The roles of replication-transcription conflict in mutagenesis and evolution of genome organization

Jeremy W. Schroeder*[1], T. Sabari Sankar*[2], Jue D. Wang*[1], Lyle A. Simmons*[3]

[1] Department of Bacteriology, University of Wisconsin-Madison
[2] Indian Institute of Science Education and Research Thiruvananthapuram
[3] Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, Michigan, USA.

*Correspondence:
jwschroeder3@wisc.edu (JWS)
sabari@iisertvm.ac.in (TSS)
wang@bact.wisc.edu (JDW)
lasimm@umich.edu (LAS)

Abstract

Replication-transcription conflicts promote mutagenesis and give rise to evolutionary signatures, with fundamental importance to genome stability ranging from bacteria to metastatic cancer cells. This review focuses on the interplay between replication-transcription conflicts and the evolution of gene directionality. In most bacteria, the majority of genes are encoded on the leading strand of replication such that their transcription is co-directional with the direction of DNA replication fork movement. This gene strand bias arises primarily due to negative selection against deleterious consequences of head-on replication-transcription conflict. However, many genes remain head-on. Can head-on orientation provide some benefit? We combine insights from both mechanistic and evolutionary studies, review published work, and analyze gene expression data to evaluate an emerging model that head-on genes are temporal targets for adaptive mutagenesis during stress. We highlight the alternative explanation that genes in the head-on orientation may simply be the result of genomic inversions and relaxed selection acting on nonessential genes. We seek to clarify how the mechanisms of replication-transcription conflict, in concert with other mutagenic mechanisms, balanced by natural selection, have shaped bacterial genome evolution.

Introduction

Studies of mutagenesis in bacteria have revealed conserved biological processes influencing mutation rate across all domains of life and identified sources of bacterial evolution against host defense, peer competition, and antibiotic exposure. Historical studies of bacterial mutations by Luria and Delbrück led to a conclusive demonstration that spontaneous mutations generate the genetic diversity to allow survival of subsequent selective pressure [1]. Mutations do not arise uniformly across the genome, because multiple mechanisms for mutagenesis exist and the effect of each mechanism depends on the genomic context of a given locus [2]. Replication-transcription conflict is one mutagenic mechanism that is particularly influenced by genomic context. Replication-transcription conflict is common to prokaryotic and eukaryotic systems [for reviews, see [3–6]]. Insertions and deletions within genes, a key mutation signature of replication-transcription conflict in bacteria, are also found in cancer cells [7–9], suggesting that the mechanisms by which conflicts generate mutations may also be conserved.

Recently, the field of replication-transcription conflict has generated both excitement and contradictions. There are new developments in the contribution of conflicting interplay between replication-transcription machineries to genome evolution and genome organization. In most bacterial species, the majority of genes are encoded in the leading strand of replication such that their transcription is co-directional with DNA replication fork movement. The extent of gene strand bias differs across bacteria, with *Escherichia coli*, *Bacillus subtilis*, and *Thermoanaerobacter tengcongensis* having 55, 75, and 87 percent of genes co-directional, respectively (Fig 1A and [5]). Gene strand bias is the result of purifying selection against head-on genes, which are targets for mutagenesis and impede DNA replication when they are highly expressed (Fig 1B and [7,10–12]). Experimental evidence demonstrates that deleterious

consequences of inverting co-directionally oriented genes to be head-on include slowed replication fork progression, replication arrest, and DNA double-strand breaks/ends [7,10,13,14].

Given the deleterious consequences of head-on conflict, it is natural to wonder how bacterial genomes have not evolved to have all genes in the co-directional orientation. After all, such an arrangement of genes should increase genome stability and fitness. Is there a driving force for some genes to be oriented head-on? One possibility is that head-on orientation of stress-induced genes is favorable. Merrikh and colleagues proposed that stress-induced genes are preferentially positioned in the head-on orientation to promote replication-transcription conflict during stress, which in turn would promote adaptive mutagenesis of stress-induced genes [15–19]. Alternatively, the prevalence of stress-induced genes in the head-on orientation may simply be the result of two well-established phenomena: the occurrence of genomic inversions and selection against head-on genes. This alternative model is supported by recent work documenting the types of mutations generated by head-on conflict, the mechanisms of mutagenesis in head-on genes, and the evolutionary forces driving strand-biased gene distribution [7,20–23]. In this review, we present the evidence raised to support each view. We first briefly review evolutionary work pertaining to head-on genes, then discuss the mechanisms of mutagenesis of head-on genes, both dependent and independent of conflict, and finally examine stress-induced genes specifically. We hope to clarify what has been learned about replication-transcription conflict as it pertains to mutagenesis and evolution of head-on genes and strand-biased gene distribution.

