


Divergent Evolution of Mutation Rates and Biases in the Long-Term Evolution Experiment with *Escherichia coli*

Rohan Maddamsetti ^{1,*} and Nkrumah A. Grant^{2,3,4}

¹Department of Biomedical Engineering, Duke University

²BEACON Center for the Study of Evolution in Action, Michigan State University

³Department of Microbiology and Molecular Genetics, Michigan State University

⁴Program in Ecology, Evolutionary Biology and Behavior, Michigan State University

*Corresponding author: E-mail: rohan.maddamsetti@gmail.com.

Accepted: 21 August 2020

Abstract

All organisms encode enzymes that replicate, maintain, pack, recombine, and repair their genetic material. For this reason, mutation rates and biases also evolve by mutation, variation, and natural selection. By examining metagenomic time series of the Lenski long-term evolution experiment (LTEE) with *Escherichia coli* (Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. 2017. The dynamics of molecular evolution over 60,000 generations. *Nature* 551(7678):45–50.), we find that local mutation rate variation has evolved during the LTEE. Each LTEE population has evolved idiosyncratic differences in their rates of point mutations, indels, and mobile element insertions, due to the fixation of various hypermutator and antimutator alleles. One LTEE population, called Ara+3, shows a strong, symmetric wave pattern in its density of point mutations, radiating from the origin of replication. This pattern is largely missing from the other LTEE populations, most of which evolved missense, indel, or structural mutations in *topA*, *fis*, and *dusB*—loci that all affect DNA topology. The distribution of mutations in those genes over time suggests epistasis and historical contingency in the evolution of DNA topology, which may have in turn affected local mutation rates. Overall, the replicate populations of the LTEE have largely diverged in their mutation rates and biases, even though they have adapted to identical abiotic conditions.

Key words: experimental evolution, metagenomics, mutation.

Significance

Bacteria often evolve elevated mutation rates during adaptation to challenging environments. Less is known about how mutation rates vary over the chromosome, and how those local biases evolve during adaptive evolution. To answer this question, we analyzed metagenomic data from an ongoing experiment with *Escherichia coli* in which 12 replicate populations of bacteria, started from a single clonal strain in 1988, were allowed to evolve for more than 30 years. We find that each replicate population has a different genomic distribution of observed mutations, indicating that local mutation rates have evolved idiosyncratically, even though each population has adapted to the same laboratory conditions. Intriguingly, our results indicate that adaptive mutations that change DNA topology may also affect local mutation rates.

Introduction

Loci that modify DNA repair and recombination modify the evolutionary process. Therefore, one might ask whether natural selection adaptively tunes mutation and recombination

rates. This idea—that second-order selection *adaptively* modifies the evolutionary process itself—is debated (Tenaillon et al. 2001; Lynch et al. 2016). Nonetheless, populations of *Escherichia coli*, engineered to have constitutive sexual recombination and elevated mutation rates, adapt faster than

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

control populations in the laboratory (Cooper 2007; Peabody et al. 2016, 2017).

To study second-order selection on mutation rates, one can use experimental evolution. By running experiments in which replicate populations evolve under controlled conditions, with different starting mutation rates, one can ask whether particular mutation rates are favored over others (Chao et al. 1983; Loh et al. 2010; Sprouffske et al. 2018). Here, we use metagenomic time series data from the Lenski long-term evolution experiment (LTEE) with *E. coli* to study how mutation rates evolve in real time.

In the LTEE, 12 populations of *E. coli*, descended from a common ancestral strain, have adapted for more than 73,000 generations to carbon-limited minimal media. Six of the populations are labeled Ara+, whereas the other six are labeled Ara–, based on the presence or absence of an evolutionarily neutral arabinose marker (Lenski et al. 1991). The LTEE populations are strictly asexual. Some populations have evolved defects in DNA repair which vastly increase their point mutation rates. The causative hypermutator alleles likely went to fixation by linkage with highly beneficial mutations, rather than being beneficial per se (Sniegowski et al. 1997; Tenaillon et al. 2016). We refer to the LTEE populations that have evolved large increases in point mutation rates as “hypermutator populations,” and refer to the others as “nonmutator populations.”

Molecular evolution in the hypermutator populations of the LTEE is dominated by “genetic draft,” in which large numbers of nearly neutral passenger mutations hitchhike with a small number of beneficial driver mutations (Neher 2013). This phenomenon has obscured the genomic signatures of adaptation in those populations (Tenaillon et al. 2016; Couce et al. 2017; Good et al. 2017; Maddamsetti et al. 2017). In this regime, also called “emergent neutrality” (Schiffels et al. 2011), the evolutionary dynamics inferred from whole-population samples of the hypermutator populations (Good et al. 2017) provides good data on mutation rates and biases, even though natural selection drives the dynamics. Here, we examined LTEE metagenomics data (Good et al. 2017) for mutation rate variation and biases over the chromosome (Foster et al. 2013; Paul et al. 2013; Jee et al. 2016; Niccum et al. 2019).

Results

Cumulative Number of Observed Mutations in Each Population Reveals Dynamics Caused by Both Hypermutator and Antimutator Alleles

We examined the number of observed mutations over time in each LTEE population (figs. 1 and 2, supplementary figs. S1–S3, [Supplementary Material](#) online). These results show that mutation rates have evolved idiosyncratically over the LTEE. Figure 1A shows the number of point mutations over time in

each population. The rate of observed point mutations decreased in three of the six hypermutator populations (Ara–2, Ara+3, and Ara+6). The decrease in the rate of molecular evolution in these populations was previously ascribed to the evolution of antimutator alleles (Tenaillon et al. 2016; Good et al. 2017). Although antimutator alleles of *mutY* compensating for defects in *mutT* have been reported in Ara–1 (Wielgoss et al. 2013), the change in slope observed at 40,000 generations in Ara–1 is subtle compared with the slope changes in Ara–2, Ara+3, and Ara+6.

Figure 1B shows the number of observed indel mutations over time in each population. Five of the six point-mutation hypermutator populations also show an indel hypermutator phenotype. These five populations all evolved defects in mismatch repair (MMR) ([table 1](#) and [fig. 4](#)). The exception is Ara–1, which evolved a frameshift *mutT* allele ([table 1](#) and [fig. 3](#)) that induces a high point mutation rate, absent a corresponding indel hypermutator phenotype.

The hypermutator dynamics in Ara–2 are particularly striking. An antimutator allele eventually fixes, and reverts both the point and indel hypermutator phenotype back to ancestral or near ancestral levels ([fig. 1A](#) and [B](#)). The hypermutator phenotype is caused by phase variation of a (TGGCGC)₃ repeat in *mutL* ([table 1](#)). Reversions to the triplet state reverse the hypermutator phenotype. The number of new point and indel mutations in Ara–2 ([supplementary figs. S1 and S2, Supplementary Material](#) online) fluctuates with the allele frequency dynamics of this *mutL* repeat ([fig. 4](#)). Although fixations are usually irreversible in large asexual populations, phase variation is an exception: polymerases often slip on repetitive sequences, causing those repeats to expand or contract at relatively high rates (Moxon et al. 2006).

At first glance, [figure 1B](#) seems to show that Ara+6 fixed a mutation reverting the indel hypermutator phenotype. However, a close examination of the indel mutation rate and allele frequency dynamics in Ara+6 reveals that a super-hypermutator clade evolved within the first 1,000 generations ([supplementary fig. S2, Supplementary Material](#) online). Additional evidence for the super-hypermutator clade comes from the evolution and extinction of an A:T→G:C and G:C→A:T hypermutator phenotype ([fig. 2](#)) that parallels the evolution of the indel hypermutator phenotype. This super-hypermutator clade carries a frameshift allele of the MMR gene *mutS* ([table 1](#) and [fig. 4](#)), is distinguished by marker alleles of the nucleotide excision repair genes *uvrA* and *uvrB* ([fig. 3](#)), and persists at low frequency until going extinct by 20,000 generations ([figs. 3 and 4, supplementary fig. S2, Supplementary Material](#) online). The majority clade in Ara+6 evolved a mutation in *mutT* at 4,750 generations ([table 1](#) and [fig. 3](#)) that causes a point mutation hypermutator phenotype without causing an indel hypermutator phenotype. The coexistence of clades with different hypermutator phenotypes, and the eventual extinction of the super-

