Laboratory Evolution Experiments Help Identify a Predominant Site of *ΔrnhA-dnaA* Constitutive Stable DNA Replication Initiation

Author list and Affiliations

Reshma T Veetil[1,2], Nitish Malhotra[1], Akshara Dubey[1], Aswin Sai Narain Seshasayee[1,*]

1. National Centre for Biological Sciences, Tata Institute of Fundamental Research, Gandhi Krishi Vigyan Kendra, Bangalore, Karnataka 560065, India.
2. School of Life science, The University of Trans-Disciplinary Health Sciences & Technology (TDU), Bengaluru, Karnataka 560064, India

*aswin@ncbs.res.in

Abstract

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 *ΔrnhA* background. This phenomenon is called constitutive stable DNA replication (cSDR). 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 use laboratory evolution experiments of *ΔrnhA-dnaA* followed by deep sequencing to show that DNA replication preferentially initiates at a site ~0.6 Mb clockwise of *oriC.* Initiation from this site 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 however 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.

Introduction

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 (Mott and Berger 2007). Replication proceeds bidirectionally outwards of *oriC* before terminating at a locus positioned diametrically opposite to *oriC* on the circular chromosome (Duggin and Bell 2009)

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 (Rocha 2004). The positioning of such genes close to *oriC* is conserved, and more so in fast growing bacteria (Couturier and Rocha 2006; Khedkar and Seshasayee 2016). Repositioning of such genes away from *oriC* or on the lagging strand can be detrimental to fitness, especially in nutrient rich conditions (J. D. Wang, Berkmen, and Grossman 2007; Bryant et al. 2014; Srivatsan et al. 2010).

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 (Supplementary table 1). Mitochondria are not known to use *oriC*-DnaA-based DNA replication initiation (Clayton 1982; Yasukawa and Kang 2018). In *E. coli* 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) (Tokio Kogoma 1997) .

Two broad types of SDR - each with its own set of genetic requirements - have been described. Inducible SDR (iSDR) requires the SOS DNA damage response (T. Kogoma, Torrey, and Connaughton 1979; Tokio Kogoma 1997) Constitutive SDR (cSDR) is activated by processes that stabilise RNA-DNA hybrids or R-loops (Tokio Kogoma 1997). These include inactivation of (a) RnhA, the RNA-DNA hybrid nuclease RNaseHI (Ogawa et al. 1984); (b) RecG, a helicase for RNA-DNA hybrids (Hong, Cadwell, and Kogoma 1995) and (c) the topoisomerase I TopA, which results in hyper negative supercoiling and elevated occurrence of RNA-DNA hybrids (Martel et al. 2015). Excessive R-loops have also been proposed to occur in strains defective for Rho-dependent transcription termination (J. Gowrishankar and Harinarayanan 2004; J. Gowrishankar, Leela, and Anupama 2013; Harinarayanan and Gowrishankar 2003; Raghunathan et al. 2018), though to our knowledge Rho-dependent transcription termination has not been 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 cSDR(Raghunathan et al. 2019). We note here that DNA replication by SDR is under normal conditions sub-optimal relative to canonical DNA replication. At least one report has described nSDR, as a non-*oriC*, non-DnaA dependent mechanism of chromosome replication

employed by *E. coli* cells transiently during the stationary phase (Hong, Cadwell, and Kogoma 1996). nSDR may be a manifestation of cSDR.

In this paper, we focus on *ΔrnhA* induced cSDR in *ΔdnaA* 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? The Kogoma group, employing traditional marker frequency analysis (MFA), had identified five '*oriK*' loci at which replication might initiate (de Massy, Fayet, and Kogoma 1984). 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, Maduike 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 Maduike et al. study mapped within the terminus of replication (Maduike et al. 2014). 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 (Nishitani, Hidaka, and Horiuchi 1993). However, both Maduike et al. and Nishitani et al. agree that the terminus sites identified in their studies are not bonafide *oriK* sites (Maduike et al. 2014; Nishitani, Hidaka, and Horiuchi 1993). In the Maduike 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 (Maduike et al. 2014). 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 (Nishitani, Hidaka, and Horiuchi 1993). Gowrishankar has synthesised these arguments (Jayaraman Gowrishankar 2015), and in conjunction with his lab's finding that RNA-DNA hybrids can occur throughout the chromosome (Leela et al. 2013), 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 (Brochu et al. 2018).

Here we attempt to answer the question of the existence of preferred *oriK* sites by taking the position that peak identification in high resolution MFA studies of cSDR is complicated by the slow growth phenotype of the parent strain, which results in weak origin to terminus copy number gradients. We address this using laboratory evolution experiments, generating suppressors that can generate strong copy number gradients even under the cSDR regime, while also identifying a principle underlying the suppression of the slow growth phenotype of cSDR.

Results and Discussion

ΔrnhA-dnaA strain of *E. coli* K12

The gene *rnhA* encodes the RNaseHI nuclease that removes RNA-DNA hybrids. The *ΔrnhA* mutant displays cSDR and therefore suppresses the lethality of *ΔdnaA* and *ΔoriC* mutants (Ogawa et al. 1984). We obtained a *ΔrnhA* single deletion mutant and a *ΔrnhA-dnaA-pHYD2388* ($dnaA^+lacZ^+$)

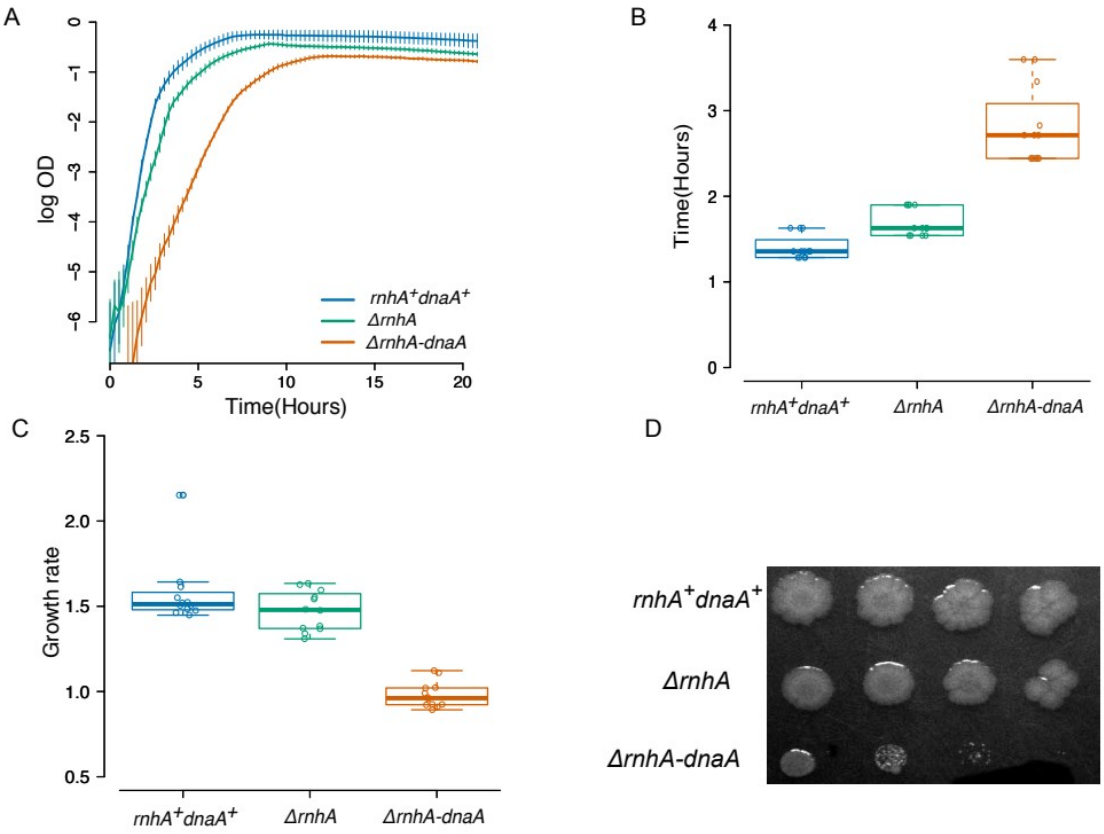

Figure 1: *ΔrnhA-dnaA* shows reduced growth in LB media. (A) growth curves of *rnhA*$^+$*dnaA*$^+$, *ΔrnhA*, and *ΔrnhA-dnaA* in LB at 37°C, 200 rpm. X-axis indicates time and Y-axis indicates $\log_2 OD_{600}$. (B) and (C) box plots for lag time and growth rate followed by each strain respectively. *ΔrnhA-dnaA* shows reduced growth rate and extended lag phase compared to *rnhA*$^+$*dnaA*$^+$ (P<<<0.001 , Wilcoxon test, one-tailed). (D) Spotting assay for *rnhA*$^+$*dnaA*$^+$, *ΔrnhA*, and *ΔrnhA-dnaA* 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 .

mutant of *E. coli* K12 (MG1655) from Prof. J. Gowrishankar's lab (Raghunathan et al. 2019). To obtain *ΔrnhA-dnaA*, we plated overnight cultures of *ΔrnhA-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 *ΔrnhA-dnaA* strain.

