Comparative Genomics of Interreplichore Translocations in Bacteria: A Measure of Chromosome Topology?

Supriya Khedkar*,†,1 and Aswin Sai Narain Seshasayee*,1
*National Centre for Biological Sciences, GKVK (Gandhi Krishi Vignana Kendra), Bellary Road, Bangalore 560065, Karnataka, India and †SASTRA (Shanmugha Arts, Science, Technology and Research Academy) University, Thanjavur 613401, Tamil Nadu, India

ABSTRACT Genomes evolve not only in base sequence but also in terms of their architecture, defined by gene organization and chromosome topology. Whereas genome sequence data inform us about the changes in base sequences for a large variety of organisms, the study of chromosome topology is restricted to a few model organisms studied using microscopy and chromosome conformation capture techniques. Here, we exploit whole genome sequence data to study the link between gene organization and chromosome topology in bacteria. Using comparative genomics across ~250 pairs of closely related bacteria we show that: (a) many organisms show a high degree of interreplichore translocations throughout the chromosome and not limited to the inversion-prone terminus (*ter*) or the origin of replication (*oriC*); (b) translocation maps may reflect chromosome topologies; and (c) symmetric interreplichore translocations do not disrupt the distance of a gene from *oriC* or affect gene expression states or strand biases in gene densities. In summary, we suggest that translocation maps might be a first line in defining a gross chromosome topology given a pair of closely related genome sequences.

Chromosome compaction is a necessary feature of all living cells including bacteria. It is intricately linked to chromosome topology and the selective pressures introduced by gene organization along the genome sequence.

In many bacteria, nucleoid-associated proteins (Dillon and Dorman 2010) and topoisomerases (DiNardo *et al.* 1982; Pruss *et al.* 1982) establish local DNA geometries, which include bending (Stella *et al.* 2010), bridging (Dame *et al.* 2006; Dupaigne *et al.* 2012; Liu *et al.* 2010; Maurer *et al.* 2009), wrapping, and relative over-/underwinding of

[1]Corresponding authors: National Centre for Biological Sciences, Tata Institute of Fundamental Research, GKVK, Bellary Road, Bangalore 560065, Karnataka. India. E-mail: aswin@ncbs.res.in; and National Centre for Biological Sciences, Tata Institute of Fundamental Research, GKVK, Bellary Road, Bangalore 560065, Karnataka, India. E-mail: supriyak@ncbs.res.in

DNA. These not only help compact DNA, but also play a role in constraining replication (Gruber *et al.* 2014; X. Wang *et al.* 2014a), recombination (Esposito and Gerard 2003; Shanado *et al.* 2001), and transcription (Browning *et al.* 2010; Dillon and Dorman 2010).

Zooming out, two broad chromosome topologies have been discerned across model bacteria. First, in *Escherichia coli*, genetic, recombination-based contact maps of the chromosome showed the presence of four macrodomains, defined by the preferential occurrence of recombination events within rather than across these macrodomains (Valens *et al.* 2004). Chromosome topology in *E. coli* is further characterized by the localization of the origin of replication (*oriC*) at midcell and the transverse spread of the left and right replichores in opposite directions toward the cell poles (Nielsen *et al.* 2006; Wang *et al.* 2006). This is referred to as the "*transverse*" chromosome topology.

On the other hand, high-resolution 5C and Hi-C chromosome contact maps of *Caulobacter crescentus* present a different picture by revealing a "*longitudinal*" topology (Le *et al.* 2013; Umbarger *et al.* 2011). In the *C. crescentus* chromosome, symmetric contacts (defined as contacts between loci equidistant from *oriC*) between the two replichores (two arms of the chromosome extending from *oriC* at one pole to the terminus at the other pole) could be discerned, possibly resulting in a helical structure (Le and Laub 2014); such contacts are rare in the *E. coli* chromosome (Nielsen *et al.* 2006; Wang *et al.* 2006). Similarly,

recent microscopy experiments showed that the two replichores of *Mycobacterium smegmatis* colocalize throughout replication (Santi and McKinney 2015). For both the topologies, distinguished by the presence or absence of close interreplichore contacts, it is apparent that *oriC* and the terminus (*ter*) rarely contact each other. Finally, the chromosome of *Bacillus subtilis* has been shown to oscillate between these two conformations (X. Wang *et al.* 2014b).

The molecular mechanisms and players underlying the establishment of a defined chromosome topology are only beginning to be understood. These include DNA-binding proteins such as MatP (Mercier *et al.* 2008), the SMC complex (Le *et al.* 2013), nucleoid-associated proteins (Umbarger *et al.* 2011; Wang *et al.* 2011), and noncoding RNAs (Qian *et al.* 2015). Chromosome topologies might also be influenced by active cellular processes such as DNA replication and transcription. In *E. coli*, replication has been implicated in shaping the topology of the chromosome (Cagliero *et al.* 2013). Modeling of the bacterial chromosome using tools from polymer science has indicated that large switches in chromosome topology can occur by the mere repositioning of *oriC* or *ter* sites within the cell (Junier *et al.* 2014). In *C. crescentus*, highly expressed genes act as topological domain boundaries on the chromosome (Le *et al.* 2013).

An important question then, is whether chromosome topology has any functional relevance. In other words, is chromosome topology under evolutionary selection? Whereas local DNA geometries, including the writhe of DNA around a nucleoid-associated protein, has definite effects on transcription and DNA repair (Dillon and Dorman 2010; Dorman 2013), whether the overall shape of the chromosome has any function, beyond being a solution for compacting DNA within the confines of a cell, is unclear. Long-range interactions between chromosomal segments might enable coexpression of genes encoded in these segments (Wang *et al.* 2016). The collapse of transcriptionally silent genes by the global repressor H-NS into a few foci in *E. coli* (Wang *et al.* 2011) – akin to the spatial clustering of distinct heterochromatin sequences in the eukaryotic nucleus − indicates conservation of this topological characteristic across phyla; whether this is a requirement for transcriptional silencing remains an open question. This point, however, must be tempered by the fact that fluorescent tags used for observing the foci may actually be the cause of some of the clustering effects (S. Wang *et al.* 2014). Further, Esnault *et al.* (2007) have revealed that the effects of DNA inversions on growth are dependent on their impact on chromosome topology; *i.e.*, inversions that span multiple topological macrodomains of the chromosome are more detrimental than those that are limited to a single domain, thus suggesting that chromosome topology might be under evolutionary selection.

While the study of aspects of bacterial chromosome topology and of its evolution is still nascent (Lagomarsino *et al.* 2015), we have a substantially deeper understanding of chromosome architecture in terms of how genes are organized on the bacterial chromosome (Rocha 2008). Depending on the time spent per cell cycle by a bacterial cell in replicating the genome, there is a clear difference in gene dosage between *oriC* and *ter* (Block *et al.* 2012; Schmid and Roth 1987). This is particularly apparent in fast-growing bacteria like *E. coli*, where *oriC* fires multiple times per cell cycle (Cooper and Helmstetter 1968). As a potential consequence of this critical difference in gene dosage between *oriC* and *ter*, essential and highly expressed genes are encoded close to *oriC* (Rocha 2004). In contrast, many stress-responsive (Sobetzko *et al.* 2012) and horizontally acquired genes are encoded around *ter* (Zarei *et al.* 2013). Finally, a majority of bacterial genes – particularly those that are essential – are encoded on the leading strand (Rocha and Danchin 2003), consistent with the idea that head-on collision between the two polymerases is mutagenic and detrimental (Mirkin and Mirkin 2005).

