Mutator genomes decay, despite sustained fitness gains, in a long-term experiment with bacteria

Alejandro Couce[a,1], Larissa Viraphong Caudwell[b,1], Christoph Feinauer[d], Thomas Hindré[b], Jean-Paul Feugeas[a], Martin Weigt[c], Richard E. Lenski[d,e], Dominique Schneider[b], and Olivier Tenaillon[a,2]

[a]Unité Mixte de Recherche 1137, Infection, Antimicrobiens, Modélisation, Evolution, INSERM, Université Paris Diderot, 75018 Paris, France; [b]Unité Mixte de Recherche 5525, Technologies de l'Ingénierie Médicale et de la Complexité – Informatique, Mathématiques et Applications Grenoble, Université Grenoble Alpes, CNRS, 38000 Grenoble, France; [c]Laboratoire de Biologie Computationnelle et Quantitative, Sorbonne Universités, Université Pierre et Marie Curie, Université Paris 06, CNRS, Institut de Biologie Paris Seine, 75005 Paris, France; [d]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824; and [e]BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, MI 48824

Edited by James J. Bull, The University of Texas at Austin, Austin, TX, and approved September 1, 2017 (received for review April 8, 2017)

Understanding the extreme variation among bacterial genomes remains an unsolved challenge in evolutionary biology, despite long-standing debate about the relative importance of natural selection, mutation, and random drift. A potentially important confounding factor is the variation in mutation rates between lineages and over evolutionary history, which has been documented in several species. Mutation accumulation experiments have shown that hypermutability can erode genomes over short timescales. These results, however, were obtained under conditions of extremely weak selection, casting doubt on their general relevance. Here, we circumvent this limitation by analyzing genomes from mutator populations that arose during a long-term experiment with *Escherichia coli*, in which populations have been adaptively evolving for >50,000 generations. We develop an analytical framework to quantify the relative contributions of mutation and selection in shaping genomic characteristics, and we validate it using genomes evolved under regimes of high mutation rates with weak selection (mutation accumulation experiments) and low mutation rates with strong selection (natural isolates). Our results show that, despite sustained adaptive evolution in the long-term experiment, the signature of selection is much weaker than that of mutational biases in mutator genomes. This finding suggests that relatively brief periods of hypermutability can play an outsized role in shaping extant bacterial genomes. Overall, these results highlight the importance of genomic draft, in which strong linkage limits the ability of selection to purge deleterious mutations. These insights are also relevant to other biological systems evolving under strong linkage and high mutation rates, including viruses and cancer cells.

genetic draft | hypermutability | GC content | experimental evolution | selection

Bacterial genomes have a remarkable variety of constitutions and sizes. The GC content ranges from <15 to >75% (1, 2), and the largest bacterial genomes are almost two orders of magnitude bigger than the smallest ones (3, 4). What factors determine this variation is a long-standing question in evolutionary biology, whose answer requires understanding the contributions of random mutation and genetic drift versus selection in shaping these genomic features. Mutational biases arise naturally from both the physicochemical properties of DNA (5) and the machinery involved with its replication, maintenance, and repair (6). Prominent biases include the higher frequency of deletions compared with insertions (7), and the overall predominance of GC to AT transitions among point mutations (8, 9). Despite these mutational pressures, large GC-rich genomes are widespread across bacterial phyla, which points to strong forces driving genomes away from their mutational equilibrium. Many adaptive explanations have been suggested, including biosynthetic costs (10, 11) and the greater stability of GC-rich sequences under high-temperature (12), oxidizing (13), and UV irradiation conditions (14).

Whatever the particular selection pressures, the efficiency of natural selection in counteracting mutational pressures is largely determined by two factors: population size and sexual recombination. It has long been understood that the fixation probabilities of beneficial, neutral, and deleterious mutations are more similar in small populations than in large ones, owing to the increased role of random genetic drift relative to selection in small populations (15). Differences in lifestyle naturally give rise to variation in population size and thereby the importance of drift. A textbook example is the extreme bottlenecks experienced by obligate endosymbionts, which exhibit highly reduced and AT-rich genomes (16)—the signatures expected when mutation dominates selection. While the importance of genetic drift has long been acknowledged, recent developments have also emphasized the effect of the extreme linkage disequilibrium in asexual organisms in limiting the role of selection (17). In the absence of recombination, highly beneficial mutations headed to fixation can drag along nonadaptive variation, a process dubbed "genetic draft" (18). Another consequence of extreme linkage is Hill–Robertson (or clonal) interference (19), whereby multiple beneficial mutations that arise in the same or different backgrounds compete for fixation, leading to the loss of many adaptive variants.

Besides relaxing selection, increasing the mutation rate is another way of altering the balance in favor of mutational biases. Genome-wide increases in the mutation rate of 10- and even 100-fold are readily caused by loss-of-function mutations in

Significance

Bacterial genomes are extremely diverse in size and composition. Biologists have long sought to explain such variability based on present-day selective and mutational forces. However, mutation rates can change dramatically over time, and experiments with hypermutable bacteria show that their genomes rapidly decay when propagated under the near absence of selection. Whether selection can prevent this decay is unclear. Here, we document the rapid genome decay of hypermutable bacteria even during tens of thousands of generations of sustained adaptation to a laboratory environment. These findings suggest the need to reexamine current ideas about the evolution of bacterial genomes, and they have implications for other hypermutable systems such as viruses and cancer cells.

Author contributions: D.S. and O.T. designed research; A.C., L.V.C., and O.T. performed research; A.C., L.V.C., C.F., T.H., J.-P.F. M.W., R.E.L., D.S., and O.T. contributed new reagents/analytic tools; A.C., L.V.C., M.W., D.S., and O.T. analyzed data; and A.C., L.V.C., R.E.L., and O.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: All of the new data and analysis scripts used in this study will be deposited in the Dryad Digital Repository upon acceptance (doi: 10.5061/dryad.sq67g).

[1]A.C. and L.V.C. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: olivier.tenaillon@inserm.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1705887114/-/DCSupplemental.

genes that encode the proteins involved in DNA replication, maintenance, and repair (6). Bacterial populations in laboratory evolution experiments often evolve hypermutable phenotypes (20–22). Hypermutators are also frequently observed among clinical isolates (23, 24), where they are considered a risk factor for drug therapy failure (25, 26). Hypermutability is also thought to be prevalent among obligate endosymbionts, because their genomes typically lack many of the mutation-preventing genes (16). More generally, bioinformatics evidence suggests that many other bacteria have repeatedly experienced periods of hypermutability during their evolutionary histories (27–32). These pulses of high mutation rate might leave a long-lasting signature in genomes, which complicates the interpretation of patterns based solely on the genetic and environmental circumstances of extant lineages. A key unresolved question concerns the potential for hypermutability to affect genomic features over long evolutionary timescales. A partial answer is provided by mutation accumulation experiments (MAEs), in which bacteria are repeatedly propagated through single-cell bottlenecks. Using this approach, Nilsson et al. (33) estimated that a mutator strain of *Salmonella* that was defective in DNA mismatch repair (MMR) can lose as much as 1 Mbp in as few as 365,000 generations. In another study, Lind and Andersson (34) projected that 500,000 generations might be enough time to reduce the GC content of mutators by as much as 1%, which was later confirmed by similar studies (35, 36). These estimates, however, were obtained from experiments designed to essentially eliminate the action of natural selection. Thus, it remains unclear whether these results can be extended to circumstances where selection is active and powerful. Here, we address this issue by analyzing genome sequence data from the *Escherichia coli* Long-Term Evolution Experiment (LTEE).