The *ΔrnhA* 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 (Figure 1). The *ΔrnhA-dnaA* double deletion mutant showed a more severe growth defect in LB, displaying an extended lag phase and a reduced maximal growth rate (Figure 1). Consistent with previous results (J. U. Dimude et al. 2015) indicating elevated SOS response in cSDR, *ΔrnhA-dnaA* cells are longer than *ΔrnhA* and *rnhA*$^+$*dnaA*$^+$ (Supplementary figure 1)

In the rest of this manuscript, we use '*ori*' as an umbrella term, when required, to refer to all sites at which replication initiates: this may include *oriC* itself or *oriK* sites at which cSDR initiates. The terminus is a more complex sequence with multiple, directional replication termination motifs at which the Tus protein traps moving replication forks; we use the generic term '*ter*' to refer to the locus bounded by these termination motifs.

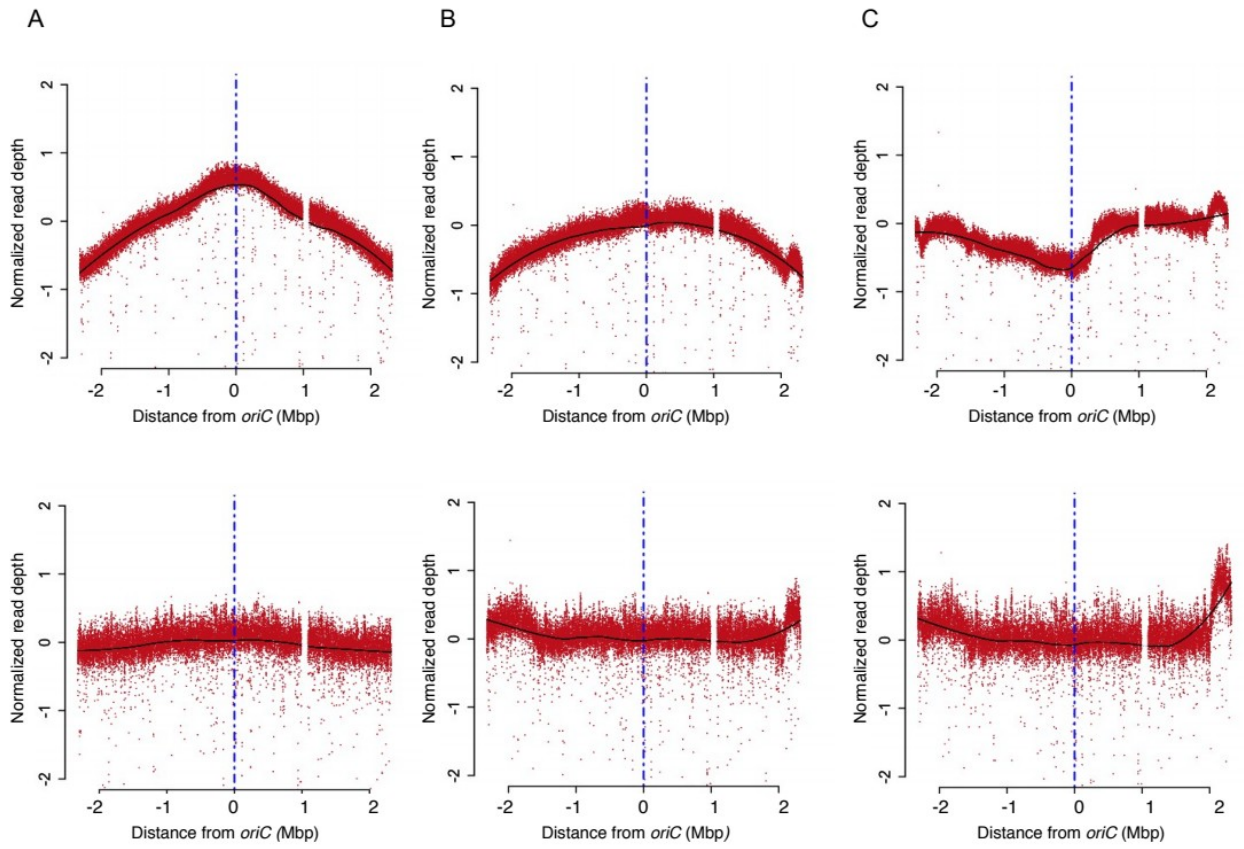

Figure 2: Deep sequencing based MFA plots for *rnhA+dnaA+, ΔrnhA* and *ΔrnhA-dnaA.* The upper panels show the MFA plots for (A) *rnhA*$^+$*dnaA*$^+$, (B) *ΔrnhA* and (C) *ΔrnhA-dnaA* at the exponential phase of growth and the lower panels show the same for the stationary phase. The X-axis represents 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 methods).

Next Generation Sequencing based MFA of *ΔrnhA* and *ΔrnhA-dnaA*

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 (Rocha 2004). 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. 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, including those cited in various places in this manuscript, 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-dnaA* strains of *E. coli* grown to exponential phase - corresponding to the culture's highest growth rate - 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*, 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 2A). The corresponding plot for stationary phase cells was relatively flat (Figure 2A lower panel).

In *ΔrnhA*, in which both *oriC*-DnaA-dependent replication initiation and cSDR are active, we observed a prominent peak at *oriC* (Figure 2B). 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 2B 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. The pattern observed here differs from that reported by Maduike et al. in that Maduike et al. do not report asymmetry in the copy number profile of *ΔrnhA* either side of *oriC* (Maduike et al. 2014). The asymmetry around *oriC* that we observe is similar however to that reported by Dimude et al. for *ΔrnhA* (J. U. Dimude et al. 2015).

The strongest peak in the exponential phase copy number plot for *ΔrnhA-dnaA* was within *ter,* wherein the pattern observed was similar to that in *ΔrnhA* but more prominent (Figure 2C). 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 around 0.6 Mb away. 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 Maduike et al and by Dimude et al. (Maduike et al. 2014; J. U. Dimude et al. 2015).

Maduike et al. had described a few peaks in their MFA data, which showed bumps in our data as well (Supplementary Table 2, Supplementary Figure 2). The most prominent peak in the Maduike et al. dataset, as well as ours, was within *ter*. Using MFA analysis of *Δtus-rnhA-dnaA*, which abolished this peak, Maduike et al. argued that the *ter* peak did not represent an *oriK* site, but the trapping of forks originating outside *ter* (Maduike et al. 2014). 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 2C lower panel).

Maduike et al. observe a peak ~0.6 Mb clockwise of *oriC* (Maduike et al. 2014). We henceforth refer to this site as *oriK45* for it being located ~4.5 Mb along the genome sequence of *E. coli* K12 MG1655 (Genbank ID NC_000913 version 3). In the figures presented by Dimude et al., exponential phase copy number data when normalised by stationary phase data in fact ranks the *ter* peak below the flat regime clockwise of *oriK45* (J. U. Dimude et al. 2015). This may be consistent with our observation that the *ter* peak is prominent in stationary phase as well. Maduike et al. wonder whether *oriK45* is an *oriK* site or a representation of replication initiation at fork reversals initiated by head-on collisions between the DNA polymerase and RNA polymerases transcribing the four rRNA operons encoded in the 0.6 Mb region separating *oriC* and *oriK45* (Maduike et al. 2014) If *oriK45* were the only cSDR initiation site and if we could assume bidirectional fork movement from this site, the flatness of the graph clockwise towards *ter* can be explained by the slow growth phenotype of the bacterial population. The gentle decline clockwise of *ter* towards *oriC* can be a result of a presumed partial rate of escape of the fork from trapping at *ter*. The sharp decline

