, the realisation that replication initiation by DnaA is sensitive to inhibition of translation resulted in the discovery of non-*oriC*, non-DnaA dependent “Stable DNA Replication” (SDR) (43)

Multiple broad types of SDR - each with its own set of genetic requirements - have been described. Inducible SDR (iSDR) requires the SOS DNA damage response (43, 44). Constitutive SDR (cSDR) is activated by processes that stabilise RNA-DNA hybrids or R-loops, such as the inactivation of (a) *rnhA* which encodes for the RNA-DNA hybrid nuclease RNaseHI (49); and *topA* which encodes for the enzyme topoisomerase I.

RnaseH1 is an endonuclease enzyme which is directly involved in the disruption of R-loops. Lack of topoisomerase I results in hyper negative supercoiling of DNA and as a result of this, there will be an elevated occurrence of R-loop structures (50). Inactivation of RecG, a protein with helicase activity for RNA-DNA hybrids along with its function in DNA recombination is also known to induce cSDR in *E.coli* (56–59). Most recently, it has been also reported that the deficiency of DNA methylase Dam also activates cSDR and this mechanism is known to be independent of RnaseH1 function (54). We note here that Stable DNA Replication (SDR) is generally repressed under fast growing conditions and get activated upon stress such as starvation and DNA damage (43).

The important question in cSDR is: where on the chromosome does cSDR initiate? To answer this, here we focus on R-loop mediated $\Delta rnhA$ induced cSDR in $\Delta dnaA$ mutants of *E. coli* K12. The gene *rnhA* encodes the RNaseH1 nuclease enzyme that carries out endonucleolytic cleavage of RNA from RNA-DNA hybrids. Therefore, it serves as a specificity factor blocking replication from sites other than *oriC* in normal cells. Deletion of the *rnhA* gene will lead to the formation of stable R-loops in *E.coli* where DNA replication can initiate (80). These R-loop mediated replication initiation sites are known as *oriK* sites to distinguish them from *oriC*.

Identifying potential sites of origin of replication (*oriK*) in a cSDR regime has been an open question for over 30 years. Earlier, traditional *Marker frequency analysis* (MFA) (63) and screening of chromosomal fragments capable of autonomous replication (65) helped to identify various *oriK* sites on *E.coli* chromosome. Previous studies on identifying sites of cSDR initiation using high resolution Marker frequency analysis (MFA) have been complicated by the slow growth state of the mutant, which makes copy number gradients difficult to establish (64). Therefore, the problem of identifying predominant cSDR initiation sites as a potential origin of replication remains unsolved.

In this study, using high resolution genomics data we identified few regions of *E.coli* chromosome as potential cSDR replication initiation sites. This chapter will broadly discuss the locations of such initiation sites in comparison with previous observations and the challenges of identifying and confirming a chromosomal region as a replication

initiation site in the cSDR regime. Apart from this, it is also important to understand how a bacterial cell with dispersed origins of replication affect the chromosome organization and gene expression. Does inactivation of *oriC* leads to loss of gene expression homeostasis? In answering these questions, we will be able to understand how the canonical replicative structure of the chromosome impacts growth and global gene expression patterns in *E.coli*.

3.2 Results

3.21 Genetic characterization of $\Delta rnhA$ - $\Delta dnaA$ strain of *E.coli* K12

The $\Delta rnhA$ mutant which lacks RNaseH1 nuclease displays cSDR and therefore suppresses the lethality of $\Delta dnaA$ and $\Delta oriC$ mutants. Wild type (denoted as $rnhA^+dnaA^+$) strain mentioned in this study is a derivative of *E.coli* K12 MG1655 strain in which 95 nonessential genes (From [*mmuP*] to [*mhpD*], including the *lac* operon) had been deleted (67). To get around the lethality of *dnaA* deletion, the construction of a $\Delta rnhA$ - $\Delta dnaA$ strain was performed in Prof. Jayaraman Gowrishankar's lab at CDFD by a novel approach explained here in detail (Figure 3.1) (54).

In order to accomplish *dnaA* gene deletion, a single copy number shelter plasmid pHYD2388 encoding a *dnaA* allele (from *S. Enterica*) along with *lacZ* gene was transformed in to $rnhA^+dnaA^+$ strain. The transformed colonies were selected using X-gal based blue-white screening method where the plasmid bearing strains will appear in blue (*lacZ*⁺). Next, the chromosomal *dnaA* gene copy was deleted from the plasmid bearing strain using single step gene deletion method. This is followed by deletion of the *rnhA* gene (using P1 phage-mediated transduction for a *rnhA*::Kan strain obtained from Keio collection) to construct a double deletion mutant $\Delta rnhA$ - $\Delta dnaA$ -pHYD2388 ($dnaA^+lacZ^+$) strain. This bacterial strain is capable of activating both *oriC* and R-loop mediated DNA replication in cells. We obtained a $\Delta rnhA$ single deletion mutant, and the previously

Figure 3.1

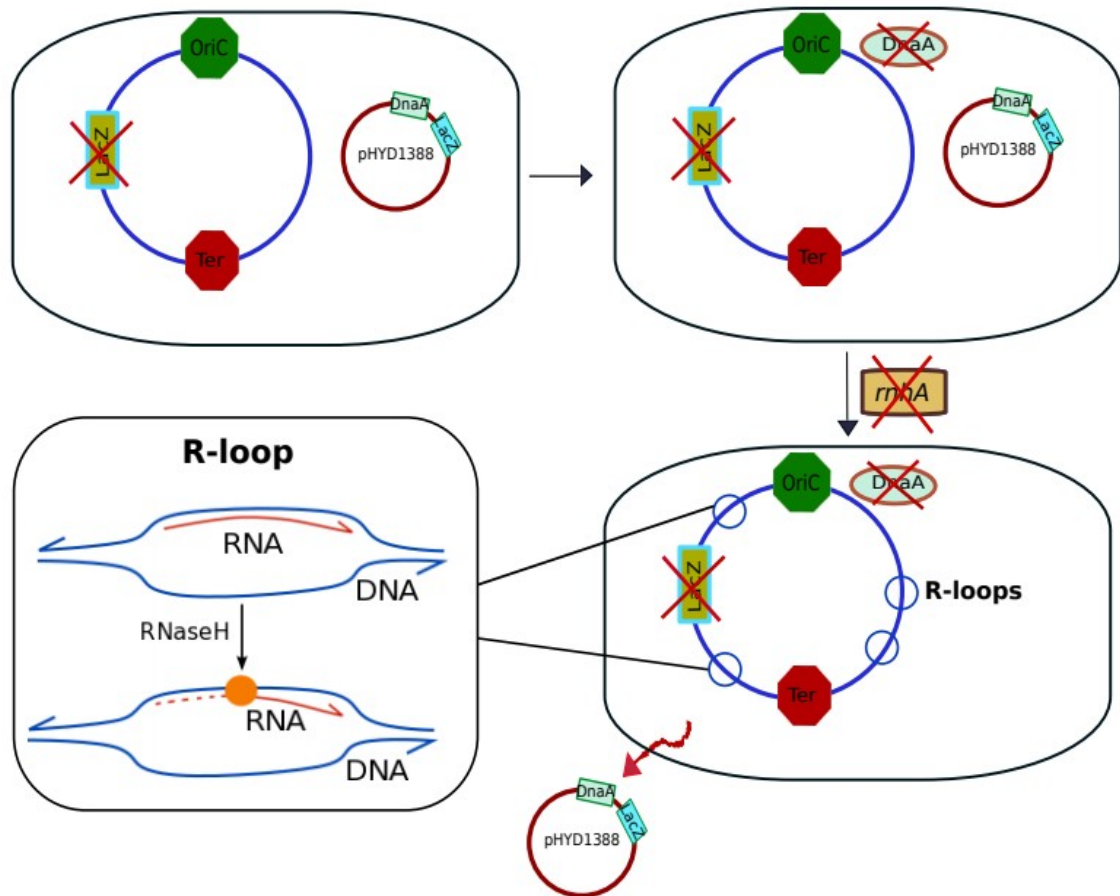


Figure 3.1: Schematic representation of the process of construction of $\Delta rnhA$ - $\Delta dnaA$ strain in *E. coli* K12 MG1655.

mentioned $\Delta rnhA-\Delta dnaA$ -pHYD2388 ($dnaA^+ lacZ^+$) mutant of *E.coli* K12 (MG1655) from Shankar lab at CDFD to proceed further experiments.

To obtain $\Delta rnhA-\Delta dnaA$ colonies, we plated overnight cultures of $\Delta rnhA-\Delta dnaA$ -pHYD2388 ($dnaA^+ lacZ^+$) on X- gal agar plates. Spontaneous loss of the $dnaA^+$ pHYD2388 plasmid produced white colonies ($dnaA^- lacZ^-$), which we selected and propagated as the $\Delta rnhA-\Delta dnaA$ ($dnaA^- lacZ^-$) strain (Figure 3.2A). Deletion of $dnaA$ and $rnhA$ genes in these mutants were confirmed using PCR method using gene specific external primers (Figure 3.2B). Whole genome sequencing and mapping of reads to reference genome of *E.coli* K12 MG1655 further confirmed gene deletions.

Figure 3.2

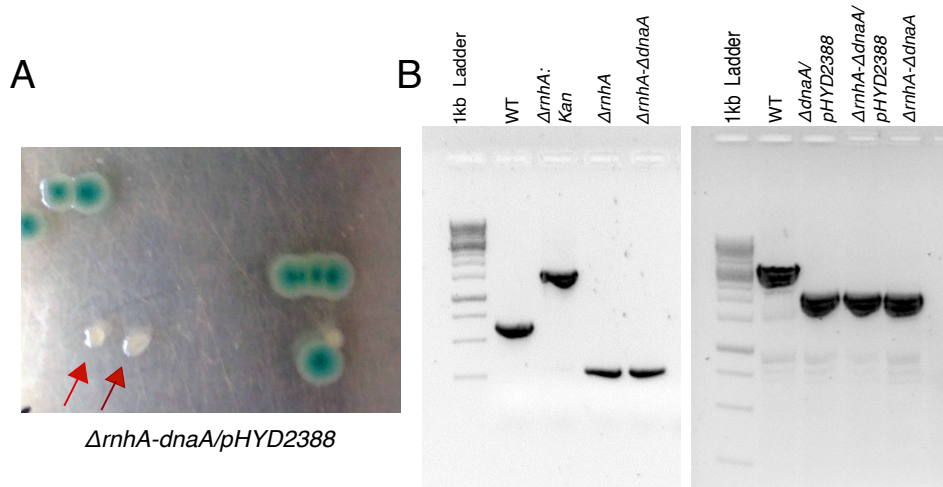


Figure 3.2: Genetic characterization of $\Delta rnhA-\Delta dnaA$ strain of *E.coli* K12: (A) colony image of $\Delta rnhA-\Delta dnaA/pHYD2388$ strain of *E.coli* on LB agar plates. white colonies (pointed by red arrows) represents the loss of plasmid pHYD2388 and thus referred as $\Delta rnhA-\Delta dnaA$ mutant. (B) Gel images after PCR amplification of *rnhA* and *dnaA* genes respectively. ccxc

3.22 Disruption of canonical replication pathway in *E.coli* K12 MG1655 alters growth and cell division

To understand the impact of *oriC*-DnaA independent replication on growth, we performed growth curve experiments in rich and minimal media conditions. The $\Delta rnhA$ single mutant, in which both *oriC*-DnaA dependent replication initiation and cSDR should be active, showed a slight growth defect in LB when compared to the corresponding $rnhA^+dnaA^+$ strain. The $\Delta rnhA-\Delta dnaA$ double deletion mutant showed a more severe growth defect in LB, displaying an extended lag phase and a reduced maximal growth rate (Figure 3.3). A similar trend was followed by these strains in minimal media conditions supplied with glucose and/ amino acids (Figure 3.4). The slow growth of $\Delta rnhA-\Delta dnaA$ colonies and lack of growth in higher dilutions on Luria agar plate exhibit the stress phenotype of the strain (Figure 3.3D).

In addition to this, $\Delta rnhA-\Delta dnaA$ strain also showed longer cell size compared to $\Delta rnhA$ and $rnhA^+dnaA^+$ cells ($P \lll 0.001$, Wilcoxon test). Cell size varies from $2\mu\text{m}$ to $30\mu\text{m}$ in a population of $\Delta rnhA-\Delta dnaA$ mutant (Figure 3.5). DAPI staining of $\Delta rnhA-\Delta dnaA$ cells showed the presence of multiple nuclei in longer cells which suggests that cell division is impaired (Figure 3.6). These results correlate with previous reports which indicate elevated SOS response in cSDR (60) and assume that DNA breaks in this mutant might play a role in cell size heterogeneity. Together, these observations suggests that perturbing single origin based replicative structure of the chromosome alters growth and cell division of fast growing bacteria.

Figure 3.3

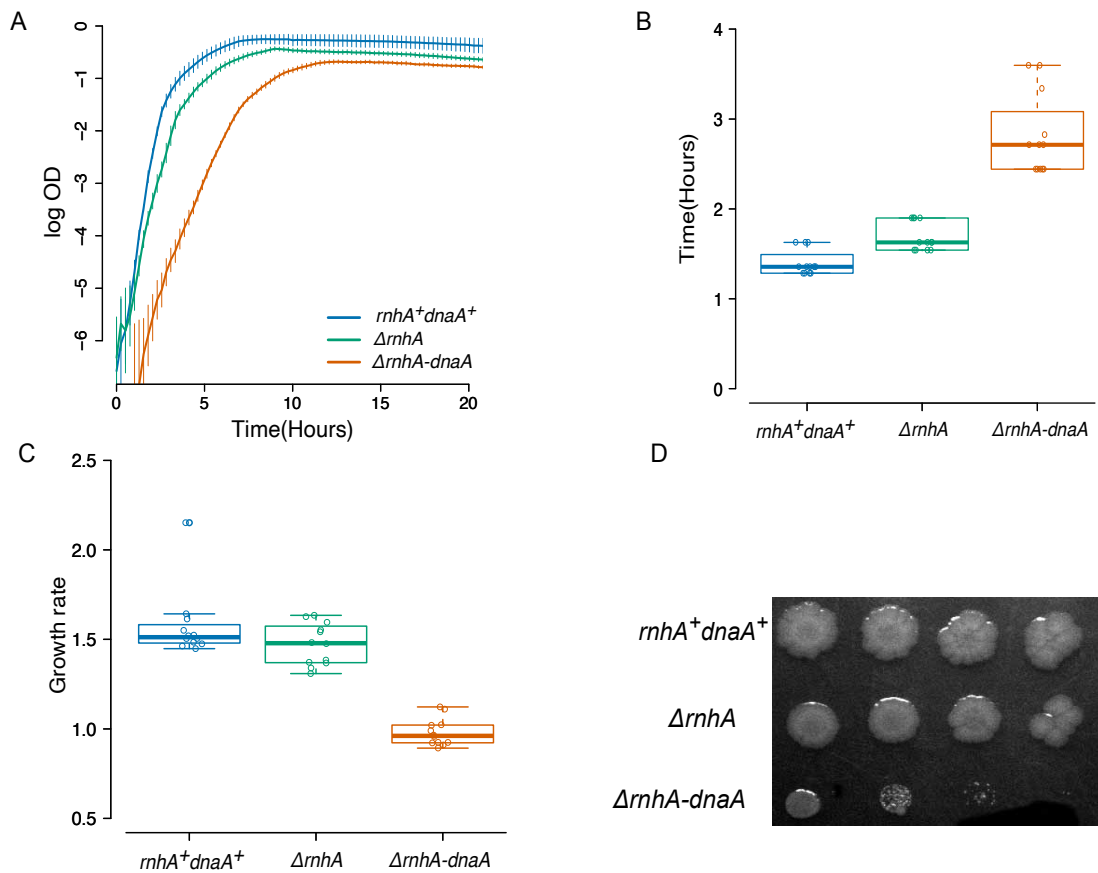


Figure 3.3: $\Delta rnhA-\Delta dnaA$ strain shows reduced growth in LB medium. (A) Growth curves of *rnhA⁺dnaA⁺*, $\Delta rnhA$, and $\Delta rnhA-\Delta dnaA$ strains in LB at 37°C and 200 rpm. The x axis indicates time, and the y axis indicates $\log_2 OD_{600}$. (B and C) Box plots for lag time and growth rate followed by each strain, respectively. The $\Delta rnhA-\Delta dnaA$ strain shows reduced growth rate and extended lag phase compared to the *rnhA⁺dnaA⁺* strain ($P_{\ll} -1$, Wilcoxon test, one-tailed). (D) Spotting assay for *rnhA⁺dnaA⁺*, $\Delta rnhA$, and $\Delta rnhA-\Delta dnaA$ strains using different dilutions of cultures (left to right, 10^{-3} , 10^{-4} , 10^{-5} , and 10^{-6}) in Luria agar plates incubated at 37°C

Figure 3.4

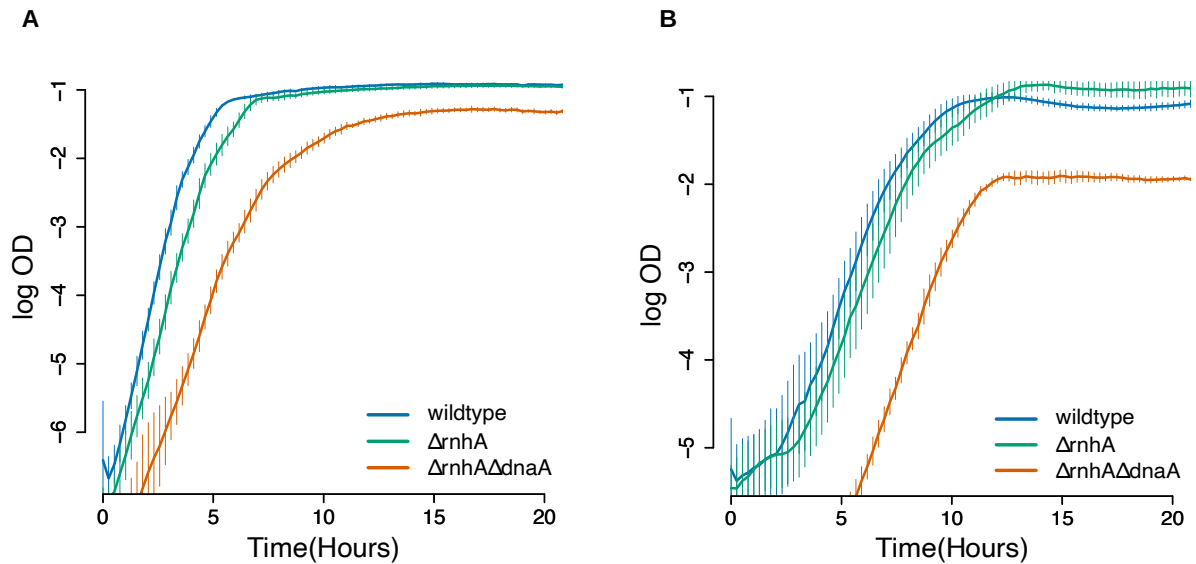


Figure 3.4: $\Delta rnhA\text{-}\Delta dnaA$ shows reduced growth in M9 minimal media conditions. (A) Growth curves of $rnhA^+dnaA^+$, $\Delta rnhA$, and $\Delta rnhA\text{-}\Delta dnaA$ in M9 media conditions supplemented with glucose and casaminoacids at 37°C, 200 rpm. X-axis indicates time and Y-axis indicates \log_2 OD₆₀₀ Values. (B) Growth curves of same strains in M9 minimal media supplemented with only glucose at 37°C, 200 rpm.

Figure 3.5

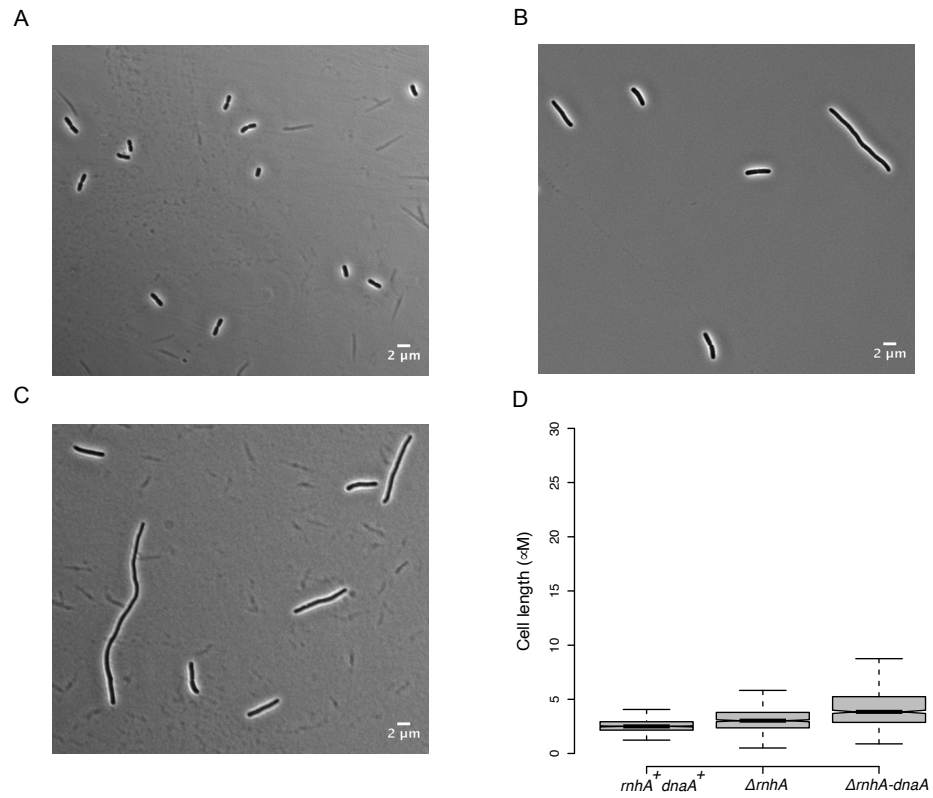


Figure 3.5: 60x phase contrast microscopy images of (A) *rnhA⁺dnaA⁺*, (B) $\Delta rnhA$ and (C) $\Delta rnhA-\Delta dnaA$ cells. (D) Box plot representing distributions of cell length for the same strains (calculated using Oufiti for ~ 600 cells). The cell lengths of both mutants $\Delta rnhA$ and $\Delta rnhA-\Delta dnaA$ are significantly greater than that of *rnhA⁺dnaA⁺* ($P \lll -1$, Wilcoxon test).

Figure 3.6

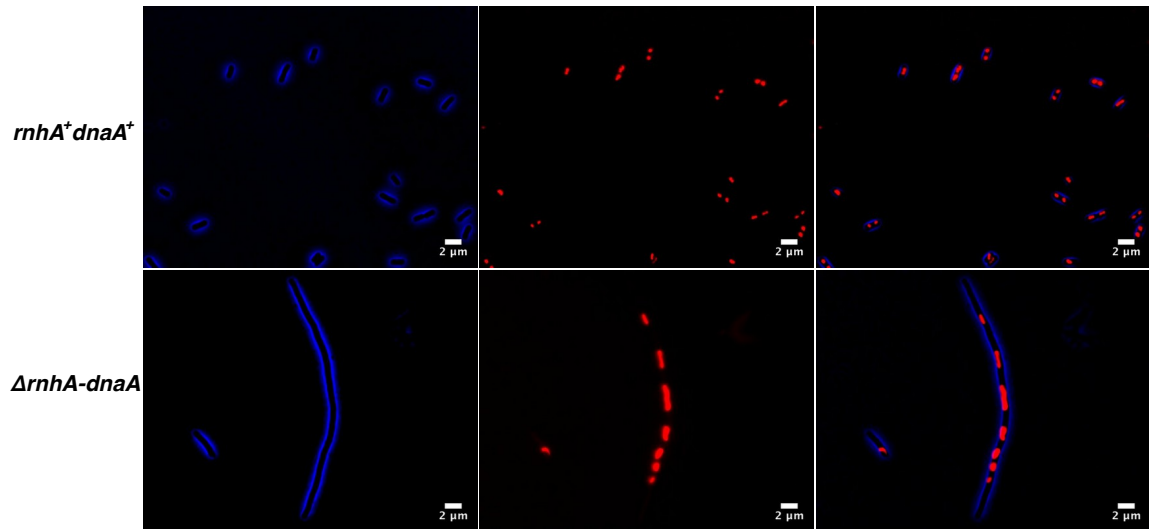


Figure 3.6: 60x phase contrast, DAPI staining, and the merged images of *rnhA⁺dnaA⁺* and $\Delta rnhA\text{-}\Delta dnaA$ cells.

