Controlling Genome Topology with Sequences that Trigger Post-replication Gap Formation During Replisome Passage: The *E. coli* RRS Elements

Phuong Pham[1], Elizabeth A. Wood[2], Emma L. Dunbar[2], Michael M. Cox[2] and Myron F. Goodman[1]

[1] Departments of Biological Sciences and Chemistry, University of Southern California, Los Angeles, CA 90089-2910, [2] Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706-1544

ABSTRACT

We report that the *Escherichia coli* chromosome includes novel GC-rich genomic structural elements that trigger formation of post-replication gaps upon replisome passage. The two nearly perfect 220 bp repeats, designated Replication Risk Sequences or RRS, are each 650 kbp from the terminus sequence *dif* and flank the Ter macrodomain. RRS sequence and positioning is highly conserved in enterobacteria. At least one RRS appears to be essential unless a 200 kbp region encompassing one of them is amplified. The RRS produce lagging strand ssDNA gaps, ≤2000 bp long, upon replisome passage, regardless of genome placement. Deletion of both RRS elements has substantial effects on global genome structure and topology. We hypothesize that RRS elements serve as topological relief valves during chromosome replication and segregation. There have been no screens for genomic sequences that trigger transient gap formation. Functional analogs of RRS could be widespread, possibly including some enigmatic G-quadruplexes.

INTRODUCTION

The *E. coli* chromosome is maintained as a highly condensed nucleoprotein structure called a nucleoid. Condensation of the nucleoid arises from several factors (reviewed in [1]). These include nucleoid-associated proteins and structural maintenance of chromosome proteins that bind and condense DNA at both local and global levels [2]. Additionally, DNA supercoiling arising from processes such as RNA transcription contribute to nucleoid condensation [3].

Nucleoid condensation in *E. coli* is neither random nor uniform. The *E. coli* chromosome is divided into 31 self-interacting segments called chromosome interaction domains or CIDs [4-10]. At least some of the CID boundaries correspond to highly transcribed genes. In addition to CIDs, the genome is further structured into several macrodomains with boundaries that prevent the diffusion of supercoiling and restrict interaction between macrodomains [6-10]. Nothing is known about the nature of the structure that defines macrodomain boundaries. One macrodomain (Ter) encompasses a million or more bp surrounding the replication terminus defined by the sequence *dif* [11,12].

In addition to genes that encode functional proteins and RNA molecules, the *E. coli* genome includes a variety of genomic elements that facilitate some aspect of genome structure and function. These include the replication origin, *ori* (cite); the replication terminus region with its *ter* and *dif* sites (cite); and about 300 bacterial interspersed mosaic elements (BIMEs). One family of BIMES functions as efficient binding sites for DNA gyrase [13].

There have been multiple efforts to screen the *E. coli* genome for repeated sequences of many types [14-18]. Most of the targeted repeats were transient gene duplications that are unstable and ultimately eliminated by recombination, inverted and tandem intergenic repeats that often have some role in gene expression, transposons, simple sequence repeats, genes with multiple genomic copies, and CRISPR repeats. We report here on a repeated genomic sequence that does not fit into any of these categories.

We recently developed a new method to examine protein binding to single-stranded DNA on a genomic scale, designated ssGAP-Seq [19]. In applying it to RecA and SSB binding to ssDNA, we identified two sites in the *E. coli* genome where both proteins were deposited at elevated rates on the lagging strand template. At the ends of both sites, we identified 220 bp and GC-rich repeats. These are positioned symmetrically about the terminal *dif* site, each 650 kbp distant from *dif*. When cloned, the repeats cannot be sequenced by standard commercial sequencing unless protocols with extra heat are applied. In the genome, the repeats are causing post-replication gap formation as replisomes pass. We have therefore named them Replication Risk Sequences or RRS [19]. The RRS appeared briefly in a 1999 screen for genomic repeats [14] and designated TRIP1 (for TRIPle; a third partial repeat was identified to make 3). However, no functional significance was attached to the repeats and they have been otherwise overlooked.

The RRS are highly conserved in both positioning and sequence among the enterobacteria [19], although some species have one or more additional RRS within their genomes. The conservation and effects on replication suggest a sequence of genomic functional significance. The present study introduces important properties and genomic effects of the RRS elements.

RESULTS

Structure and positioning of RRS elements. The RRS elements are 220 bp in length, positioned in the *E. coli* genome near the *dusC* and *lysO* genes (Figure 1A). These sites are each approximately 650 kbp from the *dif* sequence that helps define replication termination. The RRS are GC-rich and have the potential to form extensive secondary structure (Figure 1B). They were originally defined as sites that anchored a prominent peak in RecA and SSB deposition on single-stranded DNA [19]. The RecA is deposited uniquely on the lagging strand template, in a region of approximately 2 kbp (Figure 1C). Deposition is greatest near the RRS element, and trails off from that point. This behavior suggests that DNA polymerase subunits replicating the lagging strand are halted at the RRS, disengage, and re-initiate downstream, leaving a gap into which SSB and/or RecA is often loaded.

The positions of these sequences, near the reported boundaries of the Ter macrodomain [11,12], suggested an important function. We therefore constructed several strains in which one or both of the RRS were deleted or in which RRS elements were placed in other genomic locations.

RRS triggers lagging strand gap formation wherever they occur. A series of strains were constructed in which RRS elements were placed near the *fabR*, *djlC*, and *dinI* genes, singly or together (see locations in Figure 1A). All of these strains, with one, two, or three new RRS, were subjected to ssGAP-seq to determine where RecA and SSB were deposited on ssDNA. In every case, when a new RRS was positioned in the genome, a new peak of RecA and SSB deposition appeared at that location (see results for EAW1736, 1765, and 1768 in Figure 1C and Supplementary Figure S1). In each case, the RRS was oriented the same way as the natural RRS sequences are oriented relative to the direction of replisome travel. The results indicate that the RRS will function as lagging strand gap-triggering elements wherever they are positioned in the genome.

RRS elements are conditionally essential for genome function. Deletion of one or the other natural RRS does not lead to any measurable cell stress or growth defect. However, we have been unable to construct a clean deletion of both RRS. Multiple attempts to delete the RRS near *dusC* in a strain lacking the *lysO* RRS resulted in colonies that grew as well as wild type cells. However, in all cases, deep sequencing revealed that these strains have an amplification of the 200 kbp genomic region between *insH7* and *insH8* (Supplementary Figure S2). This region encompases the *dusC* RRS element. In some of the colonies, an RRS remained, presumably within one of the amplified sequences. In other colonies, including those shown in Figure 2, the RRS was no longer detectable either during PCR amplification or in Illumina sequence reads. We conclude that at least one RRS is essential unless its function is somehow compensated for with this genome rearrangement.