Main Text

An evolutionary explanation for the existence of head-on genes

Merrikh and colleagues proposed an adaptive hypothesis to explain the occurrence of genes in the head-on orientation [17,19]. Using comparative genomics, they presented several lines of evidence they propose support positive selection for head-on oriented genes. First, estimating nucleotide substitutions of several genes, they showed that nonsynonymous substitutions (dN) were higher in head-on compared to co-directional genes while synonymous substitutions (dS) showed no difference. Therefore, the ratio of dN to dS, which is an indicator of selective pressure on protein-coding genes, was higher for head-on genes [17,19]. Second, dN/dS ratios that are significantly greater than one provide evidence that a gene is under positive selection [24], and the authors showed that genes with dN/dS modestly above one from multiple bacterial species were slightly more enriched in head-on genes compared to co-directional genes [17]. Finally, they proposed that convergent amino acid substitutions, a key criterion of adaptive evolution, occurred more frequently in head-on genes [19]. Based on these observations, Merrikh and colleagues proposed that head-on genes experience positive selection for increased mutation rate caused by head-on conflicts.

Though the adaptive hypothesis is intriguing, the notion is contradicted based on the theory of mutation-selection balance [20,23]. First, the increased dN observed for lagging strand-encoded genes can be explained by either adaptive evolution or relaxed purifying selection. In this case, purifying selection is the selection against deleterious mutations. Co-directional genes are

enriched with essential genes and house-keeping genes, and are therefore subject to stronger purifying selection than head-on genes. Nonsynonymous substitutions in co-directional genes are selected against, causing lower dN and dN/dS in co-directional genes. The finding that the mean dN/dS ratio of head-on genes is slightly higher than that of co-directional genes [17,19], while remaining well below one, provides strong support that head-on genes are under relaxed purifying selection compared to co-directional genes. Second, although the percentage of genes with dN/dS greater than one, i.e., under positive selection, is higher among genes in the head-on orientation, in absolute numbers, more co-directional genes counted by Merrikh and Merrikh have dN/dS greater than one (289 of 15627 total co-directional genes versus 234 of 9757 total head-on genes) [17]. Thus, on a gene-by-gene basis there is little evidence to suggest that genes with dN/dS greater than one receive additional adaptive benefits if they are head-on. Therefore, there remains poor evidence to suggest positive selection for head-on genes. Finally, Chen and Zhang [20] show that convergent substitution, a key criterion of adaptive evolution [25], is not observed to be higher in head-on genes as claimed in [19].

Instead of the head-on orientation being positively selected, gene strand bias can represent the equilibrium distribution of head-on and co-directional genes that arises due to a balance of random genetic inversion and purifying selection against head-on genes (Fig 1B and [20]). Therefore, despite negative selection against head-on genes in general, genes that remain head-on are those whose head-on conflict is either less costly to bacterial fitness, or that have not yet been purified from the genome. In agreement with this theory, the strength of selection against the head-on orientation depends on the type of gene in which conflict occurs. For example, inverting the ribosomal RNA operons to the head-on orientation has extremely deleterious consequences because they are highly expressed essential genes, and their head-on transcription disrupts replication fork progression [4,10,11,26]. Therefore, the extreme consequences of head-on conflict at inverted ribosomal RNA operons explain why all ribosomal RNA operons are oriented co-directional to replication [27]. Mutation-selection balance may also explain differences in gene strand bias between different bacteria. Inverting genes from co-directional to head-on results in a much stronger fitness defect in *B. subtilis* [10] compared to comparable inversions in *E. coli* [28,29]. Accordingly, *B. subtilis* has higher bias for co-directional genes (75 percent co-directional) than *E. coli* (55 percent co-directional) [5]. Thus, natural selection favors genes to be co-directionally transcribed [6], but due to unavoidable inversion events the presence of head-on genes is inevitable.

Are there sufficient inversion events to support the mutation-selection balance model for gene strand bias? Bacteria often contain a circular chromosome with a single origin of replication and two replication forks, each replicating half a chromosome (replichore). Any inversion event contained within a replichore will change the relative direction of replication-transcription conflict for all genes contained within the inversion. Inversions within a replichore would rarely be detected at the population level, and this is likely due to the fitness cost of having many genes oriented head-on [10,30,31]. Bioinformatic studies in many bacteria reveal that symmetric inversions around the origin of replication frequently occur [32,33]. This further suggests that non-symmetric inversions may be more common than currently realized, but only symmetric inversions are pervasively retained through evolution due to their low fitness cost. In studies selecting for inversion within a gene bearing 12- or 23-bp regions of inverted homology, Miller

