Non-Random Inversion Landscapes in Prokaryotic Genomes Are Shaped by Heterogeneous Selection Pressures

Jelena Repar1,2 and Tobias Warnecke*1,2
1Molecular Systems Group, MRC London Institute of Medical Sciences (LMS), London, United Kingdom
2Institute of Clinical Sciences, Molecular Systems Group, Institute of Clinical Sciences (ICS), Faculty of Medicine, Imperial College London, London, United Kingdom

*Corresponding author: E-mail: tobias.warnecke@imperial.ac.uk.
Associate editor: Katja Nowick

Abstract

Inversions are a major contributor to structural genome evolution in prokaryotes. Here, using a novel alignment-based method, we systematically compare 1,651 bacterial and 98 archaeal genomes to show that inversion landscapes are frequently biased toward (symmetric) inversions around the origin–terminus axis. However, symmetric inversion bias is not a universal feature of prokaryotic genome evolution but varies considerably across clades. At the extremes, inversion landscapes in Bacillus–Clostridium and Actinobacteria are dominated by symmetric inversions, while there is little or no systematic bias favoring symmetric rearrangements in archaea with a single origin of replication. Within clades, we find strong but clade-specific relationships between symmetric inversion bias and different features of adaptive genome architecture, including the distance of essential genes to the origin of replication and the preferential localization of genes on the leading strand. We suggest that heterogeneous selection pressures have converged to produce similar patterns of structural genome evolution across prokaryotes.

Key words: inversion, replication, mutation, selection.

Introduction

In both eukaryotes and prokaryotes, genome architecture and its evolution are frequently non-random (Hurst et al. 2004; Rocha 2008). A fundamental question in this regard is whether non-random genome organization is brought about by biased mutational processes, for example relating to the dynamics of recombination, or by selection. In prokaryotes, many aspects of non-random genome organization have been plausibly attributed to the latter. This includes the clustering of functionally related genes into operons and the enrichment of essential genes on the leading strand of replication, where they avoid head-on collisions between active DNA and RNA polymerases (Rocha and Danchin 2003; Rocha 2004; Flynn et al. 2010). In addition, highly expressed genes (rRNA, tRNA, ribosomal protein genes) are commonly found near the origin of replication (ori) in both bacteria and archaea (Couturier and Rocha 2006; Andersson et al. 2010; Pelve et al. 2012), consistent with selection for elevated dosage: sequences near ori replicate earlier and are therefore transiently present in higher copy number, a phenomenon exacerbated in fast-growing organisms where multiple rounds of replication can be initiated concurrently.

For other aspects of prokaryotic genome evolution, it has been more difficult to pin down whether non-random genome structure is brought about by selection, biased mutational processes or a combination of the two. This notably includes the incidence pattern of large-scale inversions, which constitute a major source of structural diversity in prokaryotic genomes (Hughes 2000; Belda et al. 2005). Curiously, inversions in prokaryotes appear to be predominantly symmetric (fig. 1A); that is, their end points are approximately equidistant from the origin of replication, generating conspicuous X patterns (fig. 1B) in whole-genome alignments (Eisen et al. 2000; Read et al. 2000; Tillier and Collins 2000; Suyama and Bork 2001). Inversions symmetric to the origin-terminus (ori–ter) axis were initially observed in a few closely related bacterial genomes, including pairs of Chlamydia, Mycobacterium, and Helicobacter spp. (Eisen et al. 2000; Read et al. 2000; Tillier and Collins 2000; Suyama and Bork 2001), and have subsequently been highlighted in multiple other genome comparisons, notably involving γ-proteobacteria (Yersinia [Darling et al. 2008], Blochmannia [Gil et al. 2003], Buchnera [Moran and Mira 2001; Silva et al. 2001], Pseudomonas [Worning et al. 2006]) and Bacilli [Lactobacillus (Canchaya et al. 2006), Bacillus (Worning et al. 2006), Streptococcus (Nakagawa et al. 2003)], but also the single-origin archaeae Pyrococcus (Zivanovic et al. 2002). These observations have given rise to the notion that biased inversion landscapes are a prevalent feature of prokaryotic genome evolution.

Both non-random mutational processes and selection have been suggested as potential drivers of biased inversion landscapes. Regarding the former, it has been proposed that symmetric inversions might be favored by the layout of bacterial replication. As sister replisomes progress at approximately the same speed after setting out together from the origin of replication, homologous recombination across
replicores would lead to symmetric or quasi-symmetric inversions if single-stranded DNA, present in the wake of one of the sister replicosomes, is used as a template for illegitimate recombination following the generation of a double-strand break in the vicinity of the other replicosome (Tillier and Collins 2000; Makino and Suzuki 2001).