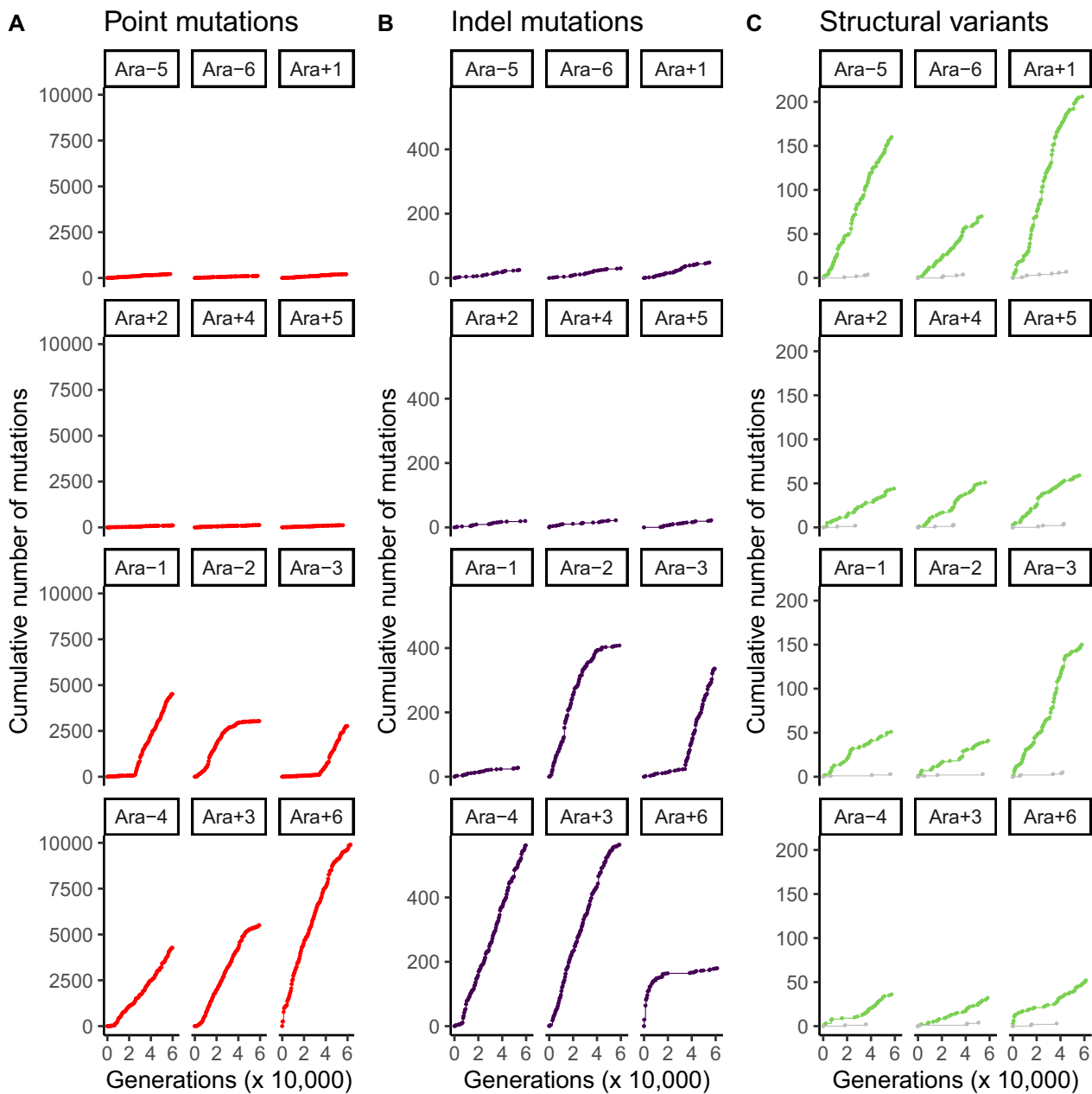


Fig. 1.—Divergent evolution of mutation rates in the LTEE. Each panel shows the cumulative number of observed mutations, subdivided by mutation class, over time in each LTEE population. The top six panels show the nonmutator LTEE populations, and the bottom six panels show the hypermutator LTEE populations. (A) Point mutations are shown in red. (B) Indel mutations are shown in purple. (C) sv associated with transposons are shown in green, whereas those that are not associated with transposons are shown in gray.

hypermutator clade, most reasonably explains the loss of the indel hypermutator phenotype from Ara+6.

Figure 1C shows the number of observed structural mutations over time. As described in the original report for this data set (Good et al. 2017), structural mutations (or structural variants, sv) are defined by junctions between two distinct locations in the reference genome. The vast majority of these structural mutations are caused by insertion sequence (IS)

transpositions. Three of the canonical nonmutator populations (Ara–5, Ara–6, and Ara+1) show an IS hypermutator phenotype. The IS hypermutator phenotype in Ara+1 was reported previously (Papadopoulos et al. 1999; Tenailon et al. 2016). In contrast, only one of the canonical hypermutator populations, Ara–3, shows an IS hypermutator phenotype. The rate of observed structural mutations in Ara–3 shows three different slopes. Ara–3 evolved an IS

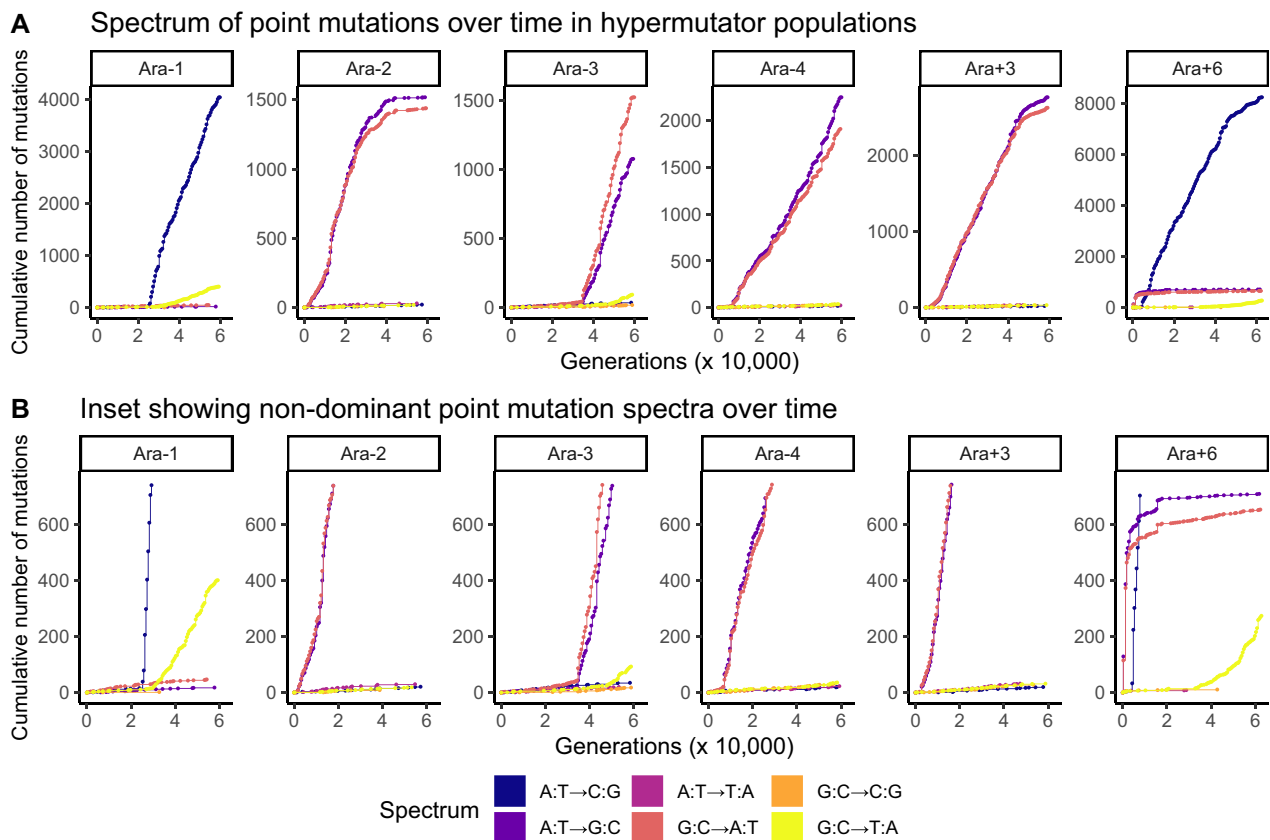


Fig. 2.—The dynamics of hypermutator and antimutator alleles affect the spectrum of observed point mutations over time. (A) Spectrum of point mutations over time in the hypermutator LTEE populations. (B) Inset figure showing nondominant point mutation spectra over time in the hypermutator LTEE populations.

hypermutator phenotype very early in the LTEE. Around 30,000 generations, the IS rate intensifies, either due to genetic evolution, or as a consequence of stress induced by the citrate metabolic innovation that evolved around that time (Blount et al. 2012, 2020). Finally, the IS rate decreases around 45,000 generations. More than 100 mutations go to fixation in the selective sweep at 45,000 generations in Ara-3, including mutations in the DNA repair genes *recR*, *recE*, *ligA*, *uvrA*, and *ybaZ*. The distinct IS rates observed in Ara-3 may, in part, reflect clonal interference between deeply diverged, competing lineages in that population (Blount et al. 2012; Leon et al. 2018), especially if those lineages have different IS transposition rates.