and co-workers estimated inversion rates to be $4*10^{-9}$ to $1*10^{-7}$ per generation [34,35]. Inversion rates will depend on many factors, including the length of homology at the inversion break-points [36]. For comparison, the rate of base-pair substitution in bacteria varies around $3*10^{-10}$ per nucleotide replicated per generation, depending on the type of substitution and sequence context [22,37]. Interestingly, both the Zhang group and the Merrikh group, although disagreeing in the numbers, provide evidence for many inversion events resulting in co-directional to head-on changes of gene orientation [17,21]. In summary, considering the potential rate of chromosomal inversions, in both methodological and theoretical terms the mutation-selection balance model best explains the existence of head-on genes [20,21,23].

Mutation signatures of head-on conflict: deleterious or beneficial?

Various reporters have been employed to study the mutagenic consequences of replication-transcription conflicts. Mutation assays based on resistance to rifampin or trimethoprim [7,10], and reversion of auxotrophic mutants to prototrophy [19], have all shown that mutation reporter genes in the head-on orientation have modestly higher mutation rates than their co-directional counterpart in *B. subtilis*. These similar results have been interpreted differently. The Merrikh group interpreted the increased rate of reversion of auxotrophic mutants to prototrophy to support their hypothesis favoring positive selection for head-on genes [18,19]. However, their assays begin with a non-functional mutant gene and select for specific reversions to restore a functional state. During natural evolution, head-on genes do not begin with a non-functional state wherein the only direction for fitness to go is up. Rather, under more natural conditions, mutagenesis comes with many risks to gene fitness. In addition, reversion assays only select for specific nonsynonymous base substitutions at few positions, making them poorly suited to study the impact of conflicts on mutagenesis. In contrast, forward mutation assays select for all loss-of-function mutations in a gene and are thus more helpful in revealing the multiple types of mutations that could inactivate a gene.

Using a forward mutation assay based on trimethoprim resistance, Sankar *et al.* comprehensively identified mutation signatures of conflicts [7]. Among a wide spectrum of loss-of-function mutations that were obtained, insertions/deletions (indels) and base substitutions at the promoter were demonstrated as the major mutation signatures resulting from replication-transcription conflicts. Indels were proximal to the sites where collisions occurred and spanned from cis-regulatory elements to coding regions. The vast majority of indels are deleterious, resulting in loss of gene function. Therefore, while there is no evidence that conflicts primarily generate beneficial mutations, there is conclusive evidence that conflicts generate deleterious mutations.

The forward mutation assay also detected hundreds of different base substitutions throughout the coding region and at the promoter. Mutation rate in the coding sequence did not increase when head-on transcription was induced compared to the uninduced condition [7]. Therefore, most base substitutions in the coding sequence likely arise independently of head-on conflict. In contrast, promoter substitutions appeared as a hotspot with a strongly elevated rate upon induction of head-on transcription [7]. This hotspot is identical in nature and position to a promoter hotspot previously identified in a head-on oriented mutation reporter in *E. coli* [38].

This striking similarity in mutagenic signature identified in two phylogenetically distant bacteria (*E. coli* and *B. subtilis*), using two different reporter systems, supports the widespread existence of promoter substitutions due to head-on replication-transcription conflict.

Promoter substitutions due to head-on conflict suggest another origin of gene strand bias. A comparative genomic analysis demonstrated that long bacterial operons are strongly biased to the leading strand and occur sparsely in the head-on orientation, especially when the operons are conserved [39]. A single promoter mutation due to head-on conflict could inactivate every gene in a multi-gene operon, causing a strong loss of fitness. Therefore, promoter mutations caused by head-on conflicts are more likely to be deleterious than beneficial, and thus convincingly explain the evolutionary pressure against head-on genes, especially for essential and highly expressed genes.

Recently, Lang and Merrikh argued against mechanistic evidence bolstering the evolutionary model of selection against head-on genes [15]. In their review, the authors conclude, mistakenly in our opinion, that promoter base substitutions due to head-on conflicts are a consequence of mutational bias introduced by toxic selection on trimethoprim [15]. We disagree with this interpretation for several reasons. First, the fundamental tenet of the fluctuation test as devised by Luria and Delbrück is that selection for mutants is separate from generation of spontaneously-occurring mutations during growth [1,2]. In the Luria and Delbrück experiment, phage-resistant mutant cells arose prior to exposure to phage, and not in response to phage infection [1]. Likewise, mutations resulting from replication-transcription conflicts were revealed by trimethoprim selection and were not generated in response to the drug, as demonstrated by the Luria-Delbrück distribution of mutants per culture [7]. Hence, mutations identified after selection for trimethoprim resistance represent authentic signatures of replication-transcription conflicts.