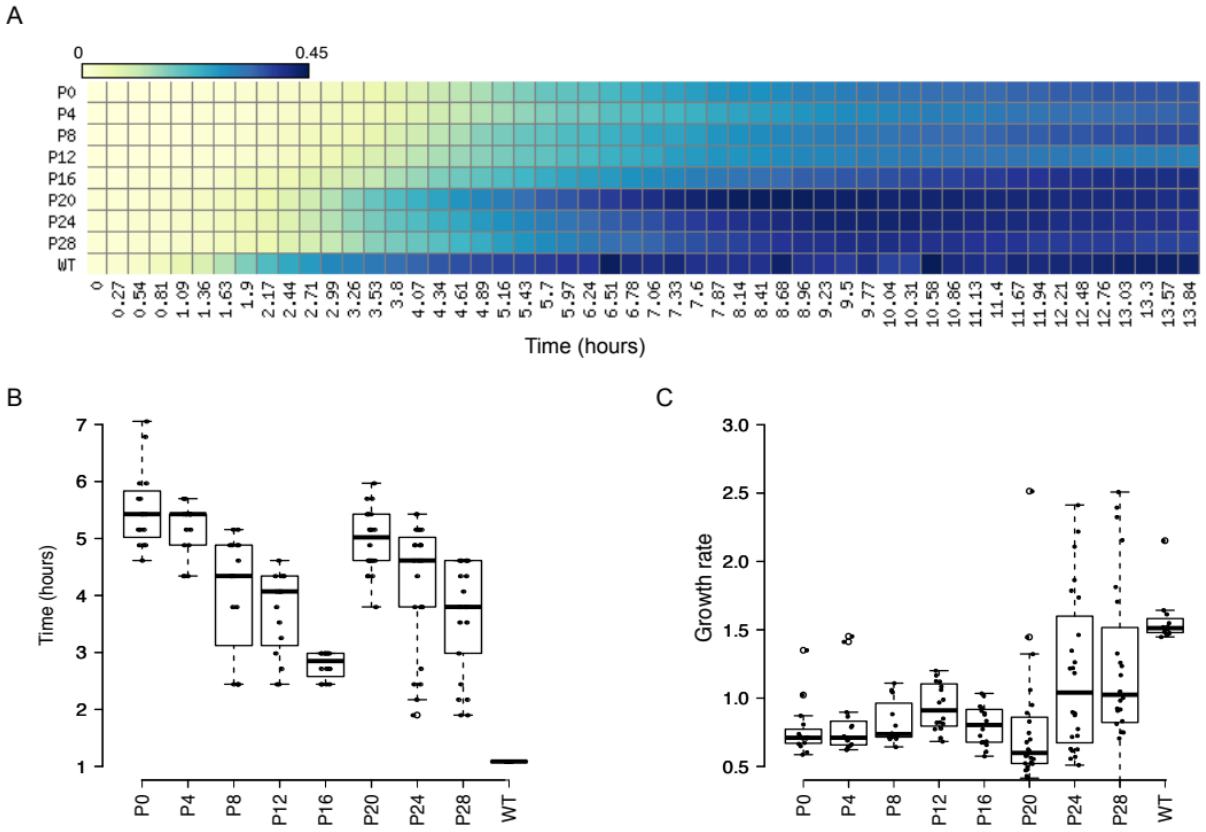

Figure 3: Growth characteristics of evolved mutants: (A) Heat map representing growth of an independently evolved population of *ΔrnhA-dnaA* from passage 0 (P0) to passage 28 (P28) based on OD measurements. X-axis shows time in hours, Y-axis shows the number of passage and the colours represent mean OD values as indicated in the colour bar. (B) and (C) box plots for lag time and growth rate followed by all populations respectively. Passage 28 population shows a significantly greater growth rate than that of parental (P0) strains (P<<< 0.001, Wilcoxon test, one tailed).

counter-clockwise from *oriK45* towards *oriC* may be a consequence of fork loss from head-on replication-transcription conflicts at rRNA operons. However, there is also a copy number peak clockwise of *ter*. However, we observe this peak in the stationary phase data as well, indicating that this again may not be a growth-related replication initiation locus.

Whether *oriK45* is a genuine replication initiation site, whether it is indeed a 'preferred' site, and whether other minor peaks around the chromosome can represent substantial *oriK*s remain complicated to answer with the present dataset. This is at least in part because of the slow growth phenotype of the mutant which ensures that there is hardly any *ori-ter* copy number gradient even during periods of its highest growth rate.

Laboratory evolution experiments and suppressors of the growth defect of *ΔrnhA-dnaA*

To obtain cSDR strains that grow fast and therefore display strong *ori-ter* gradients, we performed laboratory evolution experiments in which *ΔrnhA-dnaA* was iteratively diluted into fresh LB and grown to saturation. We used eight independent lines, each derived from a single *ΔrnhA-dnaA*

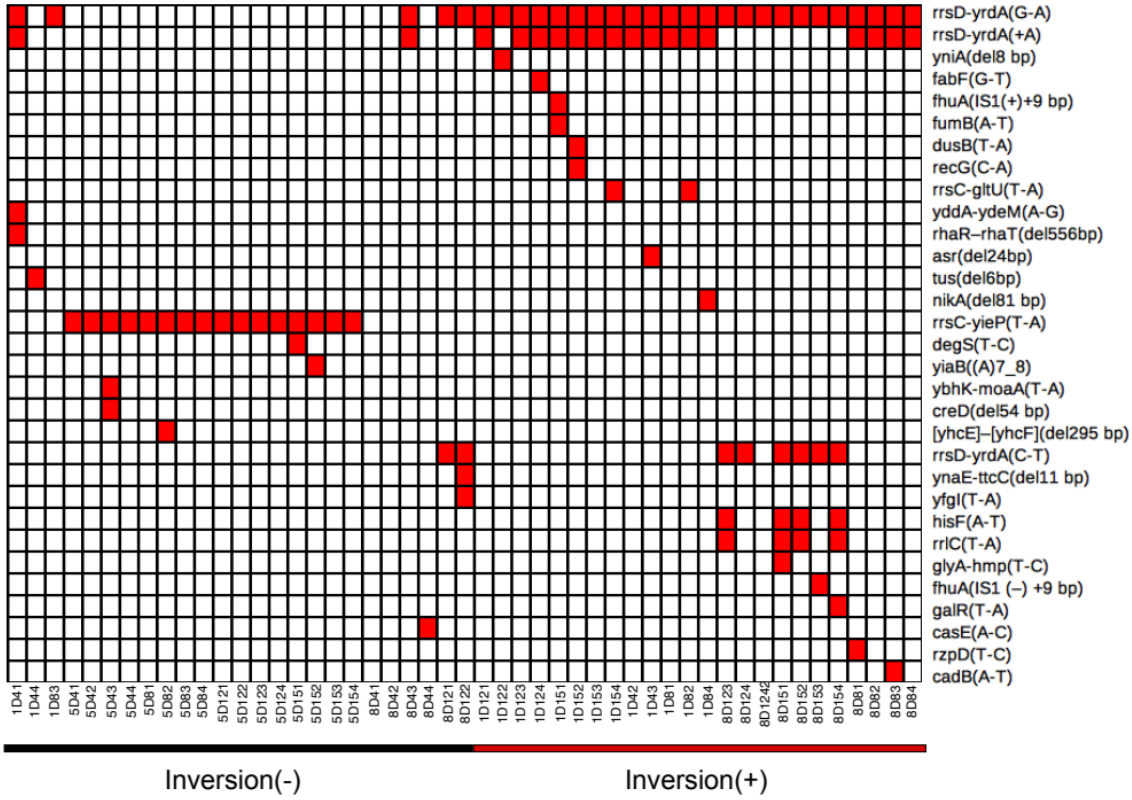

Figure 4: Unique mutations in suppressor mutants. Heatmap representing unique mutations (100% frequency) in all independent colonies sequenced generated using matrix2png. Colour represents the presence of a mutation in the respective gene shown on Y-axis. X-axis represents sample IDs of suppressor mutants evolved from three independent populations. Presence and absence of chromosomal inversions are represented using a red and black lines respectively.

colony to 36 rounds of dilution and growth, corresponding to an estimated 288 generations. Over time, the growth of the population substantially improved (Figure 3 and supplementary figure 3).

We plated aliquots of the culture after each day and noticed the presence of colonies that were visibly larger than those of the parent *ΔrnhA-dnaA*. We randomly picked 60 colonies of varying sizes - sampling across 3 independently evolved populations and 5 time-points - and subjected their genomic DNA to Illumina sequencing. Similar to our sequencing runs with the parent *ΔrnhA-dnaA*, *ΔrnhA* and $rnhA^{+}dnaA^{+}$, 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 Materials and Methods), and found that this ratio ranged between 1.04 and 2.6 (Supplementary table 3). At the lower end, a few colonies showed gradients not too different from the *ΔrnhA-dnaA* parent. The steepest gradients approached, but rarely matched that of $rnhA^{+}dnaA^{+}$.