The LTEE provides a unique opportunity to examine the degree to which hypermutability can affect genomic patterns in the presence of sustained adaptive evolution. Twelve populations of *E. coli* B were started from the same ancestor and have been independently propagated for more than 60,000 generations in a minimal glucose-limited medium (37–39). Competition assays show that fitness has been steadily increasing over the course of the entire experiment (37, 38), while genome sequencing demonstrates on-going genetic evolution (39). Most relevant to this study, 6 of the 12 populations evolved hypermutable phenotypes at different times, including some in the first 3,000 generations (21, 39, 40). After the mutation rate increased, mutator lineages experienced a slight acceleration in their rate of adaptation (37, 38). Complete genome sequences at multiple time points are available for all of the populations (39), providing a unique opportunity to directly assess the competing effects of hypermutability and selection on genome evolution. Moreover, the study of *E. coli* provides three advantages. First, we can use methods that quantify the expected fitness effects of mutations to infer the contribution of selection (41). Second, we can refine our analysis of the genes involved in adaptation to the LTEE conditions by incorporating gene expression data and the results of transposon-mediated mutagenesis. Third, we can apply our analytical framework to other *E. coli* genomes that evolved under strongly contrasting regimes, including MAEs and evolution in nature (Fig. 1), to provide alternative contexts for comparison with the LTEE results. This approach allows us to distinguish the relative contributions of selection and mutation in shaping mutator genomes over tens of thousands of generations of adaptation.

Results

An Analytical Framework for Detecting Selection and Mutational Biases. The core data used in this study are single nucleotide variants (SNVs) in protein-coding genes from complete genome sequences obtained from three different sources: the LTEE, MAEs, and natural isolates (Fig. 1). Our aim is to identify the factors that best predict the likelihood of observing a given SNV. As explained below, we considered factors associated with mutational biases (e.g., neighboring bases) and with selection (e.g., gene essentiality). We assessed the contribution of these factors by logistic regressions and analyses of variance with single or multiple factors. Values are reported as odds ratios, which express the odds of observing a particular mutation given a specific condition compared with the odds of the same mutation occurring in the absence of that condition.

As factors associated with mutational biases, we considered the six possible types of base changes and the effects of the local sequence context. The six mutator populations of the LTEE show marked biases toward specific transitions and transversions (37, 40). Two populations have mutations affecting the oxidized guanine (GO) system, which results in an elevated rate of AT to CG transversions; the other four are defective in the MMR system, causing a preponderance of AT to GC and GC to AT transitions (Fig. 2). While these biases are specific to mutators, general biases also arise from the local properties of DNA sequences (35, 36) including, for instance, the propensity of CpG dinucleotides to generate GC to AT transitions (42) and the fact

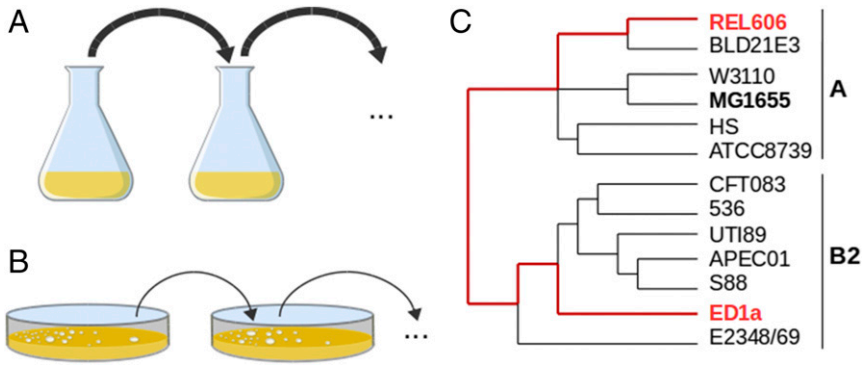

Fig. 1. Three evolutionary regimes from which the *E. coli* genomes analyzed in this study were derived. (*A*) Hypermutable populations that arose during the LTEE experienced high mutation rates and strong selection. (*B*) MAEs conducted with hypermutable strains experienced high mutation rates but weak selection. (*C*) Natural divergence: Strains diverged for millions of years in natural environments (before being isolated) experienced low mutation rates and strong selection. The diagram shows the relationships between strains belonging to two of several major *E. coli* phylogroups. REL606 is a derivative of *E. coli* B and the ancestor of the LTEE. MG1655, which belongs to the same group, is a derivative of *E. coli* K-12 and the ancestor of the MAEs analyzed in this study. The genome of ED1a is more distant from REL606 and was chosen for analyzing mutational differences that arose over long periods in nature. The red branches connecting them represent an estimated divergence time of >20 million years.

that consecutive pyrimidines are hotspots for UV-mediated dimerization (43). For simplicity, we restricted our analysis to the immediate neighborhood of the mutated base (*Materials and Methods*). To make meaningful comparisons, all of our analyses are restricted to SNVs arising from the three types of mutations that are elevated in the LTEE mutators. Mutator populations of the same type (GO- or MMR-defective) were combined to increase statistical power.

To capture the effects of selection, we included information from three types of proxies. First, we considered whether or not a given mutation resulted in an amino acid substitution. The assumption is that synonymous substitutions are selectively neutral or nearly so, although there are exceptions (44). Second, we examined whether mutations occurred in genes that are highly expressed, essential, or both. High expression and essentiality are associated with stronger purifying selection and hence slower protein evolution (45). For expression, we analyzed previously collected microarray-based transcriptomic data for *E. coli* B REL606 under the LTEE growth conditions and for *E. coli* K-12 MG1655 in rich LB medium (46). Also, highly expressed genes tend to have more highly optimized codon usage (47), and so we considered the codon adaptation index (48) of genes with mutations. For essentiality, we compiled a list of putatively essential genes by applying insertion sequencing (INSeq) (49) to a library of >100,000 independent insertion mutants propagated for 1 d under the conditions of the LTEE. In total, we identified 544 ORFs whose disruption was either lethal or strongly deleterious under the LTEE conditions (Table S1, Figs. S1–S3, and *SI Materials and Methods*).

Third, we sought to obtain an indirect estimate of the fitness effects of nonsynonymous substitutions using different measures of amino acid interchangeability. One metric is Grantham's distance of physicochemical difference (50), which has the highest predictive power of mutation effects among several distance matrices (51). We also used two alignment-based computational methods, including a generic Independent Sites Model (IND) and a Direct Coupling Analysis (DCA) (40). Alignment-based methods infer the effect of a substitution by quantifying statistically its degree of conservation across homologous sequences. The IND model considers each site independently, as in popular methods such as SIFT (52) and PolyPhen (53). In contrast, DCA includes information about multiple residues, thus capturing epistatic interactions, and has been shown to outperform context-independent methods when applied to exhaustive libraries of single-protein mutants (54). To test whether DCA improved performance at a genome-wide scale, we applied IND and DCA to >3,000 protein domains covering >1.5 Mb of the coding genome (*SI Materials and Methods*). The native sequence is expected to be highly optimized for most proteins, and therefore the accuracy of the methods can be compared by assessing how much their predictions deviate from optimality. As shown in Fig. 3, DCA predicts that most amino acid residues of the native sequences are optimal: 81% are deemed best and 9% are deemed second best, and only 4% of the possible amino acid changes are predicted to produce a protein with increased fitness. Importantly, DCA was performed without using *E. coli* genomes in the protein alignment. In contrast, the IND model implied greater deviations from optimality: Only 41% of the native amino acids residues are deemed best, and 15% of the possible amino acid changes are predicted to increase fitness. Nevertheless, both methods agreed in predicting that the vast majority of mutations would be deleterious (Fig. 3*C*).