3.23 *oriC* independent DNA replication leads to the disruption of gene gradient and activation of multiple *oriK* sites

The doubling time of *E. coli* in LB is 2-3 times less than the time required to replicate its chromosome. To account for this, chromosome replication initiates more than once per cell cycle. Thus, even in an unsynchronised population of normally growing and replicating *E. coli* cells, the copy number of *oriC* proximal regions is higher than that of *ter* proximal loci (16). A copy number gradient, decreasing smoothly from the origin towards the terminus, when averaged across an unsynchronised population, is established. The slope of this gradient is proportional to growth rate. Recent studies, have measured the copy number of different loci on the chromosome at high length resolution by subjecting genomic DNA isolated from exponentially growing cells to deep sequencing or next generation sequencing (NGS).

We isolated genomic DNA from *rnhA⁺dnaA⁺*, *ΔrnhA* and *ΔrnhA-ΔdnaA* strains of *E. coli* grown to exponential phase - corresponding to the culture's highest growth rate (μ_{\max}) - in LB. We sequenced the DNA libraries prepared from these samples to an average coverage of ~200x on the Illumina platform. As controls, we sequenced DNA isolated from stationary phase populations. For *rnhA⁺dnaA⁺*, we observed a copy number gradient decreasing from *oriC* towards *ter* as expected, symmetrically on either side of *oriC*, such that the number of reads mapping around *oriC* was ~ 2.3 fold higher than that around *ter* (Figure 3.7A). The corresponding plot for stationary phase cells was relatively flat (Figure 3.7A lower panel).

In *ΔrnhA*, in which both *oriC*-DnaA-dependent replication initiation and cSDR are active, we observed a prominent peak at *oriC* (Figure 3.7B). This peak declined smoothly in the counterclockwise direction towards *ter*. Immediately clockwise of *oriC* was a dip, followed by a sharp short rise to ~0.5 Mb clockwise of *oriC* and then a smooth decline towards *ter*. The gradient in copy number from *oriC* towards *ter* was only slightly less (*oriC:ter* ratio = 1.8) than that for *rnhA⁺dnaA⁺*. The plot for stationary phase cells was flat over most of the chromosome in comparison (Figure 3.7B lower panel). Within *ter*, we observed a sharp peak, which was retained at least qualitatively in stationary phase as well, suggesting that this peak is not fully a reflection of ongoing replication.

In line with its slow growth, *ΔrnhA-ΔdnaA* showed a flat curve with a few peaks which are candidates for *oriK* sites suggesting that the gene dosage is completely demolished (Figure 3.7C). The strongest peak in the exponential phase copy number plot for *ΔrnhA-ΔdnaA* was within *ter*, wherein the pattern observed was similar to that in *ΔrnhA* but more prominent (Figure 3.7C). The copy number declined smoothly clockwise of *ter*, reaching a trough at around *oriC*. We observed a sharp increase in copy number clockwise of *oriC*, reaching a crest at a distance of around 0.5-0.6 Mb which we call as *oriK45* for it being located at approximately 4.5 Mb into the genome sequence of *E. coli K12 MG1655*. The plot then remained flat clockwise till *ter*. The control stationary phase plot was flat except within *ter*. The overall profile was similar to that obtained by Maduiké *et al.* and by Dimude *et al.* (60, 64).

Figure 3.7

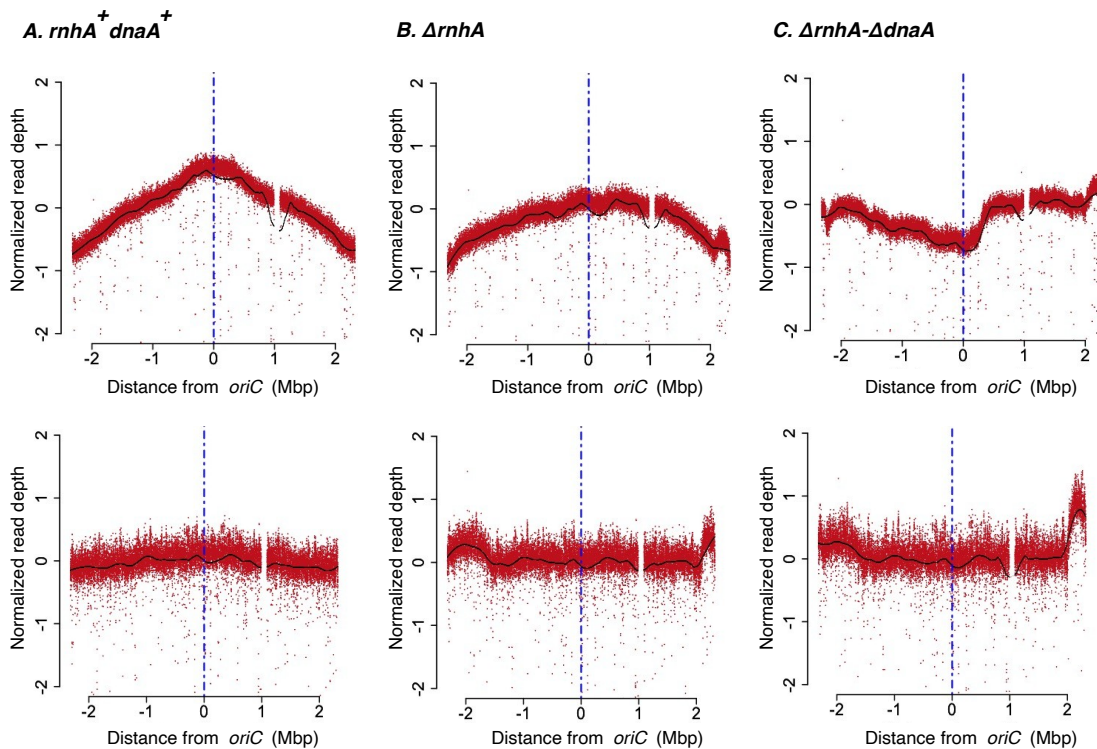


Figure 3.7: Deep sequencing based MFA plots for $rnhA^+dnaA^+$, $\Delta rnhA$ and $\Delta rnhA-\Delta dnaA$. The upper panels show the MFA plots for (A) $rnhA^+dnaA^+$, (B) $\Delta rnhA$ and (C) $\Delta rnhA-\Delta dnaA$ at the exponential phase of growth and the lower panels show the same for the stationary phase. The X-axis represents the distance of a locus either side of $oriC$ (in Mbp), with $oriC$ itself being the centre (blue vertical line). The Y-axis represents the \log_2 values of frequency of reads divided by the mode of the distribution of read counts (see chapter 2 Materials and methods).

Maduiké *et al.* had described a few peaks in their MFA data, which showed bumps in our data as well (Table 3.1 and Figure 3.8). The most prominent peak in the Maduiké *et al.* Dataset (64), as well as ours, was within *ter*. Using MFA analysis of $\Delta tus\text{-}\Delta rnhA\text{-}\Delta dnaA$, which abolished this peak, Maduiké *et al.* argued that the *ter* peak did not represent an *oriK* site, but the trapping of forks originating outside *ter* region (64). Our stationary phase data, which retains the *ter* peak, once again argues against this being an *oriK* site active only in growing cell populations (Figure 3.7C lower panel).

3.24 The effects of *oriC* independent DNA replication on gene expression states

What are the effects of cSDR on gene expression - as measured by global patterns along the length of the chromosome, and signatures on pathways related to DNA replication, repair and transcription? Towards answering this question, we performed exponential phase transcriptome analysis of $rnhA^+dnaA^+$, $\Delta rnhA$, and $\Delta rnhA\text{-}\Delta dnaA$ using RNA-seq.

Both $\Delta rnhA$ and $\Delta rnhA\text{-}\Delta dnaA$ induced large changes in gene expression when compared to $rnhA^+dnaA^+$. 600 genes were up-regulated and 543 down-regulated by a \log_2 fold change of 1.5 or above in $\Delta rnhA\text{-}\Delta dnaA$. The corresponding numbers for $\Delta rnhA$ are 472 and 360. Nearly 75% of all genes induced in $\Delta rnhA$ were also induced in $\Delta rnhA\text{-}\Delta dnaA$; the proportion for down-regulated genes being ~80%. Despite the overlap in these gene lists, the magnitude of differential expression was in general less in $\Delta rnhA$ than in $\Delta rnhA\text{-}\Delta dnaA$ ($P < 10^{-10}$, Wilcoxon signed rank test comparing magnitudes of differential expression). Functional classification of differentially expressed genes using Clusters of Orthologous Groups (COGs) showed an enriched up-regulation ($P < 0.01$, Fisher's Exact test) of various classes of genes including replication recombination and repair genes, ion transport and metabolic genes and translation genes. (Table 3.2). On the other hand, cell motility, energy production and conversion, carbohydrate transport and metabolism genes showed a significant down-regulation in $\Delta rnhA\text{-}\Delta dnaA$.

Genes encoding several members of the SOS response, including the cell division inhibitor *SulA*, error prone polymerases DinB and UmuC, RuvB and C are up-regulated in

Figure 3.8

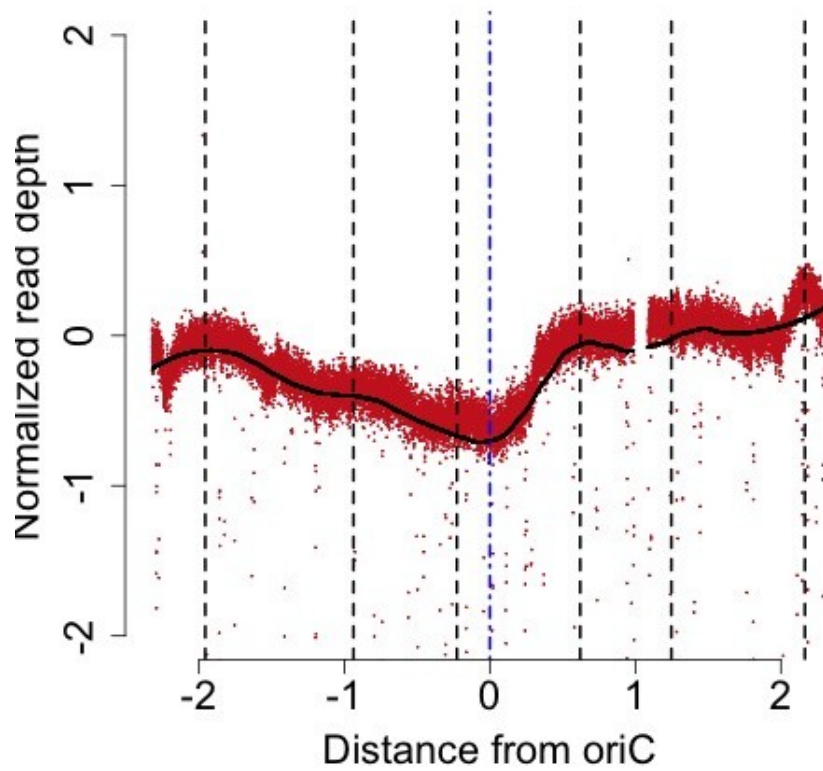


Figure 3.8: MFA plot of $\Delta rnhA\text{-}\Delta dnaA$ strain at exponential phase of growth. X-axis represents chromosome coordinates centered around *oriC* and Y-axis represents the frequency of reads normalized to the mode of the distribution. The dotted blue line represents *oriC* position and the dotted black lines represent predicted *oriK* positions.

both $\Delta rnhA$ and $\Delta rnhA-\Delta dnaA$. Upregulation of SulA, an inhibitor of FtsZ (a cell division protein forms a Z ring at the cell division site) and induction of other SOS response genes in $\Delta rnhA$ and $\Delta rnhA-\Delta dnaA$ explains the longer cell phenotype observed for these strains. *dinF*, the SOS inducible gene that also confers protection against oxidative stress was induced in both the mutants. Other signatures for an oxidative stress response included the induction of *sufB-E*, whose protein products are involved in iron-sulfur cluster biogenesis under oxidative stress (81). Very few members of the general stress response (~6%; under-represented when compared to Sigma70 targets, $P = 4 \times 10^{-6}$, Fisher's Exact Test), defined as targets of Sigma38 (RpoS), were induced.

We also observed an up-regulation of *holB* and *hold*, encoding the delta-prime and the epsilon subunits respectively of the replicative DNA polymerase III. This might in part be consistent with the SOS response, in light of the evidence that induction of SOS responsive DNA polymerases can be lethal in a genetic background that is defective for Hold (82). The gene *topA*, encoding topoisomerase, which can decrease R-loop formation presumably through its DNA relaxing activity, is also up-regulated.

Overall, there is a gradient - decreasing from *oriC* towards *ter* - in the fold change in gene expression between *rnhA⁺dnaA⁺* and $\Delta rnhA-\Delta dnaA$. In other words, genes that are proximal to *oriC* are more strongly down-regulated in $\Delta rnhA-\Delta dnaA$ when compared to *rnhA⁺dnaA⁺* (Figure 9A). Taken together, altered gene expression profiles of $\Delta rnhA-\Delta dnaA$ shows that the replicative structure of the chromosome largely plays a role in maintaining the gene expression homeostasis in *E.coli*.

Figure 3.9

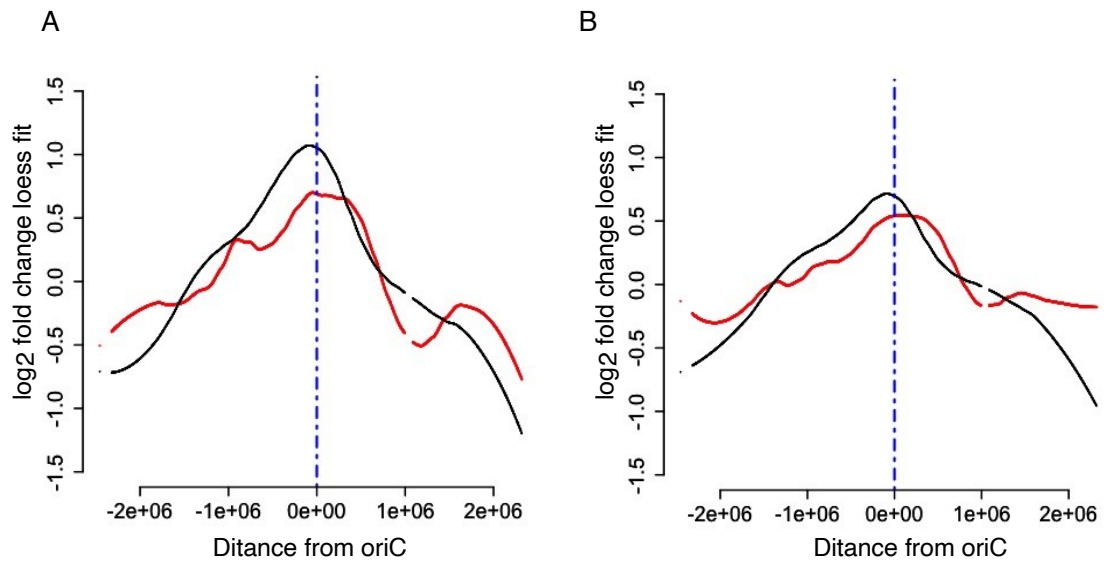


Figure 3.9: Effect of DNA copy number on gene expression: Plots showing the trend followed by the smoothed log₂fold change values for all genes in comparison with $\Delta rnhA-\Delta dnaA$ strain for gene expression and DNA copy number. (A) $rnhA^+dnaA^+/\Delta rnhA-\Delta dnaA$, (B) $\Delta rnhA/\Delta rnhA-\Delta dnaA$. X-axis represents positions centered around *oriC* and Y axis represents loess fit values of log₂ fold change. Red lines represent gene expression and black lines represent DNA copy number for the same strain.

3.3 Discussion

Single origin centric DNA replication is a highly evolved and optimised way of duplicating the bacterial genome. Understanding its importance in maintaining chromosomal architecture has been an open question for the last few decades. In the past, multiple strategies had been used to understand the importance of positioning of *oriC* on the chromosome with respect to gene expression. This includes placing *oriC* sequence at different positions on the chromosome (28, 83) as well as observing gene expression changes by introducing same gene copy at different locations on the chromosome (21, 22). Also, using comparative genomics approaches it has been shown that the positioning of highly expressed essential genes closer to origin is more conserved in fast-growing bacteria compared to slow growing ones (17, 18). In this study, we chose a unique strategy to understand the importance of replicative structure of the chromosome and this is done by inactivating the *oriC* based DNA replication system.

Inactivation of single origin based DNA replication system and allowing bacterial cells to initiate chromosomal replication from discrete regions on the chromosome capable of forming R-loops helped us to understand how important is the highly conserved replicative structure of *E.coli* chromosome in maintaining fast growth. Our results show that the loss of canonical DNA replication pathway alters growth and cell division in minimal and rich media conditions. We also observed a disruption of gene dosage effect in *E.coli* cells which also led to changes in gene expression profiles throughout the chromosome. This includes down regulation of genes encoded near *oriC* such as genes involved in ribosomal biogenesis and carbon metabolism. Therefore, these observations indicate that the perturbation of canonical replication system leads to drastic changes in growth and gene expression patterns in bacteria.

Identifying potential replication initiation sites in $\Delta rnhA-\Delta dnaA$ strain using marker frequency analysis was challenging due to poor gene gradient and slow growth phenotype. However, we could identify a few prominent peaks as potential replication initiation sites from the $\Delta rnhA-\Delta dnaA$ MFA plot. Most of these peaks were appeared

closer to the positions reported by Maduiké *et al.* (64) which suggests that cSDR may prefer specific sites on the chromosome as replication initiation loci and that are not randomly distributed on the chromosome.

The strongest peak in our study, as well as Maduiké *et al.* study (64), mapped within the terminus of replication. Nishitani and colleagues cloned and screened for fragments of the *E. coli* chromosome with potential for autonomous self-replication, and thereby identified a cluster of fragments again from within the terminus (65). However, both Maduiké *et al.* and Nishitani *et al.* appear to agree that the terminus sites identified in their studies are not bonafide *oriK* sites. In the Maduiké *et al.* study, these terminus signals disappeared in a *Δtus* background in which replication forks trapped within the terminus are released (64). The authors conclude that the terminus signal may represent trapping of forks originating from initiation sites elsewhere on the chromosome. Some of the *ter* sites identified by the Horiuchi group lost their activities in *Δtus*, but others did not (65). The Horiuchi group argued that increased copy number of fragments from the terminus can be attributed to homologous recombination based events and not autonomous replication. Gowrishankar has synthesised these arguments (66), and in conjunction with his lab's finding that RNA-DNA hybrids can occur throughout the chromosome (67), presented the case that cSDR can initiate anywhere on the chromosome; individual cells can initiate replication at different sites thus generating population-level heterogeneity; and these can well explain the prominent MFA signal within the terminus. In a recent paper, Brochu *et al.* argue that *ΔtopA-topB* (more so than *ΔtopA-rnhA*) cSDR cells show a strong copy number peak within the terminus suggesting an *oriK* site here, but do not evaluate it in a *Δtus* background (62). These authors however observe that the *ter* peak is maintained in a strain with a large inversion around the *ter*, arguing against this peak being merely a consequence of replication fork trapping events. This led us to ask whether the *ter* peak is a result of fork trapping or are there one or more *oriK* sites within *ter*?

Though the *ter* peak reported by Maduiké *et al.* and Dimude *et al.* (60, 64) disappeared in a *tus* mutant (as well as in a mutant carrying a small deletion in *tus* in our study), which accounts for fork trapping, this evidence may not fully eliminate the possibility of a

relatively weak *ter oriK*. The absence of a strong *ori-ter* gradient in these slow growing *tus* mutants may always cause such a peak to be missed. Though we observe that the *ter* peak is retained in our stationary phase cells, there is still the possibility that there is still some cSDR activity from a *ter* proximal *oriK* site in these cells. These observations suggest that the presence of a *ter* proximal *oriK* site can not be ruled out even though selection for this site as a prominent replication initiation site might come with the cost of dramatically upsetting the highly favourable copy number gradient declining from *oriC* towards *ter*.

Taken together, peak identification in high resolution MFA studies of cSDR as potential replication initiation sites is complicated by the slow growth phenotype of the strain, which results in weak origin to terminus copy number gradients. Using laboratory evolution experiments and selecting for suppressor mutants with higher growth rates will help to predict peaks with high confidence as well as to understand the evolutionary strategies accessed by bacteria to compensate for the loss of canonical replication system.

Table 3.1

Strain	Identified Peaks						
$\Delta rnhA-\Delta dnaA$	531400	1449000	1969800	2988600	-	3699200	4546800
$\Delta rnhA-\Delta dnaA$ (Maduiké <i>et al.</i> 2014)	790277	1481276	1869776	-	3228978	-	4538577
$\Delta rnhA$	-	1453600	-	-	-	-	4379400
$\Delta rnhA$ (Maduiké <i>et al.</i> 2014)	-	1471624	-	-	-	-	4253823

Table 3.1: identified peak positions and comparisons: All numbers are genome coordinates (bp). The *oriC* peaks, when present, are not included in this table. Genomic coordinates mentioned for Maduiké *et al.*, 2014 is normalized to the respective positions of peaks in *E.coli* K12 MG1655 genome version NC_000913.3.

Table 3.2

<i>Arn</i><i>hA</i> Downregulated genes					
COG categories	COG & DE	COG but !DE	!COG but DE	!COG&!DE	P-value
CELLULAR PROCESSES AND SIGNALING					
[D] Cell cycle control, cell division	1	32	338	3557	0.35
[M] Cell wall/ membrane/ envelope biogenesis	12	216	327	3373	0.06
[N] Cell motility	27	84	312	3505	3.94x10 ⁻⁷ *
[O] Post-translational modification	17	124	322	3465	0.1664
[T] Signal transduction mechanisms	14	164	325	3425	0.89
[U] Intracellular trafficking and secretion	12	115	327	3474	0.74
[V] Defense mechanisms	3	46	336	3543	0.79
INFORMATION STORAGE AND PROCESSING					
[J] Translation	8	176	331	3413	0.03
[K] Transcription	17	286	322	3303	0.05
[L] Replication, recombination and repair	6	199	333	3390	0.001
METABOLISM					
[C] Energy production and conversion	42	245	297	3344	0.0004*
[E] Amino acid transport and metabolism	34	328	305	3261	0.55
[F] Nucleotide transport and metabolism	8	90	331	3499	1

[G] Carbohydrate transport and metabolism	61	308	278	3281	2.48x10 ⁻⁷ *
[H] Coenzyme transport and metabolism	4	152	335	3437	0.003
[I] Lipid transport and metabolism	5	95	334	3494	0.27
[P] Inorganic ion transport and metabolism	12	203	327	3386	0.1
[Q] Secondary metabolites biosynthesis	4	60	335	3529	0.65
POORLY CHARACTERIZED					
[R] General function prediction only	29	372	310	3217	0.34
[S] Function unknown	23	294	316	3295	0.4

<i>Arn</i><i>hA</i> Upregulated genes					
COG categories	COG & DE	COG but !DE	!COG but DE	!COG&! DE	P-value
CELLULAR PROCESSES AND SIGNALING					
[D] Cell cycle control, cell division	3	30	345	3550	1
[M] Cell wall/ membrane/ envelope biogenesis	10	218	338	3362	0.011
[N] Cell motility	13	98	335	3482	0.3
[O] Post-translational modification	9	132	339	3448	0.36
[T] Signal transduction mechanisms	12	166	336	3414	0.34
[U] Intracellular trafficking and secretion	15	112	333	3468	0.26

[V] Defense mechanisms	1	48	347	3532	0.12
INFORMATION STORAGE AND PROCESSING					
[J] Translation	35	149	313	3431	1.244X10 ⁻⁵
[K] Transcription	32	271	316	3309	0.29
[L] Replication, recombination and repair	20	185	328	3395	0.61
METABOLISM					
[C] Energy production and conversion	41	246	307	3334	0.0016
[E] Amino acid transport and metabolism	32	330	316	3250	1
[F] Nucleotide transport and metabolism	9	89	339	3491	0.85
[G] Carbohydrate transport and metabolism	29	340	319	3240	0.56
[H] Coenzyme transport and metabolism	7	149	341	3431	0.059
[I] Lipid transport and metabolism	6	94	342	3486	0.37
[P] Inorganic ion transport and metabolism	27	188	321	3392	0.06
[Q] Secondary metabolites biosynthesis	6	58	342	3522	0.82
POORLY CHARACTERIZED					
[R] General function prediction only	28	373	320	3207	0.19
[S] Function unknown	13	304	335	3276	0,00092

<i>ArrhA-ΔdnaA</i> Down-regulated genes					
COG categories	COG & DE	COG but !DE	!COG but DE	!COG&! DE	P-value
CELLULAR PROCESSES AND SIGNALING					
[D] Cell cycle control, cell division	2	31	514	3381	0.3
[M] Cell wall/ membrane/ envelope biogenesis	17	211	499	3201	0.008
[N] Cell motility	29	82	487	3330	0.00016
[O] Post-translational modification	24	117	492	3295	0.16
[T] Signal transduction mechanisms	22	156	494	3256	0.8
[U] Intracellular trafficking and secretion	15	112	501	3300	0.78
[V] Defense mechanisms	8	41	508	3371	0.52
INFORMATION STORAGE AND PROCESSING					
[J] Translation	10	174	506	3238	0.0007
[K] Transcription	34	269	482	3143	0.33
[L] Replication, recombination and repair	14	191	502	3221	0.004
METABOLISM					
[C] Energy production and conversion	62	225	454	3187	3.71x10 ⁻⁵
[E] Amino acid transport and metabolism	46	316	470	3096	0.87
[F] Nucleotide transport and metabolism	11	87	505	3325	0.65
[G] Carbohydrate	82	287	434	3125	3.92x10 ⁻⁵

transport and metabolism					7
[H] Coenzyme transport and metabolism	10	146	506	3266	0.01
[I] Lipid transport and metabolism	8	92	508	3320	0.135
[P] Inorganic ion transport and metabolism	22	193	494	3219	0.2
[Q] Secondary metabolites biosynthesis	6	58	510	3354	0.45
POORLY CHARACTERIZED					
[R] General function prediction only	60	341	456	3071	0.27
[S] Function unknown	34	283	482	3129	0.19

<i>Arn</i>A-<i>Adna</i>A Up-regulated genes					
COG categories	COG & DE	COG but !DE	!COG but DE	!COG&! DE	P-value
CELLULAR PROCESSES AND SIGNALING					
[D] Cell cycle control, cell division	2	33	425	3468	0.58
[M] Cell wall/ membrane/ envelope biogenesis	23	205	404	3296	0.82
[N] Cell motility	10	101	417	3400	0.64
[O] Post-translational modification	12	129	415	3372	0.41
[T] Signal transduction mechanisms	18	160	409	3341	0.9
[U] Intracellular trafficking and secretion	11	116	416	3385	0.47
[V] Defense mechanisms	0	47	427	3454	0.007

INFORMATION STORAGE AND PROCESSING					
[J] Translation	45	139	382	3362	7.11E-08
[K] Transcription	31	272	396	3229	0.77
[L] Replication, recombination and repair	31	174	396	3327	0.04986
METABOLISM					
[C] Energy production and conversion	27	260	400	3241	0.49
[E] Amino acid transport and metabolism	50	312	377	3189	0.06
[F] Nucleotide transport and metabolism	14	84	413	3417	0.25
[G] Carbohydrate transport and metabolism	19	350	408	3151	9.697X10 ⁻⁵
[H] Coenzyme transport and metabolism	13	143	414	3358	0.7
[I] Lipid transport and metabolism	8	92	419	3409	0.41
[P] Inorganic ion transport and metabolism	39	176	388	3325	0.000981
[Q] Secondary metabolites biosynthesis	11	53	416	3448	0.1
POORLY CHARACTERIZED					
[R] General function prediction only	38	363	389	3138	0.39
[S] Function unknown	25	292	402	3209	0.08

Table 3.2: Functional classification of genes based on COG categories. Enrichment of a functional category is marked in yellow.