We also took strain EAW1768 and deleted both of the natural RRS sequences within it. In this case, both RRS could be deleted without amplification of the *insH7-8* region (Figure 1C and 2). This result indicates that one or more of the three abnormally positioned RRS elements can compensate for the absence of the natural RRS sites.

The functional complementation of the RRS deletions appeared to be complete. The strains with both RRS deleted, compensated by either the *insH7-8* amplification or the addition of additional RRS, exhibited no growth defect (Figure 3A). In addition, none of the strains exhibited increased sensitivity to either UV irradiation or ciprofloxacin (Figure 3B).

RRS elements affect genome-wide supercoiling. To examine effects of RRS deletions on chromosome supercoiling genome-wide, we employed a new method developed by Bates and coworkers, called Psora-Seq [20]. The method measures superhelical density by binding biotin-labeled psoralen to the DNA *in vivo*, and then carrying out deep sequencing of DNA based on the extent of psoralen binding to it.

When applied to the *E. coli* genome (Figure 4), the published results [20] were largely replicated with wild type cells as well as with either of the two strains that lacked one or the other single RRS. However, when both RRS are deleted (compensated for by the *insH7-8* amplification), major changes in global genome topology are observed. The effect is seen most dramatically when results are averaged over 250 kbp windows. A large region extending between the sites of the deleted RRS exhibits a large increase in overall supercoiling. Additional effects are seen throughout the genome. When results are averaged over smaller windows (Supplementary Figures S3, S4 and S5), more details become evident. Psoralen binding profiles of single RRS deletion strains are highly correlated ($r \sim 0.8$) with the wild-type profile, whereas double Δ*RRS* deleted cells have a much lower psoralen binding correlation value ($r = 0.3$) (Figure 4A). Average psoralen binding z-scores across macrodomains vary from -0.03 to 0.02 in wild-type and singe RRS deleted cells (Figure 4B). However, much higher variation in average macrodomain psoralen binding values (from -0.07 to 0.26) was observed in double ΔRRS deleted cells with notable enrichment of bound psoralen in TER macrodomain (Figures 4B and 4C). Double RRS deletion also disrupted the observed supercoiling twin-domains signature associated with *rrn* operon transcription (Figure 4D).

DISCUSSION

The RRS is a genomic sequence that triggers gap formation on the lagging strand (Figure 5), a gap into which SSB and RecA are often loaded. This study establishes that (a) the RRS functions at any genomic location in which it is placed, (b) at least one RRS is conditionally essential (both can be deleted only if a large genomic segment is amplified), (c) RRS at other locations may substitute for the RRS at their normal genomic location, and (d) removal of both RRS has a dramatic effect on overall genomic topology.

The function of the RRS element at other genomic locations is telegraphed by the deposition of RecA protein at the new locations as seen with ssGAP-seq. We have not been able to cleanly delete both RRS and obtain a viable cell. Such deletions are always accompanied by genomic amplification of a large region encompassing one of the RRS. Determining how this amplification can compensate for the missing RRS function will require additional research. RRS at other genomic locations can compensate for deletion of both natural RRS elements. In our experiments, three additional RRS were present in the genome. Two of these are somewhat close to the natural RRS present near *lysO*. More work will be needed to determine if any one of these RRS is sufficient on its own to compensate for the nature RRS elements. The effects of deleting both RRS on genome topology is easy to document. We hypothesize that normal RRS function is related to genome topological transitions during replication termination and chromosome segregation at cell division. Formation of a transient gap would have the effect of introducing potential swivels at every phosphodiester bond in the ssDNA region. These could

relax DNA stress related to supercoiling or they might facilitate some aspect of chromosome folding.

We assume that gap formation is transient, as conversion of the gap to duplex DNA would eventually be necessary for subsequent replication cycles. The formation of a transient gap would introduce potential swivels at every phosphodiester bond that could relieve topological stress. We postulate that the RRS provide topological relief valves, facilitating the complex genomic transitions that accompany replication termination and chromosome segregation. The RRS could also be the structural manifestation of the Ter macrodomain boundary, a function that would not exclude additional roles.

Bulky DNA lesions routinely trigger the formation of lesion-containing post-replication gaps, which are generally repaired either via RecF-mediated recombinational DNA repair or translesion DNA synthesis [21]. The presence of genomic elements that trigger post-replication gap formation has not previously been considered, primarily because nobody has looked for them. One other example of such an element has appeared in the bacterial literature, a G-quadruplex-forming structure that triggers gap formation and is required for pilin antigenic variation in *Neisseria gonorrhoeae* [22,23]. In principle, any DNA sequence that can fold up quickly – a G-quadruplex or any other stable tertiary structure – as duplex DNA is unwound during replication, could potentially block full replication of the lagging strand template and create a transient gap. Sequences with the potential to form G-quadruplexes are common in eukaryotic genomes, with approximately 700,000 such sequences in the human genome [24-29]. Triggering transient gap formation, in the service of chromosome structural transitions or replication, could be a function of some or many of them, perhaps to facilitate chromosome folding or relieve topological stress.

MATERIALS AND METHODS

Materials. Wild-type MG1655 *E. coli*, isogenic RRS deletion strains EAW1706 *(ΔRRS-lysO*), EAW1708 *(ΔRRS-dusC*) and EAW1710 (double deletion *ΔRRS-lysO*, *ΔRRS-dusC*) and mutant strains containing extra RRS sequences inserted at *fabR* locus (EAW1736); at *fabR* and *djlC* loci (EAW1765), at *fabR, djlC* and *dinI* loci (EAW1768) and polyclonal chicken anti-RecA and anti-SSB antibodies were from our labs. PrecipHen (Agarose coupled Goat anti-Chicken IgY beads) was purchased from Aves Labs, inc (Davis, CA). EZ-Link Psoralen-PEG3-Biotin and Dynabeads MyOne Streptavidin T1 were from ThermoFisher. Monarch PCR & DNA clean-up kit was purchased from New England Biolabs. DNase-free RNAse A (100 mg/ml) was from Qiagen, Proteinase K, sodium metabisulfite and hydroquinone were from Sigma-Aldrich. xGen Methyl-Seq Library kit was purchased from Integrated DNA Technologies and Nextera XT library prep kit was from Illumina, Inc.

Construction of double ΔRRS deletion strain EAW1710. The region from 168bp-389bp upstream of the *lysO* gene = the *lysO* RRS was deleted from MG1655 using λRED recombination. The deletion was confirmed by PCR using primers outside the RRS and direct

sequencing of the PCR product. The strain was designated EAW 1706. The region from 281bp-58bp downstream of the *dusC* gene = the *dusC* RRS was deleted from MG1655 using λRED recombination. The deletion was confirmed by PCR using primers outside the RRS and direct sequencing of the PCR product. The strain was designated EAW1708. The KanR gene replacing the *lysO* RRS of EAW1706 was removed by FLP recombination, and P1 transduction was used to add the *dusC* RRS deletion from EAW1708. Two colonies from the transduction had different morphologies and were designated EAW1710 colony #1 and #3. Both *Δ*RRS deletions were confirmed by PCR using the same primers used to confirm the single deletions.