Signature of Selection Prominent in Divergent Strains from Nature. We first tested the power of our analytical framework to detect the signature of selection. We computed the nucleotide differences between the shared core genomes of the LTEE ancestor REL606 and ED1a, a distant commensal strain with an estimated genome-wide divergence time of >20 million years (55). Divergence between these two strains presumably occurred under many selection pressures and without any evidence of intermittent periods of hypermutability. The mutational spectrum is highly skewed in both mutator types that evolved in the LTEE (Fig. 2); therefore, we restricted our analyses to AT to CG transversions (Fig. 4) and AT to GC transitions (Fig. 5), which are the dominant mutations in the GO- and MMR-defective

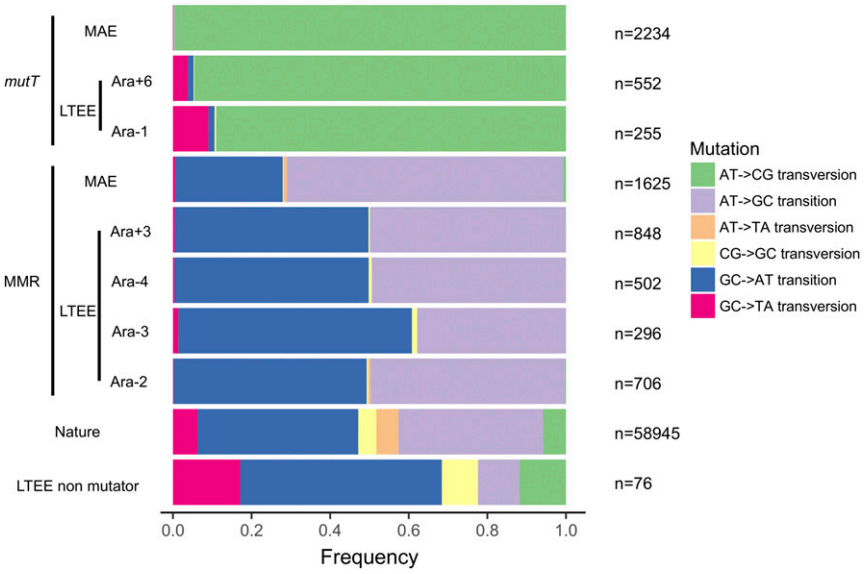

Fig. 2. Base substitution frequencies observed in genomes from the three evolutionary regimes studied here. Only synonymous mutations are included for all populations, except for those from the MAE regime, in which both nonsynonymous and synonymous mutations were included. The color code at right shows the six possible substitutions. Both mutator types show marked biases toward specific substitutions—AT to CG transversions for GO-defective (i.e., *mutT*) mutators, and AT to GC and GC to AT transitions for MMR-defective mutators—under both the MAE and LTEE regimes. The six LTEE populations that independently evolved one or the other mode of hypermutability are indicated by their population name (e.g., Ara+6). The various base substitutions are more evenly represented in the genomes derived under natural conditions. The total substitutions for each sample are indicated to the right of each distribution.

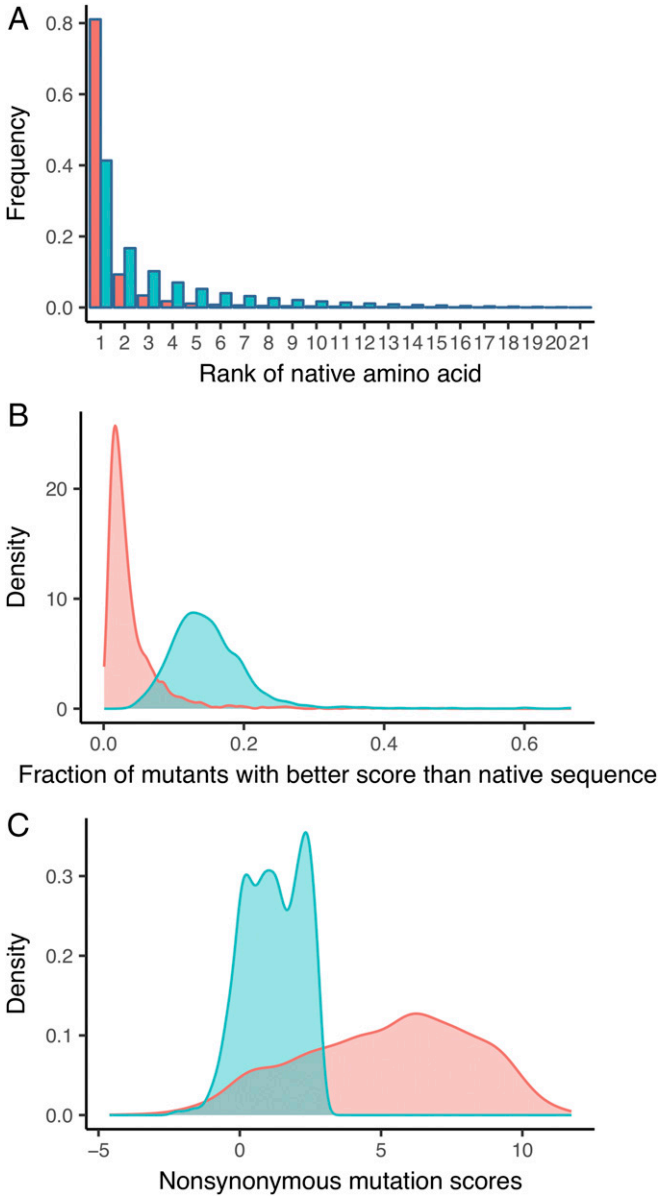

Fig. 3. Predictions of the fitness effects of nonsynonymous mutation using the nonepistatic IND (blue) and epistatic DCA (red) models. (*A*) Ranks of the native amino acid score, when comparing all 21 possible variants (including amino acid deletion) in the same position. (*B*) Distribution over all proteins of the fraction of single-residue mutants with better scores (i.e., higher predicted fitnesses) than the native sequence. For each protein, the scores for all amino acid substitutions were estimated, and the fraction of mutations that improved on the native sequence was calculated. (*C*) Distribution of scores for both methods shows that a large fraction of possible mutations produce an unlikely sequence (positive score) that probably corresponds to a low-fitness protein. Only a small fraction (negative scores) is predicted to be beneficial.

mutators, respectively. We also performed the same analyses on GC to AT transitions (the second-most dominant mutation in MMR-defective mutators), obtaining similar results (Fig. S4). Consistent with the expectation that selection was important under the natural conditions that prevailed during the divergence of these strains, the occurrence of a mutation is best explained by its effect on the amino acid sequence of the encoded protein. In particular, the likelihood of observing a synonymous mutation is 16.6 times that of a nonsynonymous mutation. By contrast, the influence of various mutational biases never reaches even a fourfold effect.