Gene organization changes as species diverge (Tamames 2001). For example, comparative genomics has shown that loci immediately adjacent to *oriC* and *ter* are prone to interreplichore inversions (Eisen *et al.* 2000; Suyama and Bork 2001). Previous anecdotal genome alignment analyses have also revealed the presence of interreplichore translocations in certain pairs of genomes; more so for distantly related organisms. These result in the classical X-shaped pattern in sequence comparison dot plots (Eisen *et al.* 2000; Koonin and Wolf 2008; Mira *et al.* 2002; Suyama and Bork 2001; Tillier and Collins 2000).

Here we interrogate the relationship between chromosome topology and gene organization across bacteria. Specifically, we ask whether we can exploit the wealth of genome sequence data to explore the utility of gene organization as a tool to infer the topology of a bacterial chromosome, thus accounting in part for the paucity of chromosome conformation data across bacteria.

MATERIALS AND METHODS

Data

We obtained genomic DNA [.fna files], RNA [.frn;.rnt files], and protein [.faa] sequence data for ~3000 completely sequenced bacteria from the National Center for Biotechnology Information (NCBI). We obtained gene coordinates, protein IDs, and Cluster of Orthologous genes (COG) annotations from .ptt files (ftp://ftp.ncbi.nlm.nih.gov/genomes/archive/old_refseq/Bacteria/).

R-factor calculation

We calculated R-factor (R_f) as a physiologically relevant measure of replication-dependent gene dosage gradient down the *oriC*−*ter* axis. This considers both replication time and minimum doubling time. It was first described by Couturier and Rocha (2006).

The R-factor (R_f) was defined as $R_f = T_R/T_D$, where T_R stands for the estimated time taken for a full round of chromosomal replication, and T_D represents the minimum doubling time. T_R was defined as the ratio of half the genome size and an average speed of DNA replication *in vivo*, which was taken to be 600 nt/sec (Reyes-Lamothe *et al.* 2008) for all bacteria. This is an oversimplification: DNA replication speeds vary across bacteria, but such data are not readily available, and the extent of its variability is unknown. T_D was obtained for ~100 bacteria from Freilich *et al.* (2009).

Distance calculation

We determined the location of a gene in the genome as its distance from *oriC* down the replichore. We first obtained *oriC* and *ter* coordinates for 1528 completely sequenced bacteria from DoriC (Gao and Zhang 2007). We calculated the shortest distance of all genes from the origin of replication and divided it by half the genome size. The normalized distances of genes from *oriC* varied from 0 to 1, 0 being close to *oriC* and 1 being close to *ter*.

Dataset for pairwise comparisons

We extracted a 16S rRNA sequence for 1528 completely sequenced bacteria. For organisms with multiple 16S rRNAs, we obtained the sequence of the copy closest to *oriC*. We did a 16S rRNA sequence similarity search using the Needleman−Wunsch global alignment algorithm (Needleman and Wunsch 1970), implemented in EMBOSS (EMBOSS:6.4.0.0) (Rice *et al.* 2000) using default parameters. We then considered all pairs of bacteria with greater than 97% 16S rRNA sequence identity. For further analysis we included only those pairs of genomes encoding the same number of 16S rRNA genes, as determined from the .rnt files in NCBI. This ensures that pairs of genomes used for

comparative genomics would have similar growth rates. As described in *Results*, we removed pairs involving: (a) strains of the same species; (b) redundancy; and (c) obligate parasites. Our final dataset consisted of 232 individual bacterial species forming 262 pairs of bacteria with the same 16S rRNA copy number and ≥97% 16S rRNA sequence identity.

Determination of orthologs

For the 262 pairs of bacteria, using the protein sequences (data from .faa files) we did an all *vs.* all bidirectional best-hit *phmmer* (HMMER 3.0, March 2010) search (Finn *et al.* 2011) with default parameters and an E-value cut off of 10^{-10}. The data pertaining to all the identified orthologs used in this study for 262 pairs of bacteria can be accessed here: http://bugbears.ncbs.res.in/genome_architecture/.

Correction for the effect of phylogenetic distance

The correction for the dependence of various measures of gene organization on phylogenetic distance (mentioned in the main text) was implemented as follows. We fitted the dependent variable (y) against 16S sequence similarity (x) using local polynomial regression (LOESS). The residual of this fit was calculated by subtracting individual values of y from the predicted values of y obtained after the LOESS regression. These residual values represented the component not explained by phylogenetic distance.

Assignment of chromosomal bins

We defined 25% of the chromosome centered around *oriC* as the O or origin bin. Twenty-five percent of the chromosome immediately to the right (along the text of the genome sequence) and to the left of the O bin was termed R and L, respectively. The remaining 25% of the chromosome, which includes *ter*, was termed the T bin.

Determination of functional enrichment

Functional annotations for individual orthologs were extracted from NCBI [.ptt files]. These COG annotations were used to determine enrichment of specific functions in each bin. Twenty-three COG classes (Supplemental Material, File S1) belonging to three broad functions — information storage and processing, cellular process, and signaling and metabolism – were considered. For each genome, Fisher's exact test was used to determine statistical significance for the enrichment of each COG class in any given bin. The obtained *P*-values were then corrected for multiple comparisons using Bonferroni correction and were negative log base 10 transformed.

Prediction of horizontally acquired genes

We used Alien Hunter (Release 1.7) (Vernikos and Parkhill 2006) with default parameter values to predict horizontally acquired genes. For this analysis, bacteria with guanine and cytosine (GC) content in the range of 40–60% were considered. The coordinates of predicted horizontally acquired genes obtained from Alien Hunter were further used to assign them to individual chromosomal bins. Depletion of predicted horizontally acquired genes in an individual bin was determined using Fisher's exact test adjusted for multiple testing by Bonferroni.

Analysis of C. crescentus 5C data

We obtained chromosome conformation capture data for *C. crescentus* NC_011916 from Le *et al.* (2013)(Gene Expression Omnibus — accession no. GSE45966; sample GSM1120445). From this dataset we extracted normalized contact frequency information for all bins showing interreplichore contacts.

Ortholog positions that showed interreplichore translocations in *C. crescentus* NC_011916 and its closely related bacterial pairs were pooled together. For all these translocated loci we extracted the contact frequency information from the above dataset. This formed the test dataset consisting of contact frequencies for interreplichore translocated loci. The control set included all other pairs of interreplichore loci that did not show translocations. The two distributions were compared using the Wilcoxon test.

Phylogenetic tree

The 16S rRNA sequence of the rRNA copy closest to origin of replication was obtained for all 232 bacterial species used in this study. 16S rRNA multiple sequence alignment was generated using MUSCLE (Edgar 2004), as implemented in MEGA version 6 (Tamura *et al.* 2013) using default parameters. A maximum likelihood (ML) tree was generated in MEGA version 6 using all sites in the 16S rRNA alignment. The Tamura and Nei (1993) nucleotide substitution model was used to infer the tree. The phylogenetic tree was visualized using iTOL (Letunic and Bork 2011), and growth rate and interreplichore translocation frequency were overlaid on the phylogenetic tree.