Construction of strains with extra RRS elements. The region from 168bp-434bp upstream of the *lysO* gene = extended *lysO* RRS was cloned into a plasmid next to the FRT-KanR-FRT cassette used in λRED recombination. The resulting plasmid, designated pEAW1330, was used as template in PCRs with primers designed to insert the extended *lysO* RRS+ FRT-KanR-FRT cassette 129bp upstream of the *fabR* gene start codon, 6bp downstream of the *djlC* gene stop codon, and 72bp downstream of the dinI gene stop codon. P1 transduction was used to add the RRS near *djlC* to the KanS strain with an extra RRS near *fabR*, and the KanR was removed by FLP recombinase. The extra RRS near *dinI*, and deletions of the wild-type RRS near *lysO*, then the wild-type RRS near *dusC* were made by subsequent P1 transductions. The resulting strain is designated EAW1776, and has 3 extra RRS near *fabR*, *djlC*, and *dinI*, with both wild-type RRS deleted. After each P1 transduction, the extra RRS and wild-type RRS deletions were confirmed by PCR using primers outside the insertion or deletion area and direct sequencing of the PCR product.

ssGap-seq. Measurement of the genome-wide distributions of RecA and SSB bound to ssDNA by ssGap-seq was carried out essentially as described in Pham et al [19].

Psora-seq. Measurement of whole-genome supercoiling at high resolution by Psora-seq was carried using a published protocol [20] with several modifications detailed below.

Cell preparation for Psora-seq. Cells from wild-type MG1655, mutant EAW1706 *(ΔRRS-lysO*), EAW1708 *(ΔRRS-dusC*) and EAW1710 (double deletion *ΔRRS-lysO*, *ΔRRS-dusC*) strains were grown in LB medium at 37º C. 100 ml of cells in mid-exponential phase (OD_{600}=0.2) were chilled on ice for 3 min, then harvested by centrifugation (4000 x g, 5 min) and resuspended in 10 ml of ice-cold TE buffer (10 mM Tris, pH 7.4, 1 mM EDTA) containing 5 mg/ml biotinylated psoralen (EZ-Link Psoralen-PEG3-Biotin, ThermoFisher). Cell suspensions were incubated in the dark at 4º C for 10 min, then transferred to a 100 mm polystyrene Petri dish. Crosslinking of psoralen to genomic DNA was performed by cell exposure to 365 nm UV light (4.8 J/cm^2 total energy) using a UVP CL-3000L crosslinker (UVP, Inc. Upland, California). Cells were washed with 40 ml of ice-cold TE buffer and cell pellets were stored at -80º C.

Genomic DNA purification. Psoralen-crosslinked cells were lysed with lysozyme (2 mg/ml) in 4 ml of lysis buffer (20 mM Tris, pH 8.0, 5 mM EDTA, 100 mM NaCl) for 10 min at 37º C. Cellular RNA was removed by treatment with RNAse A (100 μg/ml) for 30 min at 37º C. Cellular proteins were then digested by addition of Proteinase K (final concentration 0.5 mg/ml) and SDS (final concentration 1%) and incubated for 2 h at 37º C. Genomic DNA (gDNA) was purified by twice extracting with phenol:chloroform: isoamyl alcohol (25:24:1) and precipitated with ethanol. Purified gDNA were resuspended in 250 ml of TE buffer (10 mM Tris-HCl, pH 8.0; 0.1 mM EDTA) and stored at -20º C.

Psoralen affinity purification. Genomic DNA (120 ml volume) was sheared to an average size of 400-500 bp by sonication using a bath sonicator (Covaris S2) in a 6X16 mm microTUBE AFA fiber. The temperature of the water bath was held between 5 to 8º C. Sheared DNA was quantified by Qubit Fluorometer (ThermoFisher). For affinity purification, 50 ml of 1 mm diameter streptavidin-coupled magnetic beads (Invitrogen Dynabeads MyOne Streptavidin T1) were washed 3 times and incubated with 10 mg of sheared gDNA in 1 ml B&W buffer (5 mM Tris, pH 7.5, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween-20) for 1h at room temperature, followed by 18 h at 4º C in the dark. Unbound DNA was removed by washing 6 times with B&W buffer and 2 times with TE buffer at room temperature. Washed beads were resuspended in 100 ml of TE, and DNA was eluted by twice extracting with an equal volume of phenol:chloroform:isoamylalcohol (25:24:1). The phenol solution efficiently disrupts biotin-streptavidin interaction, allowing high DNA recovery approaching 100% at room temperature [30]. Psoralen pull-down DNA was further purified using a Monarch PCR & DNA clean-up kit (New England Biolabs), eluted in 15 ml of EB buffer (10 mM Tris, pH 8.5) and stored at -20º C.

Genome sequencing and analysis. 1 ng of psoralen pull-down DNA (or corresponding Input DNA from the same cells without pull-down) was used to prepare sequencing libraries using a Nextera XT library prep kit (Illumina, Inc) according to the manufacturer protocol. Libraries were quantified by Qubit, qPCR and subjected to Illumina's sequencing (2 x 150 bp paired-end) on a MiniSeq instrument using High Output Reagent Kits (300-cycles).

Illumina sequencing data were analyzed using the CLC Genomic Workbench (v. 23.0) (Qiagen). High quality sequencing reads were imported, 5 nt were trimmed from each end, and mapped to an MG1655 reference genome (Accession number NC_000913) using read alignment settings of 0.8 for a length fraction and 0.9 for similarity fraction. Default settings for match score, mismatch, insertion and deletion costs were used. Mapping coverages of aligned reads (1 - 3 million per sample) with detailed per-base coverage information for psoralen pull-down reads and input reads were exported from the CLC Workbench program in a tab delimited format. Genome-wide binding of psoralen was calculated as described by Visser et al. [20] as follow. A custom-written R-script was used to calculate the normalized read coverages for the psoralen pull-down and the corresponding Input samples at individual positions by taking moving averages per 1 kb window. Psoralen enrichment (z-scores) was calculated by calculating the ratios of normalized psoralen pull-down values to normalized Input values then $\log_2$ transformed [20]. Psoralen binding profiles on larger scales (5 kb, 10 kb, 25 kb, 50 kb, 100 kb, or 250 kb) were calculated similarly, using normalized read coverage values for a moving window of 5 kb, 10 kb, 25 kb, 50 kb, 100 kb, or 250 kb, respectively. To compare psoralen binding profiles of wild-type and RRS deletion strains, psoralen binding z-scores at genomic positions from independent replicates were averaged, and then used to calculate Pearson correlation coefficients (r). R-programming was used for statistical analyses and to generate psoralen binding graphs.