Chapter 4

***In vitro* Evolution Help Identify a Predominant Region of Constitutive Stable DNA Replication Initiation**

Note:

Parts of the text in this chapter is included in a manuscript *Veetil. et. al, mSphere 2020*

4.1 Introduction

In their natural niche, bacteria living within the environs of a host or habitat are often exposed to a variety of stresses such as starvation, temperature shifts, pH change, presence of antimicrobial agents and immune defences (84). The ability of microorganisms to respond to these fluctuations in their environment is an important physiological process that determines their capabilities to thrive in a variety of habitats (85). Such dynamic environments constantly select for genetic variants that are better adapted to the current conditions and thus leads to evolution.

In Bacteria, single origin based chromosomal replication process is highly conserved and considered as an optimal system to attain efficient duplication of the genetic material in minimal time. Many genome characteristics have been evolved around the single origin concept in bacteria which includes copy number effect and strand biases in gene encoding (16). Single origin dependent DNA replication being a highly optimised process of genome duplication in bacteria, lead us to ask how can bacteria compensate for the loss of canonical replicative structure of the chromosome. However, it is not clear how well a bacterium with altered replication structures would adapt to these perturbations and what would be the nature of these adaptations. To solve the puzzle, we used R-loop dependent, *oriC*-DnaA independent constitutive Stable DNA Replication (cSDR) system as a suboptimal DNA replication model of interest.

As described in Chapter 3, the deletion of *rnhA* and *dnaA* genes activate *oriC* independent, R-loop primed DNA replication in *E.coli* and as a result of this $\Delta rnhA$ - $\Delta dnaA$ strain exhibited a major growth defect along with an altered gene gradient and expression pattern. The important question to ask is how do bacteria revert these fitness defects to compensate for the loss of single origin based canonical DNA replication? In this chapter, we address this using laboratory evolution experiments of $\Delta rnhA$ - $\Delta dnaA$ strain of *E.coli*, generating suppressor mutants that can redress the growth. The nature of mutations that result in these adaptations is a question of primary interest. We anticipate

these to involve large chromosomal structural variations which can re-establish DNA copy number patterns, therefore resulting in restoring gene expression balance.

It is also important to understand where replication initiates in a $\Delta rnhA$ - $\Delta dnaA$ strain and does evolution select for any specific region on the chromosome as a preferred replication initiation site. As we mentioned in chapter 1, the attempt to identify potential *oriK* sites from $\Delta rnhA$ - $\Delta dnaA$ using high resolution MFA studies is challenging due to the slow growth of the parent strain. Therefore, conducting laboratory evolution experiments to select for genetic variants that can generate strong copy number gradients even under the cSDR regime will help us to identify genuine replication initiation sites on the chromosome.

Together, performing an *in vitro* evolution experiment on $\Delta rnhA$ - $\Delta dnaA$ strain of *E.coli* and generating suppressor mutants will help us to identify genetic changes and adaptive strategies chosen by bacteria to compensate for the loss of a highly evolved replicative structure of the chromosome.

4.2 Results

4.21 Laboratory evolution experiments of *ΔrnhA-ΔdnaA*

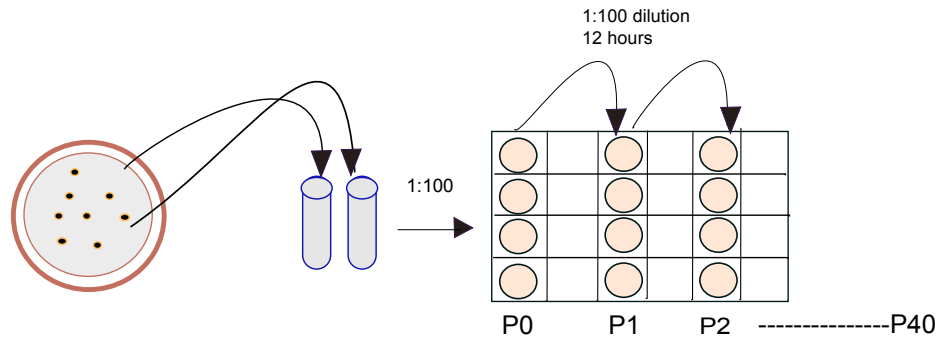
To obtain cSDR strains that grow fast and therefore display strong *ori-ter* gradients, we performed laboratory evolution experiments in which *ΔrnhA-ΔdnaA* was iteratively diluted into fresh LB and grown to saturation. We used a fast growing media condition such as LB in this case to select for suppressors at a faster rate compared to slow growing minimal media conditions used in previous studies of cSDR. Higher growth rates also produce stronger *ori-ter* gradients, which enable better *oriK* peak identification. We used 8 independent lines, each derived from a single *ΔrnhA-ΔdnaA* colony to 36 rounds of dilution and growth, corresponding to an estimated 288 generations (Schematic representation of evolution Figure 4.1A).

We plated aliquots of culture after each day and noticed the presence of colonies that were visibly larger than those of the parent *ΔrnhA-ΔdnaA* strain. From these results we observed that independently evolved populations showed an increase in colony size compared to their ancestral population (Figure 4.1B and Figure 4.2). Over time, the growth of the population substantially improved (Figure 4.3A). Calculation of maximum growth rate for these populations showed a significant increase compared to the ancestral strains (Figure 4.3B). Increase in growth also followed by a subsequent reduction of lag phase which can be a result of rapid adaptation of suppressor mutants (Figure 4.3C).

Comparing the cell length across evolved populations showed a substantial reduction in the number of longer cells during evolution (Figure 4.4A). These observations suggests that the *in vitro* evolution of *ΔrnhA-ΔdnaA* strain resulted in generation of suppressor mutants that can sustain higher growth rate in a non *oriC*-DnaA replication condition.

Figure 4.1

A



B

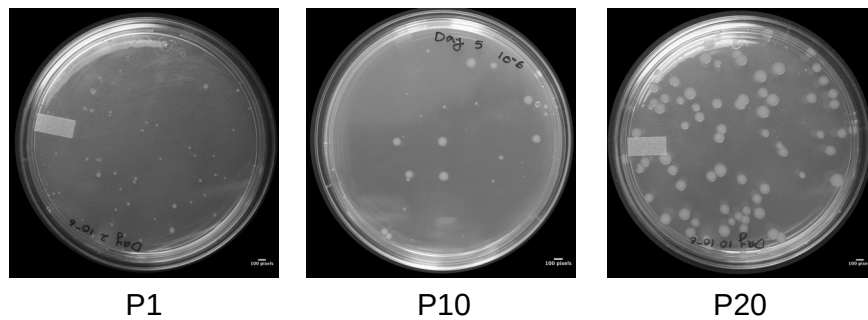


Figure 4.1: Laboratory evolution of *ΔrnhA-ΔdnaA* mutant. (A) Schematic representation of laboratory evolution experiment (P represents passage). (B) Images of LB agar plates at 10⁻⁶ dilution of cultures from different passages of evolution experiment after 30 hours of incubation at 37°C

Figure 4.2

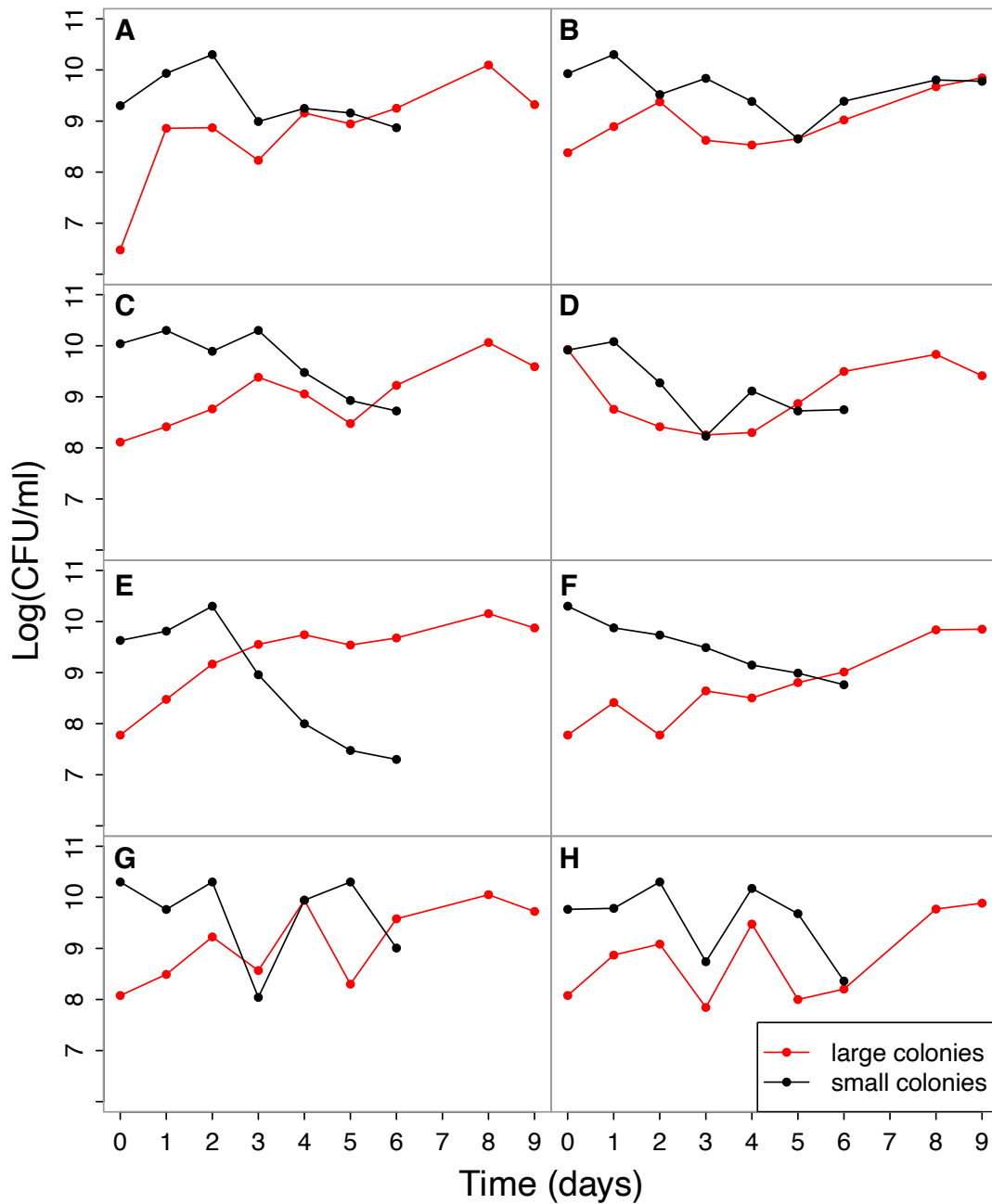


Figure 4.2: Colony count during the evolution for independent lines. Each panel represents the CFU/ml for large and small colonies counted on different days of evolution experiment for independently evolved lines.

Figure 4.3

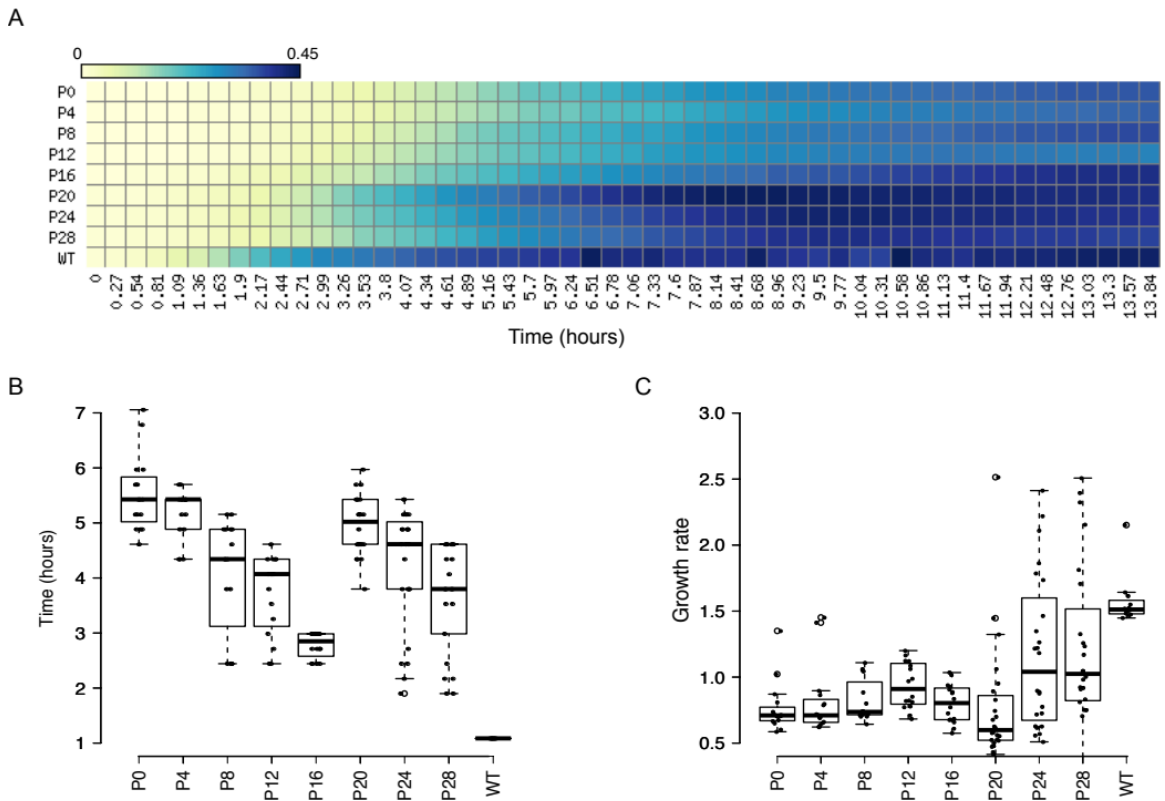


Figure 4.3: Growth characteristics of evolved mutants: (A) Heat map representing growth of an independently evolved population of $\Delta rnhA-\Delta dnaA$ from passage 0 (P0) to passage 28 (P28) based on OD measurements. X-axis shows time in hours, Y-axis shows the number of passage and the colours represent mean OD values as indicated in the colour bar. (B) and (C) box plots for lag time and growth rate followed by all independently evolved populations respectively. Passage 28 population shows a significantly greater growth rate than that of parental (P0) strains ($P \ll -1$, Wilcoxon test, one tailed).

4.22 Sequencing of suppressor mutants helps to find adaptive strategies

To identify the genetic variations in suppressor mutants, we randomly picked 60 colonies of varying sizes - sampling across 3 independently evolved populations and 5 time-points (P0, P4, P8, P16, P24, P30) including the zeroth day populations - and subjected their genomic DNA to Illumina sequencing. Similar to our sequencing runs with the parent *ΔrnhA-ΔdnaA*, we sequenced DNA isolated from mid-exponential phase. Stationary phase DNA sequencing was performed for a select few colonies based on genotypes identified from exponential phase DNA sequencing.

For all these strains, we calculated the ratio between the maxima and the minima of the mid-exponential phase copy number graphs (see chapter 2 materials and methods for more details), and found that this ratio ranged between 0.86 and 2.8 (Table 4.1). At the lower end, a few colonies showed gradients not too different from the *ΔrnhA-ΔdnaA* parent. The steepest gradients approached, but rarely matched that of *rnhA⁺dnaA⁺*.

We next used these sequencing data to identify mutations – both point variations and indels. We found several point mutations in the evolved clones, not present in the *ΔrnhA-ΔdnaA* parent (Figure 4.5). ~90% of colonies carried a mutation upstream of one of two rRNA operons, *rrnD* and *rrnC*. These mutations may have a role in regulating the transcription of these *rrn* operons since those genomic changes are located at the regulatory elements of the operon. Apart from this, most of these mutations were annotated to those proteins which are directly or indirectly involved in cSDR mechanism. One clone carried an in-frame deletion mutation in *tus* ($\Delta 6$ bp (1,684,458-1,684,463), which translates to a QSL-L variation in the amino acid sequence and lack the ter peak and exhibited similar trends to previously reported copy number plot of a *Tus* deletion strain (60) (Figure 4.4B).

We did not find any amplification, and the only deletion that was apparent in the data was an ~97 kb (*[mmuP]-[mhpD]*) deletion around the *lac* locus, which is part of the genotype

of the *rnhA*⁺*dnaA*⁺ founder strain used in this study. Apart from this, many SNPs were detected in the coding regions of various pseudogenes when mapped to the reference genome.

Figure 4.4

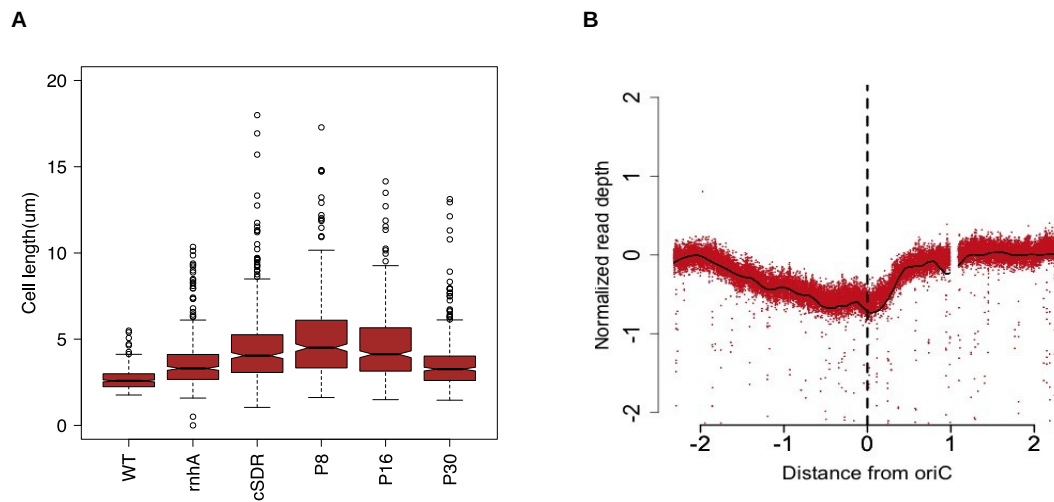


Figure 4.4: Suppressor mutants show reduced cell length A) Box plot representing distributions of cell length for strains from different passages of evolution (calculated using Outfi for ~ 600 cells). (B) MFA plot of *tus* ($\Delta 6$ bp) mutant (1D44). Black dotted line represent *oriC* position.

Figure 4.5

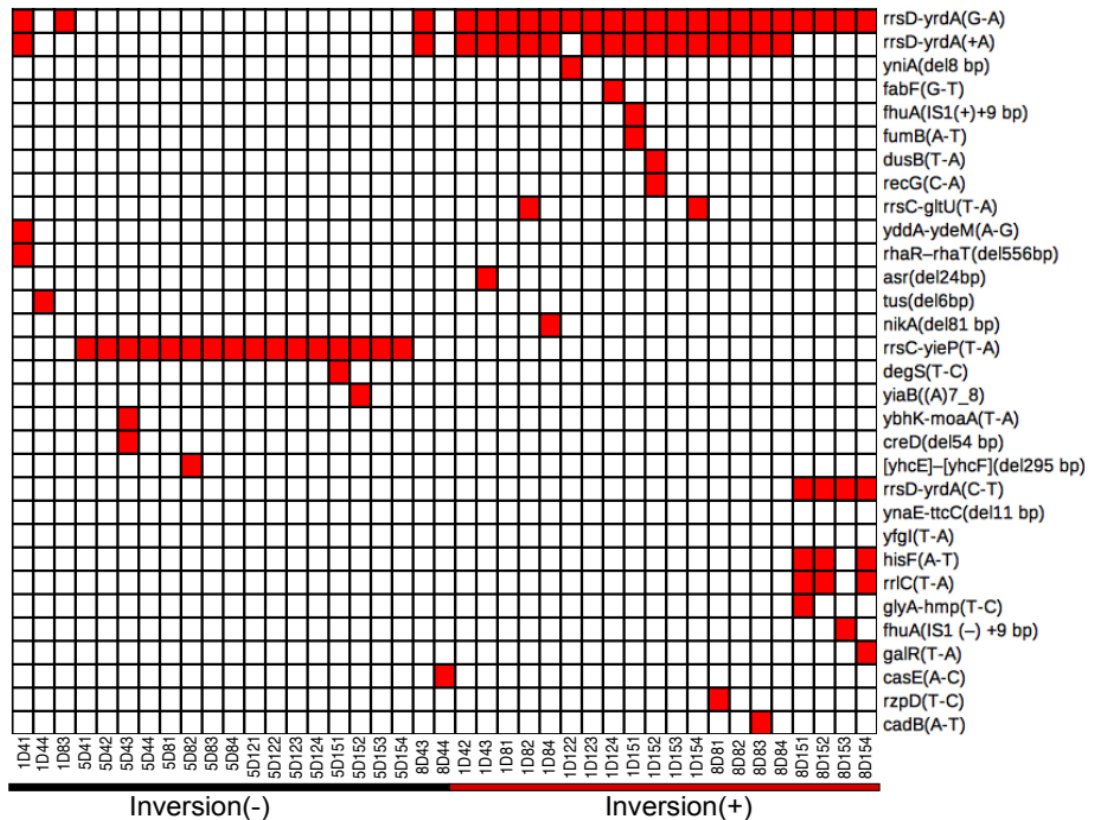


Figure 4.5: Unique mutations in suppressor mutants. Heat map representing unique mutations (100% frequency) in all independent colonies sequenced generated using matrix2png. Colour represents the presence of a mutation in the respective gene shown on Y-axis. X-axis represents sample IDs of suppressor mutants evolved from three independent populations. Presence and absence of chromosomal inversions are represented using a red and black lines respectively.

4.23 Large chromosomal inversions around *oriC* suppress the growth defect of Δ *rnhA*- Δ *dnaA*

Genomic data helps to identify chromosomal structural variations such as Large amplifications, deletions and inversions. Amplifications and deletions can be identified by sharp local increases or decreases respectively in copy number. Inversions can be detected as local flips in copy number plots of exponential phase genomic DNA sequencing data with clear *ori-ter* gradients (86).