Alternatively, the genesis of inversions is approximately random but symmetric inversions are more likely to survive purifying selection because they are, on average, less disruptive to adaptive genome architecture (Eisen et al. 2000; Tillier and Collins 2000; Mackiewicz et al. 2001). Notably, although a potentially large number of loci are translocated to the opposite replicore, they will retain their original leading/lagging strand orientation. This might be important not only to avoid replication–transcription conflicts but also for binding motifs that function in a polarized fashion, such as FtsK-orienting-polar-sequence, which facilitate FtsK translocation towards the terminus. In contrast, inversions within the same replicore inevitably result in leading/lagging strand switches. Symmetric inversions also do not alter the distance of a particular genomic origin

**Fig. 1.** Detecting symmetric inversion bias. (A) Inversions in circular prokaryotic genomes can occur with varying degrees of symmetry in relation to the ori-ter axis (grey dotted line). The furthest away the inversion axis (orange dotted line) can be from the ori-ter axis is a quarter of the genome (0.25). (B, D, F) Symmetric inversions cause X-shaped patterns in pairwise MUMmer alignments, which can be hard to discern (D, F) but are revealed as symmetry hot and cold spots when displaying $X_{i,j}$ (C, E, G). (H) Distribution of $D$ scores for Proteobacteria with experimentally determined ori positions. The black line is the log spline density fit to the whole dataset, the grey lines are log spline density fits to random 40% jackknifed samples to explore outsize influence of individual genome pairs. (I, J) $X_{i,j}$ heat maps representing a case of simulated genome divergence by random (I) and symmetric and quasi-symmetric (J) inversions.
element to ori or ter, thus avoiding potentially deleterious
changes in gene dosage and the displacement of motifs whose
function is contingent on their proximity to either ori or ter (e.g.,
DnaA boxes, parS motifs) (Hendrickson and Lawrence 2006;
Touzain et al. 2011). Finally, symmetric inversions do not alter
relative replicon length, which is important in light of exper-
imental evidence from Escherichia coli that replicon size im-
balance of more than 10% is deleterious (Essault et al. 2007).

Here, to elucidate the relative importance of selection and
mutation bias in the evolution of inversion landscapes across
prokaryotes, we first take a step back and ask: are symmetric
inversions really a universal feature of prokaryotic genome
 evolution? The tally of individual cases in the literature seems
striking, yet symmetric inversions have—with few notable
exceptions (Eisen et al. 2000; Darling et al. 2008; Khedkar
and Seshasayee 2016)—been diagnosed rather casually by
picking out apparent X patterns from pairwise whole genome
alignments by eye. This is problematic, not least because there
might be extensive reporting bias, that is, obvious symmetric
inversions are highlighted in the literature, whereas subtle
biases are missed and random inversion patterns rarely men-
tioned (Parkhill et al. 2003).

To address this issue, we developed an alignment-based
approach to systematically assess inversion symmetry be-
 tween pairs of prokaryotic genomes. Applying this approach
to 1,651 bacterial and 98 archaeal genomes, we demonstrate
that there is substantial heterogeneity in the prevalence of
symmetric inversions across prokaryotic clades. While inver-
sion landscapes in some phyla, for example Bacillus–
Clostridium, are dominated by symmetric inversions, the
propensity to invert around the ori-ter axis is much less pro-
nounced in other clades, including Bacteroidetes and
Euryarchaeota. We go on to show that putatively adaptive
features of genome architecture linked to the ori-ter axis, such
as the fraction of genes encoded on the leading strand and the
average distance of rRNA genes to the origin of replica-
cation, are predictive of symmetric inversion bias but in a clade-
specific fashion. For example, enrichment of highly expressed
informational genes near ori is predictive of symmetric inver-
sion bias in Proteobacteria but not in Actinobacteria where
we instead observe a strong correlation with relative nucleo-
tide motif abundance on the leading versus lagging strand.
We suggest that heterogeneous selection pressures have con-
volved to produce similar patterns of genome evolution.
Dissecting inversion patterns as a function of known replica-
dation dynamics in different organisms, we also suggest that
selection might act on top of and reinforce pre-existing mu-
tational biases.

Results and Discussion
To assess inversion symmetry across a large number of ge-
 nomes, we developed a simple geometric approach for pair-
wise genome comparisons that is fast, easily scalable, and
 applicable across different phylogenetic distances. Starting
from MUMmer (Kurtz et al. 2004) alignments of two ge-
 nomes, like the one shown in figure 1B, we make use of the
 fact that, if there is a single dominant axis around which
inversions have occurred between two genomes, homologous
sequence blocks will be located on one of the two diagonals
that pass through that axis, generating the familiar X pattern.
We can test how well the distribution of homologous se-
quencies in the two genomes conforms to this single-
symmetry-axis model by considering, in a strand-specific fash-
ion, the residual deviation of homologous sequence blocks
from the expected X pattern. In brief, rather than define a
likely symmetry axis (e.g., ori-ter), we systematically survey
the MUMmer alignment landscape at a resolution of 10 kb and,
for each point ij, register the strength of X-type symmetry
(Xij) by considering the position of homologous blocks with
respect to an imaginary X centered at ij. More formally, X-
type symmetry is defined as

$$X_{ij} = \frac{\sum f_m - \sum f + \sum r - \sum r_m}{\sum r_m},$$

where f and r are residual deviations from the forward arm of
the imaginary X of homologous blocks in forward and reverse
orientation, respectively, and fm and rm are residual deviations
of the same homologous blocks mirrored around the vertical
axis going through ij. The logic here (further illustrated in
supplementary fig. 1, Supplementary Material online) is the
following: An inversion around the ori-ter axis moves a given
sequence onto the other arm of the X, resulting in a homol-
ogy block with reverse-orientation. Mirroring should reverse
this step, in effect restoring collinearity in the genome align-
ment, which will minimize rm and therefore lead to a large Xij.