Second, if the argument is that promoter mutations negate the toxic nature of trimethoprim, then the same mutation should appear similarly enriched in the co-directional orientation, which was distinctly not the case. Additionally, the promoter mutant showed no growth advantage in competition with wild type and other types of mutants, suggesting that it was a *bona fide* mutational outcome of conflicts and not a result of selection bias [7]. Therefore, loss-of-function mutation assays facilitate less biased estimation of mutation rates than reversion assays and are suitable for studying the impact of replication-transcription conflict on mutagenesis.

Are mutations in head-on genes largely dependent on replication-transcription conflict?

Replication-transcription conflict is not the only mutagenic mechanism that can manifest in different mutation rates between co-directional and head-on genes. One such mechanism is differences between the accuracy of replication of the leading and lagging strands [40,41]. It is hypothesized that leading and lagging strand replication fidelity is unequal due to intrinsic differences in replication accuracy and error correction between leading and lagging strands during DNA replication, not necessarily due to replication-transcription conflict [40–43]. When a co-directional gene is inverted, its leading strand template would become the lagging strand

template and vice versa. This would have the effect of adjusting mutation rate in the inverted gene due to changes in local genome context and error correction efficiency that are independent of conflict.

What about the mutation rate in genomes with an established equilibrium of head-on and co-directional genes? These questions can be addressed by mutation accumulation studies, which bypass the limitations of mutation reporters by using whole genome resequencing to identify mutations that become fixed in a population when selection is minimized. Many research groups have performed mutation accumulation experiments over the past decade and made key findings in the evolution of mutation rates, mechanisms of GC skew, the effect of replication timing on mutagenesis, and factors associated with spontaneous mutagenesis [22,23,37,44–49]. Strikingly, Lynch and colleagues demonstrated that mutation accumulation studies have not found an association between base substitution rate and a gene's orientation relative to replication [23]. Instead, one of the strongest determinants of substitution rate was the neighboring DNA sequence context at a given genomic locus [22,45]. Importantly, the influence of neighboring nucleotide sequence context on mutagenesis operates independently of the orientation and expression of a gene. Therefore, mutation accumulation studies suggest that in bacterial genomes that have established an equilibrium of head-on and co-directional genes, replication-transcription conflict is unlikely to be a major driver of nonsynonymous mutagenesis.

Are stress-induced genes typically head-on?

Mechanistic and evolutionary evidence clearly demonstrate that the majority of head-on oriented genes are not driven to be head-on through selection for conflict-induced adaptive mutagenesis. However, there is precedence for selection causing heritable genetic change. For example, upon phage infection, CRISPR loci acquire phage DNA in the form of new spacers [50,51]. Therefore, it remains possible that specific genes are head-on to benefit from conflict-induced mutability. Merrikh and colleagues proposed that stress-induced genes benefit from mutagenesis caused by head-on conflict when their expression is induced during stress [52]. We performed several analyses to test their model.

First, we examined the coding sequences (CDSs) regulated by nine stress-associated transcription factors (Table 1). Are these stress-induced CDSs more likely to be oriented head-on, as the adaptive mutagenesis model for head-on genes suggests? No. On the contrary, all regulons except SigV contain more co-directional CDSs than head-on. For example, genes in the SigM (cell envelope stress) regulon were suggested to be preferentially in the head-on orientation [19]. However, we found that there are 70 co-directional CDSs in the SigM regulon and 23 head-on. Furthermore, the Spx (oxidative stress) and Spo0A (sporulation) regulons were enriched for co-directional CDSs (Table 1). We conclude that, in stark contrast to the prediction of the adaptive hypothesis for head-on stress-induced genes, there is no compelling evidence for selection driving stress-induced genes to be head-on.

Second, under the assumption that stress-induced genes are preferentially head-on, Lang and Merrikh suggested that the reason mutation accumulation studies have not found head-on genes to have higher mutation rates than co-directional genes is due to the relatively low-stress conditions of mutation accumulation experiments [23]. However, RNA-seq data show that head-

on CDSs occupy a similar range of expression values as co-directional CDSs (Fig 2A) [22]. To further test whether head-on genes tend to be “off” under most conditions and “on” under a select few, we analyzed *B. subtilis* transcriptomic data from [53], in which gene expression was measured under 104 conditions. We estimated each gene’s propensity for differential expression by calculating its Gini coefficient over the 104 conditions. The range of Gini coefficients was nearly identical between head-on and co-directional genes (Fig 2B). These analyses indicate that head-on genes are not particularly special in their patterns of differential expression when compared to co-directional genes.