We found chromosomal inversions around *oriC* in ~45% of the evolved colonies (Figure 4.5 and Figure 4.6). One end of these inversion was *rrnD*, located 3.42 Mb counter-clockwise of *oriC* in the reference genome of *E. coli* K12 MG1655. In ~80% of inversions, the other end was *rrnC* (3.94 Mb), (Figure 4.6A and 4.6D) and in the remaining, the second end was *rrnE* (4.2Mb) (Figure 4.6B and 4.6E). The *rrnD-rrnC* inversion (Δ *rnhA*- Δ *dnaA*-*inv*^{*rrnD-rrnC*}) measured ~0.5 Mb and the *rrnD-rrnE* (Δ *rnhA*- Δ *dnaA*-*inv*^{*rrnD-rrnE*}), ~0.8 Mb. We used long read nanopore sequencing to assemble the genome of the clone with the longer *rrnD-rrnE* inversion into just one contig *de novo*, and confirmed the presence of the inversion (Figure 4.7A).

Thus both inversions would move a set of rRNA operons from clockwise to counter-clockwise of *oriC*, and the *rrnD* operon in the opposite direction. Colonies with either inversion in the present study also carried the following mutations upstream of *rrnD*: (a) G-A (position 3,429,052) and +A (3,429,054) or (b) C-T (3,429,055) (Figure 4.5). Irrespective of the presence of the inversion, all these rRNA operons would continue to lie on the leading strand of canonical replication from *oriC*. That the fitness cost of these inversions would be minimal under conditions of normal DNA replication is also suggested by the fact that inversions bounded by at least one *oriC*-proximal rRNA operon are found in 37 other *E. coli* genomes (out of 675 considered), including another strain of *E. coli* K12 (W3110) (Figure 4.7B and Table 4.2).

Figure 4.6

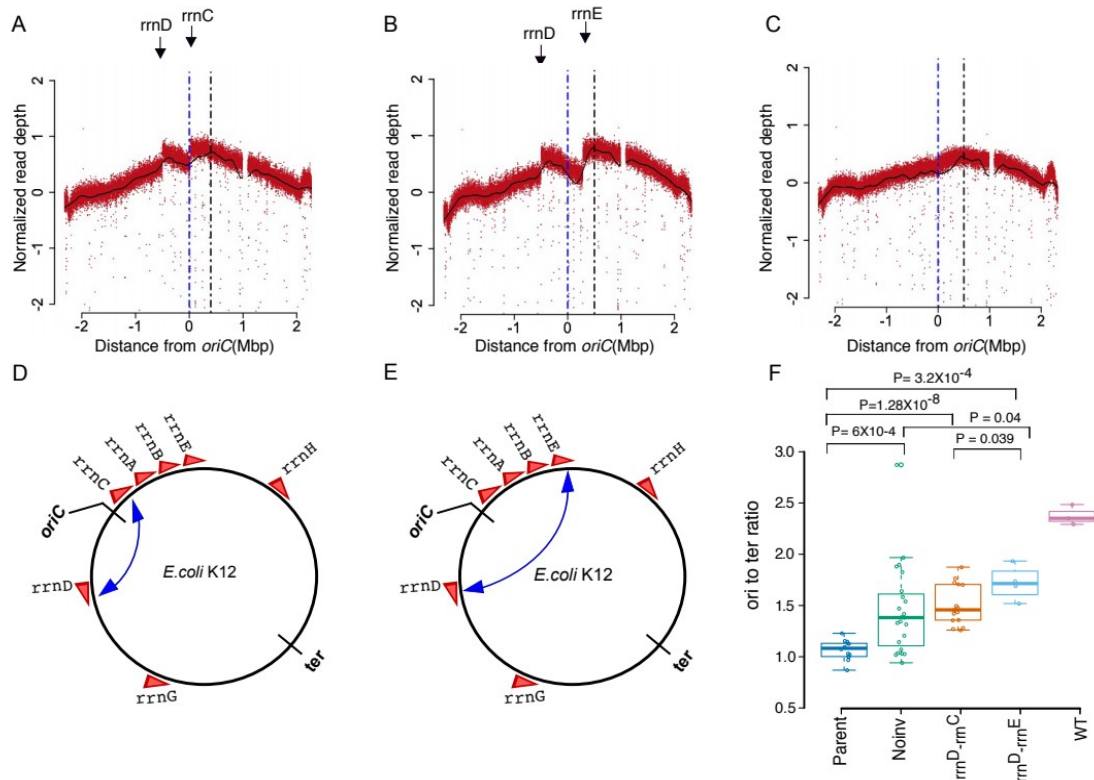


Figure 4.6: Deep Sequencing based MFA plots for suppressor mutants. (A), (B) and (C) represents MFA plots for (A) $\Delta rnhA-\Delta dnaAinv^{rrnD-rrnC}$, (B) $\Delta rnhA-\Delta dnaAinv^{rrnD-rrnE}$, (C) $\Delta rnhA-\Delta dnaA$ Noinv sequenced at the exponential phase of growth. The dotted blue line represents the *oriC* position and the black line represents the position at maxima of Loess fit value. (A) and (B) plots show the presence of different chromosomal inversions flanked by *rrn* operons (mentioned above) and the position of inversion on the chromosome has been schematically represented here (D and E). (F) box plot representing *ori*-to-*ter* ratio differences in different populations of evolved clones compared to wild type *E. coli*. (X-axis labels; Parent- $\Delta rnhA-\Delta dnaA$ strain passage 0 clones, No inv: suppressor mutants which do not show the presence of chromosomal inversion, *rrnD-rrnC*: Clones which shows presence of a chromosomal inversion from *rrnD-rrnC* ($\Delta rnhA-\Delta dnaAinv^{rrnD-rrnC}$) and *rrnD-rrnE*: Clones which shows presence of a chromosomal inversion from *rrnD-rrnE* ($\Delta rnhA-\Delta dnaAinv^{rrnD-rrnE}$).

Does chromosomal rearrangements help to restore gene dosage? We then compared the *maximum-minimum* ratios in the copy number plots of clones (not considering the peak within *ter*) with the two types of inversions and those without. For this analysis, we grouped all colonies without an inversion together, fully aware that this is a genetically heterogeneous group. Clones with the longer *rrnE-rrnD* inversion showed significantly higher maximum-minimum ratios than those with the shorter *rrnC-rrnD* inversion ($P = 0.039$, Wilcoxon test one-tailed) (Figure 4.6F). Therefore, the longer inversion appears to be a better suppressor of the growth defect of cSDR than the shorter inversion. Many clones without the inversions, including the one with the $\Delta 6$ bp inframe deletion mutation in *Tus*, showed substantially smaller maximum-minimum ratios, though a few colonies did show higher values.

4.24 *oriK45* as a preferred initiation site for cSDR in suppressors

Further, we identified the locations of the maxima of the copy number curve for the suppressors, while ignoring the *ter* peak. We noticed that these, across all cSDR strains used in this study, mapped to ~ 4.3 Mb to 4.6 Mb clockwise of *oriC*, in proximity to *oriK45* mentioned in chapter 1. Consistent with this, all suppressors showed a copy number peak at *oriK45* (Figure 4.8A, Table 4.3). This suggests that *oriK45* is a predominant site of cSDR initiation in *all* suppressors identified here.

In the strongest suppressors, we observed a strong copy number gradient peaking at *oriK45* and declining towards *ter* (figure 4.6A and 4.6B). Prediction of a single predominant *oriK* site in independently evolved clones suggest the act of a strong selection for this origin during the evolution of mutants. Also, this led to the restoration of a single origin replicative system in strongest suppressors due to the positioning of *oriK45* being very close to the *oriC*; such that the left and right replichores can travel similar distances to complete a round of DNA replication.

Figure 4.7

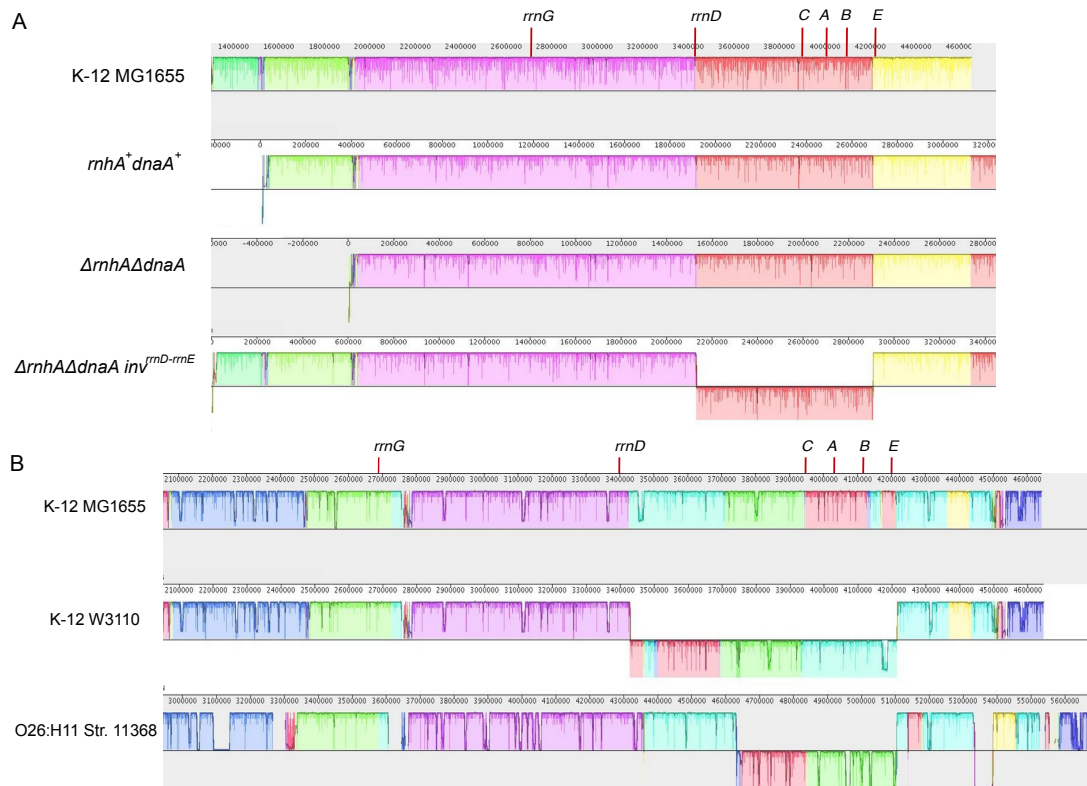


Figure 4.7: chromosomal inversions around *oriC* (A) Plot showing sequence alignment of denovo assembled contigs obtained from Nanopore sequencing for *rnhA⁺dnaA⁺*, $\Delta rnhA\text{-}\Delta dnaA$ and $\Delta rnhA\text{-}\Delta dnaA\text{ inv}^{rnhD\text{-}rnhE}$ strains with respect to *E. coli* K12 MG1655 genome. Strand shift of red color bar for the $\Delta rnhA\text{-}\Delta dnaA\text{ inv}^{rnhD\text{-}rnhE}$ shows the presence of ~ 0.8 mb chromosomal inversion. Position of *rnh* operons near *ori* region is marked. (B) Figure representing sequence alignment (using Mauve) of selected region of chromosome for three *E. coli* strains. *E. coli* K12 W3110 and *E. coli* O26:H11 str.11368 shows the presence of a chromosomal inversion around *oriC* region compared to the reference *E. coli* K12 MG1655. Strand shift for the colored bars represents chromosomal inversion. Position of *rnh* operons near *ori* region is marked.

4.25 *oriK45* is proximal to predicted R-loop forming sites

We asked whether *oriK45* is proximal to regions with high propensities to form RNA-DNA hybrids. We used a computational technique that searches for two G-rich patterns on a given DNA sequence to identify loci that have the propensity to form RNA-DNA hybrids (75, 76). This method predicted ~30 R-loop favouring sites, showing homology to at least one of the two RNA-DNA hybrid-forming sequence patterns, across the *E. coli* chromosome (Figure 4.8B). 8 of the 11 copy number bumps described by us or by Maduiké et al. for $\Delta rnhA\text{-}\Delta dnaA$ were within 200 kb of at least one of the predicted sites. This is statistically significant compared to random assignment of genome coordinates to experimentally predicted copy number peaks ($P = 10^{-5}$, Z-score, permutation test across 1,000 repetitions, one-tailed). However, only one site showed homology to both RNA-DNA hybrid-forming sequence patterns; this site is at 4.51 Mb (Figure 4.8C), within the range defined by *oriK45*.

Krishna Leela et al. (67) had identified bisulfite sensitive regions of *E. coli* chromosome and defined these as preformed R-loops. However, we did not find any statistically significant overlap of these sites with *oriK45*. Nevertheless, we found two clusters of highly bisulfite-sensitive genes in the *oriK45* region, and also observed that the R-loop forming sequence mentioned above was also highly bisulfite sensitive.

4.26 Genetic evidence for *oriK45* containing sites of cSDR initiation

Nishitani et al., while screening for genomic DNA fragments capable of autonomous replication, describe a site called *hotH*, which is at 4.55-4.56 Mb (65). However, to our knowledge, these authors did not report further exploration of the *hotH* site and focussed instead on the characterisation of the cluster of fragments from within *ter*. Among the

transposon insertions found to affect replication of $\Delta topA$ -mediated cSDR is an insertion within *fimD*, which is again in the region defined by *oriK45* (87).

Figure 4.8

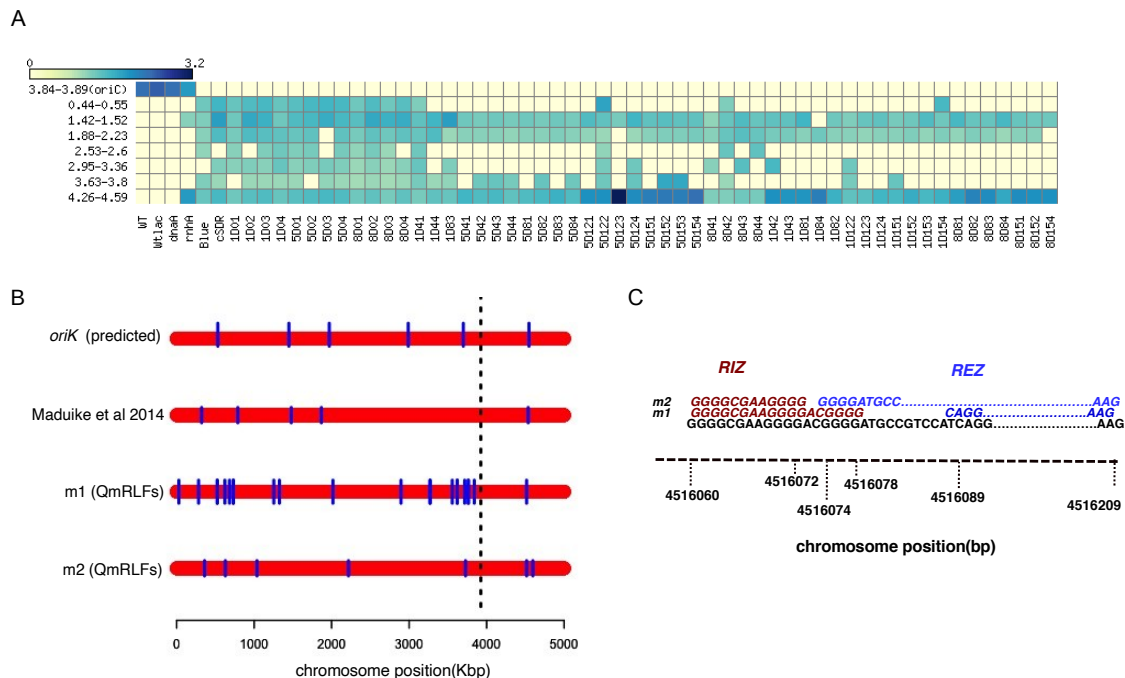


Figure 4.8: *oriK45* as a preferred initiation site for cSDR in suppressor mutants: (A) heatmap showing predicted *oriK* position ranges from marker frequency analysis across evolved strains. Y- axis represents the chromosomal positions of predicted *oriK* ranges in Mbp. Colour indicates the ratio of corresponding LOESS smoothed normalized read count to the LOESS minima of the plot at the peak. (B) plot showing positions of R-loops predicted by m1 and m2 model of QmRLFs on *E. coli* chromosome in comparison with position of predicted *oriK* sites in $\Delta rnhA\Delta dnaA$ strain from Maduiké et al. 2014 and the *oriK* sites for the same strain mentioned in this study. Each red bar represents the bacterial chromosome on which the R-loops positions are marked in blue lines. (C) Represents the sequence motif of R-loop initiation zone (RIZ) and R-loop elongation zone (REZ) predicted by m1 and m2 models of QmRLFs around 4.51mb region of *E. coli* K12 MG1655 chromosome.

To test whether *oriK45* affects the growth of $\Delta rnhA\text{-}\Delta dnaA$, we constructed a $\Delta 11.3\text{Kb}$ region (4555284: 45660615, *uxuR-yjiN*), corresponding to the restriction fragment defined by Nishitani et al. as *hotH*, in the $\Delta rnhA\text{-}\Delta dnaA\text{-pHYD2388}$ (*dnaA⁺lacZ⁺*) background (24). We measured the rate at which $\Delta rnhA\text{-}\Delta dnaA\text{-pHYD2388}$ (*dnaA⁺lacZ⁺*) and $\Delta rnhA\text{-}\Delta dnaA\text{-}\Delta hotH\text{-pHYD2388}$ (*dnaA⁺lacZ⁺*) lost the pHYD2388 plasmid. This we interpret as a measure of selection in favour of maintaining the plasmid-borne *dnaA* copy. $\sim 12\%$ of $\Delta rnhA\text{-}\Delta dnaA\text{-}\Delta hotH\text{-pHYD2388}$ (*dnaA⁺lacZ⁺*) lost the plasmid, compared to $\sim 28\%$ for the corresponding *hotH⁺* variant. This difference was statistically significant ($P = 8 \times 10^{-5}$, Wilcoxon test one-tailed) (Figure 4.9A).

The fact that $\Delta hotH$ was not lethal to $\Delta rnhA\text{-}\Delta dnaA$ suggests that replication initiation might proceed from other sites. We performed an MFA analysis of $\Delta rnhA\text{-}\Delta dnaA\text{-}\Delta hotH$ grown to mid-exponential phase and found that the peak within *oriK45* was still prominent (Figure 4.9B). This indicates that cSDR proceeds from *oriK45* despite the absence of *hotH*. Across the MFA data that we have obtained for $\Delta rnhA\text{-}\Delta dnaA$ and its suppressors, the maxima ranges from $\sim 4.26 - 4.59$ Mb on the chromosome. The top match to R-loop-forming sequences is at ~ 4.51 Mb (see previous section); *hotH* is $\sim 40\text{-}50$ kb clockwise of the above R-loop-forming sequence; *fimD*, an insertion in which has an effect on *AtopA*-mediated cSDR (87), is located between the above two sites and is closer to *hotH*. These three sites, while located in the broad region that defines *oriK45*, do not overlap. An explanation would be that *hotH* does not contain an initiation site, but improves the efficiency replication initiation from the R-loop forming sequence motif in its proximity. This however is not consistent with the findings of Nishitani et al. (65) showing that *hotH* might be capable of autonomous replication. An alternative explanation would be that *oriK45* has multiple initiation points at the same region. To test this, we deleted the chromosome region of potential R-loop forming sequence at 4.51Mbp as well as *fimD* gene independently in $\Delta rnhA\text{-}\Delta dnaA\text{-pHYD2388}$ (*dnaA⁺lacZ⁺*) background. As described earlier in this section, We measured the rate at which these strains lost the pHYD2388 plasmid. We could not detect any major changes in the

Figure 4.9

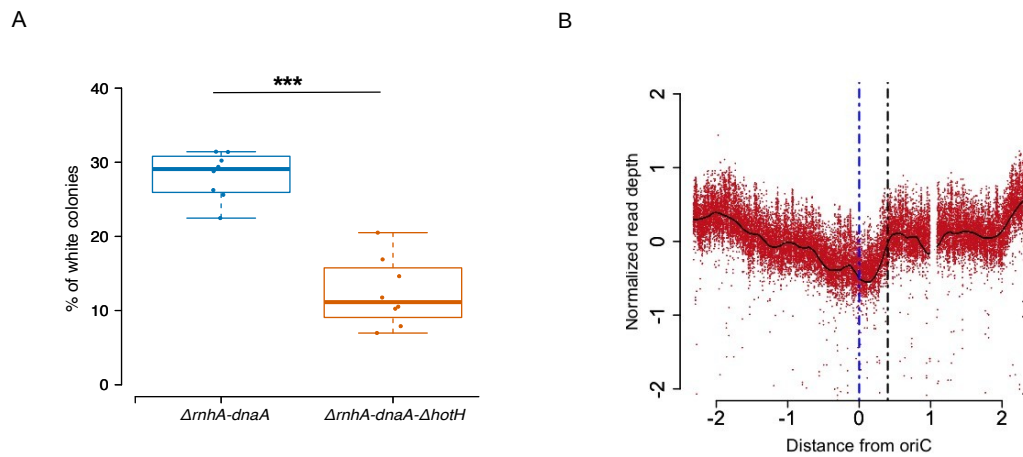


Figure 4.9: effect of *oriK45* deletion: (A) Box plot comparing the percentage of white colonies of $\Delta rnhA-dnaA$ and $\Delta rnhA-dnaA-\Delta hotH$ strains respectively (n=8). (B) Represents MFA plot $\Delta rnhA-dnaA-\Delta hotH$ strain.

percentage of white colonies produced by both mutants compared to *ΔrnhA-ΔdnaA*-pHYD2388 which again suggests that there will be multiple initiation points at the same region or this entire region increases the propensity to form R-loops at different positions.

4.27 The effects of chromosomal rearrangements on gene expression states

Large chromosomal rearrangements play an important role in long term evolution of organisms by altering gene expression and chromosome organization. These chromosomal rearrangements include events such as duplications, amplifications, inversions, deletions, and translocations (88).

We observed the presence of large chromosomal inversions in suppressor mutants isolated from two independently evolved populations in which the primary replication initiation site is *oriK45*. These mutants with chromosomal inversions showed significantly higher *ori-to-ter* ratios than those without inversion. This led us to ask to what extent the growth defects of cSDR are suppressed by the inversion around *oriC*. To answer this, we performed exponential phase transcriptome analysis of a few evolved suppressor mutants (*ΔrnhA-ΔdnaAinv^{rrnD-rrnC}* and *ΔrnhA-ΔdnaAinv^{rrnD-rrnE}*) using RNA-sequencing.

A number of genes were differentially expressed in both inversion mutants compared to *rnhA⁺dnaA⁺*. 451 genes were up-regulated and 341 down-regulated by a log (base 2) fold change of 1.5 or above in longer inversion strain *ΔrnhA-ΔdnaAinv^{rrnD-rrnE}* where as the corresponding numbers for the shorter inversion strain *ΔrnhA-ΔdnaAinv^{rrnD-rrnC}* was 513 and 456. Genes encoding SOS response shows an up-regulation in inversion mutants as similar to *ΔrnhA* and *ΔrnhA-dnaA*, whereas iron-sulfur cluster biogenesis genes shows no change in gene expression. However, we observe that there is a reduction in the number of differentially expressed genes in inversion mutants compared to their parental strain *ΔrnhA-ΔdnaA*.

Overall, the transcriptome profile of inversion mutants are more similar to the *ΔrnhA* single mutant compared to *ΔrnhA-ΔdnaA* mutant (Figure 4.10). These results show that

large chromosomal inversions help to restore the gene expression pattern to some extent in cSDR conditions.

4.28 Gene expression changes show limited but significant correlation to DNA copy number changes

Next, we compared the transcriptome profile of inversion mutants with respect to $\Delta rnhA-\Delta dnaA$ strain. At this level, the fold change in $rnhA^+dnaA^+$, in relation to $\Delta rnhA-\Delta dnaA$, shows strong similarity to that in $\Delta rnhA$ and $\Delta rnhA-\Delta dnaA-inv^{rrnD-rrnE}$ (Pearson correlation coefficient = 0.64 for both comparisons), and slightly less similar to $\Delta rnhA-\Delta dnaA-inv^{rrnD-rrnC}$ (Pearson correlation coefficient = 0.55) (Figure 4.11). These indicate that a portion of the gene expression change in $\Delta rnhA-\Delta dnaA$ relative to $rnhA^+dnaA^+$ is reversed by the longer inversion $\Delta rnhA-\Delta dnaA-inv^{rrnD-rrnE}$, and probably less so by the shorter inversion $\Delta rnhA-\Delta dnaA-inv^{rrnD-rrnC}$. Nevertheless, the magnitude of the difference in gene expression between $rnhA^+dnaA^+$ and $\Delta rnhA-\Delta dnaA$ is higher than that between the suppressors and $\Delta rnhA-\Delta dnaA$ ($P < 10^{-10}$, paired Wilcoxon test comparing magnitudes of differential expression).