This approach effectively converts the MUMmer alignment
into a matrix of symmetry hot and cold spots (fig. 1C, E, G).
Visualizing such matrices frequently reveals strong X-type
symmetry (henceforth simply referred to as symmetry),
including in genomes that are substantially diverged so that
homologous sequences are sparse and in genomes that
exhibit high rates of rearrangement and where sym-
metry would have likely escaped visual detection (fig.
1D–G). When simulating inversion events between two
dummy genomes, very similar patterns are evident when
rearrangements are restricted to symmetric or quasi-
symmetric inversions (fig. 1) whereas the tell-tale sec-
ondary diagonal is not present when no symmetry con-
straints are imposed and inversions can occur at random
throughout the genome (fig. 1).

Up to this point we have identified points of inversion
symmetry in an unbiased fashion without regard to genomic
landmarks such as ori or ter. To test empirically whether inver-
sion symmetry around the ori-ter axis is prevalent across
prokaryotic genomes, we calculated the distance D between
the point of maximal symmetry [max(Xij)] in a given
MUMmer alignment and ori or ter (whichever is closer).
First, we considered a small set (N = 19) of proteobacterial
genomes with experimentally determined origins (supple-
mental table 1, Supplementary Material online). Making
all pairwise comparisons that satisfy a lenient set of minimum
homology criteria (see “Materials and Methods” section) we
find a clear departure from the null model of random inver-
sions. D is strongly shifted towards smaller values, indicative
of inversions largely happening around the ori-ter axis (fig. 1H).
Variable Prevalence of Symmetric Inversions Across Prokaryotic Phyla

The location of ori can be robustly predicted in many prokaryotic genomes based on a set of hallmark features, including the location of DnaA boxes and strand biases in nucleotide composition that change sign around ori, reflecting divergent mutational biases associated with leading/lagging strand replication. We therefore extended our analysis to include genomes with computationally predicted origins (see “Materials and Methods” section). In total, we inferred symmetry scores and calculated D values for 528396 (2645) pairwise comparisons between 1651 bacterial (98 archaeal) genomes with predicted ori coordinates (supplementary table 2, Supplementary Material online). We then considered the distribution of D values aggregated at the phylum level. We find strong symmetric inversion bias in multiple bacterial phyla, including those that have shaped our understanding of replication-associated genome architecture—Proteobacteria and Bacillus–Clostridium (fig. 2)—in line with previous observations from individual genome pairs and a more systematic comparison of recently diverged genomes (Khedkar and Seshasayee 2016). However, ori-ter biased inversion landscapes are not universal and the degree of bias varies considerably between clades. For example, the rearrangement landscapes of Bacteroidetes–Chlorobi–Fibrobacteres (BCF) and Tenericutes show much weaker bias toward symmetric inversions than other clades, and we detect no significant inversion bias in archaea with a single origin of replication (fig. 2). Heterogeneity in symmetric inversion bias is also evident at lower taxonomic levels, as illustrated for Proteobacteria in figure 2.

D values support symmetric inversion bias across a remarkably broad range of divergence levels. However, weaker ori-ter symmetry signals are generally observed for very closely and distantly related genomes, as exemplified in figure 3A by the phylum Bacillus–Clostridium. The variability in symmetry scores over evolutionary time might be the consequence of limited rearrangement signal at short evolutionary distance and lower signal-to-noise ratio when divergence levels are high. To rule out that heterogeneity in symmetric inversion bias between clades is the consequence of differential sampling of divergence levels, we subsampled from within each individual phylum to match a common distribution of 16S rRNA distances (see “Materials and Methods” section). Taking phylogenetic distances into account suggests that we might overestimate the extent of symmetric inversion bias in Chlamydiae. However, globally, the results demonstrate that differences in D score distributions between clades are robust (fig. 3B, supplementary fig. 2, Supplementary Material online).