Conclusions

Genome organization has been shaped by selective forces that minimize replication-transcription conflict, resulting in enriched co-directionality between replication and transcription. Although the hypothesis that positive selection promotes head-on orientation of stress-induced genes to increase conflict and promote adaptive mutation is tempting [19], evolutionary [20,21,23] and mechanistic analyses [7,38] suggest this is not the case. Together, the mutation-selection balance theory and conflict-induced mutation signatures strongly support the evolutionary model that head-on genes are a result of inevitable gene inversion [20,31,36], to be later purified from the genome due to increased deleterious mutation rate and their greater propensity to block DNA replication [7,10,13,14,20,38]. Genes remaining head-on tend to be those in which purifying selection is relaxed. The distribution of stress-induced genes is not significantly different between head-on and co-directional orientations. Hence, the adaptive evolution model of head-on genes is poorly suited to explain the presence of genes on the lagging strand and still requires compelling empirical support. Genomes in their current state have likely evolved to an equilibrium distribution of co-directional and head-on genes which is maintained by selective pressure against deleterious consequences resulting from genome rearrangements.

Acknowledgements

This work was supported by National Institutes of Health grant R35 GM131772 to LAS, National Science Foundation grant MCB 1714539 to LAS, R35 GM127088 from NIGMS to JDW, the Howard Hughes Medical Institute Faculty Scholars Award to JDW, and DBT/Wellcome Trust India Alliance Fellowship grant IA/I/18/2/504037 to TSS. The authors have no conflict of interest to declare.

Methods

Transcriptomic microarray data were downloaded from the url http://genome.jouy.inra.fr/basysbio/bsubtranscriptome. The intercept $\log_2$(signal) for each gene and the effect of each of the 104 conditions tested in [53] on each gene’s $\log_2$(signal) was inferred. To avoid obtaining many false positive effects a Bayesian model with a shrinkage prior known as the Finnish Horseshoe prior [54] was employed.

Gene enrichment analysis was performed using regulon annotations hosted at the Subtiwiki database [55]. A chi-square test for independence between gene regulation by each regulator and gene direction relative to DNA replication was used to calculate the p-values in Table 1.

A detailed description of our methods, including all R and python code, and all the data required to reproduce our analysis, can be found in the gitub repository hosted at https://github.com/jadewanglab/2020_PGen_analysis.git.

References

1. Luria SE, Delbruck M. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. Genetics. 1943;28: 491–511.
2. Schroeder JW, Yeesin P, Simmons LA, Wang JD. Sources of spontaneous mutagenesis in bacteria. Crit Rev Biochem Mol Biol. 2017;2017/11/08: 1–20.
3. Hamperl S, Cimprich KA. Conflict Resolution in the Genome: How Transcription and Replication Make It Work. Cell. 2016;167: 1455–1467.
4. Brewer BJ. When polymerases collide: replication and the transcriptional organization of the E. coli chromosome. Cell. 1988;53: 679–686.
5. Rocha EP. The replication-related organization of bacterial genomes. Microbiology. 2004;150: 1609–1627.
6. Rocha EPC. The organization of the bacterial genome. Annu Rev Genet. 2008;42: 211–233.
7. Sankar TS, Wastuwidyaningtyas BD, Dong Y, Lewis SA, Wang JD. The nature of mutations induced by replication-transcription collisions. Nature. 2016;535: 178–181.
8. Imielinski M, Guo G, Meyerson M. Insertions and Deletions Target Lineage-Defining Genes in Human Cancers. Cell. 2017;168: 460–472.e14.
9. Letouzé E, Shinde J, Renault V, Couchy G, Blanc J-F, Tubacher E, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. Nat Commun. 2017;8: 1315.
10. Srivatsan A, Tehranchi A, MacAlpine DM, Wang JD. Co-orientation of replication and transcription preserves genome integrity. PLoS Genet. 2010;6: e1000810.
11. French S. Consequences of replication fork movement through transcription units in vivo. Science. 1992;258: 1362–1365.
12. Kim N, Abdulovic AL, Gealy R, Lippert MJ, Jinks-Robertson S. Transcription-associated mutagenesis in yeast is directly proportional to the level of gene expression and influenced by the direction of DNA replication. DNA Repair . 2007;6: 1285–1296.
13. Vilette D, Ehrlich SD, Michel B. Transcription-induced deletions in plasmid vectors: M13 DNA replication as a source of instability. Mol Gen Genet. 1996;252: 398–403.

14. Mirkin EV, Castro Roa D, Nudler E, Mirkin SM. Transcription regulatory elements are punctuation marks for DNA replication. Proc Natl Acad Sci U S A. 2006;103: 7276–7281.