We next used these sequencing data to identify mutations - both point variations including indels, as well as structural variations such as large amplifications, deletions and inversions. Large 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 (Skovgaard et al. 2011). We found several point mutations in the evolved clones, not present in the *ΔrnhA-dnaA* parent (Figure 4 and Supplementary figure 4). ~90% of colonies carried a mutation upstream of one of two rRNA operons, *rrnD* and *rrnC.* One clone carried an in-frame deletion mutation in *tus (*Δ6 bp (1,684,458-1,684,463), which translates to a QSL-L variation. 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 (Raghunathan et al. 2019).

We found inversions around *oriC* in ~50% of the evolved colonies (Figure 4 and Figure 5). One end of these inversions was *rrnD*, located 3.42 Mb counterclockwise of *oriC* in the reference genome of *E. coli* K12 MG1655. In ~80% of inversions, the other end was *rrnC* (3.94 Mb), and in the remaining, the second end was *rrnE* (4.2Mb). The *rrnD-rrnC* inversion ($\Delta rnhA\text{-}dnaA\text{-}inv^{rrnD\text{-}rrnC}$) measured ~0.5 Mb and the *rrnD-rrnE* ($\Delta rnhA\text{-}dnaA\text{-}inv^{rrnD\text{-}rrnE}$), ~0.8 Mb (Figure 5). 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 (Supplementary figure 5A).

Thus both inversions would move a set of rRNA operons from clockwise to counterclockwise of *oriC*, and the *rrnD* operon in the opposite direction. 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) (Hill and Harnish 1981)(Supplementary figure 5B and Supplementary table 4). 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).

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 population. Clones with the longer *rrnE-rrnD* inversion showed significantly higher maximum-minimum ratios than those with the shorter *rrnC-rrnD* inversion (P = 0.02, Wilcoxon test one-tailed) (Figure 5F). 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 Δ6 bp inframe deletion mutation in Tus, showed substantially smaller maximum-minimum ratios, though a few colonies did show higher values.

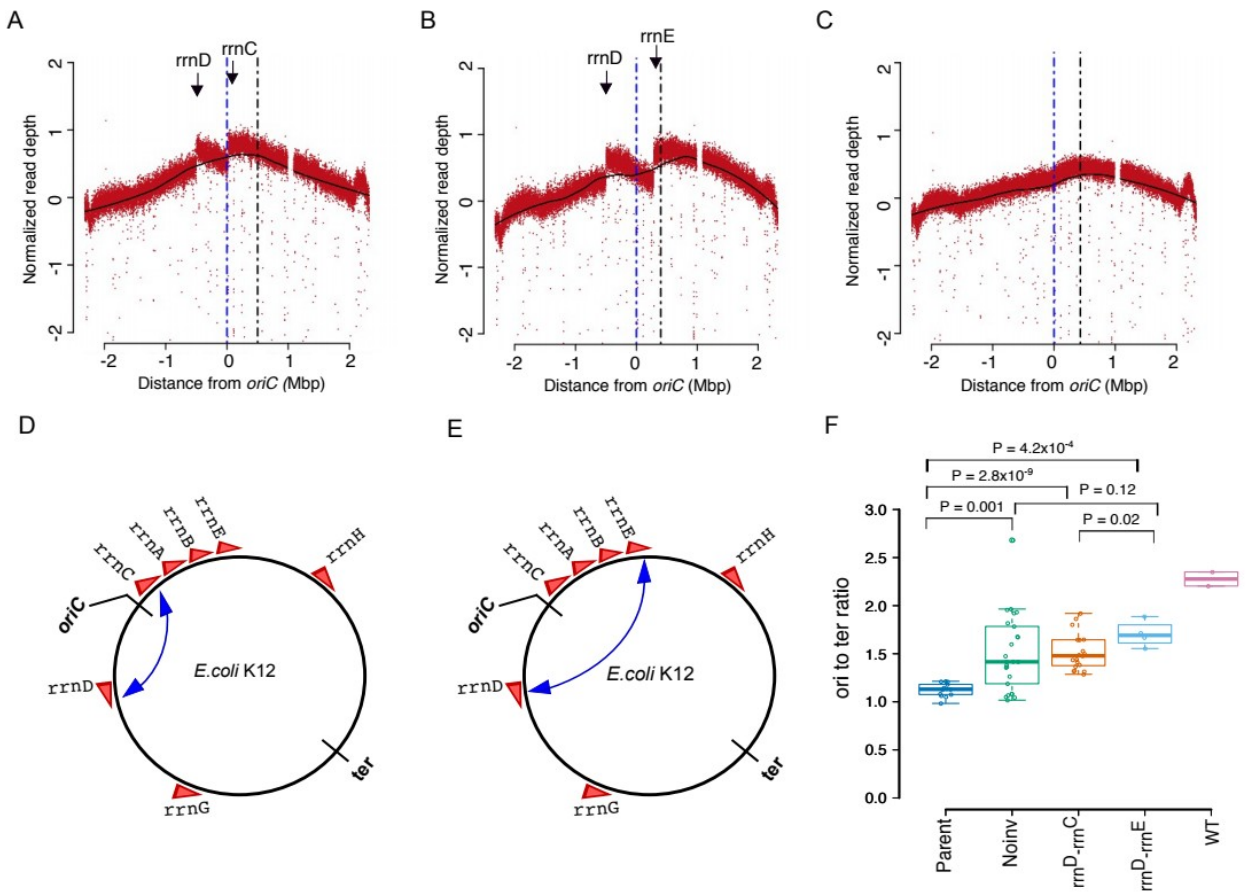

Figure 5*:* Deep Sequencing based MFA plots for suppressor mutants. (A), (B) and (C) represents MFA plots for (A) *ΔrnhA-dnaAinv*$^{rrnD\text{-}rrnC}$, (B) *ΔrnhA-dnaAinv*$^{rrnD\text{-}rrnE}$, (C) *ΔrnhA-dnaANoinv* sequenced at the exponential phase of growth. The dotted blue line represents the *oriC* 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-*ΔrnhA-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* (*ΔrnhA-dnaAinv*$^{rrnD\text{-}rrnC}$) and *rrnD-rrnE*: Clones which shows presence of a chromosomal inversion from *rrnD-rrnE*(*ΔrnhA-dnaAinv*$^{rrnD\text{-}rrnE}$).

oriK45 as a preferred initiation site for cSDR in suppressors

We identified the locations of the maxima of the copy number curve for the suppressors, while ignoring the *ter* peak. We noticed that these mapped to ~4.3 Mb - 4.6 Mb clockwise of *oriC*, in proximity to *oriK45* (Figure 6A). Consistent with this, all suppressors showed a copy number peak at *oriK45* (Figure 6A, Supplementary figure 6 and 7 and supplementary table 5). In the strongest suppressors, we observed a strong copy number gradient peaking at *oriK45* and declining towards *ter.* The peak in *ter* was computationally detected in all suppressors. However, this peak was weak in two of the suppressors. One of these contained the Δ6 bp in frame deletion mutation in Tus, and displayed a copy number pattern similar to that observed for *Δtus* by Maduike et al., (Maduike et al. 2014) indicating that the mutation observed here causes loss of function. This strain did show a slight copy number peak at *oriK45*, but being a relatively weak suppressor does not permit a more confident assignment.