A small, but statistically significant portion of the difference in gene expression can be explained by differences in DNA copy number – a consequence of differences in maximal growth rates – as measured by NGS sequencing of matched exponential phase genomic DNA samples (Pearson correlation coefficient ~ 0.2 , $P < 10^{-10}$). These correlations between DNA copy number and RNA-seq based gene expression fold changes increase to over 0.75 in all comparisons when gene expression data are smoothed by LOESS, which averages out local variation in expression levels (Figure 4.12).

The movement of the origin of replication to *oriK45*, and the large inversion, might affect the macrodomain structure of the chromosome (31), as well as supercoil gradients (39). *oriK45* would be located at the right extreme of the *ori* macrodomain. The left end of the larger inversion is within a non-structured region of the chromosome, whereas the right end is within the *ori* macrodomain, and such an inversion could have consequences to cell

Figure 4.11

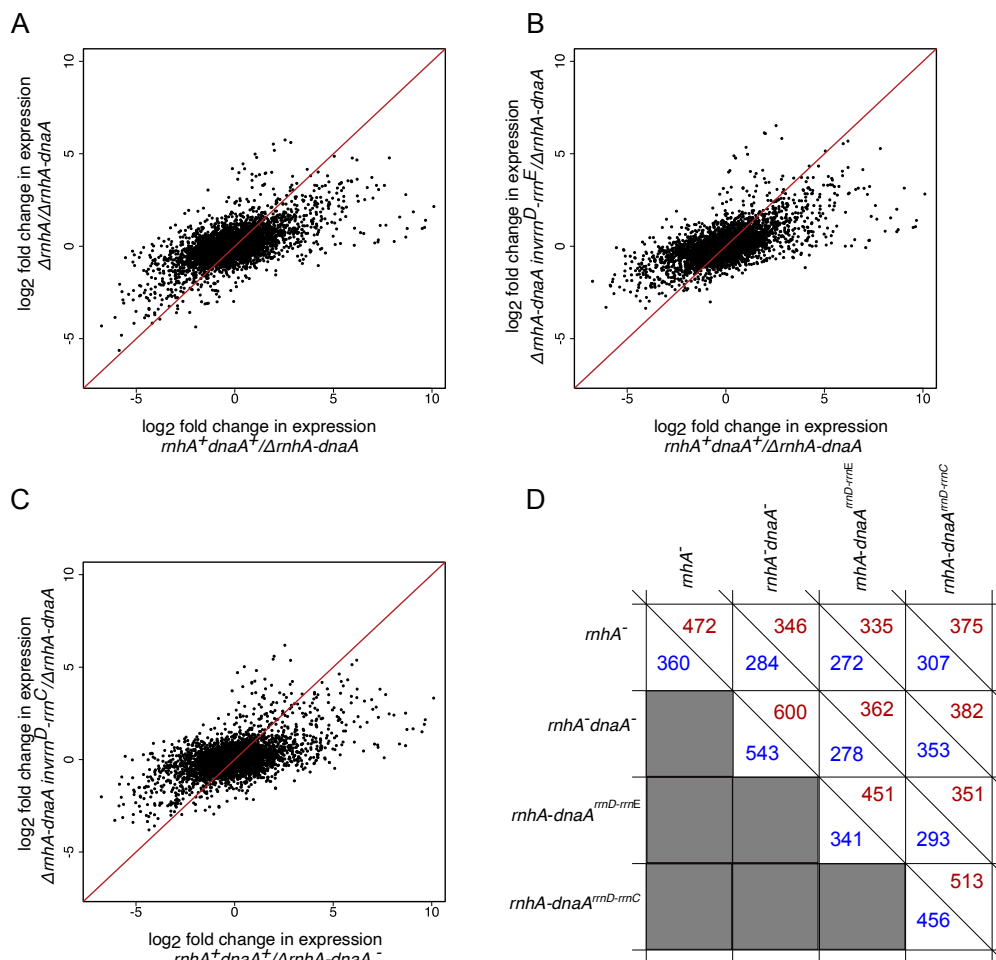


Figure 4.11: *oriC* independent DNA replication and global gene expression changes: Scatterplots representing correlation of log₂ fold change in gene expression for different conditions, compared to $\Delta rn h A - \Delta dna A$ strain. (A) $\Delta rn h A$ vs $rnh A^+ dna A^+$ (B) $\Delta rn h A - \Delta dna A inv r r m D - r r m E$ vs $rnh A^+ dna A^+$ and (C) $\Delta rn h A - \Delta dna A inv r r m C - r r m E$ vs $rnh A^+ dna A^+$. The pearson correlation values for (A), (B), (C) are 0.638, 0.639, and 0.553 respectively. (D) Plot representing the number of up-regulated (red) and down-regulated (blue) genes for all strains compared with $rnh A^+ dna A^+$.

physiology as well as gene expression (34). What the precise effect of these chromosome structure parameters are on the transcriptional profile is, in the absence of chromosome conformation data under cSDR, is not clear at the moment.

To conclude, overall gene expression changes along the chromosome are weakly correlated with the distance of a gene from *oriC* (and *oriK45*) and changes in DNA copy number. Gene expression changes that occur in $\Delta rnhA\text{-}\Delta dnaA$ relative to $rnhA^+dnaA^+$ are partly compensated by inversion containing suppressors.

4.29 The inversions reduce replication-transcription conflicts at rRNA loci but not at essential mRNA genes

How does chromosomal rearrangements helps to restore the gene expression patterns? To understand the impact of inversions on transcription-replication collisions, we calculated a fractional score for the occurrence of head-on collisions for genes on the lagging strand with respect to replication from *oriC* or *oriK45* using RNA sequencing data (see materials and methods). This score was lowest at 0.31 for $rnhA^+dnaA^+$. This increased to 0.67 in $\Delta rnhA\text{-}\Delta dnaA$, but was reduced to 0.39 in the suppressor $\Delta rnhA\text{-}\Delta dnaA\text{inv}^{rrnD\text{-}rrnE}$ (Table 4.4). This effect was the strongest when only rRNA genes (5S rRNA, which is not depleted as part of the RNA prep experiment) were considered. Despite the large decrease in replication-transcription conflict in the inversion-containing suppressors, the activation of the SOS response in cSDR is not reversed, even at a quantitative level; this requires further investigation.

Curiously however, clashes appeared to increase for mRNA genes, including essential genes; it must however be noted that the expression levels of mRNA genes would only be a fraction of rRNA levels. Therefore, it appears that any suppression in the growth defect may arise from a reversal of increased replication-transcription conflicts at rRNA loci, notwithstanding any effect on essential or non-essential mRNA genes.

Figure 4.12

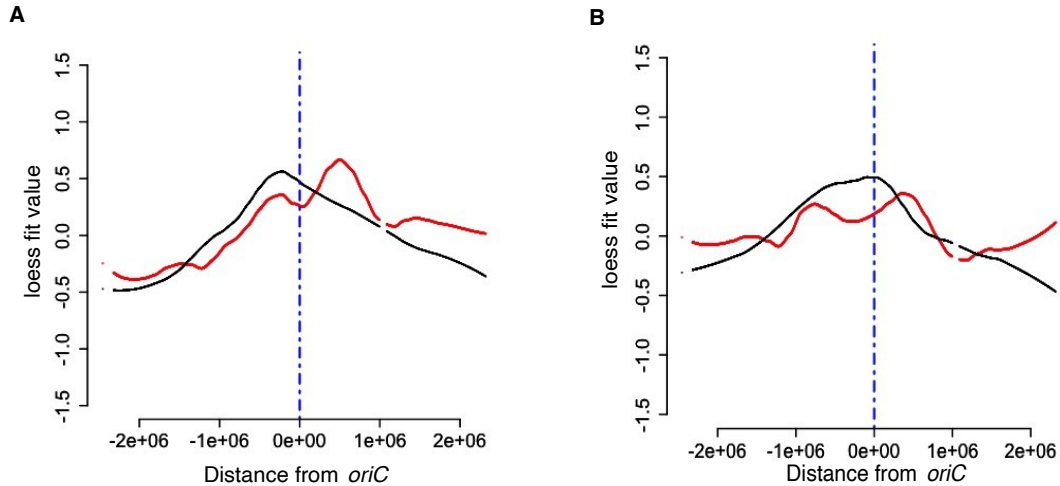


Figure 4.12: Effect of DNA copy number on gene expression: Plots showing the trend followed by the smoothed log₂ fold change values for all genes in comparison with *ΔrnhA-ΔdnaA* strain for gene expression and DNA copy number. (A) *ΔrnhA-ΔdnaAinv^{rmD-rmE}/ΔrnhA-dnaA*, and (B) *ΔrnhA-ΔdnaAinv^{rmD-rmC}/ΔrnhA-ΔdnaA*. X-axis represents positions centered around *oriC* and Y axis represents loess fit values of log₂ fold change. Red lines represent gene expression and black lines represent DNA copy number for the same strain.

4.3 Discussion

Laboratory evolution of *ΔrnhA-ΔdnaA* strain resulted in a rapid generation of suppressor mutants with a substantial increase in growth rate. Genomic analysis of these evolved mutants indicate that under *ΔrnhA-ΔdnaA* cSDR, there is a strong selection which favours preferential replication initiation from *oriK45*, located ~0.4-0.7 Mb clockwise of *oriC*. *oriK45* is a broadly-defined region, and spans an ~300 kb region across the samples analysed here. The precise location of one or more initiation sites within *oriK45* is unknown, and may be beyond the capabilities of MFA experiments in unsynchronised populations. The top homology to R-loop-forming sequences is at ~4.51 Mb; *hotH*, previously shown to be capable of autonomous replication (65), is ~40-50 kb clockwise of the above R-loop-forming sequence; *fimD*, an insertion in which has an effect on *ΔtopA*-mediated cSDR (87), is located between the above two sites and is closer to *hotH*. These observations along with our deletion data suggest that there are multiple possible replication initiation sites at *oriK45* region of the chromosome.

Detection of *oriK45* peak in all mutants exhibits a strong selection for this region as a preferable replication initiation site on *E.coli* chromosome. The preference for *oriK45* can be due to multiple reasons, including the high propensity to form R-loops. However, one of the important element for *oriK45* selection is the position of this site being closer to *oriC* position on the chromosome which helps to maintain the copy number gradient. As mentioned before, *E.coli* chromosome has similar replicore sizes in which *oriC* and *ter* is positioned almost diametrically opposite on the chromosome. This arrangement of the chromosome will ensure that the time travelled by both replication forks will be almost equal. Therefore, *oriK45* being close to *oriC* will make sure that the replicore lengths will not differ drastically from the wild type condition.

Suppressor mutants which favours *oriK45* being a prominent replication initiation site exhibit higher *ori-to-ter* gradient in presence of a large chromosomal inversion. The restoration of the gene copy number in inversion mutants also resulted in re-balancing the gene expression patterns. This results suggest that the selection for *oriK45* along with a

large chromosomal inversion near *oriC* provides a selective advantage for the bacteria under cSDR conditions.

It has been previously shown that the orientation of the gene according to the direction of replication impact growth fitness in *E.coli* (16). To systematically address the role of strand asymmetries in fitness, Srivatsan et al. performed an experiment in which they inverted a portion of the chromosome around the *oriC* in *B. Subtilis* and measured the growth characteristics of the strain. Here the inversions reverts the orientation of several ribosomal RNA operons, such that they are now encoded on the lagging strand. In this study, they observed a significant reduction in fitness for bacteria carrying these inversions profoundly during rapid growth in rich media conditions than during relatively slower growth in minimal media (27).

The chromosomal inversion reported in this study will also result in the movement of several *rrn* operons and other metabolic genes involved in faster growth. This suggests that the inversion of these genes may reduce replication-transcription conflicts during *oriK45* activation. Transcriptome analysis of these mutants validate this hypothesis by predicting a collision score for head-on events during replication. Further, Mutational analysis of inverted strains denotes the presence of SNPs in the regulatory region of an *rrn* operon which got shifted to the lagging strand of the chromosome according to the new replication initiation site *oriK45*. The effect of this mutation on the expression of *rrn* genes from that particular operon remains an open question.

Together, in this study, we observe that the loss of *oriC* mediated DNA replication in a cSDR strain of *E.coli* K12 MG1655 results in a strong selection for an alternative replication initiation site (*oriK45*) located closer to *oriC* on the chromosome. Along with this, a large chromosomal inversion of the genome helped the bacteria to restore the gene expression status of the cell. These results indicate that the orientation of highly expressed genes located closer to the origin of replication is important to be co-directional such that the replication-transcription conflicts are minimal. Thus, the replicative structure of the chromosome plays an essential role in shaping the chromosome organization in bacteria especially in fast growth conditions (16).

Table 4.1

Strain ID	<i>Ori-to-ter</i> ratio
<i>Parental strains</i>	
1D0_1	1.13
1D0_2	1
1D0_3	1.09
1D0_4	1.09
5D0_1	1.13
5D0_2	1.15
5D0_3	1.07
5D0_4	1.15
8D0_1	1.03
8D0_2	0.97
8D0_3	0.87
8D0_4	1
Δ rnhA- Δ dnaA	1.02
Δ rnhA	1.79
Δ rnhA- Δ dnaA/pHYD2388	1.22
GJ13519	2.48
K12 MG1655	2.35
<i>Suppressor mutants</i>	
1D4_1	1.02
1D4_2	1.73
1D4_3	1.69

1D4_4	1.03
1D8_1	1.51
1D8_2	1.45
1D8_3	0.94
1D8_4	1.93
1D12_2	1.28
1D12_3	1.27
1D12_4	1.35
1D15_1	1.42
1D15_2	1.49
1D15_3	1.25
1D15_4	1.47
5D4_1	1.41
5D4_2	1.33
5D4_3	1.38
5D4_4	1.20
5D8_1	1.34
5D8_2	1.38
5D8_3	1.47
5D8_4	1.31
5D12_1	1.68
5D12_2	1.82

5D12_3	2.68
5D12_4	1.67
5D15_1	1.82
5D15_2	1.88
5D15_3	1.89
5D15_4	1.96
8D4_1	1.14
8D4_2	1.02
8D4_3	1.03
8D4_4	1.07
8D8_1	1.44
8D8_2	1.87
8D8_3	1.76
8D8_4	1.36
8D15_1	1.70
8D15_2	1.70
8D15_4	1.72

Table 4.1: *Ori-to-ter* ratios for all strains calculated from MFA plots. To read the sample ids, first number represents the evolution lane(1,5,8), followed by D letter and the number which represents Day of evolution(0, 4, 8, 12, 15), and last number represents the number of colony selected from that population. For example 5D0_3 means 3rd colony selected from the 0th day of evolution of evolution lane number 5.

Table 4.2

GCF_001021635.2_ASM102163v2	NZ_CP006834.2 <i>Escherichia coli</i> APEC O2-211 chromosome
GCF_002220215.1_ASM222021v1	NZ_CP022393.1 <i>Escherichia coli</i> strain E62 chromosome
GCF_003052645.1_ASM305264v1	NZ_CP028702.1 <i>Escherichia coli</i> strain J53 chromosome
GCF_000499485.1_MYMC4100	NZ_HG738867.1 <i>Escherichia coli</i> str. K-12 substr. MC4100
GCF_003966445.1_ASM396644v1	NZ_AP018802.1 <i>Escherichia coli</i> E2863 DNA
GCF_000091005.1_ASM9100v1	NC_013361.1 <i>Escherichia coli</i> O26:H11 str. 11368 DNA
GCF_003018715.1_ASM301871v1	NZ_CP027331.1 <i>Escherichia coli</i> strain 2013C-3277 chromosome
GCF_002211725.1_ASM221172v1	NZ_CP022154.1 <i>Escherichia coli</i> strain ABWA45 chromosome
GCF_003112145.1_ASM311214v1	NZ_CP028110.1 <i>Escherichia coli</i> O121 str. RM8352 chromosome
GCF_002879975.1_ASM287997v1	NZ_CP025747.1 <i>Escherichia coli</i> strain ML35 chromosome
GCF_001420935.1_ASM142093v1	NZ_CP013029.1 <i>Escherichia coli</i> strain 2012C-4227
GCF_003052665.1_ASM305266v1	NZ_CP028703.1 <i>Escherichia coli</i> strain ME8067 chromosome
GCF_002716885.1_ASM271688v1	NZ_CP015244.1 <i>Escherichia coli</i> O91 str. RM7190 chromosome
GCF_003018795.1_ASM301879v1	NZ_CP027352.1 <i>Escherichia coli</i> strain 2012C-4606 chromosome
GCF_003017805.1_ASM301780v1	NZ_CP027325.1 <i>Escherichia coli</i> strain 2013C-4830 chromosome
GCF_002741215.1_ASM274121v1	NZ_CP024239.1 <i>Escherichia coli</i> O15:H11 strain 90-9272

	chromosome
GCF_001901105.1_ASM190110v1	NZ_CP010196.1 <i>Escherichia coli</i> strain M9
GCF_000010745.1_ASM1074v1	NC_013353.1 <i>Escherichia coli</i> O103:H2 str. 12009 DNA
GCF_001901425.1_ASM190142v1	NZ_CP010240.1 <i>Escherichia coli</i> strain C7
GCF_002012025.1_ASM201202v1	NZ_CP018970.1 <i>Escherichia coli</i> strain Ecol_542 chromosome
GCF_000010245.2_ASM1024v1	NC_007779.1 <i>Escherichia coli</i> str. K-12 substr. W3110 DNA
GCF_003112185.1_ASM311218v1	NZ_CP028116.1 <i>Escherichia coli</i> O26 str. RM8426 chromosome
GCF_003966465.1_ASM396646v1	NZ_AP018808.1 <i>Escherichia coli</i> E2865 DNA
GCF_003018495.1_ASM301849v1	NZ_CP027552.1 <i>Escherichia coli</i> strain 2015C-4498 chromosome
GCF_000258025.1_ASM25802v1	NC_017660.1 <i>Escherichia coli</i> KO11FL
GCF_000148605.1_ASM14860v1	NC_017632.1 <i>Escherichia coli</i> UM146
GCF_001901215.1_ASM190121v1	NZ_CP010221.1 <i>Escherichia coli</i> strain M19
GCF_003018035.1_ASM301803v1	NZ_CP027390.1 <i>Escherichia coli</i> strain 2015C-4944 chromosome
GCF_002057355.1_ASM205735v1	NZ_CP020107.1 <i>Escherichia coli</i> strain 13E0767 chromosome
GCF_002796445.1_ASM279644v1	NZ_CP024889.1 <i>Escherichia coli</i> strain AR_0019 chromosome
GCF_001900535.1_ASM190053v1	NZ_CP010122.1 <i>Escherichia coli</i> strain C5
GCF_002055605.1_ASM205560v1	NZ_CP020092.1 <i>Escherichia coli</i> strain 13E0725 chromosome

GCF_003018155.1_ASM301815v1	NZ_CP027548.1	<i>Escherichia coli</i>	strain	2014C-3061
	chromosome			
GCF_003019215.1_ASM301921v1	NZ_CP027766.1	<i>Escherichia coli</i>	strain	2013C-3342
	chromosome			
GCF_003018895.1_ASM301889v1	NZ_CP027387.1	<i>Escherichia coli</i>	strain	2014C-3057
	chromosome			
GCF_003018575.1_ASM301857v1	NZ_CP027582.1	<i>Escherichia coli</i>	strain	2013C-4538
	chromosome			
GCF_002057245.1_ASM205724v1	NZ_CP020106.1	<i>Escherichia coli</i>	strain	13E0780
	chromosome			

Table 4.2: List of *E.coli* strains in which presence of a chromosomal inversion around *oriC* is observed.

Table 4.3

Sample	3.84-3.89 (oriC)	0.44- 0.55	1.42-1.52	1.88-2.23	2.53-2.6	2.95-3.36	3.63-3.8	4.26- 4.59
Parental strains								
WT	2.36	-	-	-	-	-	-	-
Wtlac	2.50	-	-	-	-	-	-	-
dnaA	2.33	-	-	-	-	-	-	-
rnhA	1.90	-	1.18	-	-	-	-	1.91
Blue	-	1.33	1.22	1.17	1.10	-	1.21	1.35
cSDR	-	1.63	1.85	1.50	-	1.19	1.05	1.54
1D01	-	1.35	1.33	1.27	1.15	1.17	-	1.33
1D02	-	1.53	1.74	1.50	-	1.21	1.07	1.48
1D03	-	1.56	1.64	1.44	1.22	1.18	1.07	1.50
1D04	-	1.34	1.34	1.41	1.32	1.36	1.08	1.34
5D01	-	1.47	1.49	1.37	1.20	1.18	1.04	1.44
5D02	-	1.56	1.56	1.40	1.24	1.20	-	1.49
5D03	-	1.63	1.73	-	-	1.25	1.09	1.56
5D04	-	1.56	1.55	1.43	1.25	1.21	1.07	1.50
8D01	-	1.33	1.47	1.26	1.11	1.08	1.02	1.32
8D02	-	1.50	1.76	1.42	1.26	1.18	1.04	1.43
8D03	-	1.31	1.68	1.31	1.05	1.06	1.02	1.27
8D04	-	1.48	1.70	1.42	-	1.19	1.05	1.41
Suppressor strains								
1D41	-	1.22	1.41	1.35	1.31	1.35	1.25	1.26
1D44	-	-	1.60	1.51	-	-	1.04	1.42
1D83	-	-	1.83	1.09	-	1.24	1.11	1.36
5D41	-	-	1.30	1.18	-	-	-	1.61
5D42	-	-	1.28	1.15	-	-	1.34	1.52
5D43	-	-	1.29	1.16	-	-	1.38	1.58

5D44	-	-	1.29	1.12	-	-	1.24	1.39
5D81	-	-	1.29	1.16	-	-	-	1.53
5D82	-	-	1.28	1.15	-	-	1.38	1.57
5D83	-	-	1.32	1.19	-	-	-	1.67
5D84	-	-	1.27	1.15	-	-	1.32	1.49
5D121	-	-	1.51	1.09	-	-	-	2.02
5D122	-	1.80	1.39	1.09	1.11	1.25	1.29	2.02
5D123	-	-	1.28	-	-	-	-	3.17
5D124	-	-	1.61	1.07	-	1.30	1.36	2.04
5D151	-	-	1.41	1.17	-	-	-	2.22
5D152	-	-	1.44	1.19	-	-	1.75	2.27
5D153	-	-	1.42	1.20	-	-	1.74	2.22
5D154	-	-	1.43	1.25	-	-	-	2.33
8D41	-	-	1.10	1.14	-	1.18	-	1.22
8D42	-	1.20	1.38	1.36	1.36	-	1.06	1.22
8D43	-	-	1.49	1.46	-	1.44	-	1.32
8D44	-	-	1.34	1.35	1.41	-	-	1.26
1D42	-	-	1.41	1.26	-	1.58	-	2.01
1D43	-	-	1.40	1.25	-	-	-	1.95
1D81	-	-	1.63	1.18	-	-	-	1.95
1D84	-	-	-	1.30	-	-	-	2.23
1D82	-	-	1.32	1.18	-	-	-	1.67
1D122	-	-	1.23	1.16	-	1.34	1.28	1.41
1D123	-	-	1.35	1.13	-	-	-	1.48
1D124	-	-	1.39	1.18	-	-	-	1.59
1D151	-	-	1.24	1.19	-	-	1.38	1.56
1D152	-	-	1.31	1.21	-	-	-	1.70
1D153	-	-	1.49	1.13	-	-	-	1.57
1D154	-	1.57	1.29	1.22	-	-	-	1.66
8D81	-	-	1.54	1.10	-	-	-	1.86

8D82	-	-	1.35	1.21	-	-	-	2.13
8D83	-	-	1.37	1.18	-	-	-	2.06
8D84	-	-	1.55	1.06	-	-	-	1.76
8D151	-	-	1.37	1.14	-	-	-	2.06
8D152	-	-	1.34	1.15	-	-	-	2.02
8D154	-	-	1.33	-	-	-	-	2.00

Table 4.3 Chromosomal position of predicted *oriK* peaks from MFA plots. The value represents loess fit value for the maxima position of the peak identified. To read the sample ids in first column, first number represents the evolution lane(1,5,8), followed by D letter and the second number represents Day of evolution(0, 4, 8, 12, 15), and last number represents the number of colony selected from that population. For example 5D0_3 means 3rd colony selected from the 0th day of evolution of lane number 5.

Table: 4.4

Strain	Head-on-collision rate
All genes	
<i>rnhA⁺dnaA⁺</i>	0.31
<i>ΔrnhAΔdnaA</i>	0.67
<i>ΔrnhAΔdnaA^{rrnD-rrnE}</i>	0.39
Ribosomal genes	
<i>rnhA⁺dnaA⁺</i>	0.22
<i>ΔrnhAΔdnaA</i>	0.84
<i>ΔrnhAΔdnaA^{rrnD-rrnE}</i>	0.24
Protein coding genes	
<i>rnhA⁺dnaA⁺</i>	0.37
<i>ΔrnhAΔdnaA</i>	0.39
<i>ΔrnhAΔdnaA^{rrnD-rrnE}</i>	0.54
Essential genes	
<i>rnhA⁺dnaA⁺</i>	0.08
<i>ΔrnhAΔdnaA</i>	0.26
<i>ΔrnhAΔdnaA^{rrnD-rrnE}</i>	0.7

Table 4.4: Probability of head-on-collisions were predicted for 3.3Mbp to 4.25Mbp region of the chromosome which includes the inverted region of *ΔrnhAΔdnaA^{rrnD-rrnE}* strain. The values were calculated by taking the ratio of sum of RNA coverage values of all genes on the lagging strand with respect to the single predominant *ori* position to the total RNA coverage for the region ($P(\text{HO}) = \frac{\text{sum}(\text{lagging strand RNA coverage})}{\text{total RNA coverage}}$). The analysis was done for different classes of genes separately by using the functional annotations for genes from NC_000913.3(.ptt and .rnt) file.