15. Lang KS, Merrikh H. The Clash of Macromolecular Titans: Replication-Transcription Conflicts in Bacteria. Annu Rev Microbiol. 2018;72: 71–88.

16. Lang KS, Hall AN, Merrikh CN, Ragheb M, Tabakh H, Pollock AJ, et al. Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. Cell. 2017;170: 787–799 e18.

17. Merrikh CN, Merrikh H. Gene inversion potentiates bacterial evolvability and virulence. Nat Commun. 2018;9: 4662.

18. Million-Weaver S, Samadpour AN, Moreno-Habel DA, Nugent P, Brittnacher MJ, Weiss E, et al. An underlying mechanism for the increased mutagenesis of lagging-strand genes in *Bacillus subtilis*. Proc Natl Acad Sci U S A. 2015;112: E1096–105.

19. Paul S, Million-Weaver S, Chattopadhyay S, Sokurenko E, Merrikh H. Accelerated gene evolution through replication-transcription conflicts. Nature. 2013;495: 512–515.

20. Chen X, Zhang J. Why are genes encoded on the lagging strand of the bacterial genome? Genome Biol Evol. 2013;5: 2436–2439.

21. Liu H, Zhang J. No support for the adaptive hypothesis of lagging-strand encoding in bacterial genomes. bioRxiv. 2020. p. 2020.01.14.906818. doi:10.1101/2020.01.14.906818

22. Schroeder JW, Hirst WG, Szewczyk GA, Simmons LA. The Effect of Local Sequence Context on Mutational Bias of Genes Encoded on the Leading and Lagging Strands. Curr Biol. 2016;26: 692–697.

23. Lynch M, Ackerman MS, Gout JF, Long H, Sung W, Thomas WK, et al. Genetic drift, selection and the evolution of the mutation rate. Nat Rev Genet. 2016;17: 704–714.

24. Nei M, Kumar S. Molecular Evolution and Phylogenetics. Oxford University Press; 2000.

25. Christin P-A, Weinreich DM, Besnard G. Causes and evolutionary significance of genetic convergence. Trends Genet. 2010;26: 400–405.

26. Liu B, Alberts BM. Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. Science. 1995;267: 1131–1137.

27. Nomura M, Morgan EA. Genetics of bacterial ribosomes. Annu Rev Genet. 1977;11: 297–347.

28. Boubakri H, De Septenville AL, Viguera E, Michel B. The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo. EMBO J. 2010;29: 145–157.

29. De Septenville AL, Duigou S, Boubakri H, Michel B. Replication fork reversal after replication-transcription collision. PLoS Genet. 2012;8: e1002622.

30. Segall A, Mahan MJ, Roth JR. Rearrangement of the bacterial chromosome: forbidden inversions. Science. 1988;241: 1314–1318.

31. Mackiewicz P, Mackiewicz D, Gierlik A, Kowalczuk M, Nowicka A, Dudkiewicz M, et al. The differential killing of genes by inversions in prokaryotic genomes. J Mol Evol. 2001;53: 615–621.

32. Eisen JA, Heidelberg JF, White O, Salzberg SL. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biol. 2000;1: RESEARCH0011.

33. Darling AE, Miklós I, Ragan MA. Dynamics of genome rearrangement in bacterial populations. PLoS Genet. 2008;4: e1000128.

34. Slupska MM, Chiang JH, Luther WM, Stewart JL, Amii L, Conrad A, et al. Genes involved in the determination of the rate of inversions at short inverted repeats. Genes Cells. 2000;5: 425–437.

35. Schofield MA, Agbunag R, Miller JH. DNA inversions between short inverted repeats in Escherichia coli. Genetics. 1992;132: 295–302.

36. Gordon AJ, Halliday JA. Inversions with deletions and duplications. Genetics. 1995;140: 411–414.

37. Lee H, Popodi E, Tang H, Foster PL. Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. Proc Natl Acad Sci U S A. 2012;109: E2774–83.

38. Yoshiyama K, Higuchi K, Matsumura H, Maki H. Directionality of DNA replication fork movement strongly affects the generation of spontaneous mutations in Escherichia coli. J Mol Biol. 2001;307: 1195–1206.

39. Price MN, Alm EJ, Arkin AP. Interruptions in gene expression drive highly expressed operons to the leading strand of DNA replication. Nucleic Acids Res. 2005;33: 3224–3234.

40. Fijalkowska IJ, Jonczyk P, Tkaczyk MM, Bialoskorska M, Schaaper RM. Unequal fidelity of leading strand and lagging strand DNA replication on the Escherichia coli chromosome. Proc Natl Acad Sci U S A. 1998;95: 10020–10025.