Focusing on nonsynonymous mutations, the most important factor is the fitness effect as predicted by DCA. The DCA scores (measured as the interquartile odds ratio; *Materials and Methods*) indicated an eightfold effect, on average. IND scores exhibited less than half that discrimination, with a 3.7-fold effect, while Grantham's distance metric showed even less predictive power, with only a 1.5-fold effect. Gene expression and essentiality had weaker but still significant effects. Mutations in essential genes were 1.6-fold less likely to occur than those in other genes ($P < 10^{-4}$), while the codon adaptation index, which serves as a proxy for high gene expression over evolutionary time, affected the prevalence of mutations by 1.3-fold ($P < 10^{-6}$). Of course, the transcriptomic profile of ED1a under natural conditions is unavailable, so we analyzed data from cells grown in the laboratory in LB, a permissive and rich medium. Gene expression in these conditions accounted for only a 1.2-fold effect ($P < 10^{-7}$). We also performed a multivariate analysis of variance to apportion the effects of these factors in a single analysis. The fraction of the variance explained by the model attributed to the combination of synonymous versus nonsynonymous status and DCA scores was 87%, whereas the local sequence context contributed 12%. Expression level and essentiality together accounted for only ∼0.1%.

Genomes from MAEs Show No Signature of Selection. We used data from published MAEs conducted using mutator derivatives of *E. coli* K-12 MG1655 strain (35, 36), a laboratory strain derived independently from a natural isolate, but phylogenetically closely related to the *E. coli* B ancestor of the LTEE (32, 56, 57) (Fig. 1). Importantly, these mutator derivatives of *E. coli* K-12 are GO- and MMR-defective, as are the mutator lineages that evolved in the LTEE. Applying our analytical framework to data from these MAEs with mutator strains provides a baseline against which to infer the signatures of selection in the LTEE mutator populations.

MAEs provide a well-known tool to study the rates at which various types of mutations occur. An MAE involves propagating populations through severe bottlenecks of one or a very few individuals. These extreme bottlenecks purge genetic diversity, thus impeding the action of natural selection. As a consequence, deleterious mutations accumulate irrespective of their fitness effects, except for those that are lethal or nearly so (58). Indeed, the near absence of selection in the MAE experiments with the mutator derivatives of *E. coli* K-12 is evidenced by the observed ratio of nonsynonymous to synonymous mutations (dN/dS), which was very close to 1 in both cases (35, 36). Moreover, the observed SNVs reflect the expected signatures of the main mutational biases (Fig. 2): The GO-deficient mutator genomes exhibited almost exclusively AT to CG mutations (99.5%), and the MMR-deficient mutators mostly experienced the two types of transitions (97.7%). Biases related to the local sequence context of the mutated base were also detected. In MMR-deficient mutators, a G in a GGG triplet had a mutation rate ∼10-fold higher than a G in a DGD triplet (where D stands for A, C, or T) (36).

Our analytical framework reveals the almost-complete absence of the signatures of selection in these MAEs, whereas the signatures of mutational bias are conspicuous. The effects of synonymous versus nonsynonymous status, DCA scores, and expression levels were all nonsignificant in both GO- and MMR-deficient mutators (all $P > 0.18$; Figs. 4 and 5). Gene essentiality was marginally significant for AT to CG in GO-deficient mutators ($P < 0.04$) and for MMR-deficient mutators ($P < 0.02$), but not for AT to GC in MMR-deficient mutators ($P > 0.75$). Conversely, the local genetic context was important, as adjacent bases as well as some specific triplets had significant effects of up to about eightfold on the occurrence of mutations. These effects were particularly strong for

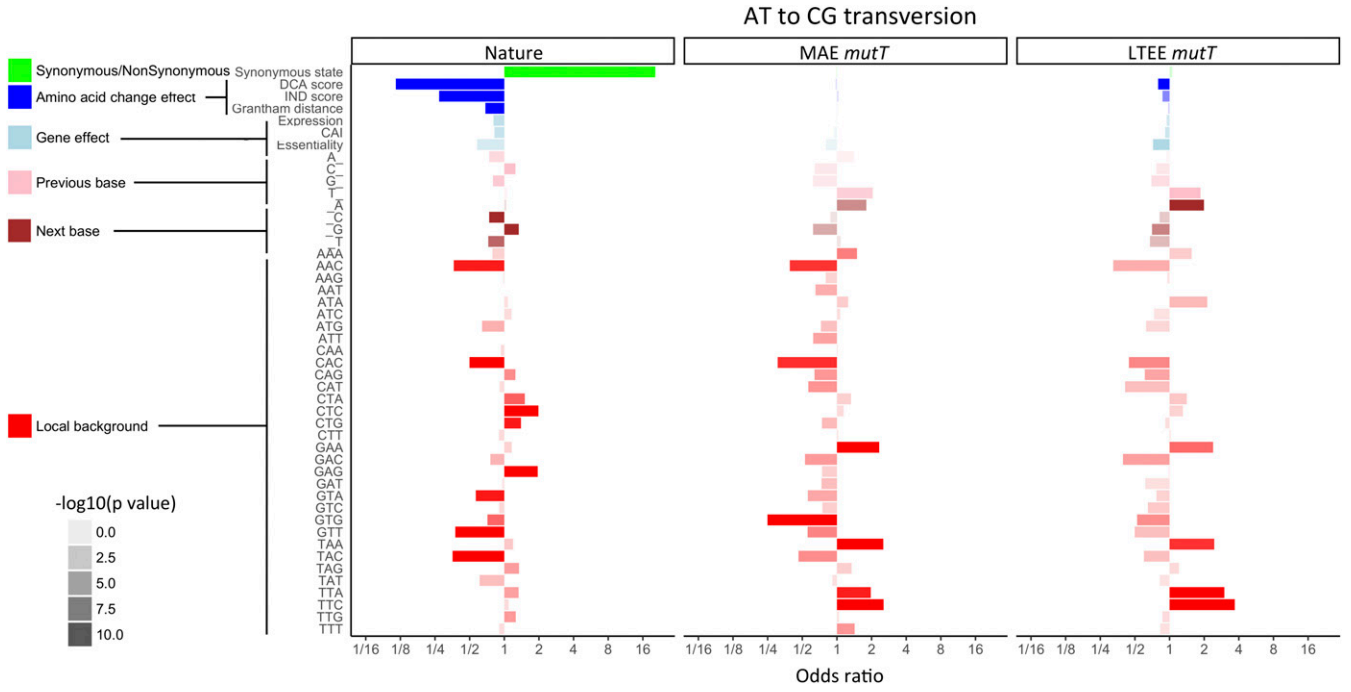

Fig. 4. Odds ratios in the three evolutionary regimes for AT to CG transversions, the typical signature of GO-defective (i.e., *mutT*) mutators. Odd ratios (shown on a log2 scale) show the importance of various factors in determining whether a particular base has experienced this type of mutation. For quantitative factors, interquartile odds ratios are shown. The green bar represents the synonymous versus nonsynonymous state. Dark blue bars show the impacts of amino acid changes based on their DCA score, IND score, and Grantham distance. Light blue bars represent factors at the gene level including gene expression level, Codon Adaptation Index (CAI), and essentiality. Pink and purple bars show the impact of the presence of a given base immediately before or after, respectively, the mutated base. Red bars represent the impact of the triplet in which the mutated base occurs. All background effects are measured on the leading strand. The intensity of each color is proportional to the log10-transformed *P* value of the test, as indicated on the scale bars. For the MAE and LTEE, only the GO-defective populations are included.

the MMR-defective mutators (Fig. 5), presumably because the typical errors made by the replicative polymerases were uncorrected without the MMR machinery (35).