Chapter 5

Conclusions and Discussion

Note:

Parts of the text in this chapter is included in a manuscript *Veetil. et. al, mSphere, 2020*

Bacterial DNA replicative structure: Perturbations and Impact

Various features of the bacterial chromosome are in evolutionary connect with its replicative structure and this is well-established in the literature (16). A most important feature that informs our interest in chromosome architecture in bacteria is the presence of a single origin of replication per circular chromosome. It has been proposed that the origin of replication – a central feature of biology – may be a highly successful horizontally-acquired selfish element (89). A single origin of replication results in a strong growth rate-dependent gene dosage gradient from the *ori-proximal* to the *ter-proximal* regions of the chromosome. So far, various experiments have investigated the role of chromosome organisation on bacterial fitness, and on the expression of specific reporters (21, 22, 27). However, it is not clear that how well a bacterium with altered replication structures would adapt to these perturbations, what would be the adaptation strategies, and what would be its impact on gene expression homeostasis and chromosome topologies. This dissertation addresses some of these fundamental problems of bacterial genome organisation and gene expression using a combination of laboratory evolution and genomics approaches.

How to perturb the replicative structure of the chromosome? This can be achieved by different ways such as (a) alternative positioning of *ori* at various locations on the genome, with or without a corresponding movement of the *ter* (b) dispersed replication initiation in strains lacking an *ori* but with suppressors impacting the formation of R-loops or recombination intermediates, which can prime replication initiation. The positioning of origin at different locations has been already addressed to an extent in the literature recently (28, 83). In this study we chose the second approach to alter the replicative structure of the chromosome.

Here, we chose *E.coli* strains which can inactivate replication from the single origin *oriC* and attain a sub-optimal replicative system with dispersed or alternative origins. These strains are capable of replicating in *oriC*-DnaA independent manner by activating R-loop

dependent DNA replication process known as cSDR (43). These strains carry a deletion of *rnhA* and *dnaA* genes which encode for an endonuclease enzyme **RNaseH1** and the canonical replication initiator protein **DnaA** respectively. Inactivation of *oriC* dependent DNA replication led to drastic changes in gene expression and gene gradient which resulted in the slow growth of the strain.

cSDR initiation sites: specific or random?

Where does cSDR initiate on the chromosome? This has been a question of interest in the field for the last few decades. Analysis of chromosomal regions showing evidence of R-loop formation, by Gowrishankar's group, showed that there may not be any hotspots for the formation of such structures, but these R-loop-forming sites are dispersed across the chromosome (43) This resulted in the idea that *oriK* sites are not discrete entities, but that replication initiation by R-loops could occur throughout the chromosome (66). However, the Marker Frequency Analysis (MFA) of *E.coli* cSDR strain in the published literature argued that there are specific sites on the chromosome which sustain R-loop formation (64). In our study, peak detection from the MFA analysis of $\Delta rnhA$ - $\Delta dnaA$ strain of *E.coli* in comparison with previous reports indicates that *oriK* sites are not randomly distributed and they re-appear at the same region of the chromosome. This observation supports the idea that cSDR is mostly the result of replication initiation occurring at a few specific sites on the chromosome rather than the result of stochastically distributed replication initiation events at numerous potential sites in the cell population. Nonetheless, identifying potential replication sites in cSDR regime has been difficult due to the slow growth of strains which results in poor copy number gradient. This problem can be solved by generating suppressor strains that can generate strong copy number gradients under the cSDR regime. Therefore, in this study, we used a laboratory evolution approach to generate suppressor mutants that can satisfy these conditions.

How do bacteria redress the fitness effects caused by the sub-optimal replication conditions? In other words, what are the principles underlying the suppression of the slow

growth phenotype of cSDR. Laboratory evolution of the *ΔrnhA-ΔdnaA* strain of *E.coli* helped us to understand the adaptive responses that redress the loss of fitness resulting from sub-optimal chromosomal replication. Our results indicate that under *ΔrnhA-ΔdnaA* cSDR, there is a strong selection for preferential replication initiation from a single origin which we call as *oriK45*, located ~0.6 Mb clockwise of *oriC*. Strong copy number gradients of evolved suppressors showed MFA plots similar to wild type conditions with a shift in the origin peak. *oriK45* site spans a broad region of the chromosome including a preferred R-loop forming sequence at ~4.51 Mb. The position of *oriK45* on the chromosome being closer to *oriC* helps the bacteria to maintain the replicore symmetry and gene expression gradients to a large extent. Even though we could detect *oriK45* in a broad region it's not clear where exactly the replication initiates at this predicted site.

There are reports of multiple discrete initiation sites within *oriK45* region (detail discussion in Chapter 4.3 section (65, 87). Constructing deletion mutants of these initiation sites, individually and/or in combinations, in *ΔrnhA-ΔdnaA* background, followed by MFA to confirm the presence or absence of the peak will be a potential future experiment to test for autonomous replication. This analysis will help to precisely locate the initiation point within the region or to find the existence of multiple initiation sites present in the same region.

Chromosomal rearrangements: A strategy to do better

Replication initiation from *oriK45* would result in head-on collisions with RNA polymerases transcribing four rRNA operons encoded between *oriC* and *oriK45*. Such head-on collisions are detrimental at least in part because of DNA topological issues that cause excessive R-loop formation in such conflict sites (90). The predominant suppressor found here would invert the DNA around *oriC* such that these four rRNA operons would now be on the leading strand of replication from *oriK45*. This would, however, place one rRNA operon now on the lagging strand. The promoter of this rRNA operon carried a mutation in the discriminator region in all inversion-carrying suppressor strains. Though

we couldn't find any significant difference in the expression levels of plasmid-borne GFP cloned downstream of the wild type *rrnD* promoter and that with the discriminator mutation (Figure 5.1), whether this mutation confers a specific ppGpp-dependent effect on gene expression in a cSDR background, and whether this affects fitness remains to be understood. Recent evidence shows that certain genes – including determinants of virulence and antibiotic resistance – over the course of evolution might have switched in the reverse direction: from leading to lagging (91). Such genes experience higher rates of non-synonymous mutations, experiencing positive selection and thereby promoting evolvability. However, this would not apply to highly expressed genes such as the rRNA genes.

In a previous study, the Sherratt lab placed a second *ori* termed *oriZ* ~1 Mb clockwise of *oriC*. This study showed that the replication initiation from *oriZ*, despite *oriZ* being positioned such that it would cause replication-transcription conflicts at rRNA operons, caused little replication or growth defects (28). However, a later attempt by Ivanova and colleagues to create a similar strain revealed a strong growth defect and also showed that mutations that allow the RNA polymerase to bypass conflicts efficiently, and those that inactivate *ter* can suppress the growth defect (29). MFA analysis of the Sherratt lab strain by Ivanova et al. indicated the presence of a large inversion, affecting several rRNA operons, which had not been detected by the Sherratt study (28). Thus, Ivanova et al. could conclude that replication-transcription conflicts are key determinants of the fitness of *E. Coli*. The inversion reported by Ivanova et al. (29) is similar to that observed in our study, except that the right end reported by the earlier study extends beyond that found by us to a position closer to that of *oriZ*. These findings are consistent with those of Srivatsan et al. who showed that a large *oriC*-proximal inversion can cause growth defects when *Bacillus subtilis* is grown in rich media (27). Contrary to these findings, Esnault *et al.* showed that inversions near *oriC* which would place 1-3 rRNA operons on the lagging strand of replication, showed little growth defect (34). That the inversion observed in our study contributes to fitness may be ascertained from the fact that the

larger inversion produces higher copy number gradients than the smaller inversion, although both strains carry the *rrnD* promoter mutation. The selective advantage conferred by the inversion also indicates that replication initiates predominantly clockwise of *oriC*, from a position that is also clockwise of the four rRNA operons that are inverted. Moreover, *oriK45* satisfies these requirements.

Structural variations around *ter* have also been found to exist in *E. coli* with a second *ori*. Dimude *et al.* placed a second *ori*, termed *oriX*, counterclockwise of *oriC*. They found that this mutant carried a ~0.8 Mb inversion spanning the *ter* (83). However, this mutant grew slowly. Since the authors did not isolate an *oriX*⁺ strain without the inversion, they were unable to directly test whether it conferred a selective advantage, even if a small one, to its parent.

Whereas the previous studies by Ivanova *et al.*, and Dimude *et al.*, (29, 83) isolated structural variations while making the parent strain, we were able to isolate our suppressors only after 4-8 days of selection in a laboratory evolution experiment.

Replication Termination: Effects on cSDR

The presence of a *ter* peak in cSDR is proposed to be the result of replication forks getting trapped from stochastically distributed non-*oriC* replication initiation events in the cell population. The disappearance of this peak in *tus* deletion mutants of cSDR strain strengthened this argument in the past (64). The authors concluded for the absence of such *oriK* activity in a sub-region in *ter* (the TerB-TerC interval) but presented candidate peaks that corresponds to *oriK* sites at certain other genomic positions located in the terminus. Dimude *et al.* and Brochu *et al.* also reported the presence of a major peak in the mid terminus region of the chromosome under different cSDR conditions (60, 62). In an *rnhA* mutant carrying a large DNA inversion in the *ter* region (62) the location of the mid-terminus peak is modified accordingly in this strain and therefore it is still located at the same position according to the reference strain. Thus, the authors suggest that this peak does not appear to be the result of replication termination events, and it is likely corresponds to an origin of replication or several such sites located close to each other in

this area. In line with these observations, our results also indicate the presence of a *ter* peak in evolved strains in which one or very few *oriK* sites are located in a specific region (*oriK45*). The fact that the peak in the *ter* region is still present in cells from the stationary phase of growth does not fully rule out the presence of *oriK* activity in this region. Unlike replication from *oriC*, replication from at least some *oriKs* could still be active in the stationary phase. In fact, cSDR was initially demonstrated in non-growing cells treated with translation inhibitors, which prevent initiation from *oriC* but not from *oriKs*. Therefore, there is a possibility that the *ter* peak may not be the result of replication fork trap events instead it may represent an *oriK* site. Nonetheless, the data presented here as well as in previously published papers is not sufficient to prove or disprove that the peak detected in MFA experiments in the *ter* region does not reflect the presence of one or more *oriK* sites. One potential future experiment could be an analysis of a strain that combines the inversion-containing suppressors isolated in our study with a *Atus* mutation. In such a mutant, a strong *oriK* site within *ter* might manifest as an obvious peak, but might appear with a cost of drastically diminishing the highly favourable copy number gradient declining from *oriC* towards *ter*.

Another major constraint of our work is its non-accountancy for the positioning of the terminus on the chromosome. How does the positioning of the terminus impact the replicative structure of the chromosome? Is *ter* position being opposite to the *oriC* region in *E.coli* a constraint to determine the *oriK* position on the chromosome during evolution? Our results indicate that *oriK45* appears to be favoured by selection, and the fact that this site is located relatively close to the canonical *oriC* may help its cause. However, this does not rule out the possibility of selection for *oriK45* as a result of the current arrangement of the replicative structure where the terminus is positioned opposite to the *oriC* region. This is yet to be tested by moving or deleting the *ter* region of the chromosome and evolving the bacteria under suboptimal replication conditions.

Gene expression and Replicative structure

The replicative structure of the chromosome is known to impact the gene expression status of the cell. The arrangement of fast growth related genes closer to *oriC* position to take advantage of gene dosage effect has been well discussed in the literature (16). It is not clear whether and how the dependence of gene expression on DNA copy number is linear and similar across genes. Here, using a global approach we show that the replicative structure of the chromosome determines the expression of certain genes according to their position and orientation on the chromosome. The cSDR strain in which the *oriC* is inactivated shows drastic down-regulation of *oriC* proximal genes which account for the loss of gene gradient. This effect has been reversed in a strong suppressor mutant with a chromosomal inversion and *oriK45* activation. This suggests that chromosome architecture itself a *regulator* of global gene expression. We also observed an up-regulation of stress response and *ter* proximal genes in parental and evolved strains, which may also correlate with the presence of more copies of DNA at the *ter* region as a result of a *ter* peak in cSDR strains. This has to be tested in a Tus deletion strain of cSDR in which the *ter* peak abolishes. From these results, we summarize that replicative structure affects the transcription state of the cell to a higher extent. Nevertheless, How does the replication-dependent gene organisation impact chromosome macrodomain organization and topological states of the genome is yet to be discovered.

Possible physiological role of cSDR

Replication of bacterial chromosomes in genetically altered cells using cSDR mechanism lacks a clear definition of its physiological role in bacteria. This led to various speculations in the field regarding the functional role of this system in the context of evolution. It has been proposed that cSDR could be a possible primitive DNA replication system in *E.coli* that was functional before the establishment of current canonical *oriC*-DnaA based replication system (43). This model also suggests the functioning of *oriC* and *oriK* systems together until the event of introduction of genes that can disrupt R-loops

which might have resulted in the repression of the less efficient *oriK* system. However, further discovery of cSDR-like activity in rapidly growing wild type cells contradicted these speculations (61). This SDR activity was observed transiently up on entry to the stationary phase under a nutritional shift-up condition and termed as “nutritional shift-up-activatable SDR,” or nSDR (61). As similar to cSDR, nSDR is completely dependent on a RecA function and independent of RecB and DnaA. The close resemblance of nSDR to cSDR reasons the possibility of nSDR being a manifestation of cSDR in wild type cells during stationary phase of growth. The proposed involvement of R-loops as replication priming sites during nSDR led to hypothesize a number of possible conditions, which could result in the stabilization of R-loops during stationary phase due to drastic changes in regulatory networks and transcription rates (43). The current limited understanding of mechanism of nSDR is insufficient to disprove any of these theories. Taken together, though cSDR may not necessarily be a physiological or natural phenomenon in *E. coli*, with the possible exception of its manifestation as nSDR in stationary phase, it has been argued that this could be a potential primordial mechanism of DNA replication initiation (43).

Gain or loss of DnaA protein in the bacterial genome might help to understand the evolution of alternate replication strategies. In an attempt to search for DnaA homologs across eubacterial strains, we could not detect a sequence homolog for this protein in 56 out of ~6000 strains analysed. Interestingly, this list included one of the *Cyanobacterium aponinum* strain. The absence of a DnaA protein in this strain can be extrapolated to further investigate the evolution of DnaA independent mitochondrial replication.

We also observed the absence of DnaA protein homolog in a mycobacterium strain *Mycobacterium abscessus* DJO-44274. *Mycobacterium abscessus* is a fast growing, pathogenic, multidrug-resistant non-tuberculous mycobacterial species (92). The DNA replication process in mycobacterial species is known to be similar to that of *E.coli*, where the *oriC*-DnaA interaction is critical for replication initiation. However, the chromosome organization and DNA replisome dynamics of *Mycobacterium* shows

significant difference from *E.coli* model. Recent microscopy experiments showed that the left and right replichores of *M.smegmatis* co-localize throughout the replication process (93), which is different from the “transverse” chromosome topology of *E.coli* (94,95). Considering the clinical relevance and unique biology of mycobacterial strains, it is important to investigate the replication related genome organization in more detail. In this context, it is curious to explore the process of DNA replication in *Mycobacterium abscessus* strain DJO-44274 (which lacks DnaA), especially, due to it’s high tolerance to antibiotics.

Conclusion

The bacterium *E. coli* can initiate replication in the absence of the replication initiator protein DnaA and / or the canonical origin of replication *oriC* in a $\Delta rnhA$ background. This phenomenon is called constitutive stable DNA replication (cSDR) (43). Whether DNA replication during cSDR initiates in a stochastic manner through the length of the chromosome or at specific sites, and how *E. coli* can find adaptations to loss of fitness caused by cSDR remain inadequately answered. We used laboratory evolution experiments of $\Delta rnhA-\Delta dnaA$ followed by deep sequencing to show that DNA replication preferentially initiates within a broad region located ~0.4-0.7 Mb clockwise of *oriC*. This region includes many bisulfite-sensitive sites, which have been previously defined as R-loop forming regions; and includes a site containing sequence motifs that favour R-loop formation. Initiation from this region would result in head-on replication-transcription conflicts at rRNA loci. Inversions of these rRNA loci, which can partly resolve these conflicts, help the bacterium suppress the fitness defects of cSDR. These inversions partially restore the gene expression changes brought about by cSDR. The inversion however increases the possibility of conflicts at essential mRNA genes, which would utilise only a miniscule fraction of RNA polymerase molecules most of which transcribe rRNA genes. Whether subsequent adaptive strategies would attempt to resolve these conflicts remains an open question. Understanding such evolutionary strategies in the

context of cSDR can provide insights into the potential causes of resistance against antibiotics that target initiation of DNA replication (96, 97).

Figure 5.2

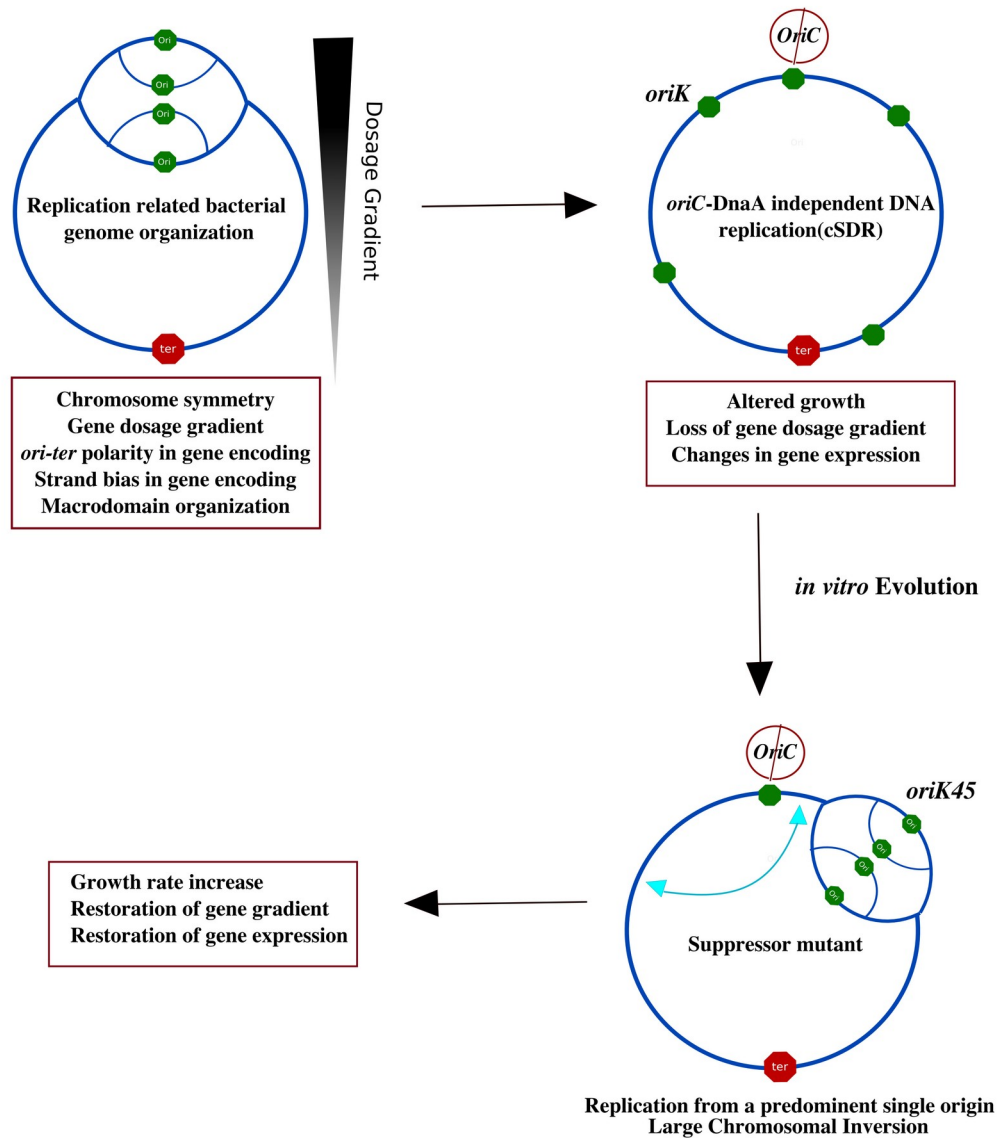


Figure 5.2: Graphical summary of the work. In this figure the circle represents *E. coli* chromosome. The green hexagons represents *ori* positions (*oriC* and *oriK*) and the red hexagons represents terminus. Cyan double arrow line denotes the presence of a large chromosomal inversion

Bibliography

1. Cooper GM, Cooper GM. 2000. *The Cell*, 2nd ed. Sinauer Associates.
2. O'Donnell M, Langston L, Stillman B. 2013. Principles and Concepts of DNA Replication in Bacteria, Archaea, and Eukarya. *Cold Spring Harb Perspect Biol* 5.
3. Ekundayo B, Bleichert F. 2019. Origins of DNA replication. *PLoS Genet* 15:e1008320.
4. Kurth I, O'Donnell M. 2009. Replisome Dynamics during Chromosome Duplication. *EcoSal Plus* 3.
5. Snedeker J, Wooten M, Chen X. 2017. The Inherent Asymmetry of DNA Replication. *Annu Rev Cell Dev Biol* 33:291–318.
6. Dewar JM, Walter JC. 2017. Mechanisms of DNA replication termination. *Nat Rev Mol Cell Biol* 18:507–516.
7. Cortez D. 2019. Replication-Coupled DNA Repair. *Mol Cell* 74:866–876.
8. Mott ML, Berger JM. 2007. DNA replication initiation: mechanisms and regulation in bacteria. *Nat Rev Microbiol* 5:343–354.
9. Hansen FG, Atlung T. 2018. The DnaA Tale. *Front Microbiol* 9:319.
10. Messer W. 2002. The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiol Rev* 26:355–374.

11. Skarstad K, Katayama T. 2013. Regulating DNA replication in bacteria. *Cold Spring Harb Perspect Biol* 5:a012922.
12. Waldminghaus T, Skarstad K. 2009. The Escherichia coli SeqA protein. *Plasmid* 61:141–150.
13. Kaguni JM. 2011. Replication initiation at the Escherichia coli chromosomal origin. *Curr Opin Chem Biol* 15:606–613.
14. Neylon C, Kralicek AV, Hill TM, Dixon NE. 2005. Replication termination in Escherichia coli: structure and antihelicase activity of the Tus-Ter complex. *Microbiol Mol Biol Rev MMBR* 69:501–526.
15. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. 2002. *Molecular Biology of the Cell*, 4th ed. Garland Science.
16. Rocha EP. 2004. The replication-related organization of bacterial genomes. *Microbiology* 150:1609–1627.
17. Couturier E, Rocha EPC. 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.
18. Khedkar S, Seshasayee ASN. 2016. Comparative Genomics of Interreplichore Translocations in Bacteria: A Measure of Chromosome Topology? *G3 Genes Genomes Genet* 6:1597–1606.

19. Sobetzko P, Travers A, Muskhelishvili G. 2012. Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proc Natl Acad Sci U S A* 109:E42-50.
20. Zarei M, Sclavi B, Cosentino Lagomarsino M. 2013. Gene silencing and large-scale domain structure of the *E. coli* genome. *Mol Biosyst* 9:758–767.
21. Schmid MB, Roth JR. 1987. Gene location affects expression level in *Salmonella typhimurium*. *J Bacteriol* 169:2872–2875.
22. Bryant JA, Sellars LE, Busby SJ, Lee DJ. 2014. Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res* 42:11383–11392.
23. Koonin EV, Wolf YI. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res* 36:6688–6719.
24. Rocha E. 2002. Is there a role for replication fork asymmetry in the distribution of genes in bacterial genomes? *Trends Microbiol* 10:393–395.
25. Mirkin EV, Mirkin SM. 2007. Replication fork stalling at natural impediments. *Microbiol Mol Biol Rev MMBR* 71:13–35.
26. Rocha EPC, Danchin A. 2003. Gene essentiality determines chromosome organisation in bacteria. *Nucleic Acids Res* 31:6570–6577.
27. Srivatsan A, Tehranchi A, MacAlpine DM, Wang JD. 2010. Co-orientation of replication and transcription preserves genome integrity. *PLoS Genet* 6:e1000810.