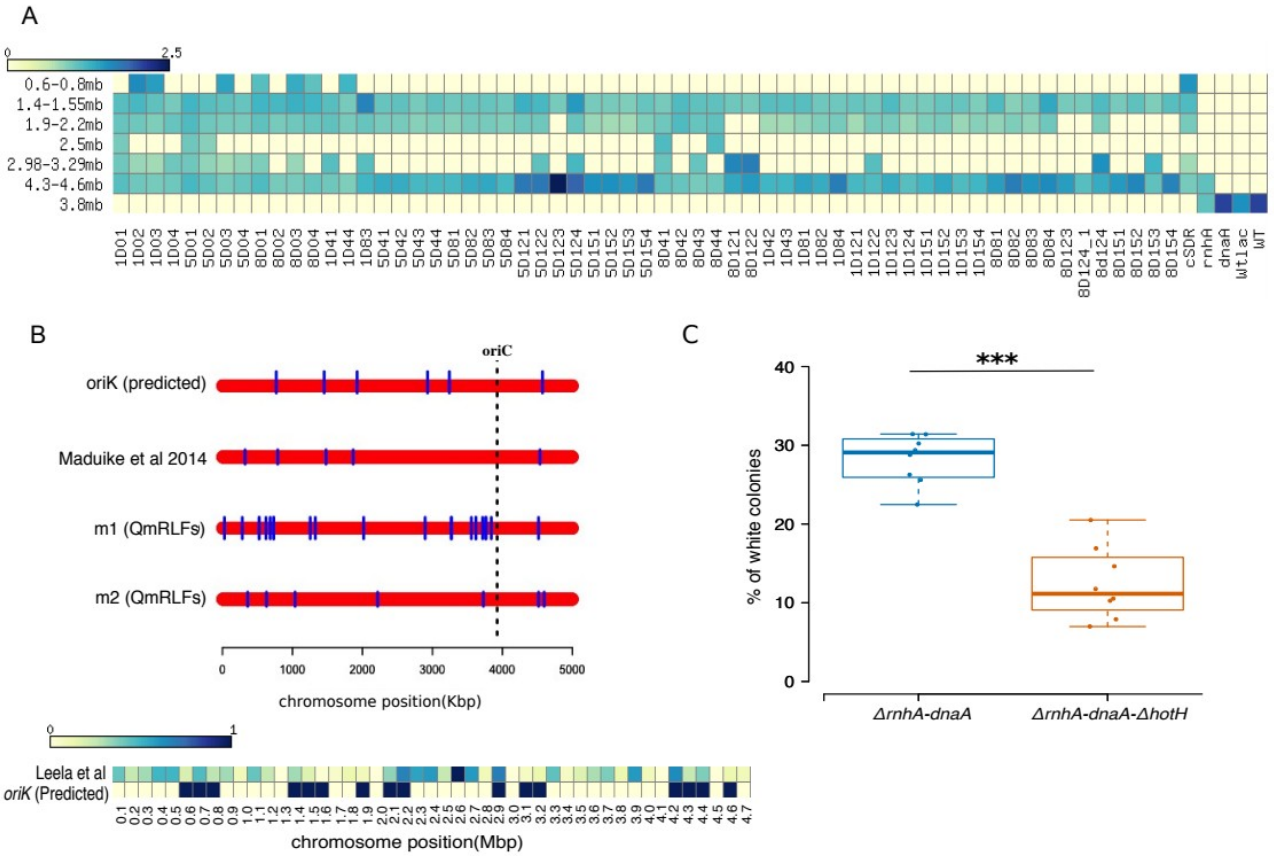

Figure 6: *oriK45* as a preferred initiation site for cSDR in suppressor mutants: (A) heatmap showing predicted *oriK* positions from marker frequency analysis across evolved strains. Y-axis represents the chromosomal positions of predicted *oriK* sites in Mbp. Colour indicates the LOESS smoothed normalised read count around 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 *ΔrnhA-dnaA* strain from Maduike 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. Heatmap represents the density of highly enriched bisulphite sensitive positions in different chromosomal regions from Leela et al. compared to predicted *oriK* sites (C) Box plot comparing the percentage of white colonies of *ΔrnhA-dnaA* and *ΔrnhA-dnaA-ΔhotH* strains respectively(n = 8).

We then asked whether *oriK45* is proximal to regions with high propensities to form RNA-DNA hybrids. Krishna Leela et al. (Leela et al. 2013) had identified bisulfite sensitive regions of *E.coli* chromosome and defined these as preformed R-loops. We notice that there are 34 highly bisulfite sensitive regions within the range of positions assigned to *oriK45* across all our cSDR strains. This is statistically significant compared to random assignment of gene coordinates to highly bisulfite sensitive regions (P = 0.005, Z-score, permutation test across 1,000 repetitions, one-tailed). *oriK* ranges predicted in our study showed significant enrichment in their proximity to highly bisulfite sensitive regions from Leela et al. 2013 (Fisher's Exact test, P = 0.02). Note however that there are many other regions that are highly bisulfite sensitive in the Leela et al. Data but not proximal to any potential *oriK* site. We then 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 (Jenjaroenpun et al. 2015; Kuznetsov et al. 2018). 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. 7 of the 10 copy number bumps described by us or by Maduike et al. for *ΔrnhA-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 6B), within the range defined by *oriK45*.

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 (Nishitani, Hidaka, and Horiuchi 1993). 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 *ΔtopA*-mediated cSDR is an insertion within *fimD*, which is again in the region defined by *oriK45* (Usongo et al. 2016).

To test whether *oriK45* affects the growth of *ΔrnhA-dnaA*, we constructed a Δ11.3Kb region (4555284: 45660615, *uxuR-yjiN*), corresponding to the restriction fragment defined by Nishitani et al. as *hotH*, in the *ΔrnhA-dnaA*-pHYD2388 (*dnaA*$^+$*lacZ*$^+$) background (Nishitani, Hidaka, and Horiuchi 1993). We measured the rate at which *ΔrnhA-dnaA*-pHYD2388 (*dnaA*$^+$*lacZ*$^+$) and *ΔrnhA-dnaA-ΔhotH*-pHYD2388 (*dnaA*$^+$*lacZ*$^+$) lost the pHYD2388 plasmid. This we interpret as a measure of selection in favour of maintaining the plasmid-borne *dnaA* copy. ~12% of *ΔrnhA-dnaA-ΔhotH*-pHYD2388 (*dnaA*$^+$*lacZ*$^+$) lost the plasmid, compared to ~28% for the corresponding *hotH*+ variant. This difference was statistically significant (P = 8 x 10^{-5}, Wilcoxon test one-tailed) (Figure 6C). This shows that the *hotH* site, corresponding to *oriK45*, confers a selective advantage to *ΔrnhA-dnaA*. The fact that this deletion was not lethal suggests that replication initiation might proceed from other sites, albeit at lower rates, in the absence of *oriK45*. In an attempted control experiment, the parental MG1655 strain (*lacZ*$^-$) rarely lost the pHYD2388 plasmid, independent of the presence or absence of the *hotH* sequence in the chromosome.

The effects of cSDR from oriK45 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? To what extent does the suppression of growth defects of cSDR by the inversion around *oriC* reverse these effects? Towards answering these questions, we performed exponential phase transcriptome analysis of *rnhA*$^+$*dnaA*$^+$, *ΔrnhA*, *ΔrnhA-dnaA*, *ΔrnhA-dnaA-inv*$^{rrnD\text{-}rrnC}$, *ΔrnhA-dnaA-inv*$^{rrnD\text{-}rrnE}$ using RNA-seq.

Both *ΔrnhA* and *ΔrnhA-dnaA* induced large changes in gene expression when compared to *rnhA*$^+$*dnaA*$^+$. 600 genes were up-regulated and 543 down-regulated by a log (base 2) fold change of 1.5 or above in *ΔrnhA-ΔdnaA.* The corresponding numbers for *ΔrnhA* are 472 and 360. Nearly 75% of all genes induced in *ΔrnhA* were also induced in *ΔrnhA-ΔdnaA*; 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 Δ*rnhA* than in *ΔrnhA-ΔdnaA* (P < 10^{-10}, paired Wilcoxon test comparing magnitudes of differential expression).