Selection Is a Prominent Driver of Symmetric Inversion Landscapes

To test whether biased inversion landscapes are likely the result of natural selection to preserve adaptive genome architecture, we considered four features of genome organization linked to the ori-ter axis and therefore potentially predictive of symmetric inversion bias: the average distance of rRNA genes to ori; the fraction of genes located on the leading strand of replication; the enrichment of highly expressed translational genes (COG J) near ori; and the biased distribution of nucleotide motifs—which may be involved in DNA replication, repair or segregation—on either the leading or lagging strand. Regarding the latter, we followed prior work (Hendrickson and Lawrence 2006) and confined analysis to octameric nucleotide motifs. We then correlated these features with the median D values of the focal genome, computed across all pairwise comparisons in which it participates. Focusing on phyla represented by at least 30 taxa, we find strong correlations with all four features in the expected direction for the phyla Bacillus–Clostridium and Proteobacteria. Genomes with higher symmetric inversion bias (lower median D) have stronger enrichments of COG J genes near the origin of replication (more negative Z scores), have rDNA genes encoded at a smaller average distance to ori; have a greater fraction of genes encoded on the leading strand of replication and exhibit stronger imbalances of nucleotide motifs (fig. 4). Surprisingly, Actinobacteria—a phylum with a stronger symmetric inversion bias than Proteobacteria—show no intra-phyllum correlation between D and either COG J gene enrichment or rDNA distance or biased gene orientation (fig. 4). However, there is a strong relationship in this clade between the degree of octamer strand bias and median D. A similar pattern is evident for the BCF phylum, where, in addition, we observe a relationship with rDNA distance. In contrast, biased distribution of genes on the leading strand is the only feature predictive of symmetric inversion bias in Tenericutes. We observe qualitatively similar patterns when considering the 16S-filtered data (supplementary fig. 3, Supplementary Material online). We also assessed whether systematic between-clade differences in ori prediction accuracy might affect our conclusions by partitioning genomes into two groups with low and high ΔGC skew. ΔGC skew is independently predictive of origin location but lower ΔGC skew values are associated with somewhat reduced prediction accuracy (supplementary fig. 4A, Supplementary Material online). As low and high ΔGC skew groups yield similar conclusions, systematic variation in ori prediction accuracy does not account for our observations (supplementary fig. 4B, Supplementary Material online). Overall, these results suggest that symmetric inversion bias is promoted by different principal selection pressures in different prokaryotic clades.

Sister Replisome Proximity Predicts Symmetric Inversion Bias at Short Evolutionary Distances

Despite inter-clade heterogeneity, the analyses above argue that selection is pervasively implicated in shaping inversion landscapes across prokaryotic evolution. But does selection operate on an initially random population of inversions or does it act to reinforce pre-existing mutational biases? In other words, is the mutational raw material already biased in favor of symmetric inversions? The sister replisome hypothesis of symmetric inversions assumes that physical proximity between sister replication forks facilitates illegitimate recombination and predicts that symmetric inversions should be
more likely in organism with higher levels of sister fork co-localization.

To test this hypothesis, we considered variability in physical proximity of sister replication forks during the replication cycle, which has been examined by fluorescence microscopy in a range of different bacteria. In some species, sister replisomes associate tightly throughout the replication cycle, notably in *Bacillus subtilis* (Lemon and Grossman 2000).

---

**FIG. 2.** Distribution of $D$ scores in different prokaryotic clades. The number of focal species and the number of eligible pairwise comparisons within each phylum or class is given in parentheses. Note that, while the focal species belongs to the indicated clade, its partner is chosen based on the presence of a minimum number of homologous blocks (see "Materials and Methods" section) and might in some instances belong to a different clade.

**FIG. 3.** $D$ scores as a function of phylogenetic distance. (A) $D$ as a function of phylogenetic distance between the aligned genome pairs in the *Bacillus–Clostridium* phylum. (B) Median symmetry scores ($D$) for each phylum considering either all pairwise genome comparisons (unfiltered) or a subset of pairwise comparisons sampled to match a common underlying distribution of 16S divergence levels (16S filtered) as described in the "Materials and Methods" section ($p = 0.8$, $P = 0.01$).
(Firmicutes) and *Pseudomonas aeruginosa* (Vallet-Gely and Boccard 2013) (γ-proteobacteria), where they co-localize at mid-cell, and in *Caulobacter crescentus* (Jensen et al. 2001) (α-proteobacteria) and *Helicobacter pylori* (Sharma et al. 2014) (ε-proteobacteria), where sister forks remain in close physical proximity when they migrate together from the cell pole to a mid-cell position. Sister replisomes are not physically tethered to each other (they move independently) but remain close. In *Mycobacterium smegmatis* (Santi and McKinney 2015) (Actinobacteria) and *Myxococcus xanthus* (Harms et al. 2013) (δ-proteobacteria) sister replisomes sporadically drift far enough apart to allow detection of independent fluorescence foci but stay within a certain distance of each other (Santi and McKinney 2015; Trojanowski et al. 2015). Finally, in *E. coli*, replisomes have been reported to localize to different cell poles after initiation of replication at mid-cell (Reyes-Lamothe et al. 2008). We find that *D* scores correlate positively with the fraction of observations in a given replisome tracking experiment that detected two foci rather than a single focus (fig. 5, \( \rho = 0.9, P = 0.08 \), Spearman’s correlation, supplementary table 3, Supplementary Material online). That is, organisms in which putative sister replisomes spend more time apart (and are thus detectable as independent foci) have a lesser tendency for symmetric inversions. In computing *D* scores here, we focus on inversions between closely (strain level) related genomes (16S rRNA divergence < 0.01) because (a) it increases the chance that we are dealing with mutational events that have not yet been eliminated by selection and (b) we can reasonably assume that co-localization dynamics in the compared taxa are the same whereas this need not be the case between more distantly related taxa, as illustrated by the different co-localization patterns of *P. aeruginosa* and *E. coli* (both γ-proteobacteria). These findings hint at the possibility that selection operates on pre-existing mutational biases to reinforce and maintain biased inversion landscapes. However, these findings should be interpreted with caution. Recent extensive replisome tracking experiments in *E. coli* suggest that finding two foci in a single cell typically...
represents a further round of replication prior to cell division rather than spatial separation of sister replisomes (Mangiameli et al. 2017). To understand differences in colocalization dynamics across species it will therefore be vital to explicitly track sister replisomes throughout the cell cycle.