The prominence of mutational biases was further indicated when computing the relative contribution of the various factors to the variance explained by a multivariate model. In striking

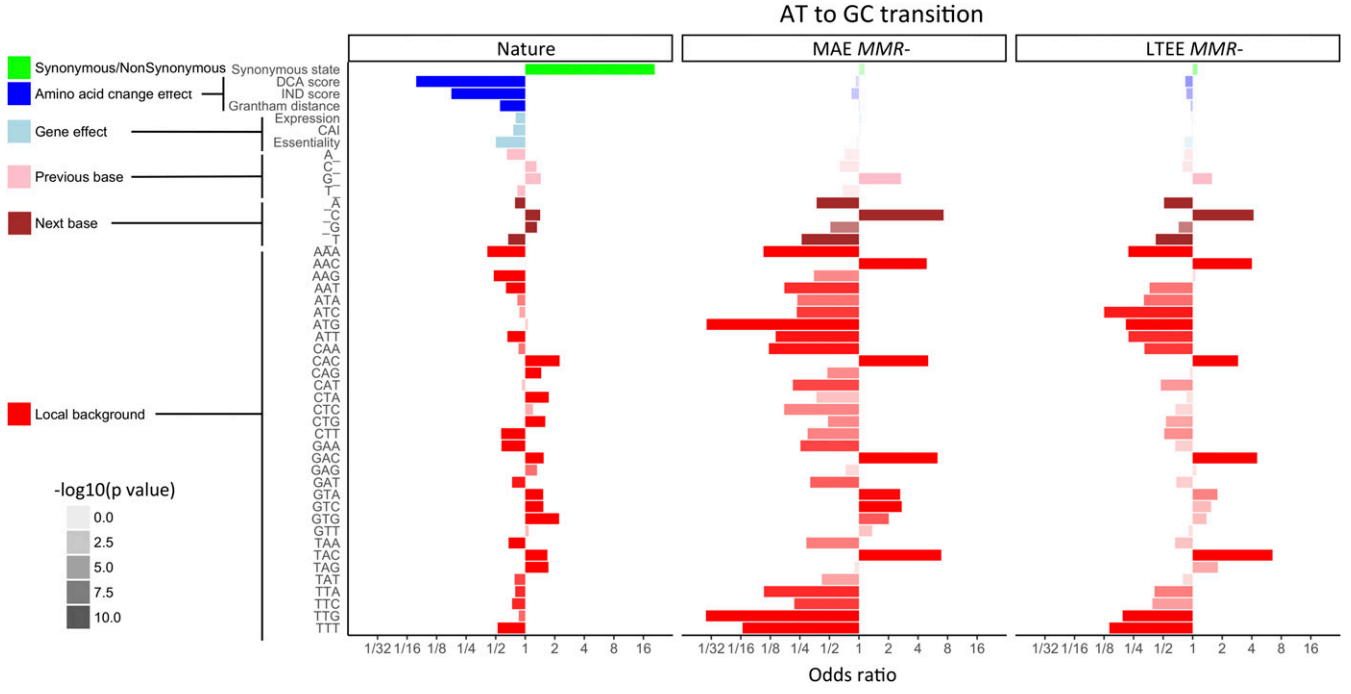

Fig. 5. Odds ratios in the three evolutionary regimes for AT to GC transitions, the most pronounced base substitution signature of MMR-defective mutators. Colors and legends are as described in Fig. 4. For the MAE and LTEE, only the MMR-defective populations are included.

contrast to the results obtained for the naturally diverged lineages, the factors related to selection explain only a tiny fraction of the variation: 2.1% for expression and essentiality, and 0.1% for synonymous versus nonsynonymous status and DCA, whereas the local genetic context explains ~97.8%.

Mutational Biases Dominate in LTEE Mutators Despite Sustained Adaptive Evolution. Having established the striking difference in genome features between lineages diverged in nature and mutator lineages subject to severe repeated bottlenecks, we applied our analytical framework to the hypermutable lineages that evolved during 50,000 generations of the LTEE. For each GO- and MMR-deficient type, we focused on the predominant nucleotide substitution (AT to CG and AT to GC, respectively), and we considered only SNVs that appeared after the emergence of the mutator phenotype. There is a striking overall similarity between the genomes derived from the MAEs and these hypermutable LTEE lineages (Figs. 4 and 5). Indeed, only two factors related to selection have significant effects (DCA score and essentiality in AT to CG transversions, Fig. 4). Although these factors are significant (both $P < 10^{-9}$), each of their effects was small at <1.5-fold. In contrast, many factors related to the local context had significant large effects, some with a greater than fourfold magnitude. In the multivariate regression analysis, variables associated with mutational biases explained fully 97.3% of the variance, whereas the selection-related variables explained only 0.4% (synonymous versus nonsynonymous status and DCA score) and 2.2% (gene expression and essentiality).

The small fraction of the genome required for growth in the simple and constant environment of the LTEE might contribute to the similarities between the mutator lineages in the LTEE and those in the MAEs. Indeed, the LTEE populations have accumulated many loss-of-function mutations affecting unused metabolic pathways (59, 60). To examine this possibility, we repeated our analysis using only the 544 essential and nearly essential genes that we identified through insertion sequencing (*SI Materials and Methods*). Consistent with the expectation of stronger purifying selection in those genes, we saw greater contributions of selection-related factors, including an increase to 13.6% in the effects of DCA score and synonymous versus nonsynonymous status on the overall variance explained. However, mutational biases still accounted for a much larger fraction of the variance at 85.6%.

To maximize sensitivity to potential signatures of selection, we then focused on nonsynonymous SNVs, which should include the vast majority of nonneutral point mutations. We investigated the relative contribution of the major factors related to selection (DCA scores) and mutation (local context) to the odds of AT to CG transversions, AT to GC transitions, and GC to AT transitions, i.e., the three most common substitutions (Fig. 6). Genomic evolution in both the LTEE and MAE mutators was dominated by mutational biases, even when the analysis was restricted to essential and nearly essential genes. We further examined the similarity between the signatures from the MAE and LTEE mutator lineages by comparing the correlations between the odds ratios across the different evolutionary regimes (Fig. 7). The odds ratios from the LTEE mutators exhibited much stronger correlations with those from MAE mutator lineages than with the lineages that diverged in nature, both for AT to CG transversions in the GO-deficient mutators and for transition mutations in the MMR-deficient mutators. The positive correlations between the LTEE mutators and the lineages diverged in nature ($r = 0.57$, $P = 0.001$, $n = 30$ for AT to CG mutations in GO-deficient mutators; $r = 0.69$, $P = 2.7 \times 10^{-10}$, $n = 64$ for transitions in MMR-deficient mutators) suggest that some mistakes caused by the oxidation of guanine and by errors of the replicative polymerases escape the action of the GO and MMR systems, respectively. Alternatively, the natural lineages may have had mutator phenotypes for some part of their histories.

