Replication-transcription conflicts promote adaptive evolution

Christopher N. Merrikh[1]*, Leonard A. Harris[1], Sarah Mangiameli[2], and Houra Merrikh[1]*

[1]Department of Biochemistry, Vanderbilt University
[2]Broad Institute

*Co-corresponding author
E-mail: christopher.merrikh@vanderbilt.edu
E-mail: houra.merrikh@vanderbilt.edu

Abstract

In 2013, we discovered that encounters between the replication and transcription machineries allow bacteria to evolve at an accelerated rate by promoting mutagenesis in lagging strand genes. Though we proposed that this process is adaptive, it is also possible that the increased mutation frequency in lagging strand genes could be the result of reduced purifying selection (neutral selection).Due to the low number of available genome sequences at the time of publication, we were unable to distinguish between these two models with a high level of confidence. Here, we utilized the wealth of newly available bacterial genome sequences to examine these two possibilities. To test the adaptive hypothesis, we analyzed convergent mutation patterns. To test the neutral hypothesis, we performed *in silico* modeling. Our results clearly demonstrate that the neutral model cannot explain the increased mutagenesis of lagging strand genes. Additionally, our evolutionary convergence data strongly support the adaptive hypothesis. We conclude that encounters between the replication and transcription machineries in lagging strand gene accelerate the discovery of beneficial mutations.

Introduction

Replication-transcription conflicts cause pervasive replisome collapse across highly divergent bacterial species[1,2]. Conflicts also occur in eukaryotes including yeast, and humans[3–6]. Hence, replication-transcription conflicts are a universal problem.

In bacteria, the majority of genes are encoded on the leading strand. A gene's orientation (encoding on the leading or lagging strand) can have major effects on cellular fitness through co-directional or head-on replication-transcription conflicts, respectively[7–11]. The two types of conflicts lead to very different outcomes. In leading strand genes, co-directional conflicts between the faster moving replisome and RNA polymerases have a modest effect on replisome stalling[12]. Conversely, in lagging strand genes, head-on conflicts can cause severe replisome stalling and DNA damage[10,11,13,14]. As a result, head-on conflicts almost certainly confer strong negative selection pressure against the majority of lagging strand alleles[8,15]. Yet in every bacterial species, 20-45% of genes remain in the lagging strand orientation[16–18]. The reasons for this phenomenon are not entirely clear.

In 2013, we made a key discovery that could at least partially explain the retention of lagging strand genes despite the detrimental outcomes of head-on conflicts. We found that lagging strand genes evolve at a faster rate than leading strand genes in *Bacillus subtilis*, both experimentally and in nature. Our findings were the first to demonstrate the existence of a temporal (transcription-dependent) and spatial (gene-specific) control mechanism for promoting evolution. Some of our results also suggested that this mechanism may be adaptive. However, because our analyses were performed using only 5 genomes, the support we presented for the adaptive hypothesis was somewhat weak. Consequently, reduced negative selection remained a valid alternative to the adaptive hypothesis.

Following our study, Chen and Zhang claimed that no support exists for the adaptive hypothesis[15]. Their argument was based on two primary forms of evidence: 1) *in silico*

simulations showing that some of the convergent mutations we identified, multi-hit site mutations (defined in Fig. 1A), could have formed by chance. They therefore accepted only the remaining parallel mutations as convergent. Though the authors ultimately found that parallel mutations are also more common in lagging strand genes, they rejected the validity of the adaptive hypothesis on the basis of a non-significant difference between the two groups (p=0.14, Figure S1). Though the issue of statistical significance could have been resolved through an investigation of additional bacterial genomes, none were available at the time.

Today many more bacterial genomes have been sequenced, allowing us to test the two models with high resolution. To this end, we conducted a new analysis of convergent evolution using 50 *B. subtilis* genomes. This increased our statistical power, and allowed us to identify extremely rare classical convergent mutations which are among the most reliable forms of evidence of evolutionary convergence (Fig 1A). Below, our new analyses show that indeed, each type of convergent mutation is more common in lagging strand genes relative to leading strand genes. We further show that their abundance is both gene length and orientation-dependent, supporting our original model that lagging strand convergent mutations arose largely through mutagenesis caused by head-on replication-transcription conflicts. Our *in silico* modeling experiments further demonstrate that the observed mutations cannot be explained by chance alone, overturning the neutral hypothesis. Finally, we repeat these analyses in two other species, and observe the same patterns of convergence. As such, we conclude that lagging strand encoding is a broadly conserved mechanism capable of benefiting the cell through head-on conflict-mediated mutagenesis and the accelerated discovery of adaptive mutations.

Materials and Methods

Bacterial genome files in were downloaded from NCBI in Genbank format and are listed in Table S1. Genomic sequences were analyzed using the program TimeZone version 1.0[19]. TimeZone conducts mutational analyses (e.g. dN, dS, dN/dS, etc.) of a subset of genes based upon only two customizable settings: the level of gene conservation in terms of 1) length and 2) amino acid content. We selected 95% conservation for both settings meaning that only genes that are highly conserved in terms of both length and amino acid content were analyzed. These are referred to as core genes, which are effectively essential in nature[20]. Core genes represent an equal fraction of leading and lagging strand genes, suggesting that the results of our analyses should be equally representative of both groups[21]. This may also help control for differences in essentiality/mutability that might otherwise, potentially, skew our results. Convergent mutations were parsed from TimeZone output files using a custom script Analysis_conv5.py.

Simulations were performed as previously described, with one exception: all simulated multi-hit sites were mutated to a second amino acid[15]. As a result, a subset of our multi-hit site mutations were identical, and thus represent parallel mutations. All code used for these simulations are publicly available at https://github.com/lh64/MultihitSimulation. All additional data is available upon request.