28. Wang X, Lesterlin C, Reyes-Lamothe R, Ball G, Sherratt DJ. 2011. Replication and segregation of an *Escherichia coli* chromosome with two replication origins. *Proc Natl Acad Sci* 108:E243–E250.
29. Ivanova D, Taylor T, Smith SL, Dimude JU, Upton AL, Mehrjouy MM, Skovgaard O, Sherratt DJ, Retkute R, Rudolph CJ. 2015. Shaping the landscape of the *Escherichia coli* chromosome: replication-transcription encounters in cells with an ectopic replication origin. *Nucleic Acids Res* 43:7865–7877.
30. Blot N, Mavathur R, Geertz M, Travers A, Muskhelishvili G. 2006. Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. *EMBO Rep* 7:710–715.
31. Valens M, Penaud S, Rossignol M, Cornet F, Boccard F. 2004. Macrodomain organization of the *Escherichia coli* chromosome. *EMBO J* 23:4330–4341.
32. Lioy VS, Cournac A, Marbouty M, Duigou S, Mozziconacci J, Espéli O, Boccard F, Koszul R. 2018. Multiscale Structuring of the *E. coli* Chromosome by Nucleoid-Associated and Condensin Proteins. *Cell* 172:771-783.e18.
33. Mercier R, Petit M-A, Schbath S, Robin S, El Karoui M, Boccard F, Espéli O. 2008. The MatP/matS site-specific system organizes the terminus region of the *E. coli* chromosome into a macrodomain. *Cell* 135:475–485.
34. Esnault E, Valens M, Espéli O, Boccard F. 2007. Chromosome structuring limits genome plasticity in *Escherichia coli*. *PLoS Genet* 3:e226.

35. Dorman CJ. 2006. DNA supercoiling and bacterial gene expression. *Sci Prog* 89:151–166.
36. Kahramanoglou C, Seshasayee ASN, Prieto AI, Ibberson D, Schmidt S, Zimmermann J, Benes V, Fraser GM, Luscombe NM. 2011. Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucleic Acids Res* 39:2073–2091.
37. Srinivasan R, Scolari VF, Lagomarsino MC, Seshasayee ASN. 2015. The genome-scale interplay amongst xenogene silencing, stress response and chromosome architecture in *Escherichia coli*. *Nucleic Acids Res* 43:295–308.
38. Scolari VF, Bassetti B, Sclavi B, Lagomarsino MC. 2011. Gene clusters reflecting macrodomain structure respond to nucleoid perturbations. *Mol Biosyst* 7:878–888.
39. Lal A, Dhar A, Trostel A, Kouzine F, Seshasayee ASN, Adhya S. 2016. Genome scale patterns of supercoiling in a bacterial chromosome. *Nat Commun* 7:11055.
40. Clayton DA. 1982. Replication of animal mitochondrial DNA. *Cell* 28:693–705.
41. Mitochondrial DNA replication in mammalian cells: overview of the pathway. - PubMed - NCBI.
42. Kogoma T. 1978. A novel *Escherichia coli* mutant capable of DNA replication in the absence of protein synthesis. *J Mol Biol* 121:55–69.
43. Kogoma T. 1997. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol Mol Biol Rev* 61:212–238.

44. Kogoma T, Torrey TA, Connaughton MJ. 1979. Induction of UV-resistant DNA replication in *Escherichia coli*: Induced stable DNA replication as an SOS function. *Mol Gen Genet MGG* 176:1–9.
45. Asai T, Sommer S, Bailone A, Kogoma T. 1993. Homologous recombination-dependent initiation of DNA replication from DNA damage-inducible origins in *Escherichia coli*. *EMBO J* 12:3287–3295.
46. Masai H, Asai T, Kubota Y, Arai K, Kogoma T. 1994. *Escherichia coli* PriA protein is essential for inducible and constitutive stable DNA replication. *EMBO J* 13:5338–5345.
47. Torrey TA, Kogoma T. 1987. Genetic analysis of constitutive stable DNA replication in *rnh* mutants of *Escherichia coli* K12. *Mol Gen Genet MGG* 208:420–427.
48. Gowrishankar J, Leela JK, Anupama K. 2013. R-loops in bacterial transcription: their causes and consequences. *Transcription* 4:153–157.
49. Ogawa T, Pickett GG, Kogoma T, Kornberg A. 1984. RNase H confers specificity in the *dnaA*-dependent initiation of replication at the unique origin of the *Escherichia coli* chromosome in vivo and in vitro. *Proc Natl Acad Sci U S A* 81:1040–1044.
50. Martel M, Balleydier A, Sauriol A, Drolet M. 2015. Constitutive stable DNA replication in *Escherichia coli* cells lacking type 1A topoisomerase activity. *DNA Repair* 35:37–47.

51. Harinarayanan R, Gowrishankar J. 2003. Host factor titration by chromosomal R-loops as a mechanism for runaway plasmid replication in transcription termination-defective mutants of *Escherichia coli*. *J Mol Biol* 332:31–46.
52. Gowrishankar J, Harinarayanan R. 2004. Why is transcription coupled to translation in bacteria? *Mol Microbiol* 54:598–603.
53. Raghunathan N, Kapshikar RM, Leela JK, Mallikarjun J, Bouloc P, Gowrishankar J. 2018. Genome-wide relationship between R-loop formation and antisense transcription in *Escherichia coli*. *Nucleic Acids Res* 46:3400–3411.
54. Raghunathan N, Goswami S, Leela JK, Pandiyan A, Gowrishankar J. 2019. A new role for *Escherichia coli* Dam DNA methylase in prevention of aberrant chromosomal replication. *Nucleic Acids Res* 47:5698–5711.
55. Cerritelli SM, Crouch RJ. 2009. Ribonuclease H: the enzymes in eukaryotes. *FEBS J* 276:1494–1505.
56. Hong X, Cadwell GW, Kogoma T. 1995. *Escherichia coli* RecG and RecA proteins in R-loop formation. *EMBO J* 14:2385–2392.
57. Lloyd RG, Rudolph CJ. 2016. 25 years on and no end in sight: a perspective on the role of RecG protein. *Curr Genet* 62:827–840.
58. Rudolph CJ, Upton AL, Briggs GS, Lloyd RG. 2010. Is RecG a general guardian of the bacterial genome? *DNA Repair* 9:210–223.

59. Midgley-Smith SL, Dimude JU, Taylor T, Forrester NM, Upton AL, Lloyd RG, Rudolph CJ. 2018. Chromosomal over-replication in *Escherichia coli* recG cells is triggered by replication fork fusion and amplified if replicore symmetry is disturbed. *Nucleic Acids Res* 46:7701–7715.
60. Dimude JU, Stockum A, Midgley-Smith SL, Upton AL, Foster HA, Khan A, Saunders NJ, Retkute R, Rudolph CJ. 2015. The consequences of replicating in the wrong orientation: bacterial chromosome duplication without an active replication origin. *MBio* 6:e01294–15.
61. Hong X, Cadwell GW, Kogoma T. 1996. Activation of stable DNA replication in rapidly growing *Escherichia coli* at the time of entry to stationary phase. *Mol Microbiol* 21:953–961.
62. Brochu J, Vlachos-Breton É, Sutherland S, Martel M, Drolet M. 2018. Topoisomerases I and III inhibit R-loop formation to prevent unregulated replication in the chromosomal Ter region of *Escherichia coli*. *PLoS Genet* 14:e1007668.
63. de Massy B, Fayet O, Kogoma T. 1984. Multiple origin usage for DNA replication in sdrA (rnh) mutants of *Escherichia coli* K-12: initiation in the absence of oriC. *J Mol Biol* 178:227–236.
64. Maduiké NZ, Tehranchi AK, Wang JD, Kreuzer KN. 2014. Replication of the *Escherichia coli* chromosome in RNase HI-deficient cells: multiple initiation regions and fork dynamics. *Mol Microbiol* 91:39–56.

65. Nishitani H, Hidaka M, Horiuchi T. 1993. Specific chromosomal sites enhancing homologous recombination in *Escherichia coli* mutants defective in RNase H. *Mol Gen Genet* 240:307–314.
66. Gowrishankar J. 2015. End of the beginning: elongation and termination features of alternative modes of chromosomal replication initiation in bacteria. *PLoS Genet* 11:e1004909.
67. Leela JK, Syeda AH, Anupama K, Gowrishankar J. 2013. Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in *Escherichia coli*. *Proc Natl Acad Sci* 110:258–263.
68. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. - PubMed - NCBI.
69. Thomason LC, Costantino N, Court DL. 2007. *E. coli* genome manipulation by P1 transduction. *Curr Protoc Mol Biol* Chapter 1:Unit 1.17.
70. Paintdakhi A, Parry B, Campos M, Irnov I, Elf J, Surovtsev I, Jacobs-Wagner C. 2016. Oufiti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis. *Mol Microbiol* 99:767–777.
71. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9:676–682.

72. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma Oxf Engl* 25:1754–1760.
73. McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* 40:4288–4297.
74. Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol Biol Clifton NJ* 1151:165–188.
75. Kuznetsov VA, Bondarenko V, Wongsurawat T, Yenamandra SP, Jenjaroenpun P. 2018. Toward predictive R-loop computational biology: genome-scale prediction of R-loops reveals their association with complex promoter structures, G-quadruplexes and transcriptionally active enhancers. *Nucleic Acids Res* 46:8023.
76. Jenjaroenpun P, Wongsurawat T, Yenamandra SP, Kuznetsov VA. 2015. QmRLFS-finder: a model, web server and stand-alone tool for prediction and analysis of R-loop forming sequences. *Nucleic Acids Res* 43:W527-534.
77. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, Di Ventra M, Garaj S, Hibbs A, Huang X, Jovanovich SB, Krstic PS, Lindsay S, Ling XS, Mastrangelo CH, Meller A, Oliver JS, Pershin YV, Ramsey JM, Riehn R, Soni GV, Tabard-Cossa V, Wanunu M, Wiggin M, Schloss JA. 2008. The potential and challenges of nanopore sequencing. *Nat Biotechnol* 26:1146–1153.

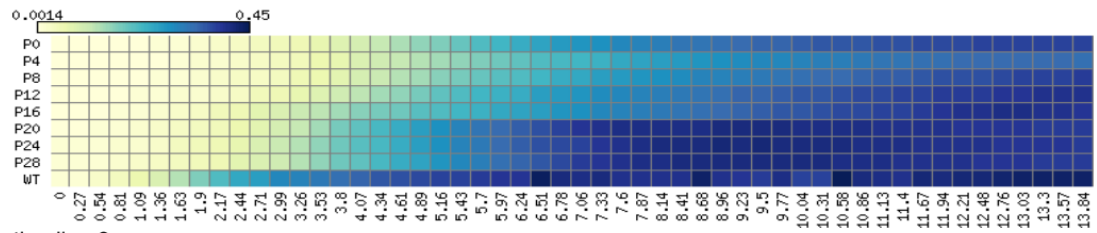
78. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 27:722–736.
79. Duggin IG, Bell SD. 2009. Termination structures in the *Escherichia coli* chromosome replication fork trap. *J Mol Biol* 387:532–539.
80. Drolet M, Brochu J. 2019. R-loop-dependent replication and genomic instability in bacteria. *DNA Repair* 84:102693.
81. Dai Y, Outten FW. 2012. The *E. coli* SufS-SufE sulfur transfer system is more resistant to oxidative stress than IscS-IscU. *FEBS Lett* 586:4016–4022.
82. Viguera E, Petranovic M, Zahradka D, Germain K, Ehrlich DS, Michel B. 2003. Lethality of bypass polymerases in *Escherichia coli* cells with a defective clamp loader complex of DNA polymerase III. *Mol Microbiol* 50:193–204.
83. Dimude JU, Stein M, Andrzejewska EE, Khalifa MS, Gajdosova A, Retkute R, Skovgaard O, Rudolph CJ. 2018. Origins Left, Right, and Centre: Increasing the Number of Initiation Sites in the *Escherichia coli* Chromosome. *Genes* 9.
84. Brooks AN, Turkarslan S, Beer KD, Lo FY, Baliga NS. 2011. Adaptation of cells to new environments. *Wiley Interdiscip Rev Syst Biol Med* 3:544–561.
85. Smits WK, Kuipers OP, Veening J-W. 2006. Phenotypic variation in bacteria: the role of feedback regulation. *Nat Rev Microbiol* 4:259–271.

86. Skovgaard O, Bak M, Løbner-Olesen A, Tommerup N. 2011. Genome-wide detection of chromosomal rearrangements, indels, and mutations in circular chromosomes by short read sequencing. *Genome Res* 21:1388–1393.
87. Usongo V, Martel M, Balleydier A, Drolet M. 2016. Mutations reducing replication from R-loops suppress the defects of growth, chromosome segregation and DNA supercoiling in cells lacking topoisomerase I and RNase HI activity. *DNA Repair* 40:1–17.
88. Raeside C, Gaffé J, Deatherage DE, Tenaillon O, Briska AM, Ptashkin RN, Cruveiller S, Médigue C, Lenski RE, Barrick JE, Schneider D. 2014. Large chromosomal rearrangements during a long-term evolution experiment with *Escherichia coli*. *mBio* 5:e01377-01314.
89. Hawkins M, Malla S, Blythe MJ, Nieduszynski CA, Allers T. 2013. Accelerated growth in the absence of DNA replication origins. *Nature* 503:544–547.
90. Topological stress is responsible for the detrimental outcomes of head-on replication-transcription conflicts | bioRxiv.
91. Merrikh CN, Merrikh H. 2018. Gene inversion potentiates bacterial evolvability and virulence. *Nat Commun* 9:4662.
92. Ryan K, Byrd TF. 2018. *Mycobacterium abscessus*: Shapeshifter of the *Mycobacterial World*. *Front Microbiol* 9.
93. Santi I, McKinney JD. 2015. Chromosome Organization and Replisome Dynamics in *Mycobacterium smegmatis*. *mBio* 6.

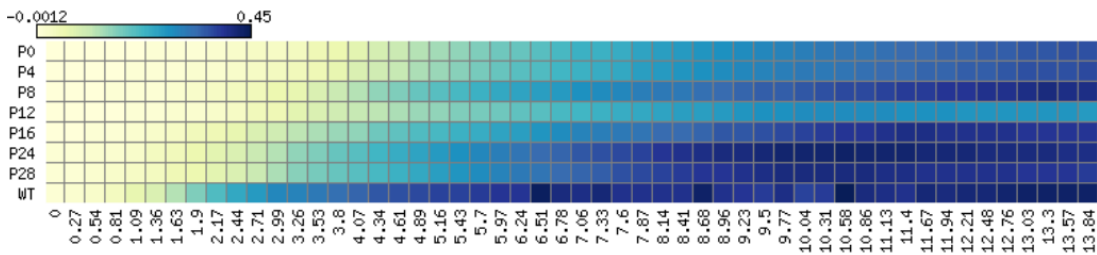
94. Wang X, Liu X, Possoz C, Sherratt DJ. 2006. The two *Escherichia coli* chromosome arms locate to separate cell halves. *Genes Dev* 20:1727–1731.
95. Nielsen HJ, Ottesen JR, Youngren B, Austin SJ, Hansen FG. 2006. The *Escherichia coli* chromosome is organized with the left and right chromosome arms in separate cell halves. *Mol Microbiol* 62:331–338.
96. Grimwade JE, Leonard AC. 2017. Targeting the Bacterial Oriosome in the Search for New Antibiotics. *Front Microbiol* 8:2352.
97. van Eijk E, Wittekoek B, Kuijper EJ, Smits WK. 2017. DNA replication proteins as potential targets for antimicrobials in drug-resistant bacterial pathogens. *J Antimicrob Chemother* 72:1275–1284.

Appendix Material

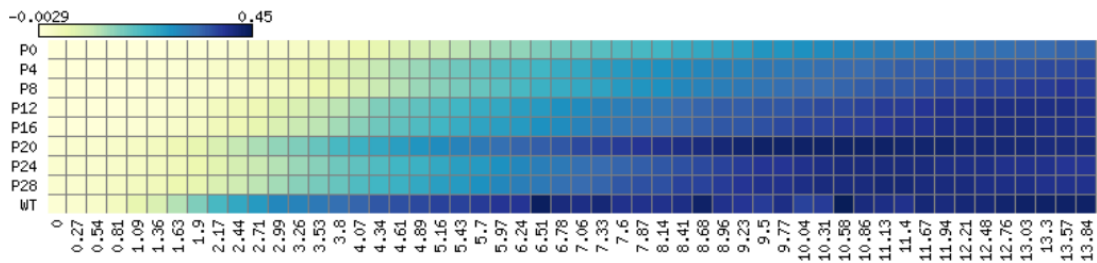
Evolution line 1



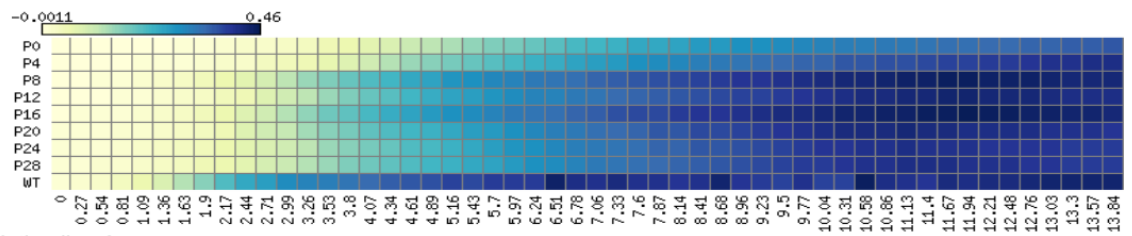
Evolution line 3



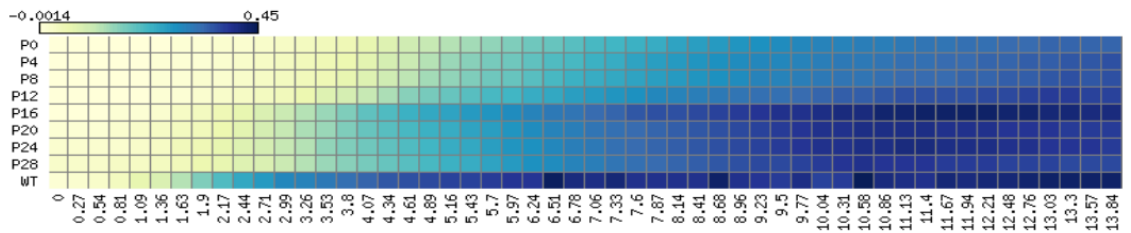
Evolution line 4



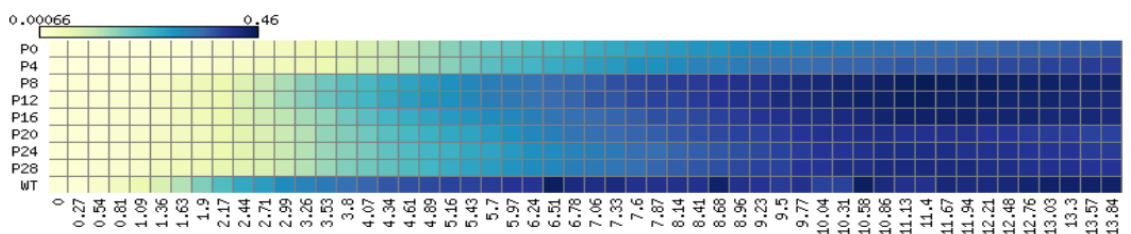
Evolution line 5



Evolution line 6



Evolution line 7



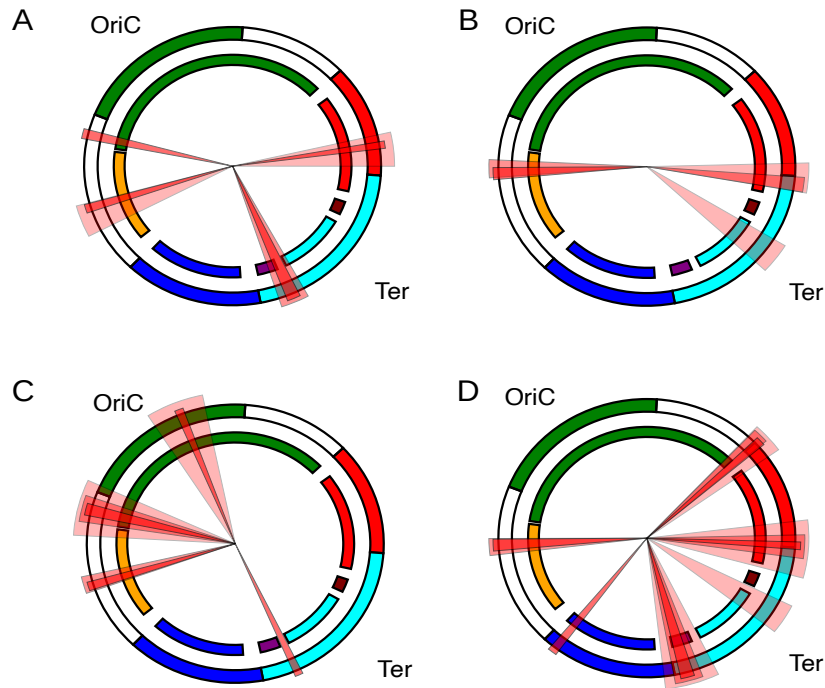


Figure A2: Clustering of differentially expressed genes on the chromosome. (A) and (C) figure represents down regulated genes of $\Delta rnhA$ and $\Delta rnhA-\Delta dnaA$ respectively. (B) and (D) figure represents up regulated genes of $\Delta rnhA$ and $\Delta rnhA-\Delta dnaA$ respectively. Each pie slice represents a chromosomal region that is statistically over-represented ($P < 0.01$) for a certain set of genes, defined here by the set of genes differentially expressed in a stated comparison. To compute whether a set of genes (G) is over-represented in a chromosomal window, the method performs randomizations where the gene order is shuffled *in silico*. For each chromosome window, the number of genes from set G is calculated from the real genome annotations as well as from the randomized sets. The difference between the two gives a measure of statistical significance. This is calculated for several window sizes. For detailed methods, see *Scolari et al. 2012*. The colours in the circle indicate chromosomal macrodomains (green: ori domain; blue: left domain; cyan: ter domain; red: right domain; white: unstructured elements). This representation is derived from the NuST server (<http://www.lgm.upmc.fr/nust/>).