We also examined the spectrum of point mutations in each hypermutator population over time (fig. 2). Ara-1 and Ara+6 show a high frequency of A:T→C:G transversion mutations, characteristic of defects in *mutT* (Tajiri et al. 1995; Fowler et al. 2003; Wielgoss et al. 2013). Ara-2, Ara-3, Ara-4, and Ara+3, which all have defects in MMR (table 1 and fig. 4), show a high frequency of A:T→G:C and G:C→A:T mutations. These findings are consistent with genomic analyses of LTEE hypermutators (Couce et al. 2017). Furthermore, Ara-1, Ara-3, and Ara+6 all show late increases in the

frequency of G:C→T:A transversion mutations, characteristic of defects in *mutY* (Tajiri et al. 1995; Fowler et al. 2003; Wielgoss et al. 2013).

In examining *mutT*, we noticed that two of the three cases of *mutT* alleles arising to high frequency in the LTEE occur on an *uvrA* background (Ara-2 and Ara+6), whereas the third, in Ara-1, occurs on an *uvrC* background (fig. 3). The *mutT* allele in Ara-2 does not cause the characteristic *mutT* A:T→C:G hypermutator phenotype found in Ara-1 and Ara+6 (fig. 2), so its association with *uvrA* may be coincidental. However, the same *uvrA* substitution that goes to fixation with *mutT* in Ara+6 also occurs in a 40,000 generation isolate from the Ara-1 population called REL10939 (Tenaille et al. 2016), which suggests that this particular *uvrA* allele may be beneficial in those contexts. Furthermore, it has been reported that *uvrA/mutT* and *uvrB/mutT* double knockouts have a substantially lower mutation rate than *mutT* knockouts, in the presence of hydrogen peroxide (Hori et al. 2007). Based on these observations, we hypothesize that the *mutT* alleles that successfully went to fixation in the LTEE may have evolved on an *uvrABC* genetic background that reduced the intensity of the *mutT* hypermutator phenotype.

Table 1.

Putative Hypermutator and Antimutator Alleles Described in the Text

Population	Gene	DNA Repair Pathway	Appearance Time (Generations)	Position (bp)	Mutation
Ara-1	<i>uvrC</i>	Oxidative damage repair	26,250	1,972,086	Q183P
Ara-1	<i>mutT</i>	Oxidative damage repair	26,250	114,034	(C) _{6→7}
Ara-1	<i>mutY</i>	Oxidative damage repair	28,750	2,988,792	L40W
Ara-1	<i>mutY</i>	Oxidative damage repair	32,250	2,989,164	L164*
Ara-2	<i>mutL</i>	MMR	2,250	4,375,786	(TGGCGC) _{3→4}
Ara-2	<i>uvrA</i>	Oxidative damage repair	12,250	4,251,585	A407T
Ara-2	<i>mutT</i>	Oxidative damage repair	13,750	114,113	R89H
Ara-2	<i>mutL</i>	MMR	*This in-frame reversion fixes at 42,250 generations	4,375,781	(TGGCGC) _{3→2}
Ara-3	<i>mutS</i>	MMR	34,750	2,753,768	Q606*
Ara-3	<i>mutY</i>	Oxidative damage repair	48,250	2,989,624	Δ1 bp
Ara-4	<i>mutL</i>	MMR	7,250	4,375,781	(TGGCGC) _{3→2}
Ara+3	<i>mutS</i>	MMR	2,750	2,752,473	+G
Ara+6	<i>mutS</i>	MMR	1,250	2,752,473	+G
Ara+6	<i>uvrA</i>	Oxidative damage repair	4,750	4,250,341	I821M
Ara+6	<i>mutT</i>	Oxidative damage repair	4,750	114,034	(C) _{6→5}
Ara+6	<i>mutY</i>	Oxidative damage repair	31,750	2,988,917	Y82D
Ara+6	<i>mutY</i>	Oxidative damage repair	49,750	2,989,297	C208W

Gene-Orientation Mutation Bias Evolves in the LTEE

Several reports indicate that mutation rates differ between the leading and lagging strands of the DNA replication bubble (Lee et al. 2012; Paul et al. 2013). Potential causes include asymmetry in nucleotide composition around the replication origin (GC skew) (Marín and Xia 2008), context-dependent mutation rates that are asymmetric around the replication origin (Sung et al. 2015), and head-on collisions between the replication and transcription molecular machinery (Paul et al. 2013). Such reports motivated us to ask whether the LTEE metagenomics data showed evidence of gene-orientation mutation biases, such that genes oriented with (or against) the leading or lagging strand of DNA synthesis have different mutation rates.

Our null expectation is that the distribution of synonymous mutations on each strand of the chromosome should be related to the amount of coding sequence on each strand (i.e., the density of genes multiplied by their length). Furthermore, the spectrum of nucleotide substitutions on each strand should reflect local G:C content in the ancestral LTEE clone REL606: for example, G:C→A:T substitutions should be more common in G:C-rich regions. Figure 5A shows this null expectation. Both the amount of coding sequence and G:C content per strand are asymmetric about the replication origin of REL606. At the replication origin, one DNA strand switches from leading to lagging, while its complement switches from lagging to leading. This switch occurs because DNA replication is bidirectional, such that two replisomes move in opposite directions from the replication origin. Even in the absence of gene-orientation mutation bias, figure 5A shows that some asymmetry in the distribution of synonymous mutations over the replication origin is expected.

The observed distributions of synonymous mutations on each strand of the chromosome are shown in figure 5B. We separately analyzed MMR- and MutT-deficient hypermutator populations. In both cases, the number of observed mutations significantly differs between genes oriented with or against the movement of the replisome, based on comparing the expected ratio of mutations to the observed ratio of mutations. The MMR-deficient hypermutator populations show significantly more gene-orientation mutation bias than expected (two-tailed binomial test: observed ratio of 2,066:2,664 mutations vs. expected ratio of 1,730,238:2,066,587 nucleotides; $P=0.0090$), whereas the MutT-deficient hypermutator populations show significantly less gene-orientation bias than expected (two-tailed binomial test: observed ratio of 947:1,033 mutations vs. expected ratio of 1,730,238:2,066,587 nucleotides; $P=0.0446$). Note that these calculations do not account for the characteristic mutation spectra of MMR- and MutT-deficient hypermutators (fig. 5B). For example, the extreme rate of A:T→C:G mutations seen in MutT-deficient hypermutators (Foster et al. 2015) should cause A:T rich genes to mutate faster than A:T poor genes.

The Genomic Distribution of Observed Mutations in Ara+3 Shows a Strong, Symmetric Wave Pattern over the Origin of Replication

Multiple studies (Sharp et al. 1989; Lang and Murray 2011; Foster et al. 2013; Dillon et al. 2018; Niccum et al. 2019) have reported correlations between local mutation rates and distance from the origin of replication. One hypermutator LTEE population, called Ara+3, shows a symmetric wave pattern reflected over *oriC* (fig. 6). Indeed, the genomic distribution of