Results

To detect signatures of convergent evolution in *B. subtilis*, we conducted a mutational analysis of 50 fully assembled genomes. This data set provided an order of magnitude more information than was used in our original analysis. To identify all point mutations that occurred since the divergence of the individual isolates, we used the program TimeZone[19]. This program is optimized for identifying recent adaptive changes in bacteria[19]. We then parsed out the convergent mutations identified by TimeZone into three groups: classical convergent, parallel, or multi-hit site mutations, and calculated their frequencies in leading or lagging strand core genes (Figure 1A, 1B). Our analysis identified a higher frequency of all three types of convergent mutations in lagging strand genes (Fig 1B). These results are highly statistically significant, strongly supporting the validity of our original analysis, as well as our inference that positive selection acts more frequently on lagging strand genes.

To determine if our findings in *B. subtilis* are indicative of a broader pattern of evolution, we conducted the same analysis in a related species, *Mycoplasma gallisepticum*, and an unrelated species from a second phylum, *Mycobacterium tuberculosis* (Fig. 1C). For each species, we observed the same trend identified in *B. subtilis*. This strongly suggests that the elevated frequency of convergent evolution is a conserved feature of lagging strand gene evolution.

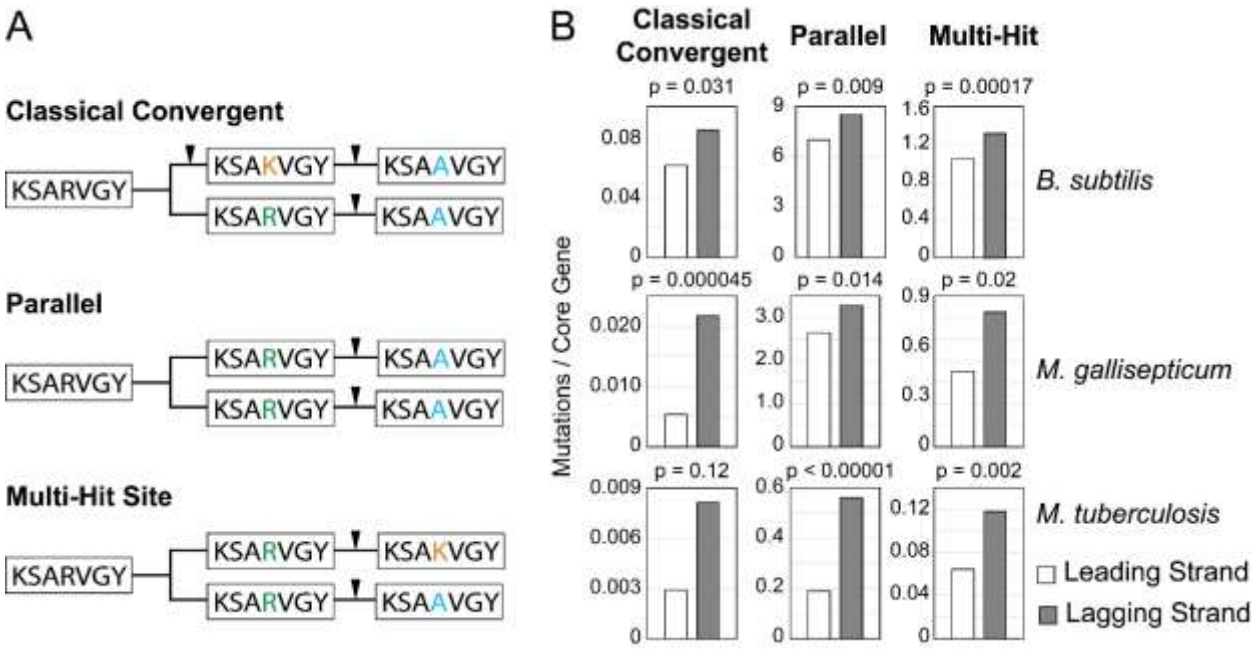


Figure 1. All three types of convergent mutations are more frequently observed in lagging strand genes. A) Three examples of convergent mutations. In each example, the amino acid sequence of a hypothetical gene fragment is shown for five species in a bifurcating lineage that evolves over time from left (older) to right (newer). Black arrows indicate the development of a nonsynonymous mutation in the displayed fragment of the example gene. B) The frequency of observed convergent mutations in either leading or

lagging strand core genes of three species. The Chi square test was used to calculate significance.

Most convergent mutations are retained due to positive selection, not chance.

Classical convergent and parallel mutations are considered standard indicators of adaptation because they are extremely unlikely to occur by chance[15,22]. However, it is theoretically possible that multi-hit site mutations could arise at random with significant frequency[15]. Accordingly, they represent a lower confidence indicator of convergence. To determine if the multi-hit sites we observed could have been produced by chance, we conducted *in silico* simulations, following the protocol established by Chen and Zhang[15]. Their method simulates a neutral reassortment of the observed non-synonymous mutations in each gene to estimate the frequency with which non-adaptive multi-hit site mutations could arise by chance. Accordingly, we drew random variable sites within each core leading or lagging strand gene with replacement until the total number of sites equaled the number of observed amino acid changes[15]. In our method, we added a second step: for any site drawn twice, the original residue was randomly mutated to one of the 19 other amino acids, yielding either a multi-hit or parallel mutation. We performed simulations for all leading and lagging strand core genes using nonsynonymous substitution data from either our original 5 genome or new 50 genome study of *B. subtilis*. We repeated these simulations 10,000 times for all leading or lagging strand core genes, yielding a distribution of values (Fig. 2). We found that the observed number of both parallel and multi-hit site mutations are greatly in excess of even the most extreme simulated data (Fig. 2). Therefore, the data strongly suggest that the multi-hit site and parallel mutations we observed in both our original and current studies could not have arisen by chance (Fig. 2). Instead, they most likely arose through positive selection, consistent with the idea that both multi-hit site and parallel mutations are indicative of evolutionary convergence.