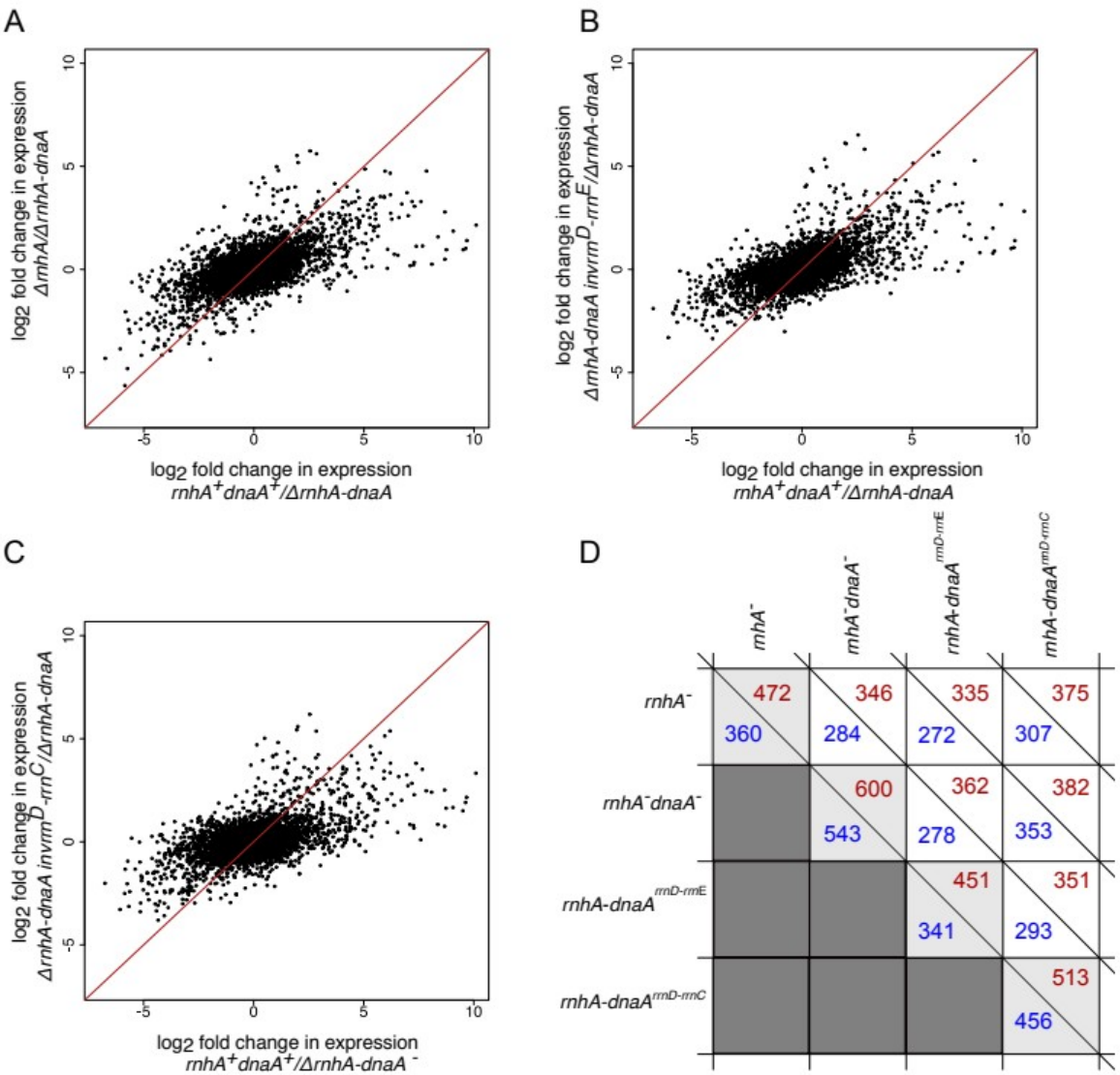

Figure 7: Scatterplots representing correlation of $\log_2$ fold change in gene expression for different conditions, compared to *ΔrnhA-dnaA* strain. (A) *ΔrnhA* vs *rnhA⁺dnaA⁺* (B) *ΔrnhA-dnaAinv*$^{rrnD\text{-}rrnE}$ vs *rnhA⁺dnaA⁺* and (C) *ΔrnhA-dnaAinv*$^{rrnC\text{-}rrnE}$ vs *rnhA⁺dnaA⁺*. The 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 *rnhA⁺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 both *ΔrnhA* and *ΔrnhA-dnaA*. *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 (Dai and Outten 2012). 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 observe 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 (Viguera et al. 2003). The gene *topA*, encoding topoisomerase, which can decrease R-loop formation presumably through its DNA relaxing activity, is also up-regulated.

We observe that several genes encoding components of the ribosome are up-regulated in the inversion mutants. At least three DEAD box RNA helicase genes (*rhlE*, *dbpA* and *srmB*) that are involved in ribosome assembly are also up-regulated. Finally, *rapA*, the gene encoding the RNA polymerase recycling factor ATPase, which is required for reloading stalled RNA polymerase is up-regulated. 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.

Overall, there is a gradient - decreasing from *oriC* towards *ter* - in the fold change in gene expression between $rnhA^{+}dnaA^{+}$ *and ΔrnhA-ΔdnaA*. In other words, genes that are proximal to *oriC* (and *oriK45*) are more strongly down-regulated in *ΔrnhA-dnaA* when compared to $rnhA^{+}dnaA^{+}$, (Supplementary figure 8). At this level, the fold change in $rnhA^{+}dnaA^{+}$, when compared to *ΔrnhA-dnaA*, shows strong similarity to that in *ΔrnhA* and $\Delta rnhA\text{-}dnaA\text{-}inv^{rrnD\text{-}rrnE}$ (Pearson correlation coefficient = 0.64 for both comparisons), and slightly less similar to $\Delta rnhA\text{-}dnaA\text{-}inv^{rrnD\text{-}rrnC}$(Pearson correlation coefficient = 0.55) (Figure 7). These indicate that a portion of the gene expression change in *ΔrnhA-dnaA* relative to $rnhA^{+}dnaA^{+}$ is reversed by the longer inversion $\Delta rnhA\text{-}dnaA\text{-}inv^{rrnD\text{-}rrnE}$, and less so by the shorter inversion $\Delta rnhA\text{-}dnaA\text{-}inv^{rrnD\text{-}rrnC}$. Nevertheless, the magnitude of the difference in gene expression between $rnhA^{+}dnaA^{+}$ and *ΔrnhA-dnaA* is higher than that between the suppressors and *ΔrnhA-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.

Therefore, 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 *ΔrnhA-dnaA* relative to $rnhA^{+}dnaA^{+}$ are partly compensated by inversion containing suppressors.

We observe little difference in gene expression change between mRNA genes on the forward and the reverse strand of the DNA between *oriC* and *oriK45*. This suggests that changes in replication-transcription conflicts have little effect on overall gene expression. 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 *ΔrnhA-dnaA*, but was reduced to 0.39 in the suppressor

ΔrnhA-dnaAinv$^{rrnD\text{-}rrnE}$ (Supplementary table 6). This effect was the strongest when only rRNA genes (5S rRNA, which is not depleted as part of the RNA prep experiment) were considered. 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.

Conclusion

Taken together, our results indicate that under *ΔrnhA-dnaA* cSDR, selection favours preferential replication initiation from *oriK45*, located ~0.6Mb clockwise of *oriC*. Replication initiation from this site would result in head-on collisions with RNA polymerases transcribing four rRNA operons encoded between *oriC* and *oriK45*. 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 GFP cloned downstream of the wildtype *rrnD* promoter and that with the discriminator mutation (Supplementary figure 9), 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.

In a previous study, the Sherratt lab placed a second *ori* termed *oriZ* ~1 Mb clockwise of *oriC*. They reported that 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 (X. Wang et al. 2011). 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 (Ivanova et al. 2015). 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 (Ivanova et al. 2015). The inversion reported by Ivanova et al. (Ivanova et al. 2015) 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*. Thus, Ivanova et al. could conclude that replication-transcription conflicts are key determinants of fitness of *E. coli*. 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 (Srivatsan et al. 2010). 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 (Esnault et al. 2007). 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. *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* (J. Dimude et al. 2018). 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., (Ivanova et al. 2015; J. Dimude et al. 2018) 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.

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 (Tokio Kogoma 1997). Further, cSDR can provide the bacterium avenues for the development of resistance against new antibiotics targeting initiation of DNA replication (Grimwade and Leonard 2017; van Eijk et al. 2017).

Materials and methods

Strains and Media conditions

Wild type($rnhA^+dnaA^+$) strain mentioned in this study is a derivative of non pathogenic *E.coli* K12 MG1655 strain mentioned as GJ13519 in (Leela et al. 2013). Gene deletions were performed using the one-step inactivation method described by Datsenko and Wanner (Datsenko and Wanner 2000) or by P1 phage mediated transduction protocol (Thomason, Costantino, and Court 2007). Growth curves were generated in 250ml flasks or 24-well plates in Luria Bertani (LB; Hi-Media, India, M575-500) broth at 37°C with shaking at 200 rpm. Optical density (OD) measurements were carried out at 600 nm (OD 600) using 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 customized R scripts.

Spotting Assay

Spotting assay was performed for all strains at μ_{max}. 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 µl spots) on LB agar plates. The plates were imaged after 30 hours of incubation at 37 °C.

Whole genome sequencing and DNA copy number analysis

For genomic DNA extraction, the 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 GenElute™ Bacterial Genomic DNA Kit (NA2120-1KT, Sigma-Aldrich) using the manufacturer's

protocol. Library preparation was carried out using Truseq Nano DNA low throughput Library preparation kit (15041757) and Paired end (2X100) sequencing of genomic DNA was performed using Illumina Hiseq 2500 platform.

The sequencing reads were aligned and mapped to the reference genome (NC_000913.3) using Burrows Wheeler Aligner (BWA) *(http://bio-bwa.sourceforge.net)* specifying alignment quality and mapping quality thresholds as 20. Read coverage across the genome was calculated for non-overlapping windows of 200nt each using customized perl scripts and the values were normalized by the mode of the distribution across these bins. The normalized values in logarithmic scale (log_2) were plotted against chromosome coordinates to get the DNA copy number plots from *ori* to *ter.* The coordinates were repositioned in such a way that the numbering starts from *oriC* position. Loess polynomial regression analysis was used for curve fitting.