Finally, it is worth noting that we performed our analyses using all observed mutations, without trying to make any distinction between those mutations that have, or have not, fixed in the population. Using all of the mutations provides greater statistical power. Also, the data were obtained by sequencing a relatively few clones from each population, making it difficult to ascertain whether a given mutation truly reached fixation. Moreover, in many cases, it is known that fixation did not occur, because at least three of the hypermutable populations harbored frequency-dependent polymorphisms for many generations, which prevented fixation at the population level (61). Nevertheless, to get an idea of the extent to which the distinction between transient and fixed mutations might be important, we repeated the analysis on the subset of mutations that were observed just once (i.e., singletons) and compared the results with those obtained for mutations that were present in the populations for at least 10,000 generations (including both fixed mutations and long-lasting polymorphisms). No differences were found between the two groups, and both were similar to the analyses that included all mutations (Figs. S5–S7).

Discussion

In this study, we examined whether hypermutability can leave a lasting signature in bacterial genomes, even when they are undergoing adaptive evolution. We developed an analytical framework to distinguish the effects of selection from mutation on genome evolution, and we applied it to genomes evolved under combinations of strong or relaxed selection with high and low mutation rates. We showed that, despite the continued action of natural selection, mutational biases dominate the signature in genomes derived from mutator populations in an LTEE with *E. coli* (Figs. 4 and 5). By comparison, the signature of selection is sufficiently weak that these adapted genomes are essentially indistinguishable from those derived from mutation accumulation experiments, which are designed to impede the action of selection. This similarity emerges despite orders-of-magnitude differences in effective population size (~1 versus $\sim 10^7$ cells) and the contrasting fitness trajectories in the two experimental settings (declining versus increasing fitness).

These results are amenable to a straightforward interpretation in the light of the recently developed theory of genetic draft (19). Experiments show that, when beneficial mutations are common and recombination is absent, the fate of most alleles is largely driven by their chance association with strongly beneficial driver mutations (62, 63). Alleles will experience random fates, increasing in frequency if they are linked to beneficial mutations, but declining to extinction otherwise. In large asexual populations, selective sweeps often involve the fixation of multiple beneficial drivers, accompanied by a cohort of passengers that may be neutral or even slightly deleterious (63). This process was called genetic draft (18) by way of comparison with genetic drift, which is the typical source of stochasticity in allele frequency that does not involve selective sweeps. While genetic draft and genetic drift operate under different regimes, both processes limit the efficacy of selection on sorting beneficial and deleterious genetic variation. When these forces are sufficiently strong, many genomic sites eventually fix random alleles, provided they are not lethal or, in the case of draft, that their fitness effects are small compared with those of driver mutations (64). This random accumulation is readily observed in MAEs, which, by design, are dominated by genetic drift. Our results show that hypermutability amplifies the effects of genetic draft, even under the strong-selection conditions of the LTEE, evidently because an elevated mutation rate enlarges the size of the passenger cohort. Our observations thus show how these two different processes—draft and drift—can lead to the same phenomenon, i.e., the near elimination of the signature of selection from bacterial genomes.

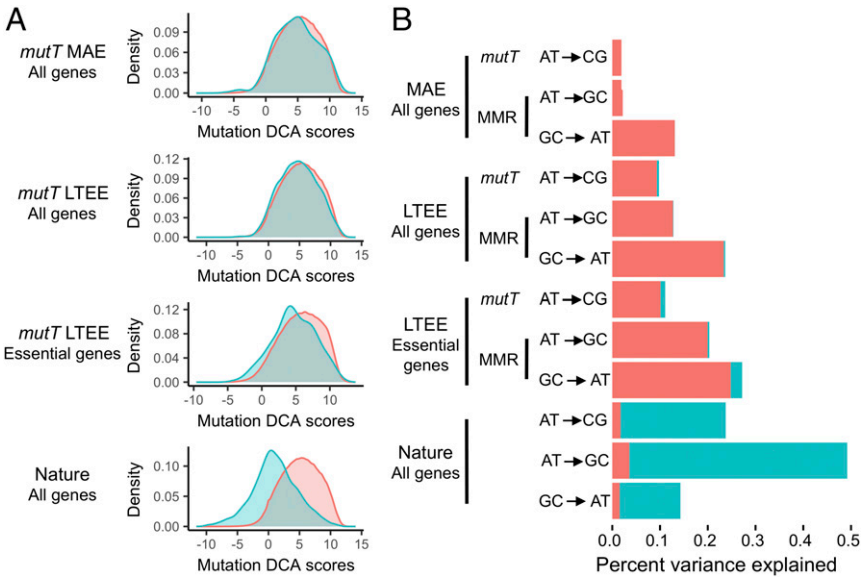

Fig. 6. Comparison of contributions of a selection-related factor (DCA score measured on nonsynonymous mutations) and local mutational effects. (*A*) Distribution of DCA scores in mutated (blue) and unmutated (red) sites. A score of *0* suggests no impact of the mutation on fitness, a positive score suggests a deleterious effect, and a negative score suggests a beneficial effect. Distributions are shown for AT to CG transversions in the case of the GO-defective (i.e., *mutT*) MAE and LTEE populations, and for all mutations in the case of the strains that diverged in nature over millions of years. For the LTEE populations, results are also shown separately for the subset of mutations that affect genes that are essential under the LTEE conditions. (*B*) Contribution of selection- and mutation-related effects to the probability of nonsynonymous sites being mutated for the three mutation types enriched in mutator populations under each of the three evolutionary regimes. For the LTEE populations, results are again shown separately for those mutations affecting essential genes. The selection-related DCA score strongly influences genomic evolution in nature, whereas mutational biases have greater influences in the MAE and LTEE populations, even when the analysis is restricted to essential genes.

Other evidence also indicates the important role of genetic draft during the LTEE. There is pervasive clonal interference in the LTEE (37, 61, 65) and other microbial evolution experiments (61–69), at least during the early stages of adaptation. On a longer timescale, both genome analyses and fitness trajectories show that the supply of beneficial mutations remained substantial even after 50,000 generations of the LTEE (37–39). Both gene-level convergence across the LTEE populations and the dynamics of genome evolution (39) also indicate that dozens of the mutations that reached high frequency were beneficial,

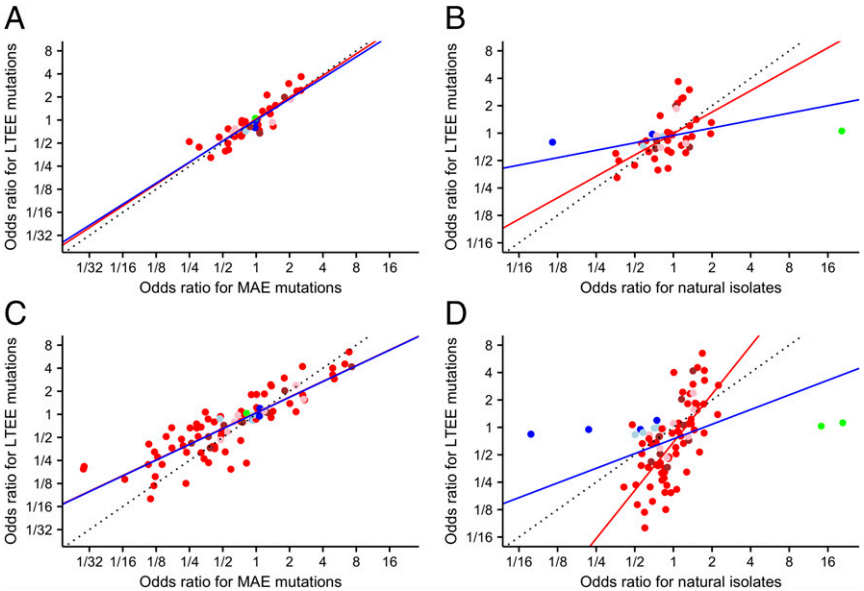

Fig. 7. Correlation between odds ratios for factors associated with mutational patterns observed for different evolutionary regimes. (*A*) Correlation between GO-defective (i.e., *mutT*) mutator populations under the LTEE and MAE regimes. (*B*) Correlation between MMR-defective mutator populations under the LTEE and MAE regimes. (*C*) Correlation between GO-defective (i.e., *mutT*) mutator populations in the LTEE and lineages that diverged for millions of years in nature. (*D*) Correlation between MMR-defective mutator populations in the LTEE and lineages that diverged for millions of years in nature. In each panel, the blue line shows the overall correlation using all factors; the red line shows the correlation restricted to the local mutational biases; and the dotted line shows the hypothetical perfect correlation. Symbol colors are the same as used in Figs. 4–6.