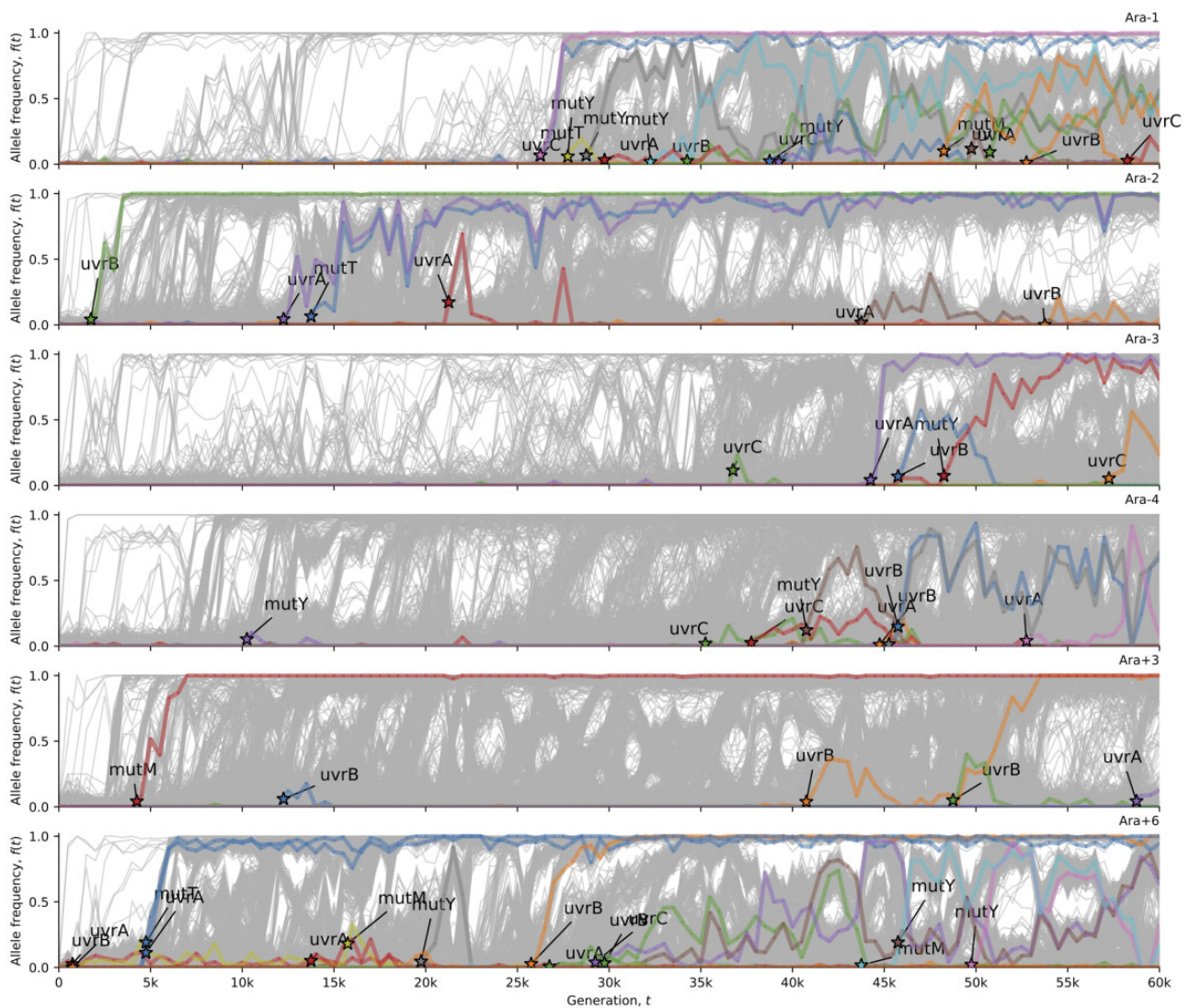


FIG. 3.—Oxidative damage repair alleles in hypermutator LTEE populations. This visualization uses computer code from Good et al. (2017). Stars indicate the time (and allele frequency) at which mutations are reliably estimated to appear in the time series. The allele frequency trajectories for all observed mutations in the hypermutator populations are shown in gray. The allele frequency trajectories of de novo mutations (excepting synonymous mutations) in oxidative damage repair genes (supplementary file 1, Supplementary Material online) are colored and labeled in each population.

observed mutations in Ara+3 is significantly different from the genomic distribution of observed mutations summed over all hypermutator populations (two-sample Kolmogorov–Smirnov test: $D=0.0567$, $P < 10^{-14}$). The wave in Ara+3 has a trough-to-peak ratio of $\sim 25:75$ (fig. 6). Excluding Ara+3, the genomic distribution of observed mutations summed over the remaining MMR-deficient LTEE populations shows a weak wave pattern, whereas the populations with defects in *mutT* shows no evidence of the wave pattern (fig. 7). The genomic distribution of observed mutations in the MMR-deficient populations (excluding Ara+3) is significantly different from the genomic

distribution of observed mutations in the *MutT*-deficient populations (two-sample Kolmogorov–Smirnov test: $D=0.040916$, $P < 10^{-9}$).

Evidence for Epistasis and Historical Contingency in the Evolution of DNA Topology

Why does a strong wave pattern only appear in Ara+3? Others have hypothesized that local chromatin structure affects local mutation rates (Foster et al. 2013; Niccum et al. 2019). Furthermore, DNA topology has evolved in parallel in the LTEE, and artificially increasing DNA supercoiling is

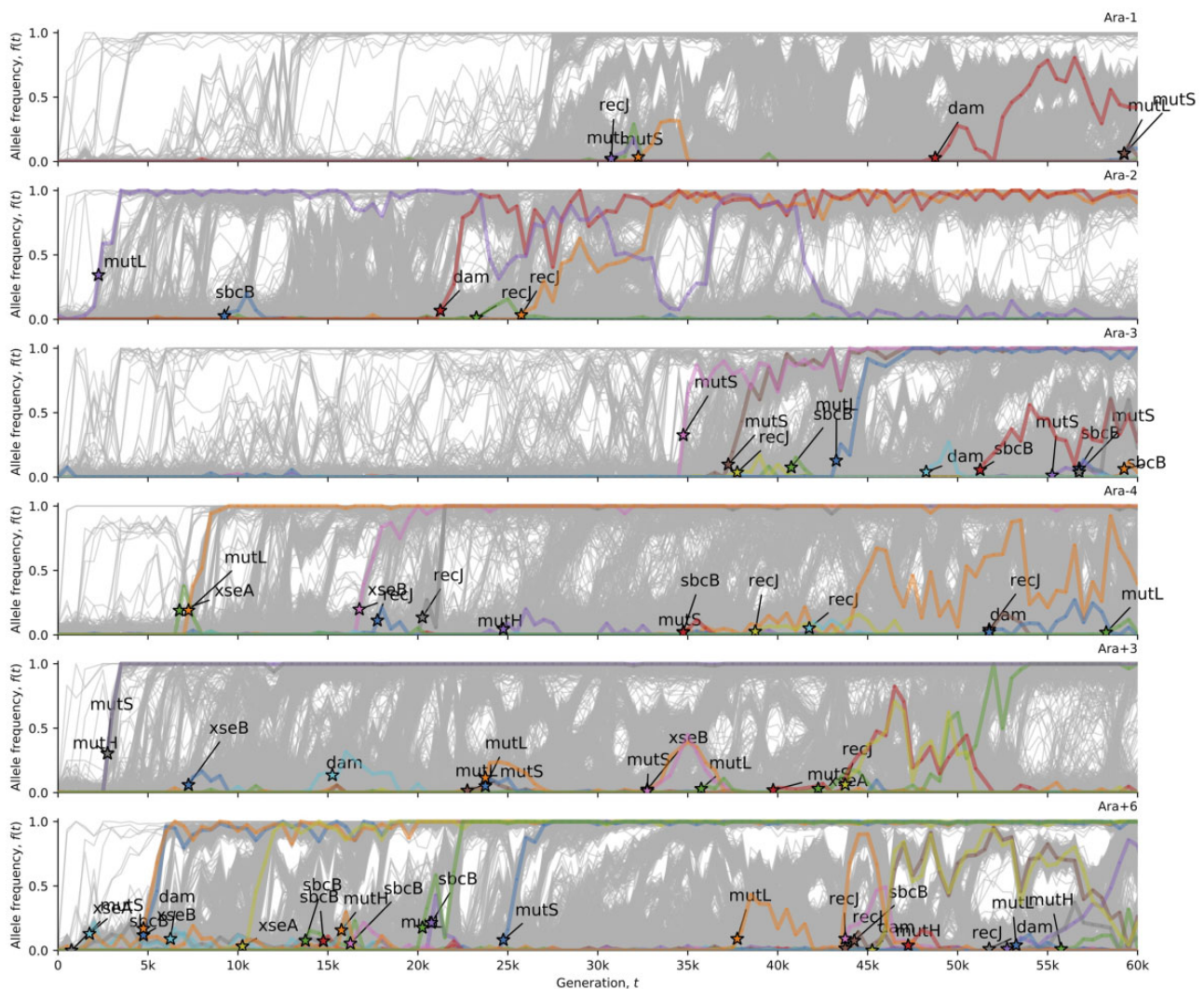


Fig. 4.—MMR alleles in the hypermutator LTEE populations. This visualization uses computer code from Good et al. (2017). Stars indicate the time (and allele frequency) at which mutations are reliably estimated to appear in the time series. The allele frequency trajectories for all observed mutations in the hypermutator populations are shown in gray. The allele frequency trajectories of de novo mutations (except synonymous mutations) in MMR genes ([supplementary file 1, Supplementary Material](#) online) are colored and labeled in each population.

beneficial under LTEE conditions (Croizat et al. 2005, 2010). Therefore, we hypothesized that mutations in genes that affect DNA topology might affect the wave pattern. To test this hypothesis, we examined the timing and distribution of mutations in *topA*, *fis*, and *dusB* (*yhdG*). We focused on these genes for several reasons. First, these loci show strong parallel evolution in the LTEE (Croizat et al. 2010). Second, introducing evolved alleles of *topA* and *fis* into the ancestral genome are sufficient to confer a fitness benefit as well as additive changes to DNA topology (Croizat et al. 2005). Finally, statistical analysis of the pattern of evolution for *dusB* and *fis* in the LTEE led to the discovery that *dusB* regulates *fis* expression (Croizat et al. 2005, 2010). We excluded synonymous mutations from this analysis. We counted both fixations and

mutations destined for extinction, because many beneficial mutations go extinct in large asexual populations due to clonal interference (Gerrish and Lenski 1998; Lang et al. 2013; Levy et al. 2015; Maddamsetti, Lenski, et al. 2015; Ba et al. 2019).