Laboratory Evolution of cSDR mutant

Laboratory evolution experiment was carried out for overnight grown cultures of eight independent isolates of *ΔrnhAΔdnaA* strain. 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 degree Celsius 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 the course of 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.

Mutational analysis and ori-to-ter ratio calculation

SNPs and indels were identified from the genome sequencing data using the BRESEQ (version 0.33.1) (Deatherage and Barrick 2014)pipeline which uses Bowtie for sequence alignment. A mutational matrix representing presence and absence of mutations were generated from BRESEQ output file using customised R scripts and heat maps were generated using Matrix2png(Pavlidis and Noble 2003). 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 scripts.

oriK peak prediction

oriK positions were predicted from the loess fitted copy number plots using custom R scripts. A position was called as *oriK* peak if it has a negative slope upto 100kbp in both directions. The predicted peak positions were normalized to a range of maximum ~0.3mb and compared across samples.

R-loop predictions using QmRLFs Finder

To predict RNA-DNA hybrids on the chromosome we used QmRLFs model (Kuznetsov et al. 2018; Jenjaroenpun et al. 2015) on *Escherichia coli* K12 MG1655 (NC_000913.3) genome with default parameters. 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.

oriK45 Deletion and Blue white screening

Appropriate dilution (10^{-6}) of Overnight cultures of *ΔrnhA-dnaA*-pHYD2388 (*dnaA*$^{+}$*lacZ*$^{+}$) and *ΔrnhA-dnaA-ΔhotH*-pHYD2388 (*dnaA*$^{+}$*lacZ*$^{+}$) were plated on M9 minimal X-gal agar plates. Plates were incubated at 37°C for 30 hours and the number of blue and white colonies appeared on these plates were counted separately and the respective percentage of white colonies were calculated.

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 Truseq RNA Sample preparation Kit without poly-A selection(NEB #E7645S/L) and single end sequencing for 50 cycles were done using Illumina HiSeq 2500 platform.

Transcriptome analysis

The sequencing reads were aligned and mapped to the reference genome (NC_000913.3) using Burrows Wheeler Aligner (BWA). 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 (verson 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 (McCarthy, Chen, and Smyth 2012). Log fold change expression values in comparison to *ΔrnhA-dnaA* 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 FDR value of 0.01 were considered for further analysis.

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.

Data availability

The genome sequence data from this work are available at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA562391.

The RNA sequence data and processed files from this work are available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE135706.

Supplementary Data

Supplementary information for this manuscript is available as Supplementary_file.pdf

Acknowledgment

We thank Prof. J. Gowrishankar for discussions, materials and comments on our manuscript. We thank Nalini Raghunathan, Sayantan Goswami and all Gowrishankar lab members at CDFD for discussions and materials. We thank Dr. Sunil Laxman (Instem, Bangalore), Dr. Dasaradhi Palakodeti (Instem, Bangalore), Prof. P V Shivaprasad (NCBS-TIFR, Bangalore), Dr. Anjana Badrinarayan (NCBS-TIFR, Bangalore) and Prof. Sandeep Krishna (NCBS-TIFR, Bangalore) for discussions and comments. We acknowledge Dr Awadhesh Pandit and Tejali Naik at the Next Generation Genomics Facility, NCBS for providing sequencing services and CIFF facility at NCBS for technical support.

Funding

This work was supported by the Wellcome Trust/DBT India Alliance Intermediate Fellowship/Grant [grant number IA/I/16/2/502711] awarded to ASNS, and CSIR Direct SRF (113205/2K17/1) fellowship to RTV.

Conflict of interest statement. None declared.

References

1. Brochu, Julien, Émilie Vlachos-Breton, Sarah Sutherland, Makisha Martel, and Marc Drolet. 2018. “Topoisomerases I and III Inhibit R-Loop Formation to Prevent Unregulated Replication in the Chromosomal Ter Region of Escherichia Coli.” *PLoS Genetics* 14 (9): e1007668.

2. Bryant, Jack A., Laura E. Sellars, Stephen JW Busby, and David J. Lee. 2014. “Chromosome Position Effects on Gene Expression in Escherichia Coli K-12.” *Nucleic Acids Research* 42 (18): 11383–11392.
3. Clayton, D. A. 1982. “Replication of Animal Mitochondrial DNA.” *Cell* 28 (4): 693–705. https://doi.org/10.1016/0092-8674(82)90049-6.
4. Couturier, Etienne, and Eduardo PC Rocha. 2006. “Replication-Associated Gene Dosage Effects Shape the Genomes of Fast-Growing Bacteria but Only for Transcription and Translation Genes.” *Molecular Microbiology* 59 (5): 1506–1518.
5. Dai, Yuyuan, and F. Wayne Outten. 2012. “The E. Coli SufS-SufE Sulfur Transfer System Is More Resistant to Oxidative Stress than IscS-IscU.” *FEBS Letters* 586 (22): 4016–22. https://doi.org/10.1016/j.febslet.2012.10.001.
6. Datsenko, K. A., and B. L. Wanner. 2000. “One-Step Inactivation of Chromosomal Genes in Escherichia Coli K-12 Using PCR Products.” *Proceedings of the National Academy of Sciences of the United States of America* 97 (12): 6640–45. https://doi.org/10.1073/pnas.120163297.
7. Deatherage, Daniel E., and Jeffrey E. Barrick. 2014. “Identification of Mutations in Laboratory-Evolved Microbes from next-Generation Sequencing Data Using Breseq.” *Methods in Molecular Biology (Clifton, N.J.)* 1151: 165–88. https://doi.org/10.1007/978-1-4939-0554-6_12.
8. Dimude, Juachi, Monja Stein, Ewa Andrzejewska, Mohammad Khalifa, Alexandra Gajdosova, Renata Retkute, Ole Skovgaard, and Christian Rudolph. 2018. “Origins Left, Right, and Centre: Increasing the Number of Initiation Sites in the Escherichia Coli Chromosome.” *Genes* 9 (8): 376.
9. Dimude, Juachi U., Anna Stockum, Sarah L. Midgley-Smith, Amy L. Upton, Helen A. Foster, Arshad Khan, Nigel J. Saunders, Renata Retkute, and Christian J. Rudolph. 2015. “The Consequences of Replicating in the Wrong Orientation: Bacterial Chromosome Duplication without an Active Replication Origin.” *MBio* 6 (6): e01294–15.
10. Duggin, Iain G., and Stephen D. Bell. 2009. “Termination Structures in the Escherichia Coli Chromosome Replication Fork Trap.” *Journal of Molecular Biology* 387 (3): 532–39. https://doi.org/10.1016/j.jmb.2009.02.027.
11. Eijk, Erika van, Bert Wittekoek, Ed J. Kuijper, and Wiep Klaas Smits. 2017. “DNA Replication Proteins as Potential Targets for Antimicrobials in Drug-Resistant Bacterial Pathogens.” *Journal of Antimicrobial Chemotherapy* 72 (5): 1275–84. https://doi.org/10.1093/jac/dkw548.
12. Esnault, Emilie, Michèle Valens, Olivier Espéli, and Frédéric Boccard. 2007. “Chromosome Structuring Limits Genome Plasticity in Escherichia Coli.” *PLoS Genetics* 3 (12): e226. https://doi.org/10.1371/journal.pgen.0030226.
13. Gowrishankar, J., and R. Harinarayanan. 2004. “Why Is Transcription Coupled to Translation in Bacteria?” *Molecular Microbiology* 54 (3): 598–603. https://doi.org/10.1111/j.1365-2958.2004.04289.x.
14. Gowrishankar, J., J. Krishna Leela, and K. Anupama. 2013. “R-Loops in Bacterial Transcription: Their Causes and Consequences.” *Transcription* 4 (4): 153–157.
15. Gowrishankar, Jayaraman. 2015. “End of the Beginning: Elongation and Termination Features of Alternative Modes of Chromosomal Replication Initiation in Bacteria.” *PLoS Genetics* 11 (1): e1004909.
16. Grimwade, Julia E., and Alan C. Leonard. 2017. “Targeting the Bacterial Orisome in the Search for New Antibiotics.” *Frontiers in Microbiology* 8 (November). https://doi.org/10.3389/fmicb.2017.02352.
17. Harinarayanan, R., and J. Gowrishankar. 2003. “Host Factor Titration by Chromosomal R-Loops as a Mechanism for Runaway Plasmid Replication in Transcription Termination-Defective Mutants of Escherichia Coli.” *Journal of Molecular Biology* 332 (1): 31–46. https://doi.org/10.1016/s0022-2836(03)00753-8.