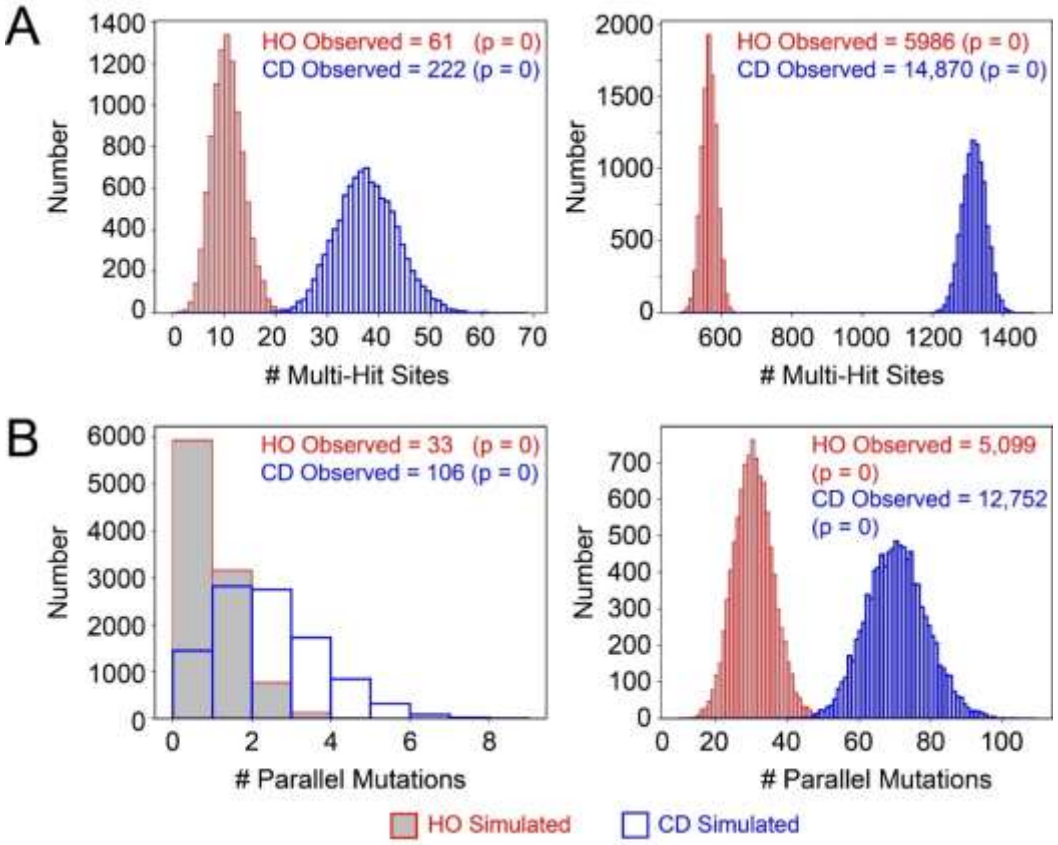


Figure 2. Observed convergent mutation frequencies cannot be explained by chance. A) The distribution of multi-hit site mutations expected under neutral selection conditions was determined by simulating a random reassortment of the observed nonsynonymous substitutions in leading strand (blue) or lagging strand genes (red) based upon the numbers identified in either our previous 5 genome analysis of *B. subtilis* (left graph), or our 50 genome analysis (right graph). This was performed for all leading or lagging strand genes x 10,000 iterations, providing a distribution for both groups. The actual observed numbers are shown above the simulated distributions. B) Lower graphs: Same as above, but for parallel mutations. See Figure S2 for the same data graphed using the previously used ratio-of-ratios method[15].

Despite having conducted an ostensibly identical simulation as Chen and Zhang, our results clearly and directly oppose their conclusion that the "observation on multi-hit site containing genes is fully expected under the neutral model"[15]. To determine how the same experiment could support opposing conclusions, we re-analyzed our raw simulation data (presented in Figure 2) using the ratio-of-ratios method employed by Chen and Zhang. Here we calculated the ratio of observed multi-hit site mutations in leading strand genes to those in lagging strand genes, and compared this number to the distribution of simulated ratios (Fig. S2). As with the previous report, we found that the observed ratio is roughly in the middle of the distribution of simulated ratios (Fig. S2, Chen and Zhang's Fig. 2). First, this shows that both simulations, and presumably the underlying data sets, are equivalent.

(Unfortunately, Chen and Zhang did not publish their raw data or computer code, and did not respond to our request for this information so this conclusion is inferred.) Second, this indicates that the authors' analysis method is the root of the conflict: Even though the raw data shows conclusively that the observed multi-hit site mutations cannot be explained by chance, the ratio-of-ratios method obfuscates this information, erroneously appearing to suggest the opposite. In fact, Chen and Zhang's method makes even parallel mutations appear to be random events which is exceptionally unlikely as discussed by the authors themselves (Supplementary Figure 3)[15]. Therefore, we conclude that the ratio-of-ratios data presented by Chen and Zhang is highly misleading, and that the conclusions of their manuscript are directly opposed by their own data. As such, the modeling data strongly support the adaptive hypothesis for lagging strand gene encoding.