All LTEE populations evolved missense, indel, or structural mutations in *topA*, *fis*, and *dusB* within the first 10,000 generations, except two: Ara+2 and Ara+3 (fig. 8). The timing and distribution of mutations in these genes across populations suggests epistasis and historical contingency (Good et al. 2017). The early arrival times for mutations in these genes suggests that there is an early, limited window of opportunity for those mutations to go to fixation. Quantitative evidence comes from Ara+3, which has no missense, nonsense, indel,

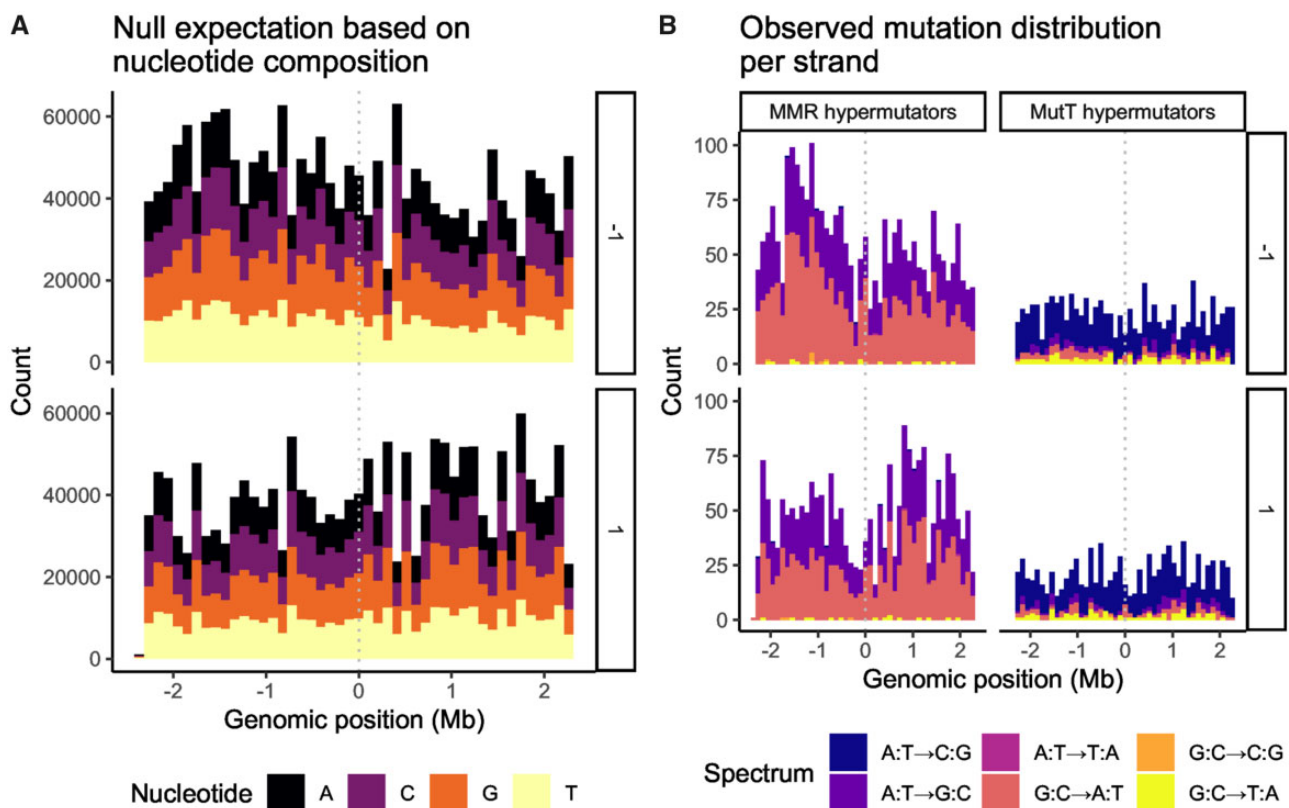


Fig. 5.—Gene-orientation mutation bias evolves in the LTEE. The x axis is the reference genome, centered on the replication origin, partitioned into 46 equally sized bins of ~100 kb. In each labeled subfigure, top and bottom panels show genes occurring on each of the two strands of the chromosome, with the arbitrary labels 1 and -1. (A) The nucleotide composition of genes on the two strands of the chromosome of the LTEE ancestral clone REL606. (B) The genomic distribution of mutations within genes, summed over MMR-deficient LTEE populations (left panel) and MutT-deficient LTEE populations (right panel).

or structural mutations in *topA*, *fis*, and *dusB* whatsoever, despite its strong hypermutator phenotype. The probability of this event is $P = (1 - (t/g))^n$, where t is the effective mutational target size, g is the length of the chromosome ($g = 4,629,812$), and n is the number of observed missense, indel, and structural mutations in Ara+3 ($n = 4,368$). Given the wave pattern in Ara+3, the effective mutational target size of *topA*, *fis*, and *dusB* could be smaller than their combined physical target size (3,861 bp), say if they occurred in the trough of the wave. To take this into account, we partitioned the chromosome into bins, counted mutations per bin, and calculated the effective mutational target size by multiplying the physical target size (length) of *topA*, *fis*, and *dusB* by the number of mutations per base pair in their respective bins. These genes are significantly depleted of mutations in Ara+3, for bin sizes ranging from 100 kb to the entire chromosome (one-tailed randomization tests with 10,000 bootstraps: $P < 0.05$ in all cases).

The distribution of synonymous mutations in *topA*, *fis*, and *dusB* across the LTEE populations is interesting (supplementary fig. S4 and Supplementary Material online). A single, synonymous A312A substitution in *dusB* went to fixation at

~4,000 generations in Ara+3, simultaneously with alleles in the MMR genes *mutS* and *mutH* that apparently caused the early hypermutator phenotype in this population. No other synonymous mutations in *dusB* are observed in Ara+3. Furthermore, there is evidence of parallel evolution at this particular position in *dusB*. The same synonymous mutation occurs in Ara+6, and another synonymous mutation, one base pair downstream in the next codon, is the only synonymous mutation in *topA*, *fis*, or *dusB* observed in Ara-2 (supplementary fig. S4, Supplementary Material online). This parallelism suggests that positive selection may be acting on these synonymous variants. Overall, it is striking how few synonymous mutations in *topA*, *fis*, and *dusB* occur in the hypermutator LTEE populations, which implies that synonymous variants in these genes may not be evolving neutrally. Indeed, STIMS (Maddamsetti and Grant 2020) finds a significant signal of purifying selection on synonymous mutations in *topA*, *fis*, and *dusB* in Ara-1 and Ara-3 (one-tailed randomization test with 10,000 bootstraps: $P < 0.0001$).