18. Hill, C W, and B W Harnish. 1981. “Inversions between Ribosomal RNA Genes of Escherichia Coli.” *Proceedings of the National Academy of Sciences of the United States of America* 78 (11): 7069–72.
19. Hong, Xiankang, G. W. Cadwell, and T. Kogoma. 1995. “Escherichia Coli RecG and RecA Proteins in R-Loop Formation.” *The EMBO Journal* 14 (10): 2385–2392.
20. Hong, Xiankang, Gregory W. Cadwell, and Tokio Kogoma. 1996. “Activation of Stable DNA Replication in Rapidly Growing Escherichia Coli at the Time of Entry to Stationary Phase.” *Molecular Microbiology* 21 (5): 953–961.
21. Ivanova, Darja, Toni Taylor, Sarah L. Smith, Juachi U. Dimude, Amy L. Upton, Mana M. Mehrjouy, Ole Skovgaard, David J. Sherratt, Renata Retkute, and Christian J. Rudolph. 2015. “Shaping the Landscape of the Escherichia Coli Chromosome: Replication-Transcription Encounters in Cells with an Ectopic Replication Origin.” *Nucleic Acids Research* 43 (16): 7865–7877.
22. Jenjaroenpun, Piroon, Thidathip Wongsurawat, Surya Pavan Yenamandra, and Vladimir A. Kuznetsov. 2015. “QmRLFS-Finder: A Model, Web Server and Stand-Alone Tool for Prediction and Analysis of R-Loop Forming Sequences.” *Nucleic Acids Research* 43 (20): 10081. https://doi.org/10.1093/nar/gkv974.
23. Khedkar, Supriya, and Aswin Sai Narain Seshasayee. 2016. “Comparative Genomics of Interreplichore Translocations in Bacteria: A Measure of Chromosome Topology?” *G3: Genes, Genomes, Genetics* 6 (6): 1597–1606. https://doi.org/10.1534/g3.116.028274.
24. Kogoma, T., T. A. Torrey, and M. J. Connaughton. 1979. “Induction of UV-Resistant DNA Replication in Escherichia Coli: Induced Stable DNA Replication as an SOS Function.” *Molecular and General Genetics MGG* 176 (1): 1–9. https://doi.org/10.1007/BF00334288.
25. Kogoma, Tokio. 1997. “Stable DNA Replication: Interplay between DNA Replication, Homologous Recombination, and Transcription.” *Microbiol. Mol. Biol. Rev.* 61 (2): 212–238.
26. Kuznetsov, Vladimir A, Vladyslav Bondarenko, Thidathip Wongsurawat, Surya P Yenamandra, and Piroon Jenjaroenpun. 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 Research* 46 (15): 7566–85. https://doi.org/10.1093/nar/gky554.
27. Leela, J. Krishna, Aisha H. Syeda, K. Anupama, and J. Gowrishankar. 2013. “Rho-Dependent Transcription Termination Is Essential to Prevent Excessive Genome-Wide R-Loops in Escherichia Coli.” *Proceedings of the National Academy of Sciences* 110 (1): 258–263.
28. Maduike, Nkabuije Z., Ashley K. Tehranchi, Jue D. Wang, and Kenneth N. Kreuzer. 2014. “Replication of the E Scherichia Coli Chromosome in RN Ase HI-Deficient Cells: Multiple Initiation Regions and Fork Dynamics.” *Molecular Microbiology* 91 (1): 39–56.
29. Martel, Makisha, Aurélien Balleydier, Alexandre Sauriol, and Marc Drolet. 2015. “Constitutive Stable DNA Replication in Escherichia Coli Cells Lacking Type 1A Topoisomerase Activity.” *DNA Repair* 35: 37–47.
30. Massy, Bernard de, Olivier Fayet, and Tokio Kogoma. 1984. “Multiple Origin Usage for DNA Replication in SdrA (Rnh) Mutants of Escherichia Coli K-12: Initiation in the Absence of OriC.” *Journal of Molecular Biology* 178 (2): 227–236.
31. McCarthy, Davis J., Yunshun Chen, and Gordon K. Smyth. 2012. “Differential Expression Analysis of Multifactor RNA-Seq Experiments with Respect to Biological Variation.” *Nucleic Acids Research* 40 (10): 4288–97. https://doi.org/10.1093/nar/gks042.
32. Mott, Melissa L., and James M. Berger. 2007. “DNA Replication Initiation: Mechanisms and Regulation in Bacteria.” *Nature Reviews Microbiology* 5 (5): 343.
33. Nishitani, Hideo, Masumi Hidaka, and Takashi Horiuchi. 1993. “Specific Chromosomal Sites Enhancing Homologous Recombination in Escherichia Coli Mutants Defective in RNase H.” *Molecular and General Genetics MGG* 240 (3): 307–314.

34. Ogawa, Tohru, Gavin G. Pickett, Tokio Kogoma, and Arthur Kornberg. 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.” *Proceedings of the National Academy of Sciences* 81 (4): 1040–1044.
35. Pavlidis, Paul, and William Stafford Noble. 2003. “Matrix2png: A Utility for Visualizing Matrix Data.” *Bioinformatics* 19 (2): 295–96. https://doi.org/10.1093/bioinformatics/19.2.295.
36. Raghunathan, Nalini, Sayantan Goswami, Jakku K. Leela, Apuratha Pandiyan, and Jayaraman Gowrishankar. 2019. “A New Role for Escherichia Coli Dam DNA Methylase in Prevention of Aberrant Chromosomal Replication.” *Nucleic Acids Research* 47 (11): 5698–5711.
37. Raghunathan, Nalini, Rajvardhan M. Kapshikar, Jakku K. Leela, Jillella Mallikarjun, Philippe Bouloc, and Jayaraman Gowrishankar. 2018. “Genome-Wide Relationship between R-Loop Formation and Antisense Transcription in Escherichia Coli.” *Nucleic Acids Research* 46 (7): 3400–3411. https://doi.org/10.1093/nar/gky118.
38. Rocha, Eduardo PC. 2004. “The Replication-Related Organization of Bacterial Genomes.” *Microbiology* 150 (6): 1609–1627.
39. Skovgaard, Ole, Mads Bak, Anders Løbner-Olesen, and Niels Tommerup. 2011. “Genome-Wide Detection of Chromosomal Rearrangements, Indels, and Mutations in Circular Chromosomes by Short Read Sequencing.” *Genome Research* 21 (8): 1388–93. https://doi.org/10.1101/gr.117416.110.
40. Srivatsan, Anjana, Ashley Tehranchi, David M. MacAlpine, and Jue D. Wang. 2010. “Co-Orientation of Replication and Transcription Preserves Genome Integrity.” *PLoS Genetics* 6 (1): e1000810.
41. Thomason, Lynn C., Nina Costantino, and Donald L. Court. 2007. “E. Coli Genome Manipulation by P1 Transduction.” *Current Protocols in Molecular Biology* Chapter 1 (July): Unit 1.17. https://doi.org/10.1002/0471142727.mb0117s79.
42. Usongo, Valentine, Makisha Martel, Aurélien Balleydier, and Marc Drolet. 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 (April): 1–17. https://doi.org/10.1016/j.dnarep.2016.02.001.
43. Viguera, Enrique, Mirjana Petranovic, Davor Zahradka, Karine Germain, Dusko S. Ehrlich, and Bénédicte Michel. 2003. “Lethality of Bypass Polymerases in Escherichia Coli Cells with a Defective Clamp Loader Complex of DNA Polymerase III.” *Molecular Microbiology* 50 (1): 193–204.
44. Wang, Jue D., Melanie B. Berkmen, and Alan D. Grossman. 2007. “Genome-Wide Coorientation of Replication and Transcription Reduces Adverse Effects on Replication in Bacillus Subtilis.” *Proceedings of the National Academy of Sciences* 104 (13): 5608–5613.
45. Wang, Xindan, Christian Lesterlin, Rodrigo Reyes-Lamothe, Graeme Ball, and David J. Sherratt. 2011. “Replication and Segregation of an Escherichia Coli Chromosome with Two Replication Origins.” *Proceedings of the National Academy of Sciences* 108 (26): E243–E250.
46. Yasukawa, Takehiro, and Dongchon Kang. 2018. “An Overview of Mammalian Mitochondrial DNA Replication Mechanisms.” *Journal of Biochemistry* 164 (3): 183–93. https://doi.org/10.1093/jb/mvy058.