including nonsynonymous and intergenic point mutations as well as insertions and deletions (39). There is no horizontal gene transfer in the LTEE populations; therefore, any beneficial mutation that rose to high frequency necessarily drafted the rest of the genome with it. In nonmutator populations, there are usually relatively few passenger mutations to be drafted (39, 61). In mutator populations, however, every new beneficial driver that escapes drift loss (while it is rare) drafts with it so many additional, nonadaptive mutations that the signatures of selection are effectively concealed. Hence, despite several dozen beneficial mutations that increase mean fitness, the genome-wide molecular evolution in mutators looks remarkably similar to what occurs during random mutation accumulation (Figs. 2 and 4–6).

In contrast to the mutator populations in both the LTEE and MAEs, the signature of selection is extremely strong when analyzing the divergence between *E. coli* strains REL606 and ED1a, which diverged for millions of years in nature. In particular, most substitutions are either synonymous or nonsynonymous with weak predicted effects on protein structure (Figs. 4–6). It seems, therefore, that purifying selection is much more efficient in nature than under the LTEE conditions. There are several possible (and not mutually exclusive) explanations for this observation. First, both strains from nature possess a complete set of functional genes for avoiding mutations, although that might not have always been the case. It remains unclear to what extent genetic draft under normal mutation rates might also mask the signature of selection; we could not test this possibility with nonmutator LTEE populations, due to the lack of statistical power. Second, fluctuating conditions in nature likely prevent the completion of many selective sweeps, especially those that are adaptive locally and for a short period but that are maladaptive across larger scales of space and time (70). Moreover, millions of years of evolution in variable environments may have promoted plastic responses that would further reduce the frequency of selective sweeps (71). Third, horizontal gene transfer can break the linkage between driver and passenger mutations, disrupting the whole process of genetic draft. Horizontal gene transfer is not infrequent in natural populations of *E. coli* species. It has been estimated that a given nucleotide difference between natural isolates is many times more likely to have resulted from recombination than from de novo mutation (56, 72).

A further issue is that past periods of hypermutability may have played an important role in shaping current-day genomes. It should be emphasized that the same conditions that give rise to genetic draft—namely, a high supply of beneficial mutations and the absence of recombination—also favor the evolution of hypermutability (19, 73). However, low mutation rates are expected to reevolve once these conditions no longer hold, provided that appropriate alleles are available. Mechanisms that would allow low mutation rates to reevolve include true reversion (74), second-site compensation (40), and horizontal gene transfer (28). In an ever-changing world, bacteria have been hypothesized to experience intermittent periods of high and low mutation rates. Support for this view comes from the discovery that some genes that affect mutation rates are enriched in recombination hotspots (27–32) as well as nucleotide repeats that could facilitate rapid changes in mutation rates (74). As shown here, even periods as short as a few decades can significantly alter genomic patterns. Moreover, genetic drift and genetic draft may act concomitantly in nature. For example, pathogen populations often face severe bottlenecks during transmission between hosts, followed by a variety of adaptive challenges, including colonization of the new host, avoidance of the immune system, and escape from antibiotic therapy (75–77).

Our work shows how genomic patterns that are sometimes attributed to weak selection and genetic drift can emerge in asexual populations even while they are rapidly adapting to their environment. This finding might explain, for instance, the much-discussed apparent contradictions in the genus *Prochlorococcus* (78). These cyanobacteria have GC-poor and reduced genomes, features typically associated with strong genetic drift. However, their ocean-wide distribution and massive populations should severely limit the effect of drift with respect to genome decay (78). It has been hypothesized that these genomic properties could have arisen as a consequence of the absence of DNA repair genes in their genomes (79), but direct measurements of *Prochlorococcus* mutation rates do not indicate a hypermutable phenotype (80). These discrepant observations might be reconciled by a scenario that combines strong genetic draft in a hypermutable progenitor of *Prochlorococcus* followed by a more recent compensation event that reestablished a low mutation rate. More generally, our results should be applicable to any large population subject to high mutation rates, absent or very infrequent recombination, and strong selection. As such, our findings may also contribute to efforts to better understand genome evolution in obligate endosymbionts (16), RNA viruses (81), and cancer cells (82).

Several important issues require further research to resolve. First, theory shows that recombination impedes both genetic draft and second-order selection for hypermutability (19, 83). The fact that both processes are affected raises the possibility that the phenomena described in this work may be highly sensitive to even small amounts of horizontal gene transfer. In principle, this issue should be amenable to study via computer simulation. Second, the fastest genome decay rates we observed in the LTEE amount to 1% change in GC content in ~400,000 generations (GO-defective strains) and 1 Mbp loss in ~600,000 generations (MMR-defective strains), which are in the range of previous estimates inferred from MAEs (33, 34). Given these rates, some present-day genomic patterns might be explained by past episodes of hypermutability, but many questions remain unanswered, including, How frequent are these episodes, how long do they typically last, and, crucially, how quickly would genomes recover after the restoration of normal mutation rates? Fittingly, the LTEE might provide insight into these questions, as several of the mutator populations are in the process of partially compensating or reverting to their low ancestral mutation rate (40). Third, it is unknown to what extent draft can mask the signature of selection at low mutation rates. For this masking to occur, the mutational cohorts that reach fixation must typically include more nonadaptive passenger mutations than beneficial drivers. Interestingly, an excess of nonadaptive passengers was recently observed during the first stages of adaptation of nonmutator yeast populations to a laboratory environment (84). This observation suggests that draft-induced decay might occur even in nonmutator asexual populations, but such decay might be prevented if the relevant timescale is too long compared with the typical frequency of recombination events. Further theoretical and empirical work is needed to gain insight into these questions.

Finally, our results also have certain implications for biotechnology. Researchers working on biotechnological applications may be tempted to use mutator strains or chemical mutagenesis to accelerate strain improvement (85, 86). Consistent with that strategy, the LTEE populations that evolved mutator phenotypes have reached slightly higher fitness levels than those that did not (37, 38). However, that increased mutability also comes with potential costs, as shown, in this study, by the tendency toward genomic decay and, in earlier studies, by reduced performance under other environmental conditions (59, 60). Therefore, unless one seeks to optimize performance under very constant and precisely controlled conditions, the accelerated rate of adaptation might not be worth the resulting genomic decay and associated loss of versatility.