We also examined the genes that encode the nucleoid-binding protein HU and the terminus-organizing protein MatP, as deletions of these loci were shown to affect the

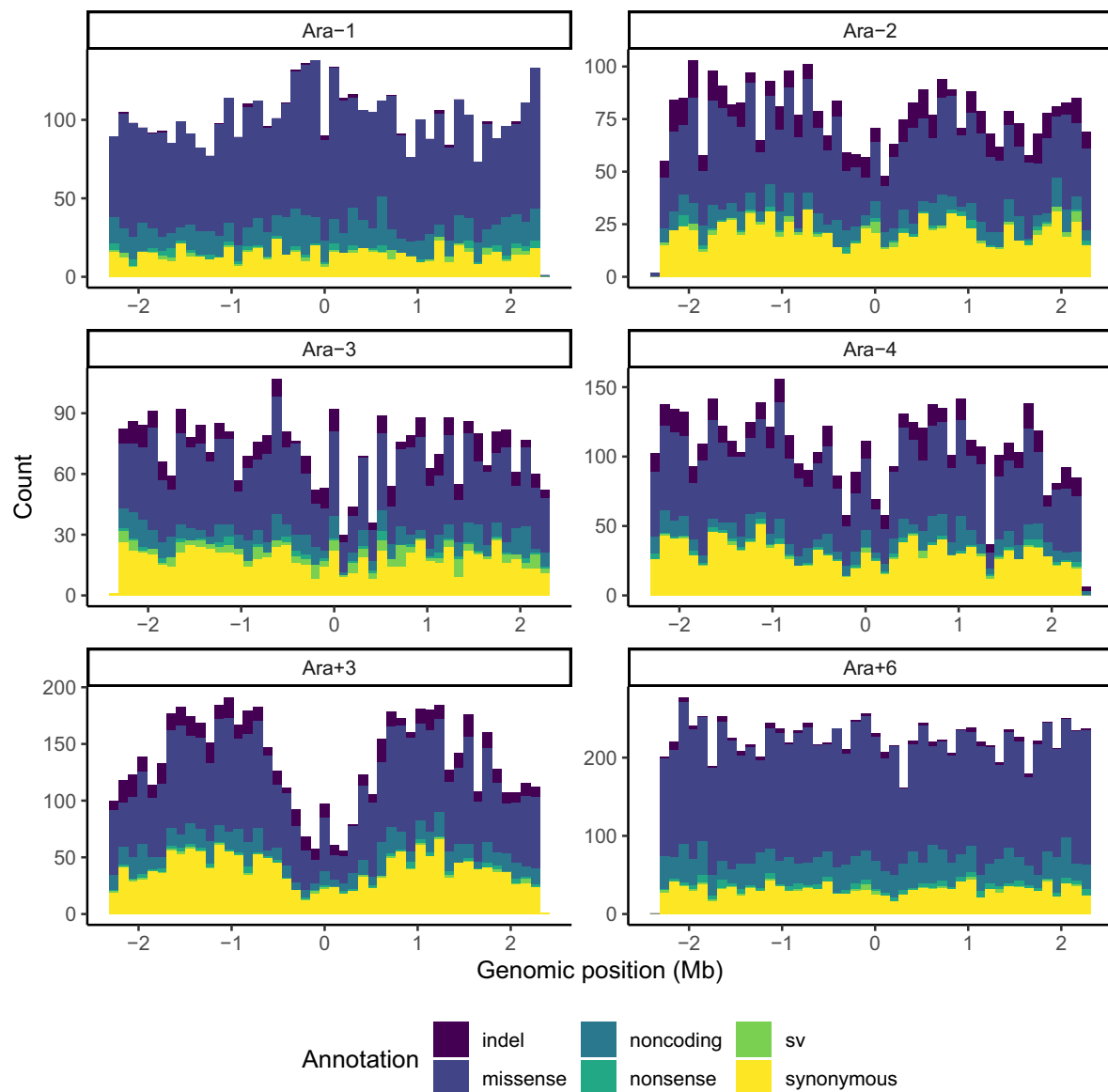


Fig. 6.—One hypermutator LTEE population, Ara+3, shows a strong wave pattern of mutation rate variation centered on the replication origin. Each panel shows the genomic distribution of mutations observed in each hypermutator LTEE population in the metagenomics data. The x axis is the reference genome, centered on the replication origin, partitioned into 46 equally sized bins of ~100 kb. Indels are in purple, missense mutations are in dark blue, noncoding mutations are blue green, nonsense mutations are sea green, sv are green, and synonymous mutations are yellow.

wave pattern (Niccum et al. 2019). Notwithstanding the relevance of HU and MatP in Niccum et al. (2019), these genes show limited evidence of parallel evolution in the LTEE (supplementary fig. S5, Supplementary Material online).

Synonymous Nucleotide Diversity in Natural *E. coli* Populations Does Not Predict Mutation Rate Variation in the LTEE

Finally, we used the LTEE metagenomic data to revisit previous work, which found that the distribution of synonymous

mutations in the LTEE does not reflect patterns of synonymous variation among natural *E. coli* isolates (Maddamsetti et al. 2015). During our reanalysis, we found a potential coding error affecting the results of the Kolmogorov–Smirnov test reported in that paper. Therefore, we used Poisson regression to ask whether the estimates of synonymous nucleotide diversity θ_s published in Martincorena et al. (2012), when treated as gene-specific estimates of the point-mutation rate per base pair, predict the distribution of synonymous mutations observed in the LTEE. A null model in which mutations occur uniformly over the chromosome (Akaike’s

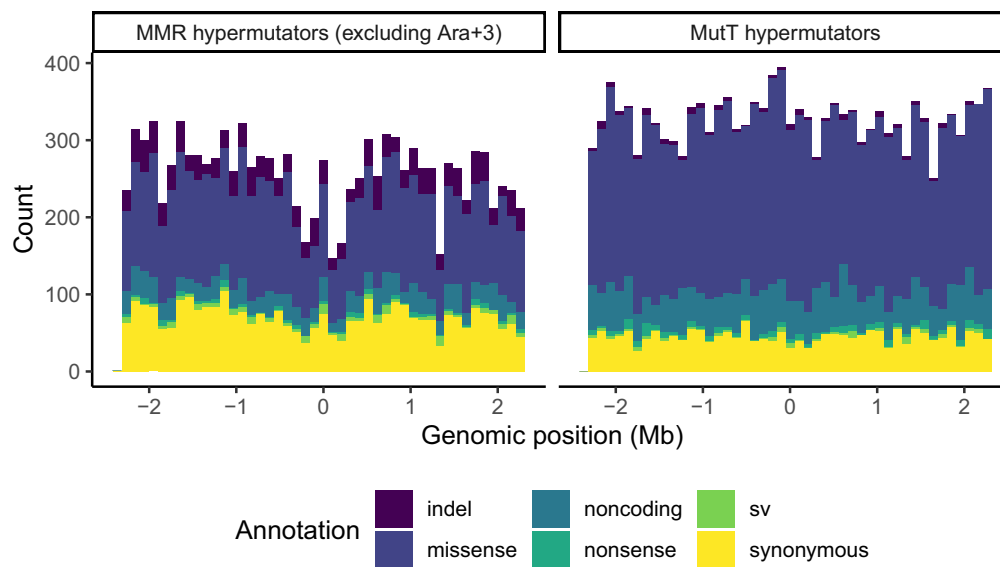


FIG. 7.—MMR-deficient LTEE populations (excluding Ara+3) show a weak wave pattern, whereas MutT-deficient LTEE populations show no wave pattern. The left panel shows the genomic distribution of mutations observed in Ara−2, Ara−3, and Ara−4. The right panel shows the genomic distribution of mutations observed in Ara−1 and Ara+6. The x axis is the reference genome, centered on the replication origin, partitioned into 46 equally sized bins of ~100 kb. Indels are in purple, missense mutations are in dark blue, noncoding mutations are blue green, nonsense mutations are sea green, sv are green, and synonymous mutations are yellow.

Information Criterion, $AIC = 8,529.6$) fits the data far better than the θ_s model ($AIC = 9,171.3$). When we fit both models to Ara+3, we again find that the null model is better than the θ_s model at predicting the observed distribution of synonymous mutations ($AIC = 2,168.2$ for null model vs. $AIC = 2,190.8$ for θ_s model). This finding validates the conclusions reported in Maddamsetti et al. (2015), despite the potential problems in that analysis.

Discussion

By examining the distribution of observed mutations over more than 60,000 generations of the LTEE (Good et al. 2017), we find that mutation rates and biases have diverged idiosyncratically, despite identical abiotic conditions. One LTEE population, Ara+3, shows strong evidence of the wave pattern in mutation rate variation. Similar patterns have been seen in mutation accumulation experiments with MMR-deficient strains of *E. coli* as well as in *Vibrio* bacteria (Dillon et al. 2018; Niccum et al. 2019). Our result shows that genomic biases in mutation rates evolve dynamically on laboratory timescales. It is likely that the identity and effects of many hypermutator and antimutator alleles in the LTEE remains unknown. For instance, we do not know what alleles, if any, cause the apparent late decrease in mutation rate seen in Ara+3. Experiments are needed, both to discover those unknown alleles, and to test for genetic interactions that modulate mutation rates in the LTEE, as we have hypothesized for alleles of *uvrABC* and *mutT*.