Randomization protocol

For each pair of bacteria showing interreplichore translocations, the assignment of an ortholog to a gene was randomized, thus creating random pairing between genes. Then D_{inter} was calculated for each gene pair in this randomized dataset. These randomized datasets were used to compute a null distribution for D_{inter} in Figure 4B and Figure 5A, and as the null expectation for the frequency of interreplichore strand flips in Figure 5B.

All the statistical tests and data visualizations were performed using *R* (unless specified otherwise).

Data availability

The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article.

RESULTS

A comparative genomic analysis of gene position conservation

In order to infer chromosome topologies and their association with gene dosage gradients down the *oriC*−*ter* axis, we established a comparative genomic framework. Our approach is based on the frequencies of translocations between distal segments of the chromosome, as defined by sequence comparisons between pairs of closely related organisms. This is based on the supposition that spatially proximal regions of the chromosome would recombine more frequently, and recombination combined with selection would result in the establishment of translocation events between such loci. This approach is similar to the experimental mapping of the topology of the *E. coli* chromosome using recombination frequencies between distal loci (Valens *et al.* 2004), with the difference that the translocations we identify are products of recombination followed by selection and probably drift over long timescales.

We obtained complete genome sequences and gene annotations for ~3000 fully sequenced bacterial genomes from the NCBI. Because gene dosage gradients down the *oriC*−*ter* axis are dependent on growth rates, we sought to annotate each organism in our collection of genomes with growth rate information. To do so, we assembled a dataset of growth rates in the form of minimum doubling times for over 100 bacteria from Freilich *et al.* (2009).

Although growth rate information is sparse relative to genome sequence data, we know that the number of rRNA genes encoded by

a bacterium is a predictor of its growth rate (Condon *et al.* 1995; Klappenbach *et al.* 2000), with fast-growing bacteria encoding more rRNA operons than slow-growing ones. Consistent with this knowledge, we observed a negative correlation ($\rho_{\text{Spearman}} = -0.76$) between minimum doubling times and rRNA copy numbers for these ~100 bacteria (Figure 1A).

Based on the fit between minimum doubling times and rRNA copy numbers, we estimated the growth rate expected for each rRNA copy number observed in our dataset. However, growth rate itself is not necessarily a strong predictor of differences in gene dosage between *oriC*-proximal and *ter*-proximal loci. Instead, a better correlate is the *R*-factor (R_f), defined as the number of replication cycles per cell division. Using a previous study from the Rocha group as a guide (Couturier and Rocha 2006), we calculated the R_f for each bacterium as described in *Materials and Methods*. We then defined organisms that are expected to initiate replication more than once per cell cycle on average ($R_f > 1$) as fast-growing, and the rest as slow-growing. We also estimated *R*-factor independently using a larger, more recent dataset (Vieira-Silva and Rocha 2010) of minimum doubling times for ~200 bacteria, and found estimates of R-factor derived from this dataset to be consistent with those used in this study (Figure S1).

We organized the ~3000 genomes into 262 genome pairs on the basis of phylogenetic distance and growth rate. We considered pairs of bacteria with ≥97% 16S rRNA sequence identity (Figure S2) and selected those with similar growth rates (same 16S rRNA copy numbers) (Figure 1B). We then removed: (a) pairs of strains of the same species, which typically showed very few long-range translocations (Figure S3); (b) redundant pairs to minimize overrepresentation of certain phylogenetic groups: for example, there are several hundred comparisons involving *E. coli* and *Salmonella*, and of these one was picked at random; and (c) pairs of obligate parasites with genomes smaller than 1.3Mb (Koonin and Wolf 2008), whose reduced genomes are known to be stable with exceptionally high gene order conservation (Tamas 2002; Mira *et al.* 2002). Thus, our final selection consisted of 232 individual bacterial species forming 262 pairs of bacteria with the same 16S rRNA copy number and ≥97% 16S rRNA sequence identity. A phylogenetic tree of these bacteria, based on their 16S rRNA sequences, is shown in Figure S4.

For all the 262 pairs of bacteria, we identified gene orthologs by bidirectional best-hit phmmer (E-value <10^{-10}). We found that ~72% (median) of genes in one organism had an ortholog in its partner genome. The percentage of genes conserved in a pair of organisms decreased with phylogenetic distance between the compared genomes and showed little correlation with growth rate (Figure S5).

For all these genes, we computed the position of each gene on its chromosome. The position of a gene was defined by its distance from *oriC* along the path of the replication fork and normalized to genome size. Positions equidistant from *oriC* on either replichore were considered equivalent in these calculations. Thus, a difference in gene position would indicate a difference in dosage, depending on the organism's R-factor. For each genome and its partner, we defined gene position vectors, in which each gene with an identifiable ortholog in its partner genome was represented by its distance from *oriC*. Then, for each genome pair, we computed the correlation coefficient between their gene position vectors (Figure 1C) and treated this correlation as an estimate of gene order conservation (ρ_{GOC}).

As expected for closely related organisms, the average gene order conservation across the 262 pairs of organisms studied here was high (median ρ_{GOC} ~0.9). Loss of gene order conservation is dependent on phylogenetic distance (Suyama and Bork 2001; Tamames 2001). Hence, we calculated residual ρ_{GOC} after correcting for the dependence of ρ_{GOC}

on phylogenetic distance (Figure S6). We found that the residual ρ_{GOC} increases with increasing R-factor (R_f) weakly but in a statistically significant fashion (P-value= 3.3×10^{-6}, Wilcoxon test; Figure 1D), consistent with earlier calculations by Rocha and colleagues (Couturier and Rocha 2006).

To identify patterns of long-range translocations, which could inform us about chromosome topologies, we took the following coarse-grained approach. We binned each chromosome into four equally sized segments. The first segment, termed 'O' for the presence of *oriC*, comprised 25% of the chromosome centered around *oriC*. The 'L' (Left) and the 'R' (Right) segments each covered 25% of the chromosome on either side of the O segment. The remaining 25%, including the terminus of replication, was termed the 'T' (Terminus) segment (Figure 2A). Each gene with an ortholog in a partner genome was assigned to one of these four bins.

Analysis of gene functions, for each bin, at a broad level using the COG annotations revealed that the only statistically significant association between gene function and gene positioning was the enrichment of genes involved in translation and ribosome biogenesis in the O (origin) bin (Figure 2B and Figure S7). In line with previous literature (Couturier and Rocha 2006), we observed that the average distance of translation and ribosome biogenesis genes from *oriC* decreases with increase in R_f (Figure S8). The enrichment of ribosome biogenesis and translation genes in the O segment was particularly apparent in fast-growing bacteria ($R_f > 1$) but not in slower growing ones (Figure 2B). In contrast, there was a significant depletion of predicted horizontally acquired genes in the O bin, in both fast- and slow-growing bacteria, and more so in the former (Figure 2C). This is consistent with the increased chance of novel gene integration disrupting essential gene functions in this region of the chromosome. These results are also in agreement with observations made in *E. coli* (Zarei *et al.* 2013) wherein horizontally acquired genes were defined using methods different from what we have used here.

Long-range interreplichore translocations

The degree of conservation – defined as the proportion of genes with an ortholog in a closely related partner genome – decreased from the O bin toward the T bin for fast-growing organisms, with slow-growing organisms showing little effect (Figure S9).