Symmetric Inversions Are Prevalent around Termini in Archaea with Multiple Origins of Replication

Importantly, sister replisome co-localization is not necessary for symmetric inversion bias to be observed. We can demonstrate this explicitly by considering inversion dynamics in archaea with multiple origins of replication. In single-origin prokaryotes, symmetry around the origin is inextricably tied to symmetry around the terminus. In multi-origin organism, on the other hand, it is possible in principle to observe symmetric inversions around a given terminus but not a neighboring origin, and vice versa. If sister replisomes traveling together from the origin provide the main mutational impetus for symmetric inversions, symmetry should be evident around origins, but not around termini of replication. Selection on the other hand, might favor symmetric inversions both around the origin and the terminus, for example to maintain a given distance to the terminus for motifs involved in DNA segregation. We therefore considered inversions around the origins and termini of replication in Sulfolobus spp. (three origins) and Halobacteria (three to four origins), the two clades where multiple complete genomes are available for comparison. In both Haloferax volcanii and Sulfolobus acidocaldarius—model representatives of these clades—origins of replication fire synchronously and evidence from S. acidocaldarius suggests that sister replisomes remain associated during replication (Gristwood et al. 2012). At the same time, despite the resemblance to a eukaryotic, multi-ori mode of chromosome replication, selection has shaped the organization of Sulfolobus genomes in a bacteria-like fashion, with essential and highly expressed genes enriched around the origins of replication (Andersson et al. 2010). We divided 15 Sulfolobus and 14 halobacterial chromosomes into ori- or ter-centered territories, and paired up orthologous territories from different taxa to investigate the presence of the X-type symmetry in each territory (see “Materials and Methods” section). Symmetric inversion bias is evident around individual origins of replication but, strikingly, also around individual termini in both clades (fig. 6). As symmetric inversions around replication termini cannot have originated from co-traveling sister replisomes, we conclude that either selection or an alternative mutational bias must account for the non-random inversion landscape.

Regarding alternative mutational biases, one intriguing possibility is that 3D chromosome topology either during or outside of replication affects inversion landscapes. Indeed, it has been suggested recently (Khodkar and Seshasayee 2016) that chromosomal arrangements within the 3D cell space might promote symmetric inversions. In most prokaryotes where chromosome conformation has been probed globally, including C. crescentus (Umberger et al. 2011; Le et al. 2013), B. subtilis (Marbouty et al. 2015), and Vibrio cholera (Marbouty et al. 2014), chromosome topology follows a longitudinal organization, the right and left replichore lying parallel to each other in the cell. If homology search following a double-strand break were biased toward regions of the genome that are close in space, a longitudinal organization would favor the emergence of symmetric inversions. Consistent with this hypothesis, Khedkar and Seshasayee (2016) find that regions of high contact density in C. crescentus are enriched for inversion breakpoints.

Although we cannot currently quantify the role mechanistic biases play in shaping symmetric inversion landscapes, this
question deserves further scrutiny, not least because mechanistic biases might also apply to the generation of inversions in eukaryotes, both during germline and somatic evolution. It is noteworthy in this regard that the 3D arrangement of chromosomes inside the cell is predictive of translocation probabilities in mammalian cells (Roukos et al. 2013). What our results do strongly suggest is that selection is the ultimate sculptor of biased inversion landscapes across prokaryotic evolution. We observe substantial variation in symmetric inversion bias across clades and this variation is linked to different aspects of functional genome organization along the ori-ter-axis. Notably different features of genome organization are predictive of inversion symmetry in different clades, suggesting that heterogeneous selection pressures can converge to produce similar patterns of genome structure evolution in different prokaryotic clades.

Materials and Methods

Data Acquisition and Processing

Genome sequences of 2,784 and 2,773 completely sequenced prokaryotes were downloaded from Refseq (ftp.ncbi.nlm.nih.gov/genomes/Bacteria, accessed 5 October 2014) and GenBank (ftp://ftp.ncbi.nlm.nih.gov/genomes/archive/old_refseq/Bacteria/, accessed 12 March 2015), respectively, along with corresponding annotations. Where multiple genome elements were present for a given taxon (secondary chromosomes, plasmids), we only considered the largest chromosome.