REFERENCES

1 Surovtsev, I. V. & Jacobs-Wagner, C. Subcellular Organization: A Critical Feature of Bacterial Cell Replication. *Cell* **172**, 1271-1293 (2018). https://doi.org:10.1016/j.cell.2018.01.014

2 Dame, R. T., Rashid, F. Z. M. & Grainger, D. C. Chromosome organization in bacteria: mechanistic insights into genome structure and function. *Nature Reviews Genetics* **21**, 227-242 (2020). https://doi.org:10.1038/s41576-019-0185-4

3 Dorman, C. J. DNA supercoiling and transcription in bacteria: a two-way street. *BMC Molecular and Cell Biology* **20**, 26 (2019). https://doi.org:10.1186/s12860-019-0211-6

4 Espeli, O. & Boccard, F. Organization of the *Escherichia coli* chromosome into macrodomains and its possible functional implications. *Journal of Structural Biology* **156**, 304-310 (2006). https://doi.org:10.1016/j.jsb.2006.07.010

5 Espeli, O. *et al.* A MatP-divisome interaction coordinates chromosome segregation with cell division in *E. coli*. *EMBO Journal* **31**, 3198-3211 (2012). https://doi.org:10.1038/emboj.2012.128

6 Cagliero, C., Grand, R. S., Jones, M. B., Jin, D. J. & O'Sullivan, J. M. Genome conformation capture reveals that the *Escherichia coli* chromosome is organized by replication and transcription. *Nucleic Acids Research* **41**, 6058-6071 (2013). https://doi.org:10.1093/nar/gkt325

7 Lioy, V. S., Junier, I. & Boccard, F. Multiscale dynamic structuring of bacterial chromosomes. *Annual Review of Microbiology, Vol 75, 2021* **75**, 541-561 (2021). https://doi.org:10.1146/annurev-micro-033021-113232

8 Macvanin, M. & Adhya, S. Architectural organization in *E. coli* nucleoid. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms* **1819**, 830-835 (2012). https://doi.org:10.1016/j.bbagrm.2012.02.012

9 Verma, S. C., Qian, Z. & Adhya, S. L. Architecture of the *Escherichia coli* nucleoid. *Plos Genetics* **15**, e1008456 (2019). https://doi.org:10.1371/journal.pgen.1008456

10 Wang, W. Q., Li, G. W., Chen, C. Y., Xie, X. S. & Zhuang, X. W. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science* **333**, 1445-1449 (2011). https://doi.org:10.1126/science.1204697

11 Mercier, R., Espeli, O. & Boccard, F. Bacterial chromosome organization. *M S-Medecine Sciences* **25**, 885-887 (2009). https://doi.org:10.1051/medsci/20092511885

12 Valens, M., Penaud, S., Rossignol, M., Cornet, F. & Boccard, F. Macrodomain organization of the *Escherichia coli* chromosome. *EMBO Journal* **23**, 4330-4341 (2004). https://doi.org:10.1038/sj.emboj.7600434

13 Espeli, O. & Boccard, F. *In vivo* cleavage of *Escherichia coli* BIME-2 repeats by DNA gyrase: genetic characterization of the target and identification of the cut site. *Molecular Microbiology* **26**, 767-777 (1997). https://doi.org:10.1046/j.1365-2958.1997.6121983.x

14 Rudd, K. E. Novel intergenic repeats of *Escherichia coli* K-12. *Research in Microbiology* **150**, 653-664 (1999). https://doi.org:10.1016/s0923-2508(99)00126-6

15 Malhotra, N. & Seshasayee, A. S. N. Replication-dependent organization constrains positioning of long DNA repeats in bacterial genomes. *Genome Biology and Evolution* **14**, evac102 (2022). https://doi.org:10.1093/gbe/evac102

16 Qian, Z. & Adhya, S. DNA repeat sequences: diversity and versatility of functions. *Current Genetics* **63**, 411-416 (2017). https://doi.org:10.1007/s00294-016-0654-7

17 Smirnov, G. B. Repeats in bacterial genome: evolutionary considerations. *Molecular Genetics Microbiology and Virology* **25**, 56-65 (2010). https://doi.org:10.3103/s0891416810020023

18 Treangen, T. J., Abraham, A. L., Touchon, M. & Rocha, E. P. C. Genesis, effects and fates of repeats in prokaryotic genomes. *Fems Microbiology Reviews* **33**, 539-571 (2009). https://doi.org:10.1111/j.1574-6976.2009.00169.x

19 Pham, P., Wood, E. A., Cox, M. M. & Goodman, M. F. RecA and SSB genome-wide distribution in ssDNA gaps and ends in *Escherichia coli*. *Nucleic Acids Research* **51**, 5527-5546 (2023). https://doi.org:10.1093/nar/gkad263

20 Visser, B. J. *et al.* Psoralen mapping reveals a bacterial genome supercoiling landscape dominated by transcription. *Nucleic Acids Research* **50**, 4436-4449 (2022). https://doi.org:10.1093/nar/gkac244

21 Cox, M. M. *et al.* Generation and repair of postreplication gaps in *Escherichia coli*. *Microbiology and Molecular Biology Reviews* **87**, e00078-00022 (2023). https://doi.org:doi:10.1128/mmbr.00078-22

22 Palmer, G. H., Bankhead, T. & Seifert, H. S. Antigenic variation in bacterial pathogens. *Microbiology Spectrum* **4**, 28 (2016). https://doi.org:10.1128/microbiolspec.VMBF-0005-2015

23 Vink, C., Rudenko, G. & Seifert, H. S. Microbial antigenic variation mediated by homologous DNA recombination. *FEMS Microbiology Reviews* **36**, 917-948 (2012). https://doi.org:10.1111/j.1574-6976.2011.00321.x

24 Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nature Reviews Genetics* **13**, 770-780 (2012). https://doi.org:10.1038/nrg3296

25 Cueny, R. R., McMillan, S. D. & Keck, J. L. G-quadruplexes in bacteria: insights into the regulatory roles and interacting proteins of non-canonical nucleic acid structures. *Critical Reviews in Biochemistry and Molecular Biology* **57**, 539-561 (2022). https://doi.org:10.1080/10409238.2023.2181310

26 Mellor, C., Perez, C. & Sale, J. E. Creation and resolution of non-B-DNA structural impediments during replication. *Critical Reviews in Biochemistry and Molecular Biology* **57**, 412-442 (2022). https://doi.org:10.1080/10409238.2022.2121803

27 Varshney, D., Spiegel, J., Zyner, K., Tannahill, D. & Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. *Nature Reviews Molecular Cell Biology* **21**, 459-474 (2020). https://doi.org:10.1038/s41580-020-0236-x