The divergence in the rates, biases, and spectra of mutations across replicate populations in this simple long-term evolution experiment makes one wonder about the scope of natural variation in mutation rates, biases, and spectra. An evolution experiment with replicate mouse microbiomes has indicated that microbial evolution in the gut is probably characterized by long-term maintenance of intraspecies genetic diversity, including mutation rate polymorphism (Ramiro et al. 2020). Phylogenomic studies have also found extensive evidence for horizontal gene transfer in DNA repair genes (Denamur et al. 2000), which suggests that polymorphism in DNA repair genes may cause extensive natural variation in mutation and recombination rates within and across bacterial (meta-) populations and communities.

We find statistical evidence for historical contingency and epistasis in the evolution of DNA topology in the LTEE, and for Ara+3 in particular. These findings suggest a relationship between local DNA topology and local mutation rate variation, consistent with the experiments reported by Niccum et al. (2019). These findings immediately suggest the need for experiments to test whether local DNA topology causes local mutation rate variation, and to test whether local DNA topology affects strand-specific and gene-orientation mutation biases.

A comparison of synonymous genetic variation estimated from natural *E. coli* isolates to the distribution of observed synonymous mutations in the LTEE confirms the conclusion of earlier work (Maddamsetti et al. 2015) using richer data,

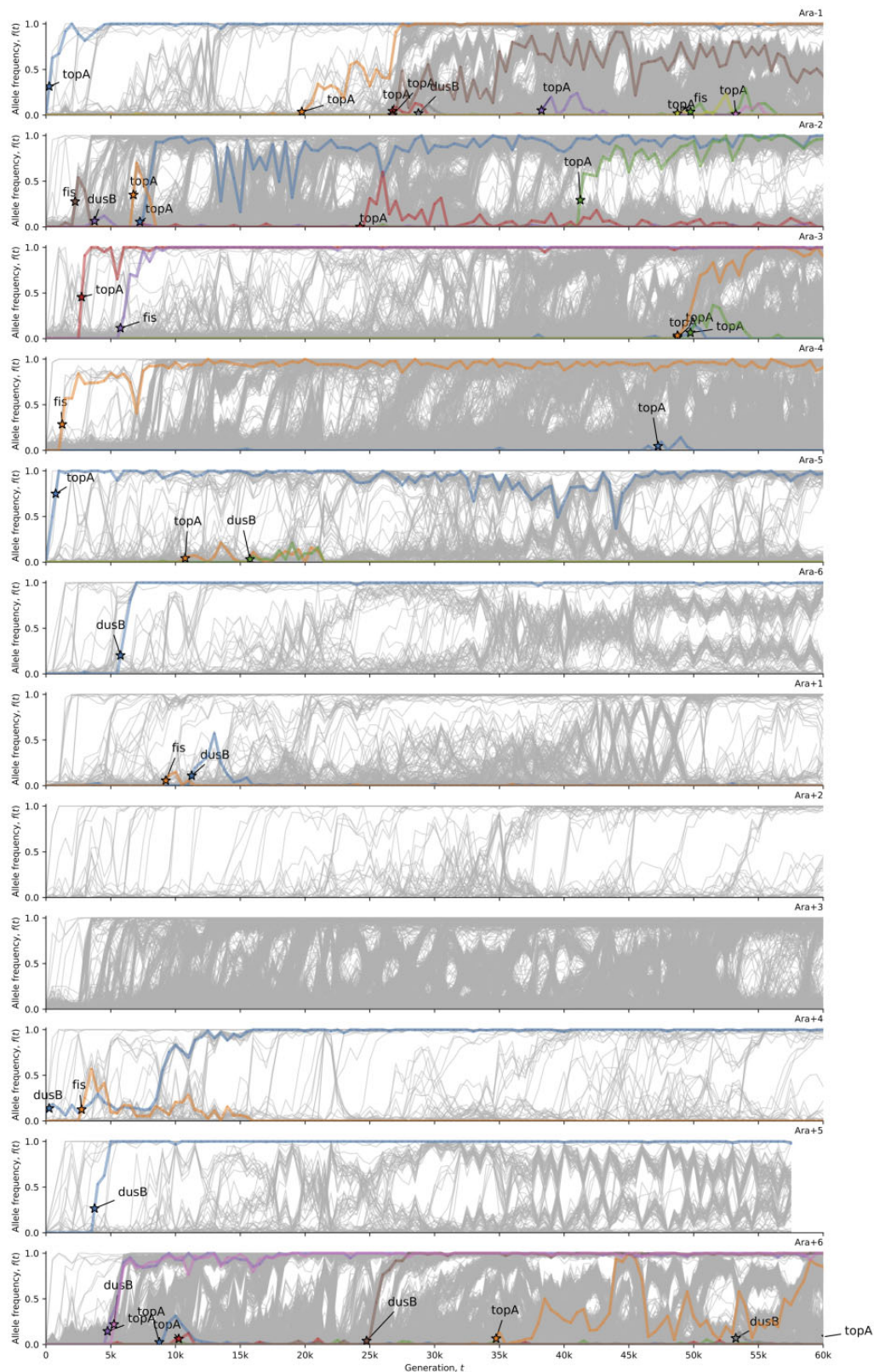


FIG. 8.—The strong wave pattern in Ara+3 anticorrelates with mutations (excluding synonymous changes) in the DNA topology genes *topA*, *fis*, and *dusB*. This visualization uses computer code written by Good et al. (2017). The allele frequency trajectories for all observed mutations in the 12 LTEE populations are shown in gray. The allele frequency trajectories of de novo mutations in *topA*, *fis*, and *dusB* (excepting synonymous mutations) are colored and labeled in each population.

and is consistent with other reports as well (Lee et al. 2012; Chen and Zhang 2013; Lynch et al. 2016). In sum, gene-specific variation in synonymous nucleotide diversity θ_s , estimated from natural isolates of *E. coli*, does not predict the genomic distribution of synonymous mutations observed in the LTEE. In any case, the other results that we have presented, in addition to prior reports (Foster et al. 2013; Paul et al. 2013; Sung et al. 2015; Jee et al. 2016; Niccum et al. 2019), strongly indicate that mutation rates vary over the *E. coli* chromosome.

These results add to the robust debate on the causes and consequences of mutation rate evolution. It is clear that a deeper understanding of the relationships among chromatin structure, genomic variation in mutation and recombination rates, and natural selection, and their consequences for short- and long-term genome evolution, will be a fruitful goal for further research.

Materials and Methods

Preprocessed LTEE metagenomic data, and associated analysis and visualization code was downloaded from: <https://github.com/benjaminhg/LTEE-metagenomic>. Analysis codes are available from: <https://github.com/rohanmaddamsetti/LTEE-purifying-selection/blob/master/mutation-rate-analysis.R> and <https://github.com/rohanmaddamsetti/LTEE-purifying-selection/blob/master/metagenomics-library.R>. We systematically examined DNA repair genes in *E. coli* (Eisen and Hanawalt 1999; Lee et al. 2016; Deatherage et al. 2018), as well as annotated DNA polymerases, and other proteins of the replisome. A table of these genes and their annotations is in [supplementary data file 1, Supplementary Material](#) online. We cross-checked the evolutionary dynamics of alleles of these genes in the LTEE metagenomic data against the observed changes in mutation rates and spectra in each LTEE population. We also examined the LTEE genomic data (Tenaillon et al. 2016) for mutations in these genes, using the R Shiny web app interface at www.barricklab.org/shiny/LTEE-Ecoli. In this manner, we curated a list of putative hypermutator and antimutator alleles in the LTEE ([table 1](#)). Those alleles, and alleles of other genes in their respective DNA repair pathways, are shown in [figures 3 and 4](#). [Figure 3](#) shows the evolutionary dynamics of alleles in genes encoding base excision repair, nucleotide excision repair, and degradation of oxidized nucleotide triphosphates. [Figure 4](#) shows the evolutionary dynamics of alleles in genes encoding DNA MMR. Data sets and analysis codes to replicate the findings and figures in this paper are available on the Dryad Digital Repository (DOI: <https://doi.org/10.5061/dryad.kpr4xh2z>).

Acknowledgments

We thank Richard Lenski and Helen Murphy for valuable discussions and advice. We thank Jeffrey Barrick for valuable

discussions and comments on an earlier version of our manuscript. We also thank Benjamin Good for making preprocessed LTEE metagenomic data and analysis scripts accessible for the research community. The LTEE that generated the bacteria we used in this study is supported by a grant from the National Science Foundation (currently DEB-1951307) to Richard Lenski and Jeffrey Barrick. N.A.G. was supported by the BEACON Center for the Study of Evolution in Action (NSF cooperative agreement DBI-0939454) and by Michigan State University.