41. Maslowska KH, Makiela-Dzbenska K, Mo J-Y, Fijalkowska IJ, Schaaper RM. High-accuracy lagging-strand DNA replication mediated by DNA polymerase dissociation. Proc Natl Acad Sci U S A. 2018;115: 4212–4217.

42. Lujan SA, Williams JS, Pursell ZF, Abdulovic-Cui AA, Clark AB, Nick McElhinny SA, et al. Mismatch repair balances leading and lagging strand DNA replication fidelity. PLoS Genet. 2012;8: e1003016.

43. Pavlov YI, Mian IM, Kunkel TA. Evidence for preferential mismatch repair of lagging strand DNA replication errors in yeast. Curr Biol. 2003;13: 744–748.

44. Dillon MM, Sung W, Lynch M, Cooper VS. Periodic Variation of Mutation Rates in Bacterial Genomes Associated with Replication Timing. MBio. 2018;9. doi:10.1128/mBio.01371-18

45. Sung W, Ackerman MS, Gout JF, Miller SF, Williams E, Foster PL, et al. Asymmetric Context-Dependent Mutation Patterns Revealed through Mutation-Accumulation

Experiments. Mol Biol Evol. 2015;32: 1672–1683.

46. Foster PL, Lee H, Popodi E, Townes JP, Tang H. Determinants of spontaneous mutation in the bacterium Escherichia coli as revealed by whole-genome sequencing. Proc Natl Acad Sci U S A. 2015;112: E5990–9.

47. Lind PA, Andersson DI. Whole-genome mutational biases in bacteria. Proc Natl Acad Sci U S A. 2008;105: 17878–17883.

48. Lee H, Doak TG, Popodi E, Foster PL, Tang H. Insertion sequence-caused large-scale rearrangements in the genome of Escherichia coli. Nucleic Acids Res. 2016;44: 7109–7119.

49. Bhagwat AS, Hao W, Townes JP, Lee H, Tang H, Foster PL. Strand-biased cytosine deamination at the replication fork causes cytosine to thymine mutations in Escherichia coli. Proc Natl Acad Sci U S A. 2016;113: 2176–2181.

50. McGinn J, Marraffini LA. Molecular mechanisms of CRISPR-Cas spacer acquisition. Nat Rev Microbiol. 2019;17: 7–12.

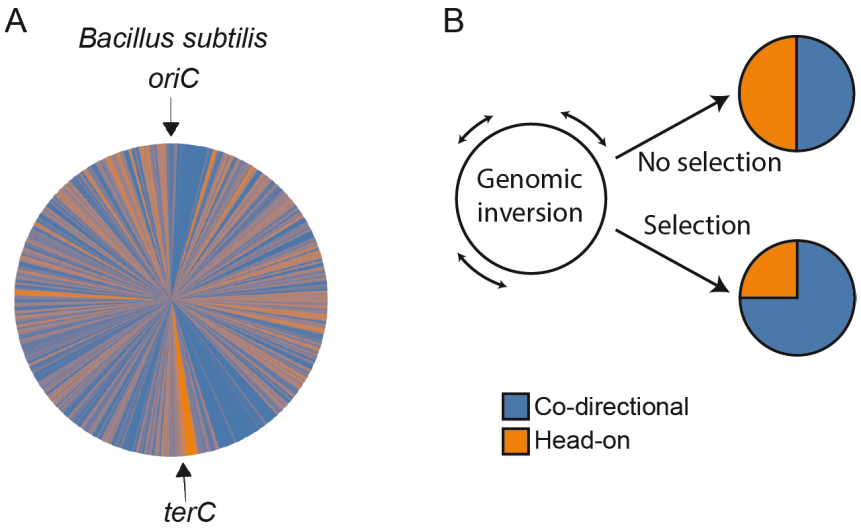
51. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007;315: 1709–1712.

52. Merrikh H. Spatial and Temporal Control of Evolution through Replication-Transcription Conflicts. Trends Microbiol. 2017. doi:10.1016/j.tim.2017.01.008

53. Nicolas P, Mäder U, Dervyn E, Rochat T, Leduc A, Pigeonneau N, et al. Condition-dependent transcriptome reveals high-level regulatory architecture in Bacillus subtilis. Science. 2012;335: 1103–1106.

54. Piironen J, Vehtari A. Sparsity information and regularization in the horseshoe and other shrinkage priors. Electron J Stat. 2017;11: 5018–5051.

55. Michna RH, Zhu B, Mäder U, Stülke J. SubtiWiki 2.0—an integrated database for the model organism Bacillus subtilis. Nucleic Acids Res. 2016;44: D654–D662.