28 Wang, G. L. & Vasquez, K. M. Dynamic alternative DNA structures in biology and disease. *Nature Reviews Genetics* **24**, 211-234 (2023). https://doi.org:10.1038/s41576-022-00539-9

29 Zyner, K. G. *et al.* Genetic interactions of G-quadruplexes in humans. *eLife* **8**, e46793 (2019). https://doi.org:10.7554/eLife.46793

30 Bearden, S., Wang, F. & Hall, A. R. Simple and efficient room-temperature release of biotinylated nucleic acids from streptavidin and Its application to selective molecular detection. *Analytical Chemistry* **91**, 7996-8001 (2019). https://doi.org:10.1021/acs.analchem.9b01873

FIGURE LEGENDS

Figure 1. RRS induces formation of genomic ssDNA on the lagging strands. A. Sketch depicting positions of RRS-*lys*O, RRS-*dusC*, OriC, *dif, Ter* sites, four macrodomains (ORI, LEFT, RIGHT, TER) and two non-structured domains (NSL and NSR) of *E. coli* chromosome. Positions of *fabR*, *djlC* and *di*nI genes are shown in green color. **B,** A putative RRS-*lysO* secondary structure. **C,** Genome-wide distribution of RecA-ssDNA in *wild-type* and mutant strains containing extra RRS sequences inserted at *fabR* locus (EAW1736); *fabR* and *djlC* loci (EAW1765); *fabR, djlC* and *dinI* loci (EAW1768). Strain EAW1776 contains three RRS elements at *fabR, djlC* and *dinI* loci, but two original RRS at *lysO* and *dusC* had been removed. Normalized coverages of RecA-ssDNA are shown as blue traces for the W-strand, and red traces for the C-strand. Green lines in graphs (value = 1) indicate the genome average coverage. Each data point represents a moving average of RecA-ssDNA in a 10 kb window. *Top inserts*, zoom-in RRS-*lysO* and RRS-*dusC* regions showing RecA-ssDNA binding on the lagging strand peaks at RRS then extends further by about 2 Kbp.

Figure 2. Genome amplification of *insH7-insH8* genomic in colonies with double *ΔRRS-lysO* and *ΔRRS-dusC* deletion. Whole genome sequencing read mappings of wild-type MG1655, single RRS deletion (*ΔRRS-lysO* and *ΔRRS-dusC*) and double *ΔRRS* strains. Surviving double *ΔRRS* deletion colonies have 2 - 3 folds amplification of a 200 kb region between *insH7* and *insH8,* encompassing *dusC* locus. Mapped reads are shown as aggregated coverage graphs with average coverage values at each genome position shown in blue, maximum coverage - dark blue and minimum coverage – light blue. Positions of -*lys*O, *dusC*, OriC, *dif* and macrodomains are shown on top. Strain EAW1776 contains three RRS elements at *fabR, djlC* and *dinI* loci, but two original RRS at *lysO* and *dusC* had been removed.

Figure 3. Deletion of both RRS (with compensation) does not affect growth or sensitivity to UV or ciprofloxacin. (A) Growth curves for all of the strains studied in Figure 2 are shown. Two colonies of EAW1710 (double Δ*RRS* deletion), each picked separately, are strains with the *insH7-8* region amplified but otherwise have no detectable RRS elements. EAW1776 has both of the natural RRS deleted but also has three additional RRS in other genomic locations as shown in Figure 1A. Panels in parts B and C show UV and ciprofloxacin sensitivities, respectively.

Figure 4. RRS elements affect genome-wide supercoiling. A, Psora-seq maps (250 kb resolution) of wild-type and RRS deletion mutants. Psoralen binding values (log2 pull-down/input) are the average of at least two independent Psora-seq experiments for mid-exponential cells. Red and blue tracks indicate supercoiling that is more negative or more positive than the genome average, respectively. Pearson correlation coefficients (*r*) are shown for comparison between psoralen binding profiles of RRS deletion mutants and wild-type. Positions OriC, *dif, lysO, dusC* and 7 ribosomal operons (*rrn*) are indicated by vertical dashed lines. **B**, Quantification of psoralen binding within macrodomains of wild-type and RRS deletion mutants. Error bars indicate ± standard deviations. **C**, Average psoralen binding in negatively and supercoiled regions of whole chromosome and TER macrodomain. Psoralen binding z scores in **B** and **C** were calculated using 1-kbp sliding windows. **D**, The absence of the twin supercoiling domain signature at *rrn* operons in double *ΔRRS* cells. Psoralen binding along an 80-Kbp region surrounding *rrnA* operon from three independent replicates for wild-type (*top graph*) and two

replicates for double *ΔRRS* cells (*bottom graph*). Sketch on top showed predicted psoralen binding profile according to a transcription-induced supercoiling twin-domain model. **E** and **F**, Most of BIME-2 sites are located between supercoiling regions of chromosome. Distribution of BIME-2 sites along *E. coli* chromosome (*top graph*) and zoom-in genomic region between 0.7 - 0.8 Mbp for wild-type and double *ΔRRS*. Positions of BIME-2 sites are indicated by green vertical dashed lines.

Figure 5. Model for how RRS triggers gap formation.