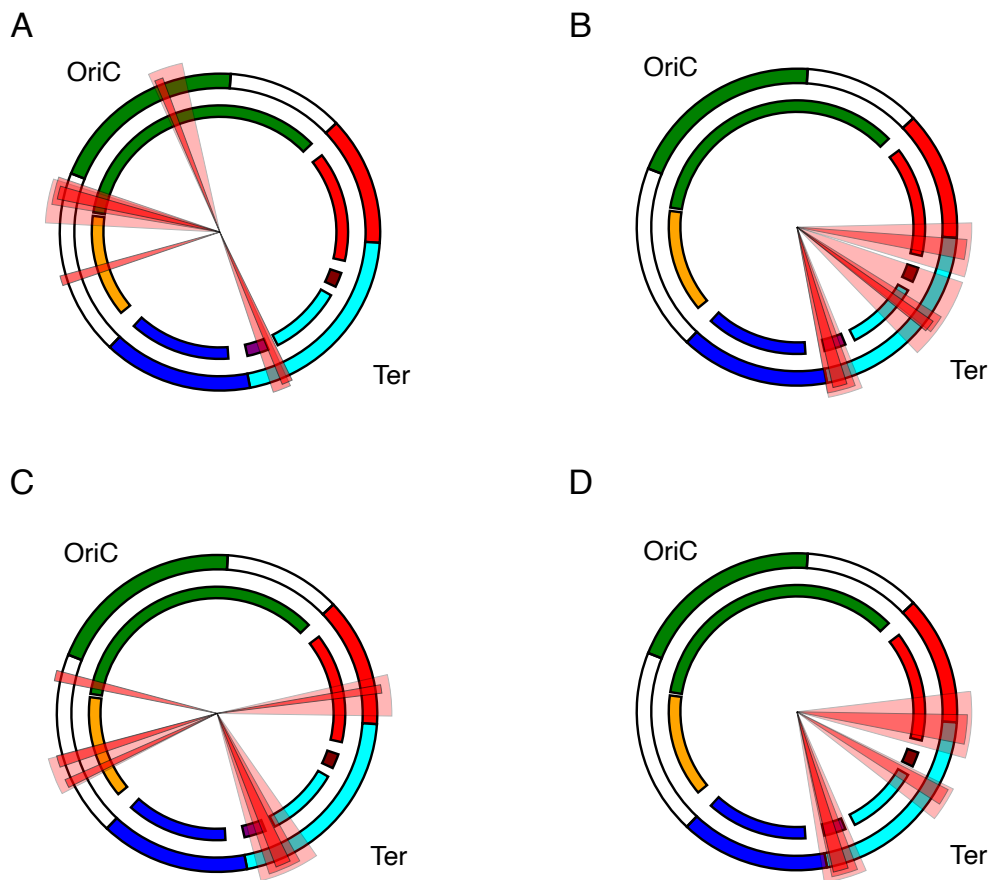
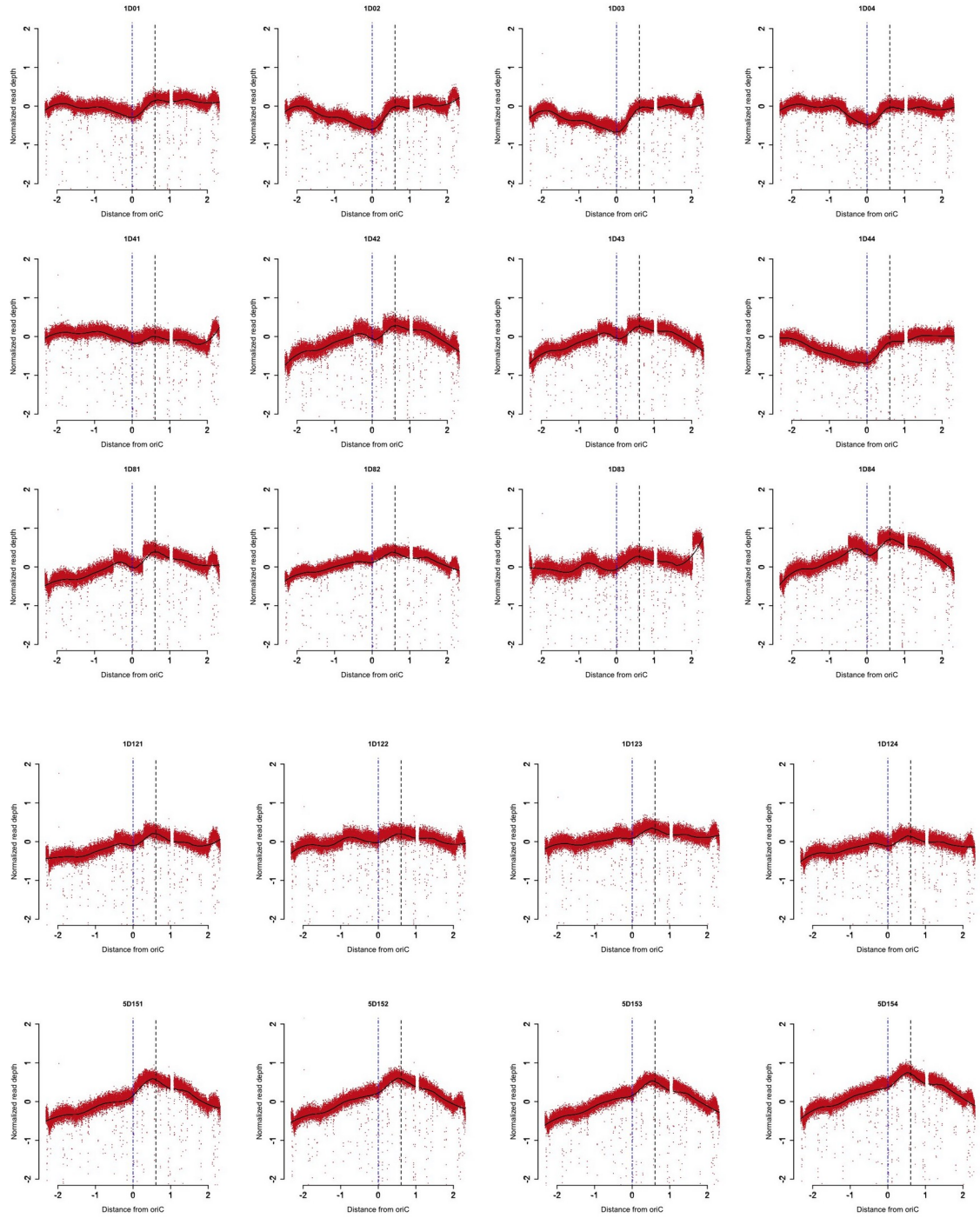
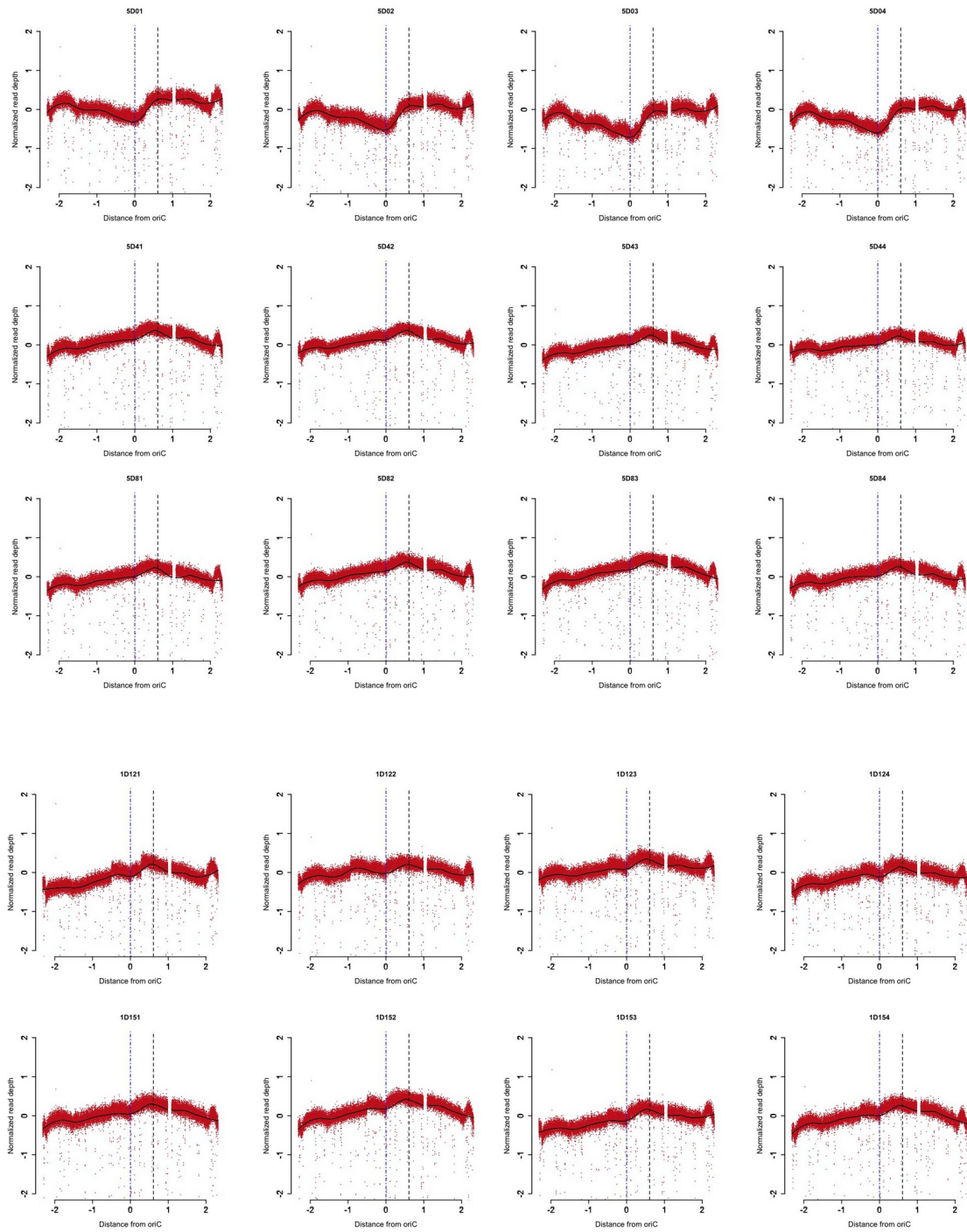


Figure A3: Clustering of differentially expressed genes on the chromosome. (A) and (C) figure represents down regulated genes of $\Delta rn h A-\Delta d n a A i n v^{r m D-r m C}$ and $\Delta rn h A-\Delta d n a A i n v^{r m D-r m E}$ respectively. (B) and (D) figure represents up regulated genes of $\Delta rn h A-\Delta d n a A i n v^{r m D-r m C}$ and $\Delta rn h A-\Delta d n a A i n v^{r m D-r m E}$ respectively. Each pie slice represents a chromosomal region that is statistically over-represented ($P < 0.01$) for a certain set of genes, defined here by the set of genes differentially expressed in a stated comparison. To compute whether a set of genes (G) is over-represented in a chromosomal window, the method performs randomizations where the gene order is shuffled in silico. For each chromosome window, the number of genes from set G is calculated from the real genome annotations as well as from the randomized sets. The difference between the two gives a measure of statistical significance. This is calculated for several window sizes. For detailed methods, see Scolari et al. 2012. The colours in the circle indicate chromosomal macrodomains (green: ori domain; blue: left domain; cyan: ter domain; red: right domain; white: unstructured elements). This representation is derived from the NuST server (<http://www.lgm.upmc.fr/nust/>).

Evolution line 1



Evolution line 5



Evolution line 8

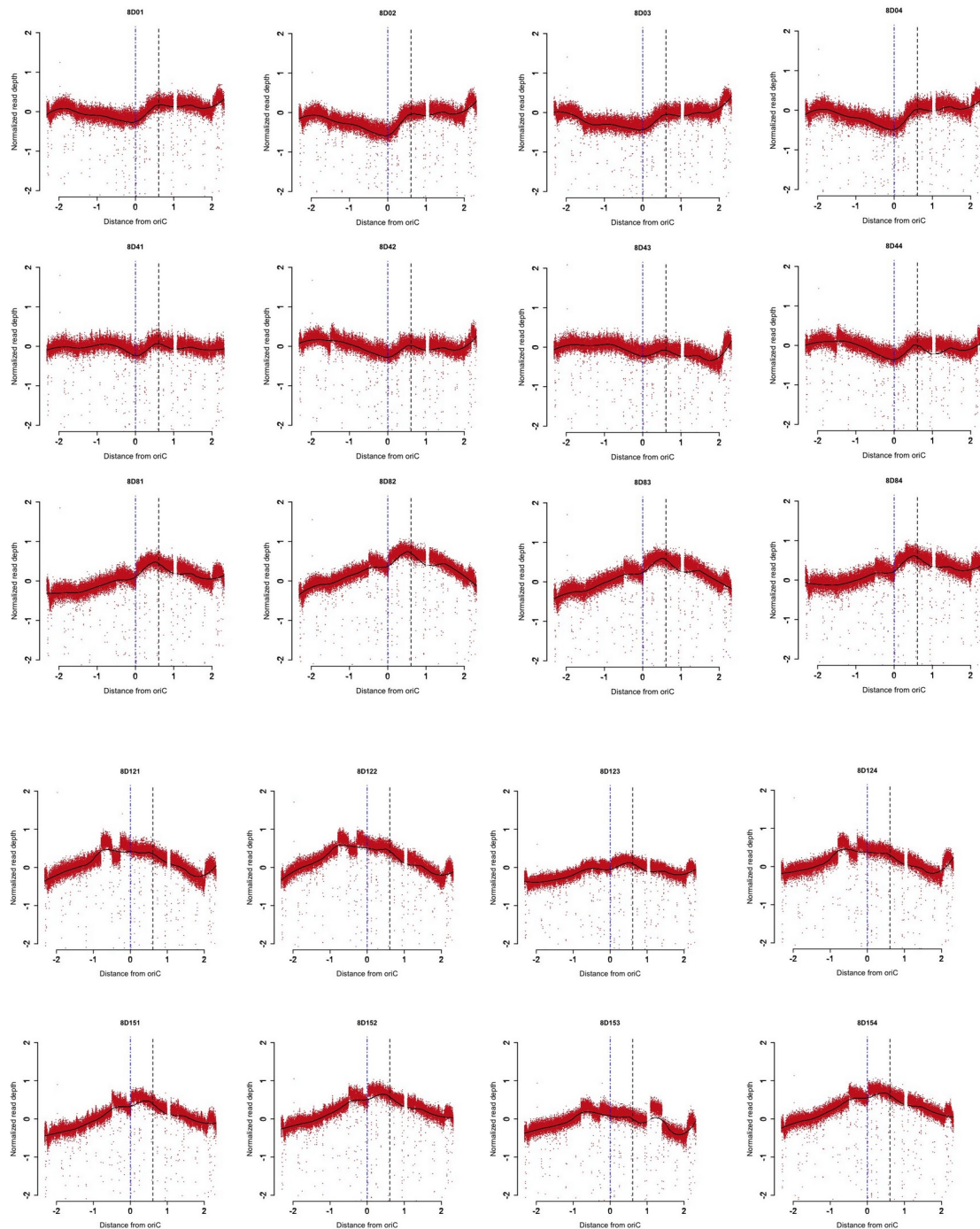


Figure A4: Deep Sequencing based MFA of suppressor mutants at the exponential phase of growth: MFA plots for all suppressor mutants sequenced from three independently evolved lines at the exponential phase of growth. The dotted blue line represents the *oriC* position and the black line represents the predicted position of *oriK45* site.

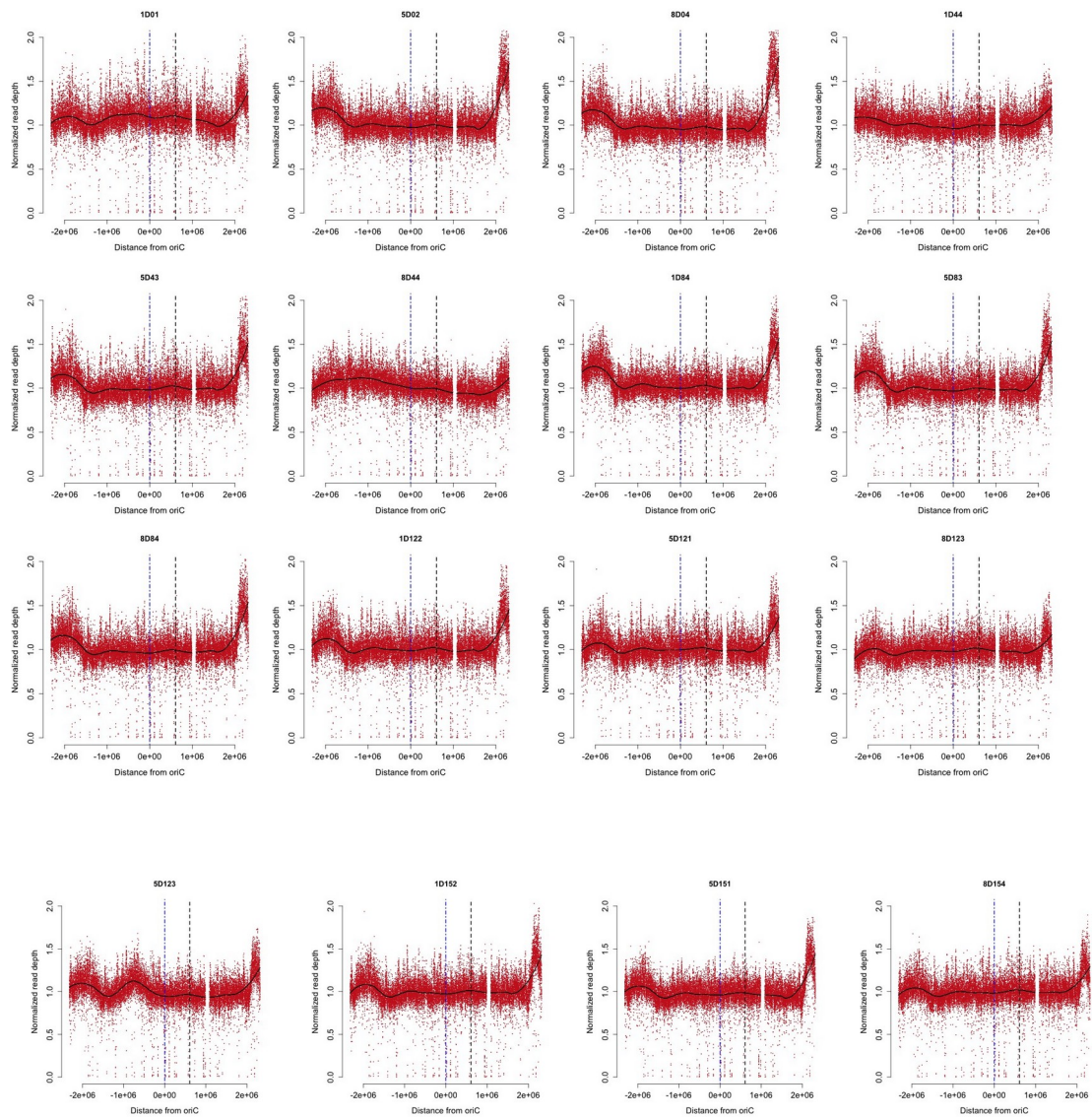


Figure A5: Deep Sequencing based MFA of suppressor mutants at the stationary phase of growth: MFA plots for all suppressor mutants sequenced from three independently evolved lines at the stationary phase of growth. The dotted blue line represents the *oriC* position and the black line represents the predicted position of *oriK45* site.

Table A1

GCF_000008885	Wigglesworthia_glossinidia_endosymbiont_of_Glossina_brevipalpis_plasmid_pWb1_DNA
GCF_000011745	Candidatus_Blochmannia_pennsylvanicus_str._BPEN
GCF_000013185	Baumannia_cicadellinicola_str._Hc_(Homalodisca_coagulata)
GCF_000022605	Blattabacterium_sp._(Blattella_germanica)_str._Bge_plasmid_pBge
GCF_000025125	Candidatus_Atelocyanobacterium_thalassa_isolate_ALOHA
GCF_000043285	Blochmannia_floridanus_complete_genome
GCF_000093065	Candidatus_Riesia_pediculicola_USDA_plasmid_pPAN
GCF_000146025	Uncultured_Termite_group_1_bacterium_phylotype_Rs-D17_plasmid_pTGRD3 DNA
GCF_000165505	Ilyobacter_polytropus_DSM_2926_plasmid_pILYOP02
GCF_000177535	Corynebacterium_resistens_DSM_45100
GCF_000179035	Mycoplasma_suis_str._Illinois
GCF_000185985	Candidatus_Blochmannia_vafer_str._BVAF
GCF_000196515	'Nostoc_azollae'_0708_plasmid_pAzo02
GCF_000203215	Mycoplasma_suis_KI3806_complete_genome
GCF_000219175	Candidatus_Moranella_endobia_PCIT
GCF_000223375	Ketogulonigenium_vulgarum_WSH-001_plasmid_2
GCF_000236405	Blattabacterium_sp._(Cryptocercus_punctulatus)_str._Cpu_plasmid_pCpu
GCF_000247565	Wigglesworthia_glossinidia_endosymbiont_of_Glossina_morsitans_morsitans
GCF_000255275	Corynebacterium_diphtheriae_PW8
GCF_000262655	Helicobacter_pylori_XZ274_plasmid_pXZ274

GCF_000287295	Candidatus_Carsonella_ruddii_HT_isolate_Thao2000
GCF_000292685	Candidatus_Portiera_aleyrodidarum_BT-B
GCF_000298385	Candidatus_Portiera_aleyrodidarum_BT-QVLC
GCF_000300035	Candidatus_Portiera_aleyrodidarum_BT-QVLC
GCF_000300075	Candidatus_Portiera_aleyrodidarum_BT-B
GCF_000304735	Borrelia_afzelii_HLJ01
GCF_000317675	Cyanobacterium_aponinum_PCC_10605_plasmid_pCYAN10605.01
GCF_000319385	Candidatus_Endolissoclinum_faulkneri_L2
GCF_000331065	Candidatus_Blochmannia_chromaiodes_str._640
GCF_000364725	Candidatus_Moranella_endobia_PCVAL
GCF_000441555	Candidatus_Proffotella_armatura_plasmid
GCF_000471965	Blattabacterium_sp._(Nauphoeta_cinerea)_plasmid
GCF_000477415	Mycoplasma_parvum_str._Indiana
GCF_000505725	Francisella_noatunensis_subsp._orientalis_LADL--07-285A
GCF_000508245	Mycoplasma_ovis_str._Michigan
GCF_000604125	Treponema_pallidum_subsp._pallidum_str._Sea_81-4
GCF_000709555	Endosymbiont_of_Llaveia_axin_axin
GCF_000767685	Corynebacterium_ulcerans_FRC11
GCF_000769635	Corynebacterium_ulcerans_strain_05146
GCF_000770175	Mycobacterium_abscessus_strain_DJO-44274
GCF_000815025	Coxiella_endosymbiont_of_Amblyomma_americanum
GCF_000827855	Candidatus_Portiera_aleyrodidarum_MED_(Bemisia_tabaci)_strain_BT-Q

GCF_000828815	Candidatus_Tachikawaea_gelatinosa_DNA
GCF_000828835	Thioploca_ingrica_DNA
GCF_000829235	Cyanobacterium_endosymbiont_of_Epithemia_turgida_isolate_EtSB_Lake_Yunoko_DNA
GCF_000953435	Candidatus_Evansia_muelleri_genome_assembly_CEM1.1
GCF_000973545	Blochmannia_endosymbiont_of_Camponotus_(Colobopsis)_obliquus_strain_757
GCF_001021025	Corynebacterium_epidermidicanis_strain_DSM_45586
GCF_001278785	Candidatus_Profftella_armatura_strain_YCPA_plasmid
GCF_001318295	Candidatus_Xiphinematobacter_sp._Idaho_Grape
GCF_001548095	Geminocystis_sp._NIES-3708_plasmid_pGM05_DNA
GCF_001548115	Geminocystis_sp._NIES-3709_plasmid_pGM3709_11_DNA
GCF_001587015	Campylobacter_jejuni_strain_OD267_plasmid_pCJDM67_S
GCF_001682195	Flammeovirga_sp._MY04_plasmid
GCF_900048035	Enterobacteriaceae_bacterium_symbiont_of_Ferrisia_virgata_isolate_GEFVIR_genome_assembly
GCF_900048045	Enterobacteriaceae_bacterium_symbiont_of_Paracoccus_marginatus_isolate_MEPMAR_genome_assembly

Table A1: List of eubacterial strains which did not have a *dnaA* homologue detected in their genomes (out of 5976 organisms).

Table A2

Strain	Genotype	Source
<i>GJ13519</i>	MG1655 $\Delta(\text{argF-lac})\text{U169}$	LBG, CDFD
<u>Derivatives of GJ13519</u>		
ΔrnhA	$\Delta\text{rnhA}::\text{FRT}$	LBG, CDFD
$\Delta\text{dnaA}/\text{dnaA}^+$	$\Delta\text{dnaA}::\text{FRT}/\text{pHYD2388}$	LBG, CDFD
$\Delta\text{rnhA-dnaA}/\text{dnaA}^+$	$\Delta\text{dnaA}::\text{FRT } \Delta\text{rnhA}::\text{FRT}/\text{pHYD2388}$	LBG, CDFD
$\Delta\text{rnhA-dnaA}$	$\Delta\text{dnaA}::\text{FRT } \Delta\text{rnhA}::\text{FRT}$	This study
<i>GJ13519/dnaA</i> ⁺	<i>GJ13519/pHYD2388</i>	This study
<i>GJ13519</i> $\Delta\text{hotH}/\text{dnaA}^+$	<i>GJ13519 } \Delta/\text{pHYD2388}</i>	This study
$\Delta\text{rnhA-dnaA}\Delta\text{hotH}/\text{dnaA}^+$	$\Delta\text{dnaA}::\text{FRT } \Delta\text{rnhA}::\text{FRT } \Delta\text{4555284: 45660615}(\text{uxuR-yjiN}) / \text{pHYD2388}$	This study
$\Delta\text{rnhA-dnaA}\Delta\text{hotH}$	$\Delta\text{dnaA}::\text{FRT } \Delta\text{rnhA}::\text{FRT } \Delta\Delta\text{4555284: 45660615}(\text{uxuR-yjiN})$	This study
<i>pUA139</i>	Low copy plasmid with fast folding GFP mut2	SAFS lab
<i>pUA139::Wt rrnD IGR</i>	<i>pUA139</i> vector carrying 598bp <i>rrsD-yrdA</i> intergenic region sequence	This study
<i>pUA139::Mut rrnD IGR</i>	<i>pUA139</i> vector carrying 598bp <i>rrsD-yrdA</i> mutant intergenic [G-A(3,429,052), +A(3,429,054)] sequence	This study
<i>pHYD2388</i>	<i>pMU575</i> derivative carrying <i>S.enterica dnaA</i> ⁺	LBG, CDFD

Table A2: Strains and plasmids used in this study

Table A3

Primer Description	Sequence 5' -----> 3'
<i>HotHKO pKD13 F</i>	GTTGACGATATTTATTTTGATGGCTATCTGTTTGATgtgtaggctggagctgct tcg
<i>HotHKO pKD13 R</i>	TTCGCTGGCTGGAGAGCGAGCATCCACTGAAAGCCAattccgggatccg tcgacc
<i>rrsD -yrdA IGR F</i>	ATTACTCGAGTCGTCAGCGAAACAGCAA
<i>rrsD-yrdA IGR R</i>	TAATAGATCTGTATGGGCGTAAAACATC

Table A3: Primers used in this study