We then asked whether genes present in a given segment of a chromosome would be conserved within the same segment in its closely related partner genome and whether this property would differ across the four segments defined here. We define P_i – the *gene position conservation* for segment i (where i stands for the four chromosomal segments O, T, R, or L) − as the proportion of genes in that segment, which are conserved in the corresponding segment in its partner genome (Figure 3A). The median P_O and P_T, for the O and the T bins, across our dataset of 262 genome pairs, was ~80%, and the distributions of P_O and P_T were tight. In contrast, P_L and P_R showed a wide distribution, with the median around 50% (Figure 3A).

Next, we asked whether a gene present in a particular bin in one organism translocates to the diametrically opposite bin (or a neighboring bin − Figure S10) in the genome of a closely related organism. We called this *translocation rate* (T_{i-j}, where i and j indicate the segments between which the translocation rate is measured) (Figure 3B). Consistent with our observation of high P_O and P_T, we found that T_{O-T} and T_{T-O} are very low with a median of ~1% (Figure 3B). This could be a possible consequence of selection acting against translocations that would drastically disrupt gene dosage, and is consistent with the observation of few contacts between *oriC* and *ter*.

We next looked at translocations between the L and the R segments (T_{L-R} and T_{R-L}), which represent interreplichore translocations
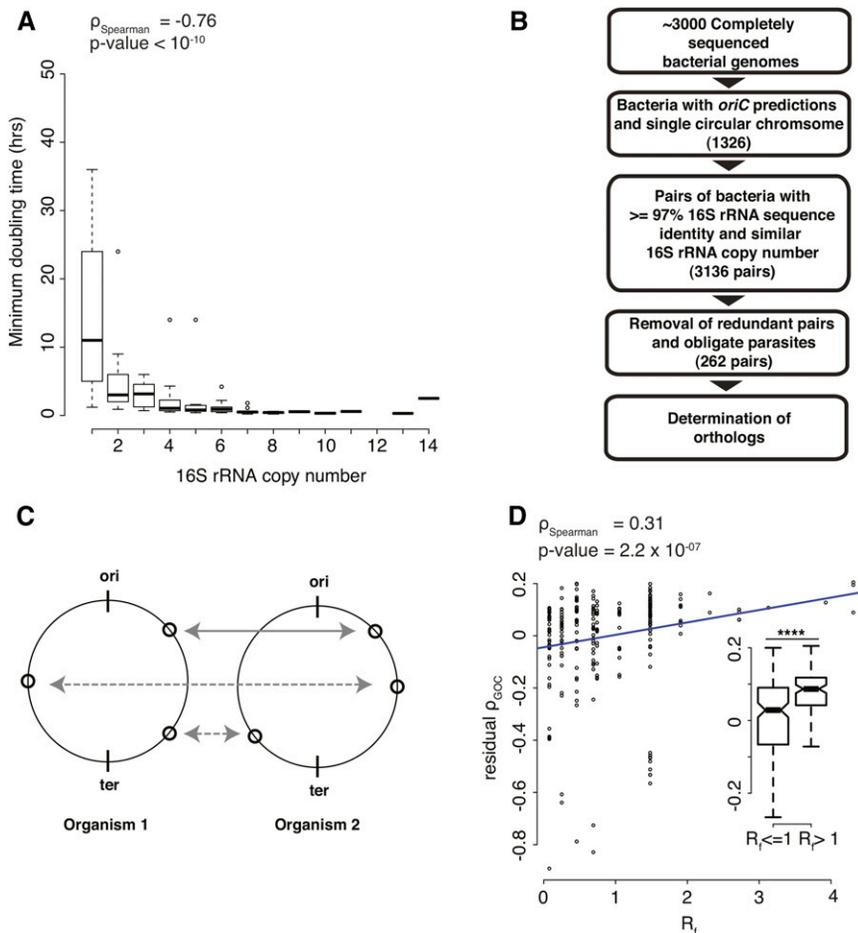
Figure 1 Comparative Genomics Framework. (A) Plot representing the relationship between minimum doubling time (hr) and 16S rRNA copy number ($\rho_{Spearman} = -0.76$, P-value $<10^{-10}$). Minimum doubling time values >50 hr are not represented in the plot. (B) Flowchart showing the selection of 262 pairs of bacteria used in this study. (C) Schematic showing the conservation of gene position (solid lines) and gene translocations (dashed lines) between pairs of bacteria. (D) Plot representing the correlation between residual ρ_{GOC} and R_f ($\rho_{Spearman} = 0.31$, P-value $= 2.2 \times 10^{-7}$). Inner panel shows a statistically significant difference between residual ρ_{GOC} for slow ($R_f \leq 1$) and fast ($R_f > 1$) growing bacteria (P-value $= 3.3 \times 10^{-6}$, Wilcoxon test). Asterisks indicate P-value $< 10^{-3}$.

(Figure 3B). For $\sim$42% of genome pairs, $T_{L-R/R-L} <0.1$; for the remaining, $T_{L-R/R-L} \sim$0.45 (median). Therefore, in the majority of genome pairs compared here, genes present in one replichore in one genome are often present in the opposite replichore in its partner genome.

Note that in identifying interreplichore translocations, by considering only the L and the R bins, we neglected 25% of each replichore immediately adjacent to *oriC* and *ter*. This was done to ensure that our observations were not skewed by previously reported high inversion rates immediately around *oriC* and *ter* (Eisen *et al.* 2000; Suyama and Bork 2001; Tillier and Collins 2000). Note, however, that the proportion of interreplichore translocations within the O and the T segments correlate well with that between the R and the L segments. Translocations appear to be more frequent in the *ter*-proximal than in the *oriC*-proximal half of the chromosome (Figure S11). The difference in translocation rates between the *oriC*-proximal and the *ter*-proximal halves of a chromosome not only suggests a complementary strategy for studying long-range translocations, but also further corroborates the interplay of chromosome topology and gene organization with the process of replication.

To ensure that the interreplichore translocations we describe are not merely a result of discrepancies in gene annotations, we predicted genes using Glimmer (Delcher *et al.* 1999) for all genomes included in this study. We compared predicted gene start positions with those reported in NCBI and found that the two datasets are in agreement with each other (Figure S12).

In summary, we show that interreplichore translocations between the L and the R bins are probably more common than could be perceived

from the literature (Eisen *et al.* 2000; Tillier and Collins 2000). Further, the variation in the distribution of L−R translocation rates is only weakly explained by phylogenetic distance (Figure S13 and Figure S14).

Interreplichore translocations as a possible readout of chromosome topology

Rates of recombination have been used as a proxy for measuring contact frequencies between distant regions of a chromosome. Together with the higher resolution analysis of contact frequencies by chromosome conformation capture methods and microscopy, these have so far underscored two broad chromosome topologies. First, in *E. coli* it has been shown that *oriC* is localized to the midcell and the left and the right arms of the chromosome stretch toward opposite poles with few interreplichore contacts (Nielsen *et al.* 2006; Wang *et al.* 2006). In line with this, we observed low interreplichore translocations ($T_{L-R/R-L} \sim$0.03) between *E. coli* and *Salmonella* (two closely related bacteria).

In contrast, in *C. crescentus* (Le *et al.* 2013; Umbarger *et al.* 2011) and *M. smegmatis* (Santi and McKinney 2015), *oriC* and *ter* are present at the two poles of the cell and the left and the right arms of the chromosome are in close contact with each other. Both these genomes showed high interreplichore translocations ($T_{L-R/R-L} \sim$0.50 and $\sim$0.70, respectively), consistent with their chromosome topology.