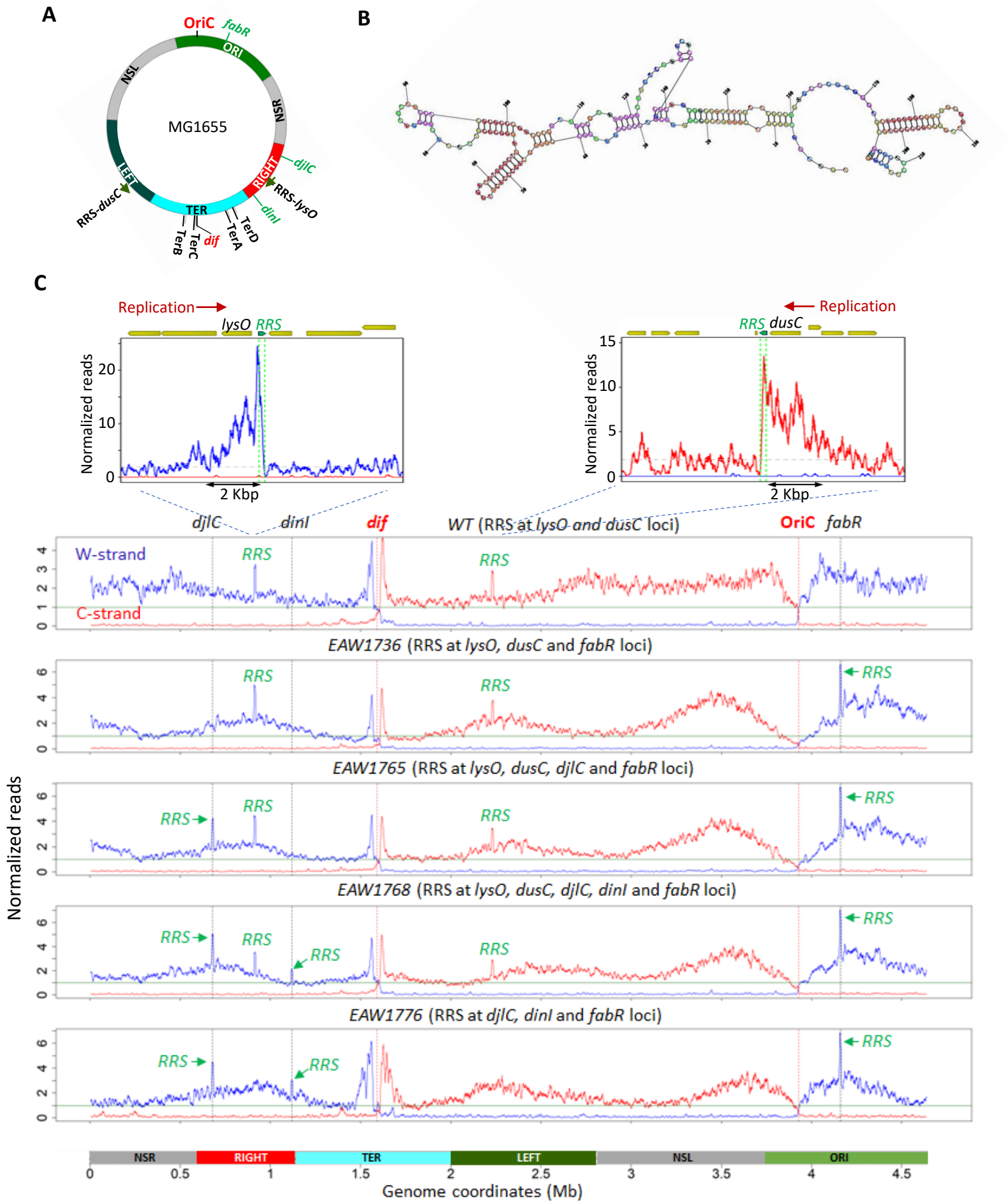

A
OriC
fabR
ORI
NSR
NSL
MG1655
djlC
RIGHT
RRS-lysO
dinI
TER
TerD
TerA
dif
TerC
TerB
LEFT
RRS-dusC
B
C
Replication
lysO
RRS
Normalized reads
2 Kbp
Replication
RRS
dusC
djlC
dinI
dif
WT (RRS at lysO and dusC loci)
OriC
fabR
W-strand
C-strand
EAW1736 (RRS at lysO, dusC and fabR loci)
EAW1765 (RRS at lysO, dusC, djlC and fabR loci)
EAW1768 (RRS at lysO, dusC, djlC, dinI and fabR loci)
EAW1776 (RRS at djlC, dinI and fabR loci)
NSR
RIGHT
TER
LEFT
NSL
ORI
Genome coordinates (Mb)

Figure 1

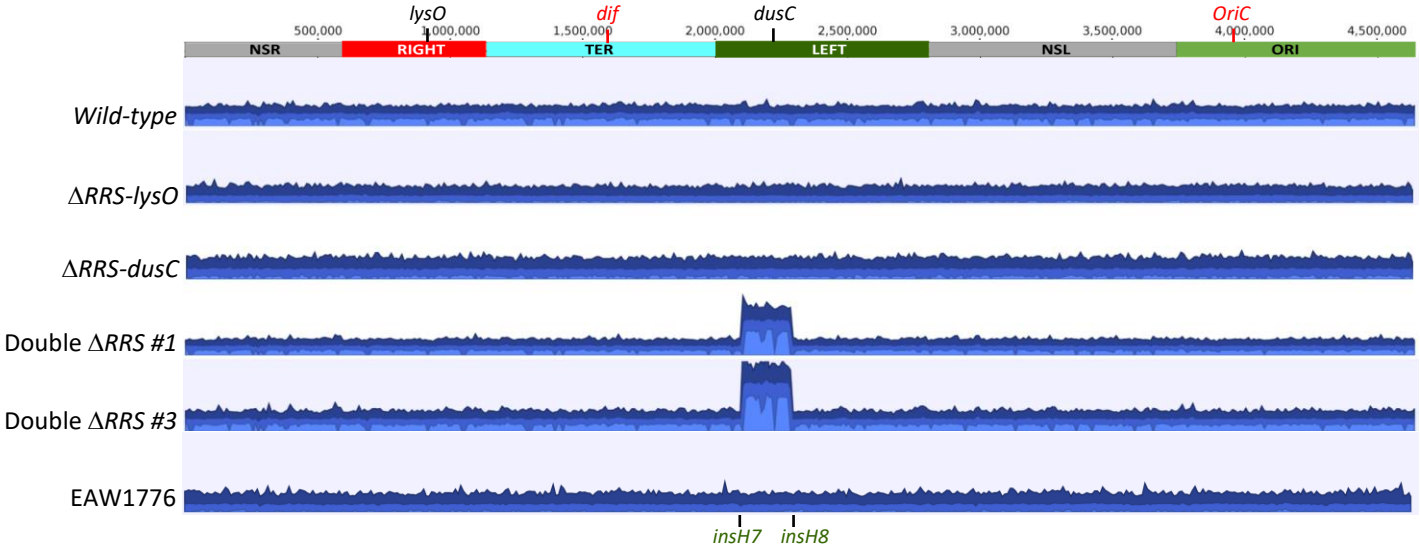

lysO
dif
dusC
OriC
500,000
1,000,000
1,500,000
2,000,000
2,500,000
3,000,000
3,500,000
4,000,000
4,500,000
NSR
RIGHT
TER
LEFT
NSL
ORI
Wild-type
ΔRRS-lysO
ΔRRS-dusC
Double ΔRRS #1
Double ΔRRS #3
EAW1776
insH7
insH8