Head-on replication-transcription conflicts increase the frequency of convergent mutations development in lagging strand genes

Our previous work indicated that head-on replication-transcription conflicts are the mechanistic basis for the increased mutation rate of head-on genes[7,21,23]. Evidence for this hypothesis includes our observation of a gene length and orientation-dependent increase in both dN/dS ratios and mutation frequency[7]. This result was consistent with the idea that head-on conflict severity should increase in direct relation to gene length, whereas co-directional conflicts (leading strand genes) should not[7]. If head-on replication-transcription conflicts are responsible for promoting the formation of the observed convergent mutations identified here, their abundance should follow the same pattern. To test this, we calculated the number of convergent mutations per core gene, then assessed the relationship with gene length (Fig. 3). We found that all three types of convergent mutations increase in frequency in a gene length-dependent manner. We also found that this effect is more pronounced in lagging strand genes, strongly suggesting that head-on conflicts are indeed responsible for the increased frequency of convergent mutations in lagging strand genes.

We then repeated these analyses in *M. gallisepticum* and *M. tuberculosis* and identified the same pattern, demonstrating that this mechanism is conserved in other species (Fig. S3). (Raw data for all three species are shown in Fig. S4.)

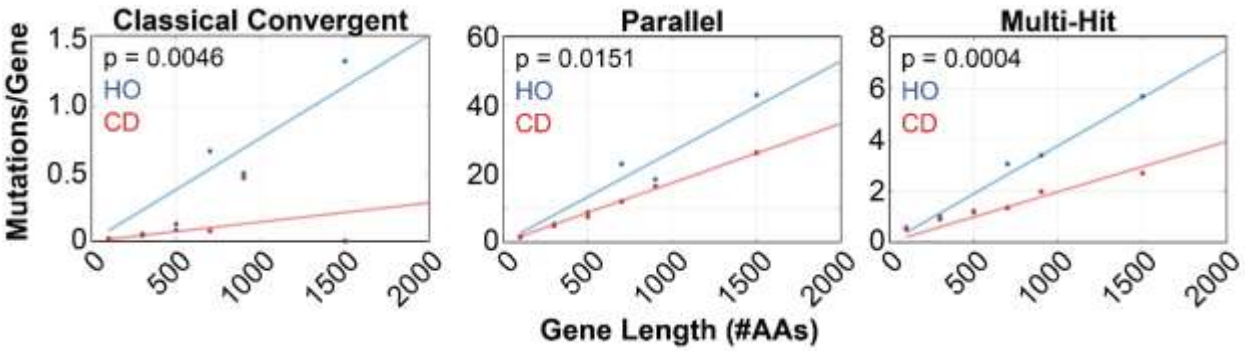


Figure 3. The frequency of convergent mutations increases in a gene length and orientation-dependent manner. The number of convergent was calculated for genes of varying length. The convergent mutation type is indicated over each graph. Genes were binned into 6 groups based upon their length in amino acids: 1) <200 AA, 2) 201-400 AA, 3) 401-600 AA, 4) 601-800 AA, 5) 801-1000 AA, 6) 1001-2000 AA. Statistical significance was calculated using the F-test (p-values are indicated in the top left). Data for *M. gallisepticum* and *M. tuberculosis* are presented in Fig. S3. Raw data for all three species are presented in Fig. S4.

Discussion

In summary, our new analyses show that lagging strand genes accumulate convergent mutations at a higher rate than leading strand genes, and that this effect is broadly conserved. Importantly, our modeling experiments demonstrate that the observed parallel and multi-hit site mutations cannot be explained by chance. This overturns the neutral hypothesis, strongly suggesting that these mutations are beneficial. Even if one disagrees with the idea that multi-hit site mutations are indicative of evolutionary convergence, the higher frequency of both parallel and classical convergent mutations in lagging strand genes is powerful and independent evidence that the lagging strand encoding can be adaptive. The gene length and orientation-dependence of these mutation patterns strongly suggests head-on replication-transcription conflicts are a key molecular mechanism driving their formation.

The results we present here are consistent with those of our previous study in which we identified a higher frequency of genes with a dN/dS ratio significantly above 1 encoded on the lagging strand six diverse species[21]. This metric represents a second standard indicator of positive selection[24]. In the same study, we identified pervasive, recent, leading-to-lagging strand gene inversion events in every bacterial species tested. (Though our method could not detect "reversion" events in which newly inverted head-on gene flipped back to the leading strand, such events almost certainly occur as part of a dynamic equilibrium.) Together, these observations provided strong support for the notion that lagging strand encoding can be beneficial.

Interestingly, both our new empirical data (Fig. 1B) and modeling experiments (Fig. 2) directly contradict the conclusions of the Chen and Zhang manuscript in which they claimed to have overturned the adaptive hypothesis[15]. In fact, this is the second time we have discovered that a critique of our work by the Zhang lab was based almost entirely on inaccurate, erroneous, or misleading data[25].

In summary, the convergent evolution analysis presented here, together with our prior investigations of gene-specific mutation rates show that lagging strand genes gain adaptive mutations at a faster rate than leading strand genes in species across phyla. In combination with our identification of pervasive leading-to-lagging strand gene inversions, our findings paint a consistent picture: lagging strand encoding for some genes can confer a net benefit to the cell through head-on replication-transcription conflicts, despite their detrimental

effects. As such, we conclude that gene orientation provides the cell with a high precision mechanism for temporally and spatially controlling adaptive evolution.

Acknowledgements

H.M. and C.M. were supported by the National Institute of Health, R01-AI-127422.