Literature Cited

- Ba ANN, et al. 2019. High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature* 575(7783):494–499.
- Blount ZD, et al. 2020. Genomic and phenotypic evolution of *Escherichia coli* in a novel citrate-only resource environment. *eLife* 9:e55414.
- Blount ZD, Barrick JE, Davidson CJ, Lenski RE. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* 489(7417):513–518.
- Chao L, Vargas C, Spear BB, Cox EC. 1983. Transposable elements as mutator genes in evolution. *Nature* 303(5918):633–635.
- Chen X, Zhang J. 2013. No gene-specific optimization of mutation rate in *Escherichia coli*. *Mol Biol Evol.* 30(7):1559–1562.
- Cooper TF. 2007. Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. *PLoS Biol.* 5(9):e225.
- Couce A, et al. 2017. Mutator genomes decay, despite sustained fitness gains, in a long-term experiment with bacteria. *Proc Natl Acad Sci USA.* 114(43):E9026–9035.
- Crozat E, et al. 2010. Parallel genetic and phenotypic evolution of DNA superhelicity in experimental populations of *Escherichia coli*. *Mol Biol Evol.* 27(9):2113–2128.
- Crozat E, Philippe N, Lenski RE, Geiselmann J, Schneider D. 2005. Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics* 169(2):523–532.
- Deatherage DE, Leon D, Rodriguez AE, Omar SK, Barrick JE. 2018. Directed evolution of *Escherichia coli* with lower-than-natural plasmid mutation rates. *Nucleic Acids Res.* 46(17):9236–9250.
- Denamur E, et al. 2000. Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell* 103(5):711–721.
- Dillon MM, Sung W, Lynch M, Cooper VS. 2018. Periodic variation of mutation rates in bacterial genomes associated with replication timing. *mBio* 9(4):e01371.
- Eisen JA, Hanawalt PC. 1999. A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat Res/DNA Repair* 435(3):171–213.
- Foster PL, Hanson AJ, Lee H, Popodi EM, Tang H. 2013. On the mutational topology of the bacterial genome. *GP* 3:399–407.
- Foster PL, Lee H, Popodi E, Townes JP, Tang H. 2015. Determinants of spontaneous mutation in the bacterium *Escherichia coli* as revealed by whole-genome sequencing. *Proc Natl Acad Sci USA.* 112(44):E5990–5999.
- Fowler RG, et al. 2003. Interactions among the *Escherichia coli* *mutT*, *mutM*, and *mutY* damage prevention pathways. *DNA Repair* 2(2):159–173.
- Gerrish PJ, Lenski RE. 1998. The fate of competing beneficial mutations in an asexual population. *Genetica* 102:127–144.
- Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. 2017. The dynamics of molecular evolution over 60,000 generations. *Nature* 551(7678):45–50.
- Hori M, et al. 2007. UvrA and UvrB enhance mutations induced by oxidized deoxyribonucleotides. *DNA Repair* 6(12):1786–1793.

- Jee J, et al. 2016. Rates and mechanisms of bacterial mutagenesis from maximum-depth sequencing. *Nature* 534(7609):693–696.
- Lang GI, et al. 2013. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* 500(7464):571–574.
- Lang GI, Murray AW. 2011. Mutation rates across budding yeast chromosome VI are correlated with replication timing. *Genome Biol Evol.* 3:799–811.
- Lee H, Doak TG, Popodi E, Foster PL, Tang H. 2016. Insertion sequence-caused large-scale rearrangements in the genome of *Escherichia coli*. *Nucleic Acids Res.* 44(15):7109–7119.
- Lee H, Popodi E, Tang H, Foster PL. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci.* 109(41):E2774–2783.
- Lenski RE, Rose MR, Simpson SC, Tadler SC. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am Nat.* 138(6):1315–1341.
- Leon D, D'Alton S, Quandt EM, Barrick JE. 2018. Innovation in an *E. coli* evolution experiment is contingent on maintaining adaptive potential until competition subsides. *PLoS Genet.* 14(4):e1007348.
- Levy SF, et al. 2015. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* 519(7542):181–186.
- Loh E, Salk JJ, Loeb LA. 2010. Optimization of DNA polymerase mutation rates during bacterial evolution. *Proc Natl Acad Sci.* 107(3):1154–1159.
- Lynch M, et al. 2016. Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet.* 17(11):704–714.
- Maddamsetti R, et al. 2015. Synonymous genetic variation in natural isolates of *Escherichia coli* does not predict where synonymous substitutions occur in a long-term experiment. *Mol Biol Evol.* 32(11):2897–2904.
- Maddamsetti R, et al. 2017. Core genes evolve rapidly in the long-term evolution experiment with *Escherichia coli*. *Genome Biol Evol.* 9(4):1072–1083.
- Maddamsetti R, Grant NA. 2020. A simple test to infer mode of selection in metagenomics time series of evolving asexual populations. *bioRxiv*. doi: 10.1101/2020.05.23.11250.
- Maddamsetti R, Lenski RE, Barrick JE. 2015. Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with *Escherichia coli*. *Genetics* 200(2):619–631.
- Marín A, Xia X. 2008. GC skew in protein-coding genes between the leading and lagging strands in bacterial genomes: new substitution models incorporating strand bias. *J Theor Biol.* 253(3):508–513.
- Martincorena I, Seshasayee AS, Luscombe NM. 2012. Evidence of non-random mutation rates suggests an evolutionary risk management strategy. *Nature* 485(7396):95–98.
- Moxon R, Bayliss C, Hood D. 2006. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu Rev Genet.* 40(1):307–333.
- Neher RA. 2013. Genetic draft, selective interference, and population genetics of rapid adaptation. *Annu Rev Ecol Evol Syst.* 44(1):195–215.
- Niccum BA, Lee H, Mohammedsmaïl W, Tang H, Foster PL. 2019. The symmetrical wave pattern of base-pair substitution rates across the *Escherichia coli* chromosome has multiple causes. *mBio* 10(4):e01226–19.
- Papadopoulos D, et al. 1999. Genomic evolution during a 10,000-generation experiment with bacteria. *Proc Natl Acad Sci.* 96(7):3807–3812.
- Paul S, Million-Weaver S, Chattopadhyay S, Sokurenko E, Merrikh H. 2013. Accelerated gene evolution through replication–transcription conflicts. *Nature* 495(7442):512–515.
- Peabody G, et al. 2016. Benefits of a recombination-proficient *Escherichia coli* system for adaptive laboratory evolution. *Appl Environ Microbiol.* 82(22):6736–6747.
- Peabody G, Li H, Kao KC. 2017. Sexual recombination and increased mutation rate expedite evolution of *Escherichia coli* in varied fitness landscapes. *Nat Commun.* 8(1):9.
- Ramiro RS, Durão P, Bank C, Gordo I. 2020. Low mutational load and high mutation rate variation in gut commensal bacteria. *PLoS Biol.* 18(3):e3000617.
- Schiffels S, Szöllösi GJ, Mustonen V, Lässig M. 2011. Emergent neutrality in adaptive asexual evolution. *Genetics* 189(4):1361–1375.
- Sharp PM, Shields DC, Wolfe KH, Li W-H. 1989. Chromosomal location and evolutionary rate variation in enterobacterial genes. *Science* 246(4931):808–810.
- Sniegowski PD, Gerrish PJ, Lenski RE. 1997. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* 387(6634):703–705.
- Sprouffske K, Aguilar-Rodríguez J, Sniegowski P, Wagner A. 2018. High mutation rates limit evolutionary adaptation in *Escherichia coli*. *PLoS Genet.* 14(4):e1007324.
- Sung W, et al. 2015. Asymmetric context-dependent mutation patterns revealed through mutation–accumulation experiments. *Mol Biol Evol.* 32(7):1672–1683.
- Tajiri T, Maki H, Sekiguchi M. 1995. Functional cooperation of MutT, MutM and MutY proteins in preventing mutations caused by spontaneous oxidation of guanine nucleotide in *Escherichia coli*. *Mutat Res/ DNA Repair* 336(3):257–267.
- Tenaillon O, et al. 2016. Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* 536(7615):165–170.
- Tenaillon O, Taddei F, Radman M, Matic I. 2001. Second-order selection in bacterial evolution: selection acting on mutation and recombination rates in the course of adaptation. *Res Microbiol.* 152(1):11–16.
- Wielgoss S, et al. 2013. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc Natl Acad Sci.* 110(1):222–227.

Associate Editor: George Zhang