Figure 2

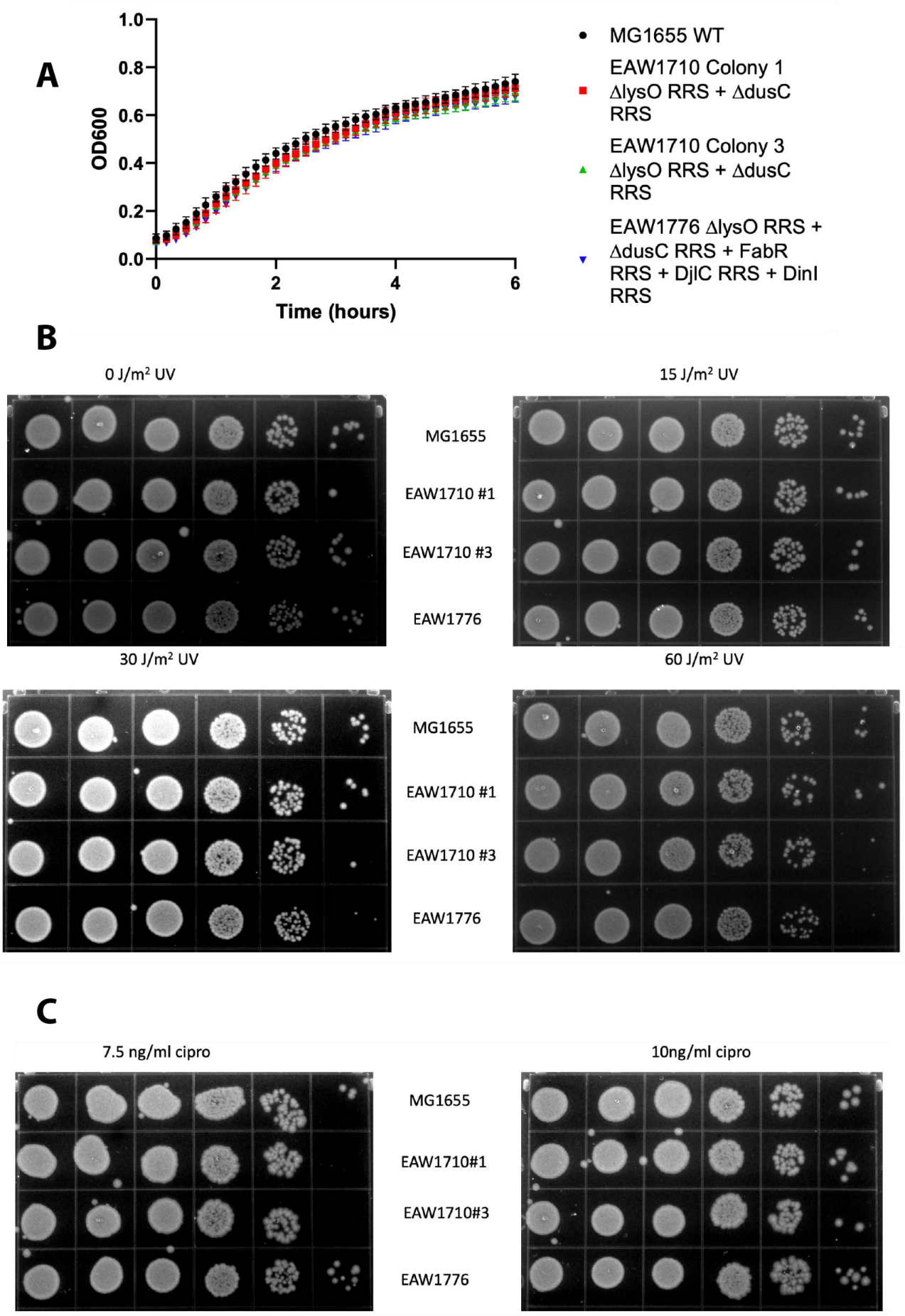
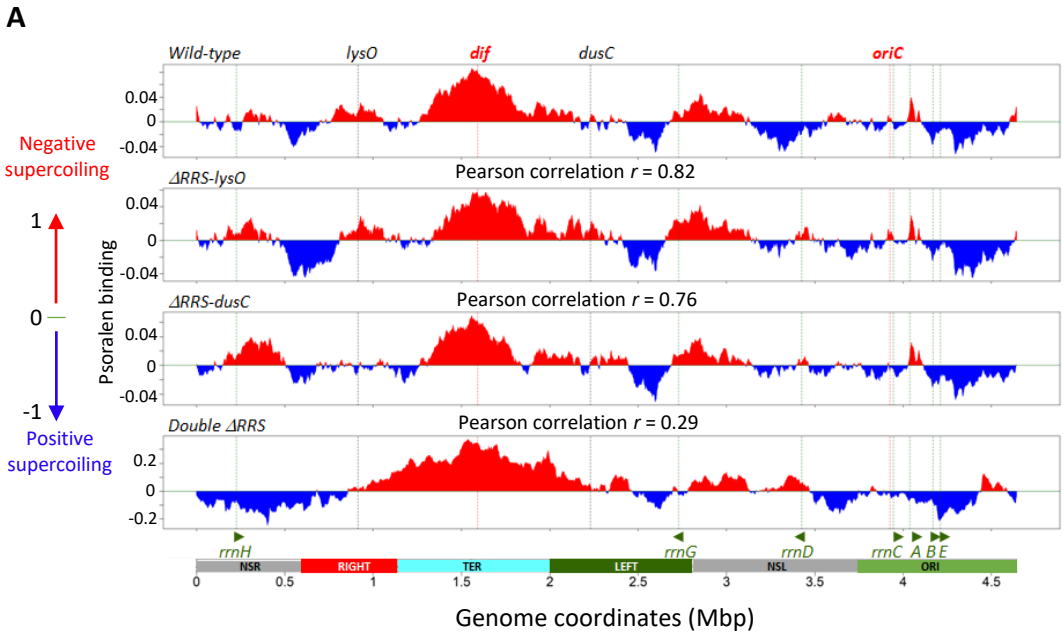
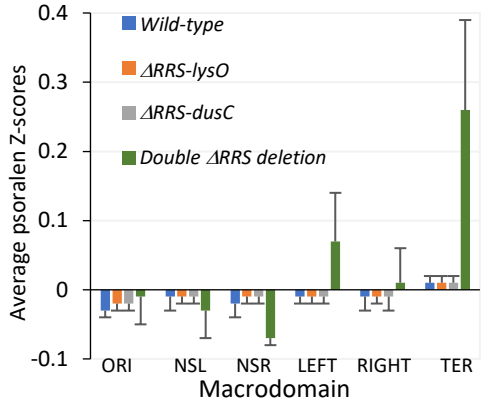

A
OD600
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
Time (hours)
MG1655 WT
EAW1710 Colony 1 ΔlysO RRS + ΔdusC RRS
EAW1710 Colony 3 ΔlysO RRS + ΔdusC RRS
EAW1776 ΔlysO RRS + ΔdusC RRS + FabR RRS + DjlC RRS + DinI RRS
B
0 J/m² UV
15 J/m² UV
30 J/m² UV
60 J/m² UV
MG1655
EAW1710 #1
EAW1710 #3
EAW1776
C
7.5 ng/ml cipro
10ng/ml cipro
MG1655
EAW1710#1
EAW1710#3
EAW1776