References

1. Pace, N. R., Olsen, G. J. & Woese, C. R. Ribosomal RNA phylogeny and the primary lines of evolutionary descent. *Cell* **45**, 325–326 (1986).
2. Danchin, A. Comparison Between the Escherichia coli and Bacillus subtilis Genomes Suggests That a Major Function of Polynucleotide Phosphorylase is to Synthesize CDP. *DNA Res.* **4**, 9–18 (1997).
3. Prado, F. & Aguilera, A. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO J.* **24**, 1267–1276 (2005).
4. Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T. & Cimprich, K. A. Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell* **170**, 774-786.e19 (2017).
5. Mirkin, E. V. & Mirkin, S. M. Mechanisms of transcription-replication collisions in bacteria. *Mol. Cell. Biol.* **25**, 888–895 (2005).
6. Hamperl, S. & Cimprich, K. A. The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA Repair* **19**, 84–94 (2014).
7. Paul, S., Million-Weaver, S., Chattopadhyay, S., Sokurenko, E. & Merrikh, H. Accelerated gene evolution through replication-transcription conflicts. *Nature* **495**, 512–515 (2013).
8. Merrikh, H., Zhang, Y., Grossman, A. D. & Wang, J. D. Replication-transcription conflicts in bacteria. *Nat. Rev. Microbiol.* **10**, 449–458 (2012).

9. Lang, K. S. *et al.* Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. *Cell* **170**, 787-799.e18 (2017).

10. Merrikh, C. N., Brewer, B. J. & Merrikh, H. The B. subtilis Accessory Helicase PcrA Facilitates DNA Replication through Transcription Units. *PLoS Genet.* **11**, e1005289 (2015).

11. Boubakri, H., de Septenville, A. L., Viguera, E. & Michel, B. The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo. *EMBO J.* **29**, 145–157 (2010).

12. Merrikh, H., Machón, C., Grainger, W. H., Grossman, A. D. & Soultanas, P. Co-directional replication-transcription conflicts lead to replication restart. *Nature* **470**, 554–557 (2011).

13. Mangiameli, S. M., Merrikh, C. N., Wiggins, P. A. & Merrikh, H. Transcription leads to pervasive replisome instability in bacteria. *eLife* **6**, (2017).

14. Million-Weaver, S., Samadpour, A. N. & Merrikh, H. Replication Restart after Replication-Transcription Conflicts Requires RecA in Bacillus subtilis. *J. Bacteriol.* **197**, 2374–2382 (2015).

15. Chen, X. & Zhang, J. Why are genes encoded on the lagging strand of the bacterial genome? *Genome Biol. Evol.* **5**, 2436–2439 (2013).

16. Kobayashi, K. *et al.* Essential Bacillus subtilis genes. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 4678–4683 (2003).

17. Rocha, E. P. C. & Danchin, A. Ongoing Evolution of Strand Composition in Bacterial Genomes. *Mol. Biol. Evol.* **18**, 1789–1799 (2001).

18. Lobry, J. R. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* **13**, 660–665 (1996).

19. Chattopadhyay, S., Paul, S., Dykhuizen, D. E. & Sokurenko, E. V. Tracking recent adaptive evolution in microbial species using TimeZone. *Nat. Protoc.* **8**, 652–665 (2013).

20. Paul, S., Bhardwaj, A., Bag, S. K., Sokurenko, E. V. & Chattopadhyay, S. PanCoreGen - Profiling, detecting, annotating protein-coding genes in microbial genomes. *Genomics* **106**, 367–372 (2015).

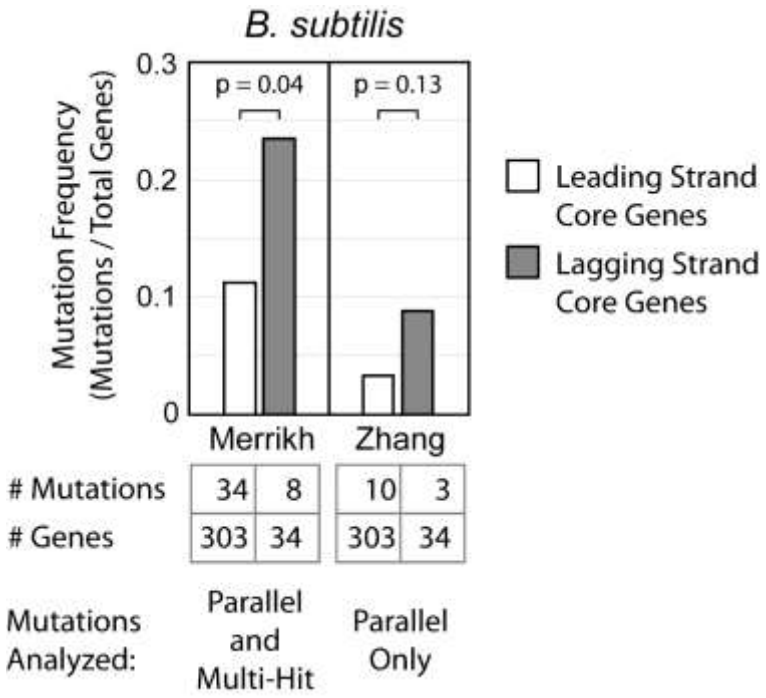
21. Merrikh, C. N. & Merrikh, H. Gene inversion potentiates bacterial evolvability and virulence. *Nat. Commun.* **9**, 4662 (2018).

22. Christin, P.-A., Weinreich, D. M. & Besnard, G. Causes and evolutionary significance of genetic convergence. *Trends Genet.* **26**, 400–405 (2010).