We assembled a dataset of Proteobacteria with experimentally determined ori position from the literature (supplementary table 1, Supplementary Material online). Where oris are defined by two flanking genes rather than a precise genome coordinate, the center position between the two genes was assigned as ori. In silico-predicted ori positions were obtained from the DoriC database (http://tubic.tju.edu.cn/doric/; last accessed April 19, 2017), which integrates information about multiple genome features (nucleotide skews, DNA box distribution, genes adjacent to candidate oris) to predict ori positions (Gao et al. 2013). Where necessary, the NCBI Genome Browser was used to map Refseq genome element identifiers from DoriC to GenBank identifiers. Only genomes with assemblies used to build the DoriC database were included in the analysis. In some instances, DoriC annotates multiple origin locations for a single bacterial chromosome. These cases might constitute genuine multi-partite origins—as observed, for example, in H. pylori (Donczew et al. 2012)—or might constitute artifacts of feature-based origin calling or represent plasmid integration events. Following manual inspection, all putative origin locations separated by 3,500 bp or less were considered to be a single origin with multiple parts, and the center of the region containing multiple parts was taken as ori. Bacterial chromosomes with multiple ori locations separated by >3,500 bp were excluded. We defined the terminus as the position half a genome length away from the origin. While not every terminus region in bacteria is located at precisely this position, there is evidence for strong long-term selection to maintain equal replicon size, notably from analyses of nucleotide skews, which suggest that replicon imbalance rarely exceeds a 60:40 split (Hendrickson and Lawrence 2006).

Assessing Symmetric Inversion Bias

To establish whether there are biases for symmetric versus asymmetric inversions in different prokaryotic clades, we first aligned all possible intra-phylum genome pairs using the “mummer” program from the MUMmer package (Kurtz et al. 2004) to detect maximal unique matches (MUMs) between genome pairs. The length of MUMs varies from a preset threshold to the length of the longest exact and unique sequence match detected in both genomes. A threshold of 20 bp for the minimal MUM length was chosen based on the recommendations of the authors of the MUMmer package as sufficiently large to avoid spurious matches, and sufficiently small to detect a number of matches in the phylogenetically distant genome pairs. We obtained almost identical results for a trial dataset (Proteobacteria with experimentally determined origins) when applying a significantly larger minimum threshold (100 bp) suggesting that spurious matches for shorter MUMs are rare and/or do not unduly affect results (supplementary fig. 5, Supplementary Material online). Only genome alignments with at least 40 MUMs in one direction and 20 MUMs in the other direction were retained for analysis.

We then considered the residual deviation of individual MUMs from the forward or reverse arm of an imaginary X spanning the alignment plot, as described in the main text, scanning the alignment space at a resolution of 10 kb, and assessing X-type symmetry ($X_{ij}$) for each point $i,j$ according to Equation 1. To place greater confidence in longer homologous blocks, all residuals were weighed by the ratio of MUM length to minimal MUM length (20 bp) and each $X_{ij}$ value was normalized by the sum total of all the weights of MUMs detected, so as to allow comparisons between genome pairs. Density plots were generated using the logspline function in R with default smoothing parameters.

Phylogenetic Controls

We obtained 16S rRNA alignments from the SILVA database v. 119.1 (Quast et al. 2013) and mapped GenBank identifiers to DoriC Refseq genomes using the NCBI Genome Browser. In case where the same taxon was associated with multiple 16S rRNA sequences, a single sequence (belonging to the largest genome element) was chosen at random. Two bacteria from the dataset with experimentally determined origins of replication were not found in the SILVA alignments (Rickettsia prowazekii and Vibrio harveyi). For these bacteria, the first 16S rRNA sequence in the GenBank file of the largest genome element was added to the alignment using the SINA aligner available on the SILVA website (https://www.arb-silva.de/; last accessed April 19, 2017). Based on the full 16S rRNA alignment, we calculate pairwise phylogenetic distances between taxa using the dnadist program from the Phylip package v. 3.695 with default settings. We used NCBI Taxonomy (https://www.ncbi.nlm.nih.gov/taxonomy; last accessed April 19, 2017) to divide species into phylogenetic clades of interest (phyla and classes). To render X-type symmetry
values comparable between clades, we randomly subsampled from genome pairs within each clade of interest to match a common template of phylogenetic distances. As the distance template, we used the distribution of phylogenetic distances within the Thaumarchaeota–Aigarchaeota–Crenarchaeota–Korarchaeota (TACK) superphylum, a clade with a relatively small number of genome pairs.