Next, we analyzed the chromosome contact map from a previously published Hi-C experiment for *C. crescentus* (Le *et al.* 2013). Pairs of loci translocated between the two replichores displayed significantly higher contact frequencies in Hi-C experiments when compared to pairs of interreplichore loci not showing translocations (Figure 4A
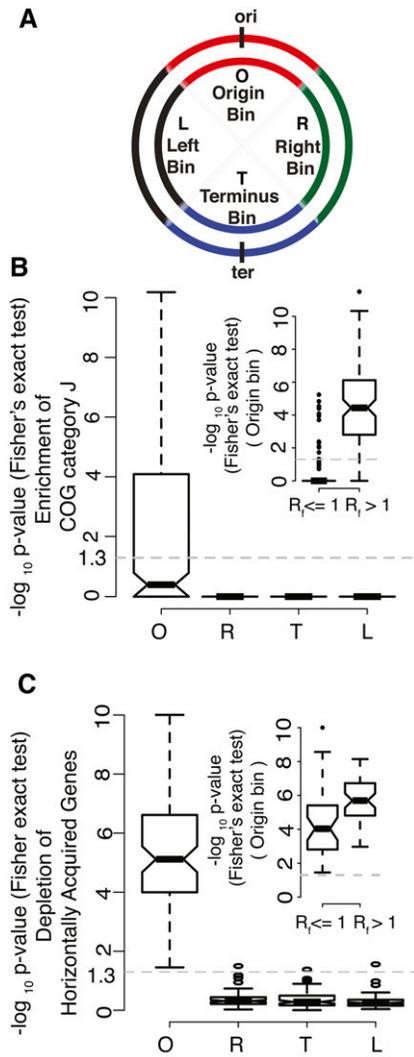
Figure 2 Chromosomal bin-based conservation of function. (A) Schematic representing equal sized bacterial chromosomal segments. Red: Origin bin (O); green: Right bin (R); blue: Terminus bin (T); black: Left bin (L). (B) Boxplot showing the enrichment of translation and ribosome biogenesis genes (COG category J) in O, R, T, and L chromosomal bins. (The dashed gray line at $y = 1.3$ corresponds to a P-value of 0.05. Values ≥ 1.3 are statistically significant.). Inner panel shows the enrichment of COG J in the O bin of fast-growing bacteria ($R_f > 1$). (C) Boxplot showing the depletion of horizontally acquired genes in different chromosomal bins. Inner panel showing the depletion of horizontally acquired genes in the O bin of both slow- ($R_f \leq 1$) and fast ($R_f > 1$) growing bacteria.

and Figure S15). These indicate that high rates of translocation are localized to spatially proximal regions of the chromosome.

Many interreplichore translocations in *C. crescentus* (Figure 4B) and in *M. smegmatis* (Figure S16A) are symmetric, *i.e.*, a gene at a certain distance d_R from *oriC* in one replichore translocates to a position d_L in the other replichore such that $D_{inter} = |d_R - d_L|$ is close to zero. This is again consistent with the topology of these chromosomes, and also with the idea that a pair of concurrent replication forks, which are typically equidistant from *oriC*, are hotspots for recombination. Such symmetric translocations are true, not only for *C. crescentus* and *M. smegmatis*, but are a general feature of genomes with high $T_{L-R/R-L}$ values (Figure 5A), but less so for genome pairs with low $T_{L-R/R-L}$ ($T_{L-R/R-L} < 0.1$)
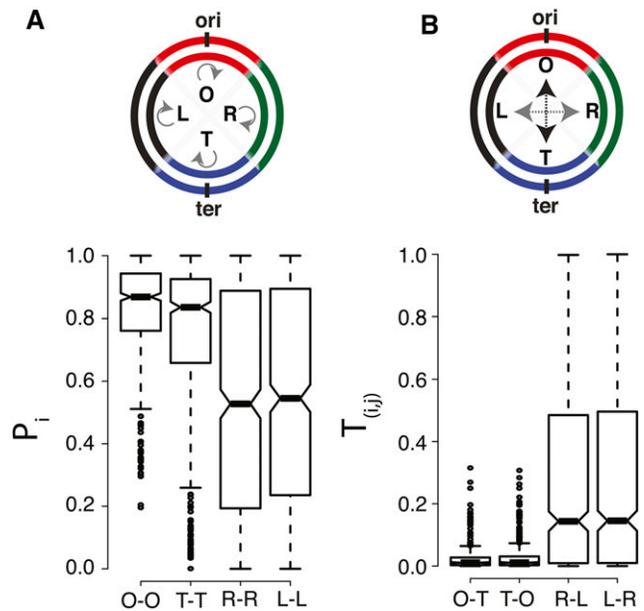
Figure 3 Dynamics of gene organization. (A) Schematic and boxplot showing the conservation of gene position (P_i) in O bin (O−O), T bin (T−T), R bin (R−R), and L bin (L−L). (B) Schematic representing O−T and T−O translocations (black arrows) and L−R and R−L translocations (gray arrows) and boxplot showing the proportion of these translocations ($T_{(i,j)}$).

(Figure S16B). These are in line with the high ρ_{GOC} values across most genome pairs tested here, indicating that interreplichore translocations rarely disrupt the relative distance of genes from *oriC*, and thereby maintain their dosage.

We also found that interreplichore translocations do not impact strand bias in gene densities. We noticed that the translocated partners of leading strand genes tend to be encoded on the leading strand of the opposite replichore. This is more prominent for organisms showing high $T_{L-R/R-L}$ as compared to low $T_{L-R/R-L}$ (Figure 5B). This difference in strand maintenance between genomes with high and low $T_{L-R/R-L}$ values might point to mechanistic differences between the two groups of organisms in the way these translocations occur.

Finally, we observed that interreplichore translocations rarely break operons. However, the contiguity of genes belonging to the same functional category is often broken by these translocations (Figure S17). In order to explicitly test the impact of these interreplichore translocations on gene expression states, we obtained publicly available microarray data for *Shewanella oneidensis MR1* (high $T_{L-R/R-L}$; $T_{L-R/R-L} \sim 0.56$) across many conditions from the M3D database (http://m3d.mssm.edu). Analysis of these microarray data showed that genes present on different replichores show a higher correlation in expression ($\rho_{Spearman} > 0.5$) when they are equidistant from *oriC*. In contrast, gene pairs showing little or negative correlations between their gene expression profiles tend to show greater differences in their distance from *oriC*. Although the effect is very weak, the trend is statistically significant in terms of the difference between medians (P-value $< 10^{-10}$, Wilcoxon test; Figure S18). This suggests that interreplichore translocations between genes equidistant from *oriC* may not disrupt their gene expression states.