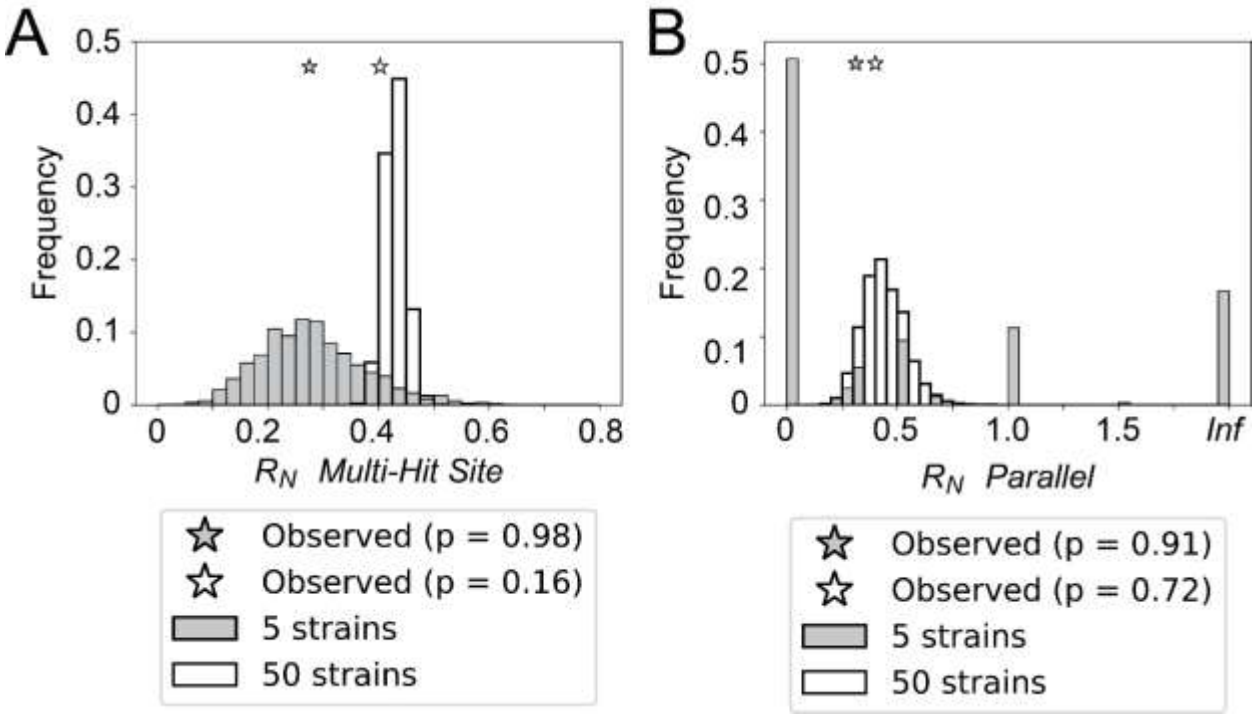
23. Million-Weaver, S. *et al.* An underlying mechanism for the increased mutagenesis of lagging-strand genes in Bacillus subtilis. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E1096-1105 (2015).

24. Kryazhimskiy, S. & Plotkin, J. B. The population genetics of dN/dS. *PLoS Genet.* **4**, e1000304 (2008).

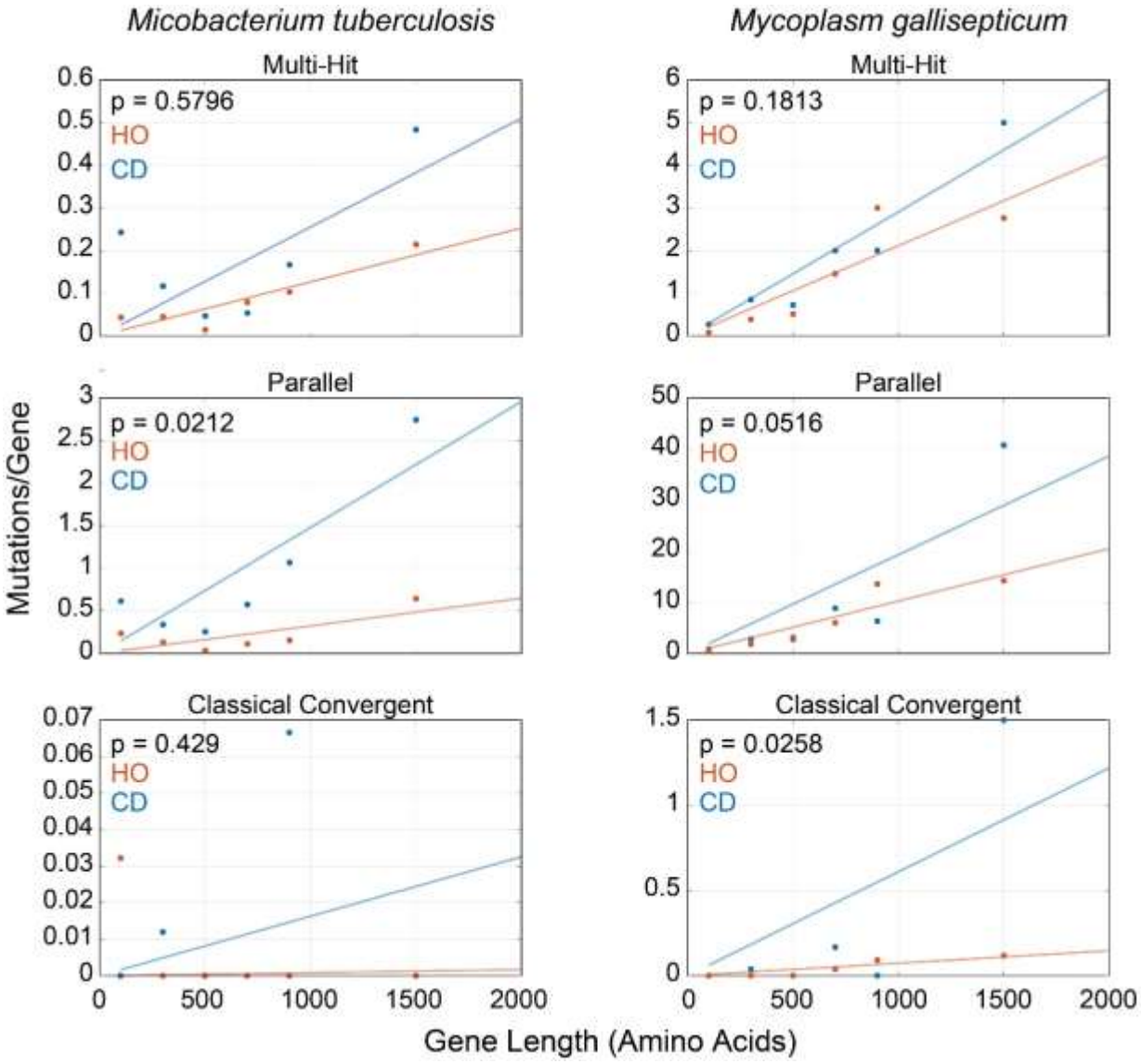
25. Merrikh, C. & Merrikh, H. The recent report by Liu and Zhang Claiming “No support for the adaptive hypothesis of lagging-strand encoding in bacterial genomes” is factually incorrect. *bioRxiv* 2020.05.26.117366 (2020) doi:10.1101/2020.05.26.117366.