Covariance between Symmetric Inversion Bias and Genome Architecture

We assessed four putatively adaptive features of genome architecture to understand whether symmetric inversion bias might be driven by selection. First, we determined the average distance of RNA genes to the origin of replication (calculated as a percentile of replicon size). Second, we computed the enrichment of protein-coding genes belonging to the COG class J (Translation, ribosomal structure, biogenesis) near ori as a Z-score, comparing the average distance of COG J genes to the average distance of all COG-annotated genes. Lower Z-scores indicate a stronger enrichment of COG J genes near ori. COG J genes are often highly expressed and essential and therefore taken to represent a class of genes previously identified as enriched near ori in some fast-growing bacteria (Rocha and Danchin 2003). Third, we determined the relative enrichment of genes on the leading versus lagging strand as an indicator of replication–transcription conflicts. The strand with more genes was considered to be the leading strand. Finally, we considered the relative enrichment of nucleotide motifs on the leading versus lagging strand. Following Hendrickson and Lawrence (2006), we focused on octamers as informative motifs, identifying the most strand-biased octamer in each genome by comparing the difference in abundance between the two replicons. All octamers with a non-AGTC base were excluded from the analysis.

Ori- and ter-Centered Territories in Archaea with Multiple Origins of Replication

Genomes of multi-ori archaea were separated into either ori-centered territories (with neighboring ters as boundaries of territories) or ter-centered territories (with neighboring oris as boundaries). Ori territories were trimmed so that both replicons were of the same length (which is not the case if origins are not equidistant). Homologous territories were defined as those carrying the most MUM hits to each other when all territory pairs between two species are compared, with an additional constraint of maximal difference in length of 10 kb. D was measured as the distance of max(\(X_{ij}\)) to the central point of the territory or to the counterpart of the central point half a territory away.

Supplementary Material
Supplementary data are available at Molecular Biology and Evolution online.

Author Contributions
J.R. carried out all analyses. J.R. and T.W. conceived the study, interpreted the data and wrote the manuscript.

Funding
This study was supported by core funding from the Medical Research Council and an Imperial College Junior Research Fellowship.

References

Andersson AF, Pelve EA, Lindeberg S, Lundgren M, Nilsson P, Bernander R. 2010. Replication-biased genome organisation in the crenarchaeon Sulfolobus. BMC Genomics 11:454.
Belda E, Moya A, Silva FJ. 2005. Genome rearrangement distances and gene order phylogeny in gamma-Proteobacteria. Mol Biol Evol. 22:1456–1467.
Canchaya C, Claesson MJ, Fitzgerald GF, van Sinderen D, O’Toole PW. 2006. Diversity of the genus Lactobacillus revealed by comparative genomics of five species. Microbiology 152:3185–3196.
Couturier E, Rocha EP. 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.
Darling AE, Mlikóš I, Ragan MA. 2008. Dynamics of genome rearrangement in bacterial populations. PLoS Genet. 4:e1000128.
Donczew R, Weigel C, Lurz R, Zakrzewska-Czerwińska J, Zawiałak-Pawlik A. 2012. Helicobacter pylori oriC—the first bipartite origin of chromosome replication in Gram-negative bacteria. Nucleic Acids Res. 40:9647–9660.
Eisen JA, Heidelberg JF, White O, Salzberg SL. 2000. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biol. 1:research0011.
Esnault E, Valens M, Espéli O, Boccard F. 2007. Chromosome structuring limits genome plasticity in Escherichia coli. PLoS Genet. 3:e226.
Flynn KM, Vohr SH, Hatcher PJ, Cooper VS. 2010. Evolutionary rates and gene dispensability associate with replication timing in the archaean Sulfolobus islandicus. Genome Biol Evol. 2:859–869.
Gao F, Luo H, Zhang C-T. 2013. DoriC 5.0: an updated database of oriC regions in both bacterial and archaeal genomes. Nucleic Acids Res. 41:D90–D93.
Gil R, Silva FJ, Zientz E, Delmott E, Gonzalez-Candelas F, Latorre A, Rausell C, Kamerbeek J, Gadou J, Hollödöbler B, et al. 2003. The genome sequence of Blochmannia floridanus: comparative analysis of reduced genomes. Proc Natl Acad Sci U S A. 100:9388–9393.
Gristwood T, Duggin IG, Wagner M, Albers SV, Bell SD. 2012. The subcellular localization of Sulfolobus DNA replication. Nucleic Acids Res. 40:5487–5496.
Harms A, Treuner-Lange A, Schumacher D, Søgaard-Andersen L. 2013. Tracking of chromosome and replisome dynamics in Myxococcus xanthus reveals a novel chromosome arrangement. PLoS Genet. 9:e1003802.
Hendrickson H, Lawrence JG. 2006. Selection for chromosome architecture in bacteria. J Mol Evol. 62:615–629.
Hughes D. 2000. Evaluating genome dynamics: the constraints on rearrangements within bacterial genomes. Genome Biol. 1:1.
Hurst LD, Pál C, Lercher MJ. 2004. The evolutionary dynamics of eukaryotic gene order. Nat Rev Genet. 5:299–310.
Jensen RB, Wang SC, Shapiro L. 2001. A moving DNA replication factory in Caulobacter crescentus. EMBO J. 20:4952–4963.
Khdaker S, Seshasayee ASN. 2016. Comparative genomics of inter-replicon translocations in bacteria: a measure of chromosome topology? G3: Genes|Genomes|Geneticscs3:6:1597–1606.
Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol. 5:R12.
Le TBK, Imakaev MV, Mirny LA, Laub MT. 2013. High-resolution mapping of the spatial organization of a bacterial chromosome. Science 342:731–734.
Lemon KP, Grossman AD. 2000. Movement of replicating DNA through a stationary replisome. Mol Cell. 6:1321–1330.
Mackiewicz P, Mackiewicz D, Kowalczyk M, Cebrat S. 2001. Flip-flop around the origin and terminus of replication in prokaryotic genomes. Genome Biol. 2:1004.
Marbouty M, Cournac A, Flot J-F, Marie-Nelly H, Mozziconacci J, Koszul R, McVeYan G. 2014. Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. *eLife* 3:e03318.