Thus, interreplichore translocations − potentially encouraged by the spatial proximity of the two replichores − neither disrupt gene dosage gradients down the *oriC*−*ter* axis nor do they impact strand
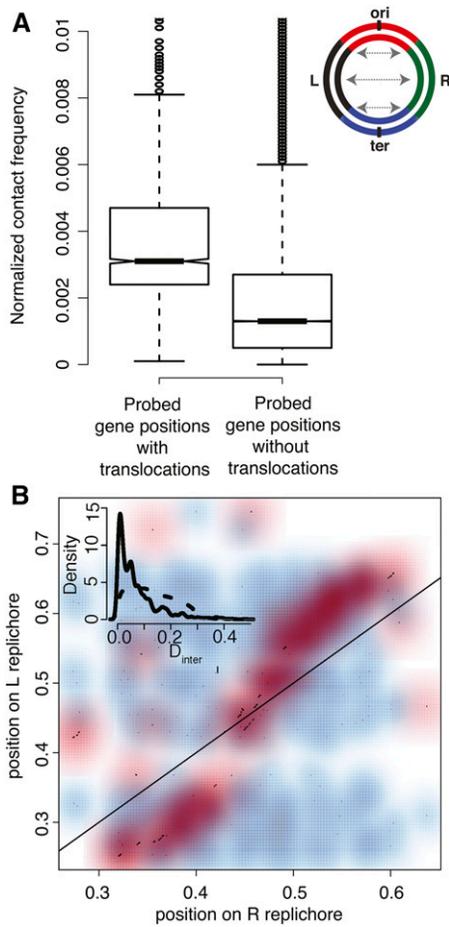
Figure 4 Interreplichore translocations and chromosomal contact points. (A) Boxplot representing the normalized interreplichore contact frequencies as derived from Le *et al.* (2013) for *C. crescentus* NA1000 (NC_011916); for all the probed gene positions with or without interreplichore translocations (*P*-value < 10^{-10}, Wilcoxon test). (B) Scatterplot representing 386 interreplichore (R → L) translocations between *C. crescentus* NA1000 (NC_011916) and *C. segnis* (NC_014100) in red and randomized dataset (as described in *Materials and Methods*) in blue. Inner panel shows the distribution of $D_{inter} = | d_{Ri/Lj} - d_{Li/Rj} |$ (solid line) for interreplichore translocations between *C. crescentus* NA1000 (NC_011916) and *C. segnis* (NC_014100). Dashed line indicates distribution of D_{inter} for randomized data (as described in *Materials and Methods*).

Figure 5 Selective value of chromosome topology. (A) Plot representing the distribution of $D_{inter} = | d_{R/L} - d_{L/R} |$ for all bacteria showing high interreplichore translocations ($T_{L-R/R-L} > 0.1$). Dashed line indicates the distribution of D_{inter} for randomized datasets (as described in *Materials and Methods*), similar to the inner panel of Figure 4B. Inner panel shows the percentage of gene pairs with $D_{inter} < 0.1$ for all bacteria showing high interreplichore translocations, alongside random expectation (determined from randomized datasets as described in *Materials and Methods*). (B) Plot representing the proportion of leading-lagging strand flips post interreplichore translocations as a function of the proportion of interreplichore translocations ($\rho_{Spearman}$ = 0.17, *P*-value < 10^{-10}) in cyan. A randomized dataset (as described in *Materials and Methods*) is shown in gray. Adjacent bloxplot represents the distribution of leading-lagging strand flips in bacteria with high and low interreplichore translocations alongside random expectation (R_m) for each dataset (D) (low $T_{L-R/R-L}$, *P*-value = 2.9 × 10^{-6}, Wilcoxon test and high $T_{L-R/R-L}$, *P*-value < 10^{-10}, Wilcoxon test). Asterisks indicate *P*-value < 10^{-3}.

biases in gene densities (Figure S22); they may have only a small effect on gene expression states.

DISCUSSION

Genome architecture broadly refers to gene organization and chromosome topology. While the wealth of genome sequence data provides valuable insights into gene organization, interrogation of chromosome topologies at high resolution requires additional experimental techniques involving chromosome conformation capture (Dekker *et al.* 2002). These techniques were deployed on a genome-wide scale initially for eukaryotes and are now being used to study prokaryotic genome structures (Cagliero *et al.* 2013; Le *et al.* 2013; Marbouty *et al.* 2014, 2015; Umbarger *et al.* 2011). These studies have been limited so far to studying the three-dimensional chromosome architectures of a few model organisms. This currently

precludes further analysis of associations between gene order and chromosome topologies.

In this study, we have used a comparative genomic framework to indicate that large-scale translocations over short phylogenetic distances could be a proxy for chromosome topologies. This is consistent with the observation, in eukaryotes, that various cancer-causing translocations in the human genome involve loci with high contact frequencies in normal tissue (Engreitz *et al.* 2012). We notice that genomes with high interreplichore translocations may also experience high rates of intrareplichore recombination. Whether this implicates symmetric interreplichore chromosomal contacts in the maintenance of gene order is hard to address with confidence (Figure S19A). It is however not clear as to what extent chromosome topology reflects phylogeny, as we observe a wide variation in the

rates of interreplichore translocations within phylogenetic classes (Figure S20). Additionally, we do not observe any association between growth rates and interreplichore translocation rates. And in light of our suggestion that high rates of interreplichore translocations may be predictive of a longitudinal chromosome topology, the growth rate of an organism may not determine its global chromosome topology (Figure S21).

At this resolution, our analysis might predict whether the two replichores of a bacterial chromosome are spatially colocalized, but does not say anything about finer aspects of its structuring, such as macro- and microdomains. For example, there are only a few long-range translocations in the genome of *E. coli*; many of these appear to be located around *ter*, which is closely organized by the binding of the protein MatP (Dupaigne *et al.* 2012) and where many targets of the nucleoid-associated protein H-NS are localized. Even the *C. crescentus* chromosome has local domain structures in the 20−400 kb size range (Le *et al.* 2013). However, these are not predicted by our translocation analysis.

Another caveat of our study is that the dynamics of chromosome topology may not be predictable. For example, recent studies in *B. subtilis* have shown that its chromosome oscillates between the longitudinal and the transverse topologies (X. Wang *et al.* 2014b). During replication *B. subtilis* is shown to have the *E. coli*-like transverse topology and the *C. crescentus*-like longitudinal topology for the rest of the cell cycle. However, *B. subtilis* shows interreplichore translocation frequency (∼0.02) similar to that of *E. coli*. This is consistent with the notion that concurrent replication forks are hotspots for recombination. Therefore, interreplichore translocations might occur during replication, and may be facilitated by close spatial contacts between the two replichores during replication. These observations hint toward the existence of a strong interplay between replication and chromosome topology. Hence, we suggest that the association between chromosome topologies and growth phenotypes, and its mechanistic and evolutionary underpinnings, should be explored in greater detail.

Recent studies have shown that genome architecture and gene expression are linked in the following ways: (a) through differential positioning of genes with respect to *oriC*, popularly referred to as the gene dosage effect (Block *et al.* 2012; Bryant *et al.* 2014; Narula *et al.* 2015; Schmid and Roth 1987); (b) through spatial clustering of nonconsecutive genes, which might enforce transcriptional coherence (Cabrera and Jin 2003; Junier *et al.* 2012; Qian *et al.* 2012; Wang *et al.* 2011); (c) through preferential binding of bacterial nucleoid-associated proteins like H-NS, which show position and gene feature-specific binding to the chromosome (Tupper *et al.* 1994; Weng and Xiao 2014); and (d) through the effect of proposed gradients of DNA superhelicity on the chromosome on gene expression (Lal *et al.* 2016; Sobetzko *et al.* 2012). However, despite our attempts to interrogate this using a compendium of microarray data for one organism, the impact of interreplichore translocations on gene expression states remains open. Along this direction, it remains to be seen whether genomes showing high interreplichore translocations would also display symmetry in nucleoid organization between the two replichores; this could also serve to minimize the impact of such translocations on gene expression. Detailed analyses of the impact of replication on gene expression and chromosome topologies are also essential, in the light of evidence that *oriC* might itself be a selfish element (Hawkins *et al.* 2013), and the ability of bacteria to undergo dispersed replication events independent of *oriC* (Bernander *et al.* 1991; Dimude *et al.* 2015; Gowrishankar 2015; Koppes 1992; McGeoch and Bell 2008).