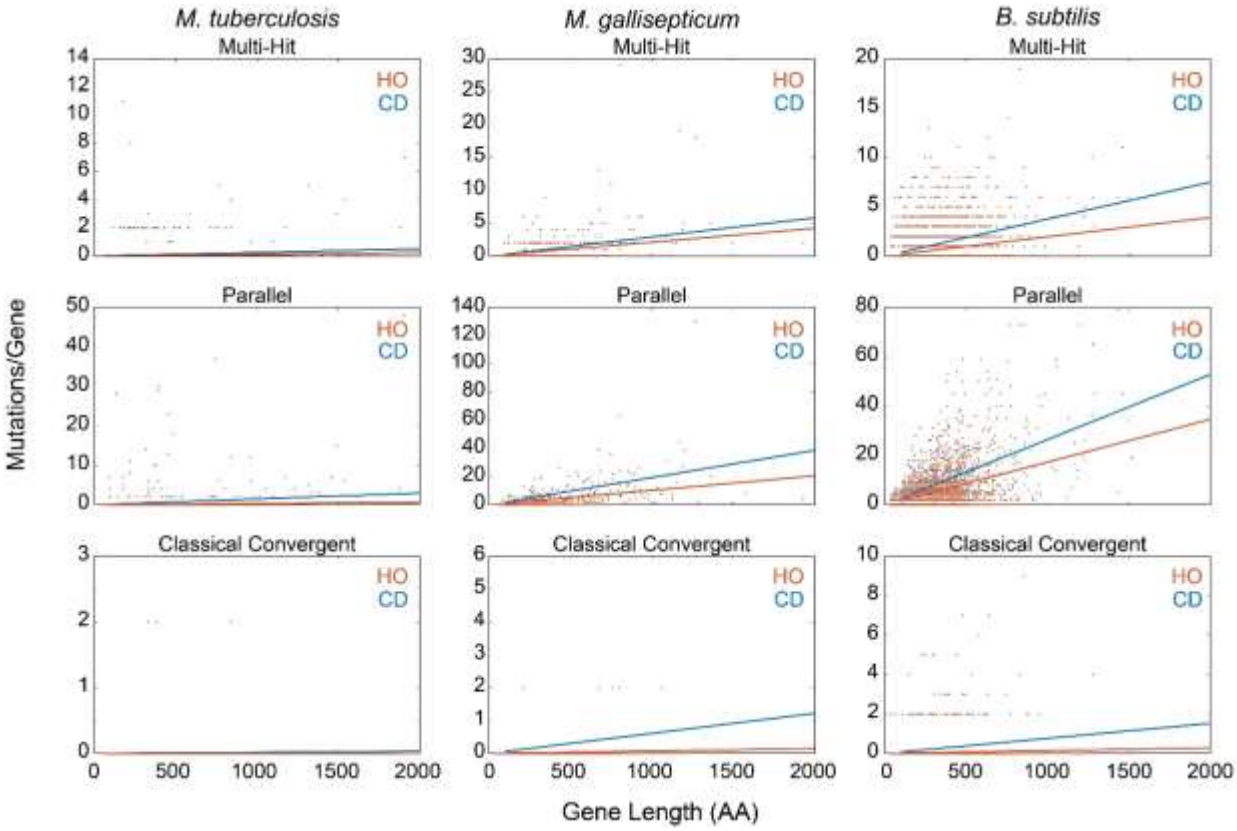
Supplemental Fig. 1. The relative frequency of convergent mutations in leading or lagging strand genes. Convergent mutation data from 5 *B. subtilis* isolates are adapted from Paul et al. Table 1, or Chen and Zhang (data are shown only in the text, page 2437, right column). Our original analysis of pooled multi-hit and parallel mutations (left side) indicated a higher frequency of convergent mutations in lagging strand core genes. This was contested by the Zhang analysis which accepted the validity of only classical convergent and parallel mutations as indicators of convergence (right side). Note that no classical convergent mutations were identified in the original data set.



Supplemental Figure 2. The ratio-of-ratios analysis method used by Chen and Zhang yields highly misleading results. Raw simulation data (presented in Fig. 2) was re-analyzed using the ratio-of-ratios method employed by Chen and Zhang (the "R_N")[15]. The R_N is a ratio representing the number of simulated lagging strand mutations divided by the number of simulated leading strand mutations. This was performed for A) multi-hit site mutations, or B) parallel mutations. We calculated the R_N distributions based upon observed non-synonymous mutation data from our original 5 genome study data (gray) or our 50 genome study data (white). In all cases, the observed ratios (stars) are not different than the simulated ratio distributions, as previously reported[15]. Critically, the likelihood that a parallel mutation could occur by chance is extremely low[15]. Therefore, parallel mutation data essentially represents a positive control:If the ratio-of-ratios method works, it should show that the observed parallel mutation ratio is significantly higher than the simulated ratio.. As shown in white in graph B, the R_N calculation fails this test by suggesting that parallel mutations could have arisen by chance. This demonstrates that the method used by Chen and Zhang is extraordinarily insensitive at best, and should not be used.