Marbouty M, Le Gall A, Cattoni DI, Cournac A, Koh A, Fiche J-B, Mozziconacci J, Murray H, Koszul R, Nollmann M. 2015. Condensin- and replication-mediated bacterial chromosome folding and origin condensation revealed by Hi-C and super-resolution imaging. *Mol Cell*. 59:588–602.

Moran NA, Mira A. 2001. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol*. 2:1.

Nakagawa I, Kurokawa K, Yamashita A, Nakata M, Tomiyasu Y, Okahashi N, Kawabata S, Yamazaki K, Shiba T, Yasunaga T, et al. 2003. Genome sequence of an M3 strain of *Streptococcus pyogenes* reveals a large-scale genomic rearrangement in invasive strains and new insights into phage evolution. *Genome Res*. 13:1042–1055.

Parkhill J, Sebaihia M, Preston A, Murphy L D, Thomson N, Harris DE, Holden MTG, Churcher CM, Bentley SD, Mungall KL, et al. 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet*. 35:32–40.

Pelve EA, Lindås AC, Knöppel A, Mira A, Bernander R. 2012. Four chromosome replication origins in the archaeon *Pyrobaculum calidifontis*. *Mol Microbiol*. 85:986–995.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 41:D590–D596.

Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, Hickey EK, Peterson J, Utterback T, Berry K, et al. 2000. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res*. 28:1397–1406.

Reyes-Lamothe R, Posoza C, Daniova O, Sherratt DJ. 2008. Independent positioning and action of *Escherichia coli* replisomes in live cells. *Cell* 133:90–102.

Rocha EPC. 2004. The replication-related organization of bacterial genomes. *Microbiology* 150:1609–1627.

Rocha EPC. 2008. The organization of the bacterial genome. *Annu Rev Genet*. 42:211–233.

Rocha EPC, Danchin A. 2003. Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nat Genet*. 34:377–378.

Roukos V, Voss TC, Schmidt CK, Lee S, Wangsa D, Misteli T. 2013. Spatial dynamics of chromosome translocations in living cells. *Science* 341:660–666.

Santi I, McKinney JD. 2015. Chromosome organization and replisome dynamics in *Mycobacterium smegmatis*. *mBio* 6:e01999–e01914.

Sharma A, Kamran M, Verma V, Dasgupta S, Dhar SK. 2014. Intracellular locations of replication proteins and the origin of replication during chromosome duplication in the slowly growing human pathogen *Helicobacter pylori*. *J Bacteriol*. 196:999–1011.

Silva FJ, Latorre A, Moya A. 2001. Genome size reduction through multiple events of gene disintegration in Buchnera APS. *Trend Genet*. 17:615–618.

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

Tillier ER, Collins RA. 2000. Genome rearrangement by replication-directed translocation. *Nat Genet*. 26:195–197.

Touzain F, Petit M-A, Schbath S, Karoui ME. 2011. DNA motifs that sculpt the bacterial chromosome. *Nat Rev Microbiol*. 9:15–26.

Trojanowski D, Gindz K, Piśko M, Holčíkova J, Skut P, Jakimowicz D, Zakrzewska-Czerwińska J. 2015. Choreography of the Mycobacterium replication machinery during the cell cycle. *mBio* 6:e01215–e01214.

Umbarger MA, Toro E, Wright MA, Porreca GJ, Bai D, Hong S-H, Ferro MI, Zhu LJ, Marti-Renom MA, McAdams HH, et al. 2011. The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. *Mol Cell*. 44:252–264.

Vallet-Gely I, Boccard F. 2013. Chromosomal organization and segregation in *Pseudomonas aeruginosa*. *PloS Genet*. 9:e1003492.

Worning P, Jensen LJ, Hallin PF, Stærfeldt HH, Ussery DW. 2006. Origin of replication in circular prokaryotic chromosomes. *Environ Microbiol*. 8:353–361.

Zivanovic Y, Lopez P, Philippe H, Forterre P. 2002. *Pyrococcus* genome comparison evidences chromosome shuffling-driven evolution. *Nucleic Acids Res*. 30:1902–1910.