435

436 **Fig 1. Genome-wide bias against head-on genes reflects negative selection.** (A) The *B.*
437 *subtilis* genome is represented to scale as a circle filled with blue or orange to represent loci of
438 the genome containing co-directional and head-on genes, respectively. The origin (*oriC*) and
439 terminus (*terC*) of replication are labelled. Strain PY79 is shown but the basic genome
440 organization is conserved among *B. subtilis* strains. (B) Schematic of random inversions of
441 genomic loci and evolutionary outcomes depending on selection. In the absence of selection
442 against head-on genes, genomic inversions over time would eventually result in a 50:50
443 distribution of head-on:co-directional genes. In the presence of selection against head-on
444 genes, gene content is biased in favor of co-directional genes. Genes that remain head-on are
445 typically those for which there is less selection against the head-on orientation.

446

447

448

449 **Table 1. Most stress-induced genes are co-directional.**

Stress-associated regulator	Number of head-on CDSs in regulon (% of 1177 total head-on)	Number of co-directional CDSs in regulon (% of 3148 total co-directional)	p-value (chi-square)
LexA	14 (1.2%)	46 (1.5%)	0.59
SigB	72 (6.1%)	149 (4.7%)	0.078
SigM	23 (2.0%)	70 (2.2%)	0.67
SigV	14 (1.2%)	7 (0.22%)	0.00013
SigX	9 (0.07%)	33 (1.0%)	0.50
SigY	0 (0%)	7 (0.22%)	0.23
SinR	13 (1.1%)	36 (1.1%)	0.96
Spo0A	25 (2.1%)	117 (3.7%)	0.012
Spx*	3 (0.25%)	38 (1.2%)	0.0069

450 *Only CDSs activated by Spx are included

451 The number of head-on and co-directional coding sequences in *B. subtilis* strain 168 regulated
452 by each transcription factor was tabulated. In addition, their percentage over the total number of
453 head-on or co-directional coding sequences is in parentheses.

454

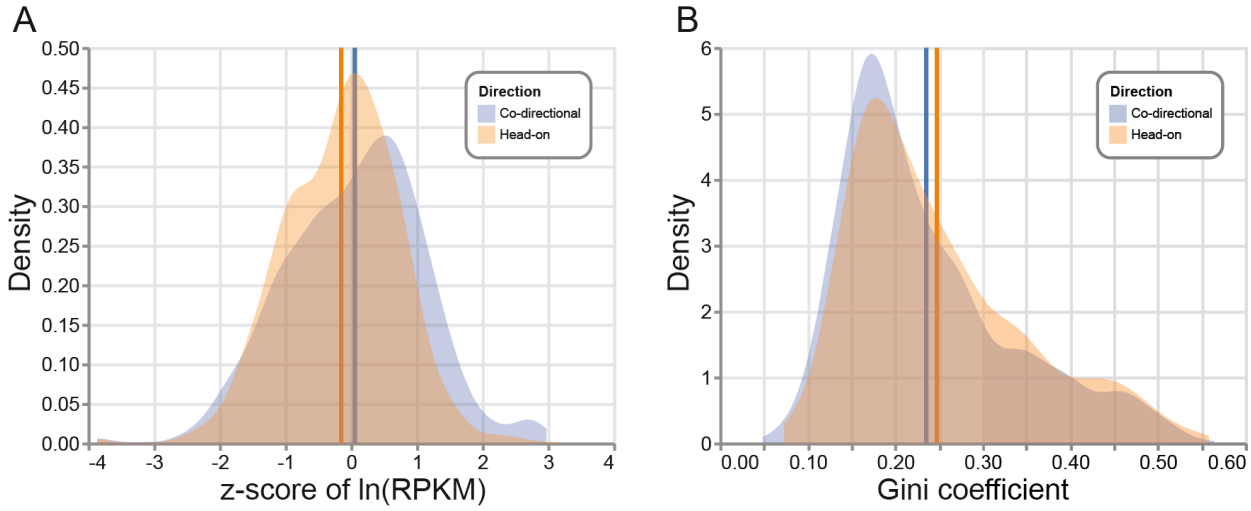


Fig 2. Head-on and co-directional genes have similar expression properties. Distributions of two metrics relating to gene expression in *B. subtilis* transcriptomic datasets. Distributions were subset by direction of transcription relative to DNA replication. Vertical lines represent the mean for each class of genes. (A) The distribution of gene expression values for coding sequences in *B. subtilis* strain PY79. The z-score was calculated for each coding sequence's natural-logarithm-transformed RPKM values prior to subsetting by direction. RNA-seq data are from [22]. (B) Transcriptomic microarray data [53] were analyzed. Each gene's Gini coefficient was calculated over 104 conditions.