Supplemental Figure 3. The gene length and orientation-dependence of convergent mutation formation is conserved in *M. tuberculosis* and *M. gallisepticum*. Similar to our findings in *B. subtilis* (Fig. 3), there is a direct correlation between the frequency of observed convergent mutations and gene length. This correlation is stronger in lagging strand genes than leading strand genes, strongly suggesting that head-on replication-transcription conflicts drive the formation of all three types of convergent mutations. Statistical significance was determined using the F-test. All raw data are presented in Figure 4.



Supplemental Figure 4. Raw data for Figure 3, and Figure S3. The frequency of the indicated type of convergent mutation was determined for all genes. The frequency for each gene was plotted individually according to its length in amino acids. Trend lines are the same as those shown in Figure 3 and Figure S3.

B. subtilis	*M. gallisepticum*	*M. tuberculosis*
GCA_000009045.1	GCA_000092585.1	GCF_000195955.2
GCA_003148415.1	GCA_900476085.1	GCF_000193185.2
GCA_000789275.1	GCA_000025365.1	GCF_000277735.2
GCA_000344745.1	GCA_008728895.1	GCF_000331445.1
GCA_002055965.1	GCA_008728935.1	GCF_000400615.1
GCA_006088795.1	GCA_000286715.1	GCF_000572195.1
GCA_006741845.1	GCA_000211545.6	GCF_000738445.1
GCA_000227465.1	GCA_000025385.1	GCF_000738475.1
GCA_011456075.1	GCA_000286795.1	GCF_000786505.1
GCA_003665215.1	GCA_004771115.1	GCF_000828995.1
GCA_002163815.1	GCA_004771095.1	GCF_000934325.3
GCA_009913275.1	GCA_000286735.1	GCF_001078615.1
GCA_009662175.1	GCA_000286675.1	GCF_001275565.2
GCA_000953615.1	GCA_008728915.1	GCF_001483905.1
GCA_009914705.1	GCA_000286695.1	GCF_001544705.1
GCA_001704095.1	GCA_000286755.1	GCF_001597645.2
GCA_004119775.1	GCA_000286775.1	GCF_001708265.1
GCA_009662195.1	GCA_000286815.1	GCF_001870145.1
GCA_009662255.1	GCA_000428645.1	GCF_001922485.1
GCA_009363835.1	GCA_001676495.1	GCF_001938725.1
GCA_004328925.1	GCA_001676575.1	GCF_002072775.2
GCA_000971925.1	GCA_001676505.1	GCF_002116755.1
GCA_009662435.1	GCA_001683635.1	GCF_002116775.1
GCA_009662415.1	GCA_001705745.1	GCF_002116795.1
GCA_000789295.1	GCA_001676515.1	GCF_002116835.1
GCA_001703495.1	GCA_001683675.1	GCF_002208235.1
GCA_003610955.1	GCA_001676525.1	GCF_002357955.1
GCA_001697265.1		GCF_002357975.1
GCA_009913515.1		GCF_002446915.1
GCA_009866865.1		GCF_002446955.1
GCA_004101365.1		GCF_002447015.1
GCA_003665195.1		GCF_002447035.1
GCA_003665255.1		GCF_002447055.1
GCA_003665395.1		GCF_002447195.1
GCA_002269195.1		GCF_002447215.1
GCA_001604995.1		GCF_002447235.1
GCA_000827065.1		GCF_002447255.1
GCA_000699525.1		GCF_002447275.1
GCA_009913535.1		GCF_002447335.1
GCA_004103535.1		GCF_002447355.1
GCA_003665275.1		GCF_002447375.1
GCA_009662215.1		GCF_002447415.1

GCA_002173615.1		GCF_002447455.1
GCA_000699465.1		GCF_002447475.1
GCA_001747445.1		GCF_002447515.1
GCA_003665235.1		GCF_002447575.1
GCA_009497815.1		GCF_002447615.1
GCA_002893805.1		GCF_002447635.1
GCA_009662275.1		GCF_002447655.1
		GCF_002447675.1
		GCF_002447695.1
		GCF_002447715.1
		GCF_002447735.1
		GCF_002447755.1
		GCF_002447775.1
		GCF_002447795.1
		GCF_002447835.1
		GCF_002447855.1
		GCF_002447895.1
		GCF_002447915.1
		GCF_002447975.1
		GCF_002447995.1
		GCF_002448015.1
		GCF_002448055.1
		GCF_002448095.1
		GCF_002448115.1
		GCF_002448135.1
		GCF_002448155.1
		GCF_002448215.1
		GCF_002886145.1
		GCF_002886165.1
		GCF_002886195.1
		GCF_002886225.1
		GCF_002886335.1
		GCF_002886405.1
		GCF_002886505.1
		GCF_002886585.1
		GCF_002886685.1
		GCF_002886775.1
		GCF_002886865.1
		GCF_002886945.1
		GCF_002887065.1
		GCF_002887145.1
		GCF_002887255.1
		GCF_002887335.1

		GCF_002975475.1 GCF_003006115.1 GCF_003006135.1 GCF_003265005.1 GCF_003287125.1 GCF_003287145.1 GCF_003287165.1 GCF_005155785.1 GCF_005156105.1 GCF_007833935.1 GCF_008761675.1 GCF_009664875.1 GCF_009730215.1 GCF_009730235.1 GCF_902459825.1

Table S1. Genomes used for mutational analyses. The NCBI Genome assembly serial numbers for all genomes analyzed by TimeZone are listed. The Genbank formatted genomes were downloaded and used as input.