Figure 3



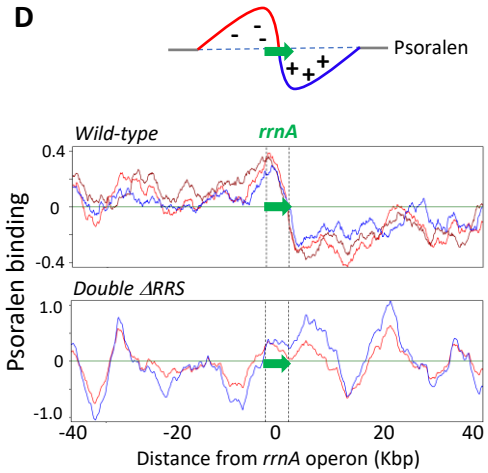


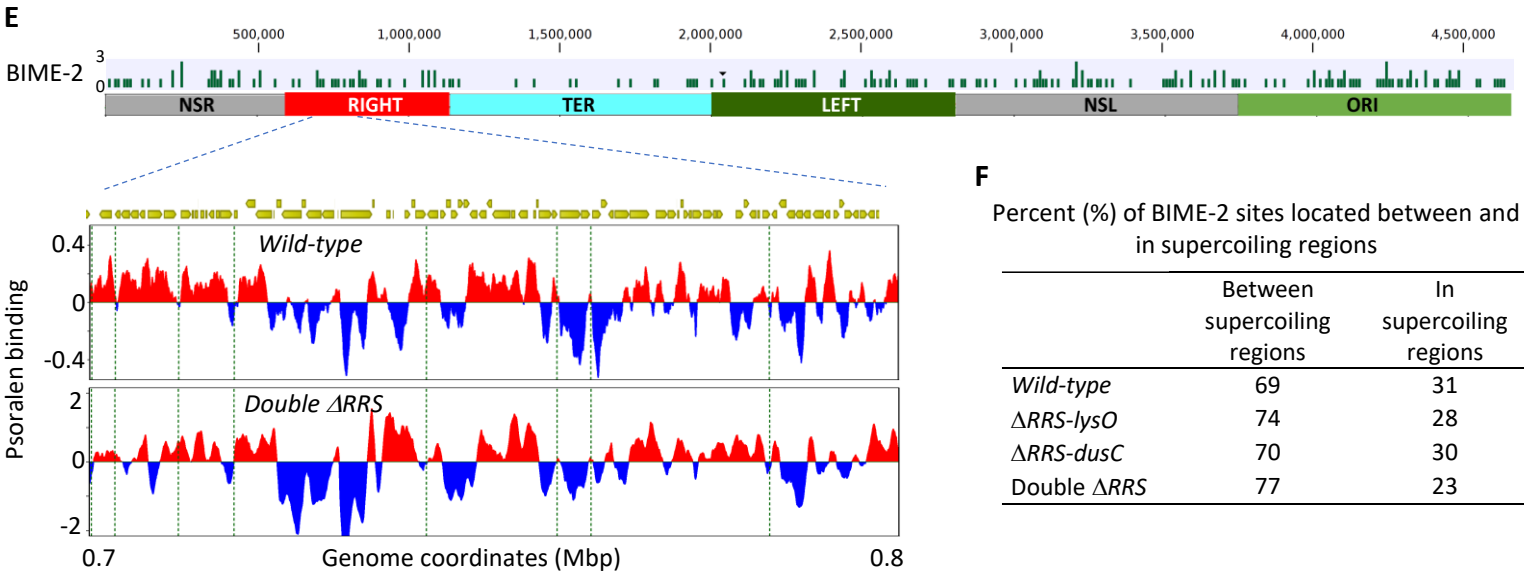
C

Average psoralen binding in negatively supercoiled and positively supercoiled regions of chromosome (Chr) and TER macrodomain*

Cells	Negative supercoiling		Positive supercoiling	
	Chr	TER	Chr	TER
Wild-type	0.16 ± 0.02	0.18 ± 0.01	-0.17 ± 0.02	-0.18 ± 0.01
ΔRRS-lysO	0.16 ± 0.01	0.18 ± 0.01	-0.17 ± 0.01	-0.19 ± 0.02
ΔRRS-dusC	0.17 ± 0.02	0.19 ± 0.03	-0.18 ± 0.02	-0.19 ± 0.03
Double *ΔRRS*	0.51 ± 0.19	0.55 ± 0.20	-0.54 ± 0.15	-0.40 ± 0.11

* Values represent Z-score means ± standard deviations of at least two replicates.





F

Percent (%) of BIME-2 sites located between and in supercoiling regions

	Between supercoiling regions	In supercoiling regions
Wild-type	69	31
ΔRRS-lysO	74	28
ΔRRS-dusC	70	30
Double *ΔRRS*	77	23

Figure 4

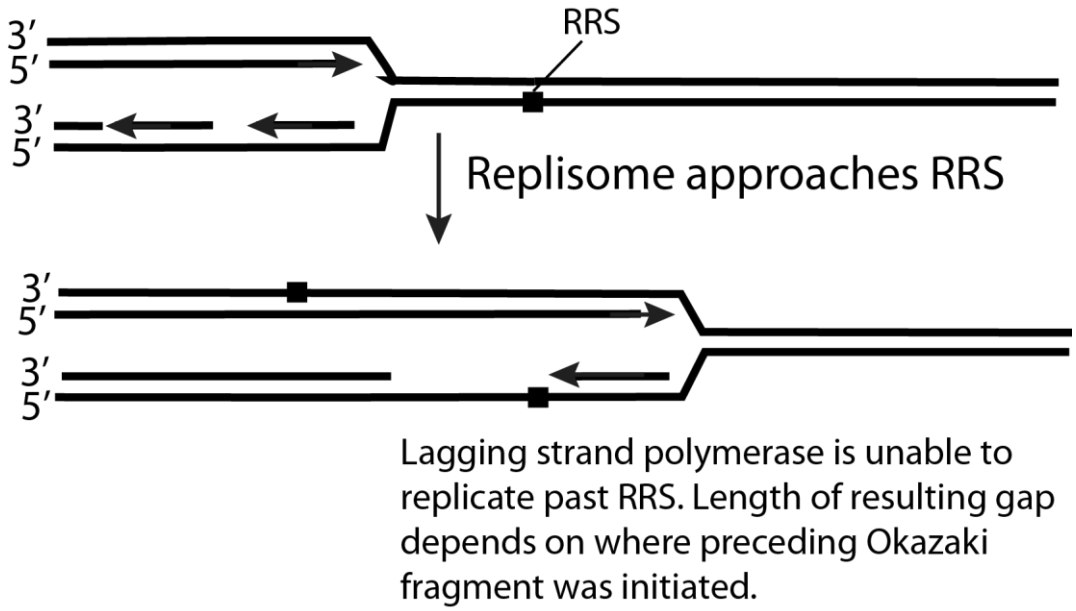

RRS
3′
5′
3′
5′
Replisome approaches RRS
3′
5′
3′
5′
Lagging strand polymerase is unable to replicate past RRS. Length of resulting gap depends on where preceding Okazaki fragment was initiated.

Figure 5