LITERATURE CITED

Bernander, R., S. Dasgupta, and K. Nordström, 1991 The E. coli cell cycle and the plasmid R1 replication cycle in the absence of the DnaA protein. Cell 64: 1145–1153.

Block, D. H. S., R. Hussein, L. W. Liang, and H. N. Lim, 2012 Regulatory consequences of gene translocation in bacteria. Nucleic Acids Res. 40: 8979–8992.

Browning, D. F., D. C. Grainger, and S. J. Busby, 2010 Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. Curr. Opin. Microbiol. 13: 773–780.

Bryant, J. A., L. E. Sellars, S. J. W. Busby, and D. J. Lee, 2014 Chromosome position effects on gene expression in Escherichia coli K-12. Nucleic Acids Res. 42: 11383–11392.

Cabrera, J. E., and D. J. Jin, 2003 The distribution of RNA polymerase in Escherichia coli is dynamic and sensitive to environmental cues. Mol. Microbiol. 50: 1493–1505.

Cagliero, C., R. S. Grand, M. B. Jones, D. J. Jin, and J. M. O'Sullivan, 2013 Genome conformation capture reveals that the Escherichia coli chromosome is organized by replication and transcription. Nucleic Acids Res. 41: 6058–6071.

Condon, C., D. Liveris, C. Squires, I. Schwartz, and C. L. Squires, 1995 rRNA operon multiplicity in Escherichia coli and the physiological implications of rrn inactivation. J. Bacteriol. 177: 4152–4156.

Cooper, S., and C. E. Helmstetter, 1968 Chromosome replication and the division cycle of Escherichia coli. J. Mol. Biol. 31: 519–540.

Couturier, E., and E. P. C. Rocha, 2006 Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. Mol. Microbiol. 59: 1506–1518.

Dame, R. T., M. C. Noom, and G. J. L. Wuite, 2006 Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. Nature 444: 387–390.

Dekker, J., K. Rippe, M. Dekker, and N. Kleckner, 2002 Capturing chromosome conformation. Science 295: 1306–1311.

Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg, 1999 Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 27: 4636–4641.

Dillon, S. C., and C. J. Dorman, 2010 Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. Nat. Rev. Microbiol. 8: 185–195.

Dimude, J. U., A. Stockum, S. L. Midgley-Smith, A. L. Upton, H. A. Foster et al., 2015 The consequences of replicating in the wrong orientation: bacterial chromosome duplication without an active replication origin. MBio 6: e01294–e15.

DiNardo, S., K. A. Voelkel, R. Sternglanz, A. E. Reynolds, and A. Wright, 1982 Escherichia coli DNA topoisomerase I mutants have compensatory mutations in DNA gyrase genes. Cell 31: 43–51.

Dorman, C. J., 2013 Genome architecture and global gene regulation in bacteria: making progress towards a unified model? Nat. Rev. Microbiol. 11: 349–355.

Dupaigne, P., N. K. Tonthat, O. Espéli, T. Whitfill, F. Boccard *et al.*, 2012 Molecular basis for a protein-mediated DNA-bridging mechanism that functions in condensation of the E. coli chromosome. Mol. Cell 48: 560–571.

Edgar, R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792–1797.

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

Engreitz, J. M., V. Agarwala, and L. A. Mirny, 2012 Three-dimensional genome architecture influences partner selection for chromosomal translocations in human disease. PLoS One 7: e44196.

Esnault, E., M. Valens, O. Espéli, and F. Boccard, 2007 Chromosome Structuring Limits Genome Plasticity in Escherichia coli. PLoS Genet. 3: e226.

Esposito, D., and G. F. Gerard, 2003 The Escherichia coli Fis protein stimulates bacteriophage integrative recombination in vitro. J. Bacteriol. 185: 3076–3080.

Finn, R. D., J. Clements, and S. R. Eddy, 2011 HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39: W29–W37.

Freilich, S., A. Kreimer, E. Borenstein, N. Yosef, R. Sharan *et al.*, 2009 Metabolic-network-driven analysis of bacterial ecological strategies. Genome Biol. 10: R61.

Gao, F., and C.-T. Zhang, 2007 DoriC: a database of oriC regions in bacterial genomes. Bioinformatics 23: 1866–1867.

Gowrishankar, J., 2015 End of the beginning: elongation and termination features of alternative modes of chromosomal replication initiation in bacteria. PLoS Genet. 11: e1004909.

Gruber, S., J.-W. Veening, J. Bach, M. Blettinger, M. Bramkamp *et al.*, 2014 Interlinked sister chromosomes arise in the absence of Condensin during fast replication in B. subtilis. Curr. Biol. 24: 293–298.

Hawkins, M., S. Malla, M. J. Blythe, C. A. Nieduszynski, and T. Allers, 2013 Accelerated growth in the absence of DNA replication origins. Nature 503: 544–547.

Junier, I., J. Hérisson, and F. Képès, 2012 Genomic organization of evolutionarily correlated genes in bacteria: limits and strategies. J. Mol. Biol. 419: 369–386.

Junier, I., F. Boccard, and O. Espeli, 2014 Polymer modeling of the E. coli genome reveals the involvement of locus positioning and macrodomain structuring for the control of chromosome conformation and segregation. Nucleic Acids Res. 42: 1461–1473.

Klappenbach, J. A., J. M. Dunbar, and T. M. Schmidt, 2000 rRNA operon copy number reflects ecological strategies of bacteria. Appl. Environ. Microbiol. 66: 1328–1333.

Koonin, E. V., and Y. I. Wolf, 2008 Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res. 36: 6688–6719.

Koppes, L. J., 1992 Nonrandom F-plasmid replication in Escherichia coli K-12. J. Bacteriol. 174: 2121–2123.

Lal, A., A. Dhar, A. Trostel, F. Kouzine, A. S. N. Seshasayee *et al.*, 2016 Genome scale patterns of supercoiling in a bacterial chromosome. Nat. Commun. 7: 11055.

Lagomarsino, M. C., O. Espéli, and I. Junier, 2015 From structure to function of bacterial chromosomes: Evolutionary perspectives and ideas for new experiments. FEBS Lett. 589: 2996–3004.

Le, T. B., and M. T. Laub, 2014 New approaches to understanding the spatial organization of bacterial genomes. Curr. Opin. Microbiol. 22: 15–21.

Le, T. B. K., M. V. Imakaev, L. A. Mirny, and M. T. Laub, 2013 High-resolution mapping of the spatial organization of a bacterial chromosome. Science 342: 731–734.

Letunic, I., and P. Bork, 2011 Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res. 39: W475–W478.

Liu, Y., H. Chen, L. J. Kenney, and J. Yan, 2010 A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. Genes Dev. 24: 339–344.

Marbouty, M., A. Cournac, J.-F. Flot, H. Marie-Nelly, J. Mozziconacci *et al.*, 2014 Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. eLife 3: e03318.

Marbouty, M., A. Le Gall, D. I. Cattoni, A. Cournac, A. Koh *et al.*, 2015 Condensin- and replication-mediated bacterial chromosome folding and origin condensation revealed by Hi-C and super-resolution imaging. Mol. Cell 59: 588–602.

Maurer, S., J. Fritz, and G. Muskhelishvili, 2009 A systematic in vitro study of nucleoprotein complexes formed by bacterial nucleoid-associated proteins revealing novel types of DNA organization. J. Mol. Biol. 387: 1261–1276.

McGeoch, A. T., and S. D. Bell, 2008 Extra-chromosomal elements and the evolution of cellular DNA replication machineries. Nat. Rev. Mol. Cell Biol. 9: 569–574.

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

Mira, A., L. Klasson, and S. G. E. Andersson, 2002 Microbial genome evolution: sources of variability. Curr. Opin. Microbiol. 5: 506–512.

Mirkin, E. V., and S. M. Mirkin, 2005 Mechanisms of transcription-replication collisions in bacteria. Mol. Cell. Biol. 25: 888–895.

Narula, J., A. Kuchina, D. D. Lee, M. Fujita, G. M. Süel *et al.*, 2015 Chromosomal arrangement of phosphorelay genes couples sporulation and DNA replication. Cell 162: 328–337.

Needleman, S. B., and C. D. Wunsch, 1970 A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48: 443–453.

Nielsen, H. J., J. R. Ottesen, B. Youngren, S. J. Austin, and F. G. Hansen, 2006 The Escherichia coli chromosome is organized with the left and right chromosome arms in separate cell halves. Mol. Microbiol. 62: 331–338.

Pruss, G. J., S. H. Manes, and K. Drlica, 1982 Escherichia coli DNA topoisomerase I mutants: increased supercoiling is corrected by mutations near gyrase genes. Cell 31: 35–42.

Qian, Z., E. K. Dimitriadis, R. Edgar, P. Eswaramoorthy, and S. Adhya, 2012 Galactose repressor mediated intersegmental chromosomal connections in Escherichia coli. Proc. Natl. Acad. Sci. USA 109: 11336–11341.

Qian, Z., M. Macvanin, E. K. Dimitriadis, X. He, V. Zhurkin *et al.*, 2015 A new noncoding RNA arranges bacterial chromosome organization. MBio 6: e00998-15.

Reyes-Lamothe, R., C. Possoz, O. Danilova, and D. J. Sherratt, 2008 Independent positioning and action of Escherichia coli replisomes in live cells. Cell 133: 90–102.

Rice, P., I. Longden, and A. Bleasby, 2000 EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. TIG 16: 276–277.

Rocha, E. P. C., 2004 The replication-related organization of bacterial genomes. Microbiology 150: 1609–1627.

Rocha, E. P. C., 2008 The organization of the bacterial genome. Annu. Rev. Genet. 42: 211–233.

Rocha, E. P. C., and A. Danchin, 2003 Gene essentiality determines chromosome organisation in bacteria. Nucleic Acids Res. 31: 6570–6577.

Santi, I., and J. D. McKinney, 2015 Chromosome organization and replisome dynamics in Mycobacterium smegmatis. MBio 6: e01999–e14.

Schmid, M. B., and J. R. Roth, 1987 Gene location affects expression level in Salmonella typhimurium. J. Bacteriol. 169: 2872–2875.

Shanado, Y., K. Hanada, and H. Ikeda, 2001 Suppression of gamma ray-induced illegitimate recombination in Escherichia coli by the DNA-binding protein H-NS. Mol. Genet. Genomics 265: 242–248.

Sobetzko, P., A. Travers, and G. Muskhelishvili, 2012 Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. Proc. Natl. Acad. Sci. USA 109: E42–E50.

Stella, S., D. Cascio, and R. C. Johnson, 2010 The shape of the DNA minor groove directs binding by the DNA-bending protein Fis. Genes Dev. 24: 814–826.

Suyama, M., and P. Bork, 2001   Evolution of prokaryotic gene order: genome rearrangements in closely related species. Trends Genet. TIG 17: 10–13.

Tamames, J. (2001)   Evolution of gene order conservation in prokaryotes. Genome Biol. 2: RESEARCH0020.

Tamas, I., 2002   50 Million Years of Genomic Stasis in Endosymbiotic Bacteria. Science 296: 2376–2379.

Tamura, K., and M. Nei, 1993   Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10: 512–526.

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar, 2013   MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30: 2725–2729.

Tillier, E. R., and R. A. Collins, 2000   Genome rearrangement by replication-directed translocation. Nat. Genet. 26: 195–197.

Tupper, A. E., T. A. Owen-Hughes, D. W. Ussery, D. S. Santos, D. J. Ferguson *et al.*, 1994   The chromatin-associated protein H-NS alters DNA topology in vitro. EMBO J. 13: 258–268.

Umbarger, M. A., E. Toro, M. A. Wright, G. J. Porreca, D. Baù *et al.*, 2011   The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. Mol. Cell 44: 252–264.

Valens, M., S. Penaud, M. Rossignol, F. Cornet, and F. Boccard, 2004   Macrodomain organization of the *Escherichia coli* chromosome. EMBO J. 23: 4330–4341.

Vernikos, G. S., and J. Parkhill, 2006   Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the Salmonella pathogenicity islands. Bioinformatics 22: 2196–2203.

Vieira-Silva, S., and E. P. C. Rocha, 2010   The systemic imprint of growth and its uses in ecological (meta)genomics. PLoS Genet. 6: e1000808.

Wang, J., X. Meng, H. Chen, C. Yuan, X. Li *et al.*, 2016   Exploring the mechanisms of genome-wide long-range interactions: interpreting chromosome organization. Brief. Funct. Genomics. DOI: 10.1093/bfgp/elv062.

Wang, S., J. R. Moffitt, G. T. Dempsey, X. S. Xie, and X. Zhuang, 2014   Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. Proc. Natl. Acad. Sci. USA 111: 8452–8457.

Wang, W., G.-W. Li, C. Chen, X. S. Xie, and X. Zhuang, 2011   Chromosome Organization by a Nucleoid-Associated Protein in Live Bacteria. Science 333: 1445–1449.

Wang, X., X. Liu, C. Possoz, and D. J. Sherratt, 2006   The two Escherichia coli chromosome arms locate to separate cell halves. Genes Dev. 20: 1727–1731.

Wang, X., O. W. Tang, E. P. Riley, and D. Z. Rudner, 2014a   The SMC condensin complex is required for origin segregation in Bacillus subtilis. Curr. Biol. 24: 287–292.

Wang, X., P. Montero Llopis, and D. Z. Rudner, 2014b   Bacillus subtilis chromosome organization oscillates between two distinct patterns. Proc. Natl. Acad. Sci. USA 111: 12877–12882.

Weng, X., and J. Xiao, 2014   Spatial organization of transcription in bacterial cells. Trends Genet. TIG 30: 287–297.

Zarei, M., B. Sclavi, and M. Cosentino Lagomarsino, 2013   Gene silencing and large-scale domain structure of the E. coli genome. Mol. Biosyst. 9: 758.

Communicating editor: S. I. Wright