Materials and Methods

Mutation List. We restricted all our genomic analyses to protein-coding regions. Mutations from the LTEE were retrieved from the 264 genomes sequenced in Tenaillon et al. (39). For the six LTEE mutator populations, we collated all of the independent point mutations that occurred along the phylogenetic branches with mutator phenotypes. Some analyses were also performed on the subset of those mutations present at 50,000 generations. Mutator populations with similar mutation-promoting defects were pooled: the two GO-defective populations, Ara−1 and Ara+6, caused by mutations in *mutT*; and the four MMR-defective populations, Ara−2, Ara−3, Ara−4, and Ara+3, caused by mutations in *mutS* and *mutL*. For MAE populations, we retrieved published sets of mutations from GO-defective and MMR-defective populations (35, 36). For mutations that distinguish strains REL606 and ED1a (Fig. 1*C*), we first found homologous gene pairs using the Usearch algorithm (87). Exclusive pairs of homologs with at least 95% similarity were retained, and the differences between the two alleles were recorded. To apply the same algorithm as used in the other cases, the mutations were oriented from the REL606 state to the ED1a state. To ensure accuracy and consistency when inferring the background context and amino acid changes that occurred, only those mutations without others in the close neighborhood were considered; specifically, no mutations in an adjacent (previous or next) base or in the same codon were tolerated. This approach filtered out about 10% of the mutations.

Computing Possible and Observed Mutations. Several parameters were recorded for each possible genomic mutation. The local background effect on mutations has been shown to be strand-dependent (88), and so the effect was computed for the leading strand. The origin and terminus were defined to be the *oriC* and *dif* sites, respectively. Local background effects included the previous base, the next base, and the triplet in which the base occurred (centered on the base in question). DCA and IND scores depend on the identification of protein domains using the Pfam database (89). The Pfam domains of *E. coli* strain MG1655 proteins are known (89) and were used to identify the Pfam domains in strain REL606. Homologous genes between MG1655 and REL606 were found using Usearch clustering, and the corresponding Pfam annotations were transferred.

To compute all possible protein-coding mutations in the genome, we wrote a custom Perl program. For every coding site in a gene, all three possible point mutations were generated, and each one's impact on the coding sequence was computed and combined with information on the affected gene, such as its expression. As a result, each coding site was represented by three lines in a large table, with each line corresponding to a specific mutation. Each mutation line included the genome position, the affected gene, its expression level, its codon adaptation index, its essentiality, the mutation type, its local background (i.e., the three bases centered on the site in question), the previous and next bases (computed on the leading strand), the affected (i.e., unmutated) codon, the resulting (i.e., mutated) codon, the synonymous versus nonsynonymous status of the mutation, the amino acid change, its effect on Grantham distance, its effect on IND score, its effect on DCA score, and, finally, whether or not this specific mutation was observed in the relevant dataset.

We computed possible mutations using only the genomic regions where mutations could be reliably called based on the information available. For the LTEE and MAEs, this strategy excluded repeated regions where mutations cannot be called based on the short reads used for genome sequencing. For the natural isolates, we used only coding regions of homologous genes conserved between the two strains. In all datasets, we excluded codons in overlapping reading frames.

Statistical Analysis. We analyzed the large mutation tables generated by the custom Perl program (described in *Computing Possible and Observed Mutations*) in the R environment for statistical computing (90). We performed Fisher's exact test and logistic regression to infer the impact of each factor on the probability of genomic sites with a particular feature being mutated in a dataset. Each type of mutation was analyzed independently. For qualitative factors (e.g., synonymous or not), a contingency table was produced, and the log2 odds-ratio of the Fisher's test and the log10 of the corresponding *P* value were retained. For quantitative or multifactorial data, a binomial logistic regression was performed. For example, to measure the contribution of the DCA score, the model could be written as Mutated ~ DCA, in which "Mutated" is the column telling whether or not the mutation has been observed and "DCA" is the column with the DCA score of each mutation. The result is a regression that expresses how the DCA score affects the logit of the probability of a site being mutated: $\ln[p/(1-p)] = a\,\mathrm{DCA} + b$. To transform the result into an odds ratio that can be compared with the odds ratio for qualitative factors, we used the regression to compute the logistic change in the probability of a site being mutated associated with a transition from the first to the third quartile of the DCA score's distribution. This method is called interquartile odds ratio.

For multivariate analysis, we also used a logistic regression, but with two or more factors in the model. Using both synonymous and nonsynonymous mutations, we could, for instance, test the model: *Mutated* ~ *Background* + *Synonymous* + *DCA* + *Expression* + *Essentiality*. In that case, rather than comparing *P* values obtained for each separate factor, we performed an analysis of variance and computed the contribution of each factor to the fraction of the variance explained by the model. The datasets we analyzed include only thousands of mutations over millions of possible mutations, and so the overall variance explained by the models is very low: between 0.5% and 2.5% for the natural strains, between 0.10% and 0.25% for the LTEE, and between 0.01% and 0.15% for the MAEs. Nevertheless, the relative contributions of the different factors are informative.

ACKNOWLEDGMENTS. We thank Harry Kemble and two anonymous reviewers for their helpful comments on the manuscript. This work was supported by the European Research Council (ERC) under the European Union's 7th Framework Program, ERC Grant 310944 (to O.T.) and Grants FP7-ICT-2013.9.6 and ICT-610427 (to D.S.); by Université Grenoble Alpes (D.S.); Centre National de la Recherche Scientifique (D.S.); National Science Foundation Grant DEB-1451740 (to R.E.L.); Agence Nationale de la Recherche Grant ANR-13-BS04-0012-01 (to M.W.); and Université Pierre and Marie Curie (C.F. and M.W.). L.V.C. was supported by a PhD Fellowship from the École Normale Supérieure Cachan.

1. McCutcheon JP, Moran NA (2010) Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biol Evol* 2:708–718.
2. Thomas SH, et al. (2008) The mosaic genome of *Anaeromyxobacter dehalogenans* strain 2CP-C suggests an aerobic common ancestor to the delta-proteobacteria. *PLoS One* 3:e2103.
3. Han K, et al. (2013) Extraordinary expansion of a *Sorangium cellulosum* genome from an alkaline milieu. *Sci Rep* 3:2101.
4. Bennett GM, Moran NA (2013) Small, smaller, smallest: The origins and evolution of ancient dual symbioses in a Phloem-feeding insect. *Genome Biol Evol* 5:1675–1688.
5. Duncan BK, Miller JH (1980) Mutagenic deamination of cytosine residues in DNA. *Nature* 287:560–561.
6. Miller JH (1996) Spontaneous mutators in bacteria: Insights into pathways of mutagenesis and repair. *Annu Rev Microbiol* 50:625–643.
7. Mira A, Ochman H, Moran NA (2001) Deletional bias and the evolution of bacterial genomes. *Trends Genet* 17:589–596.
8. Balbi KJ, Rocha EPC, Feil EJ (2009) The temporal dynamics of slightly deleterious mutations in *Escherichia coli* and *Shigella spp*. *Mol Biol Evol* 26:345–355.
9. Hershberg R, Petrov DA (2010) Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet* 6:e1001115.
10. Rocha EPC, Danchin A (2002) Base composition bias might result from competition for metabolic resources. *Trends Genet* 18:291–294.
11. Seligmann H (2003) Cost-minimization of amino acid usage. *J Mol Evol* 56:151–161.
12. Kagawa Y, et al. (1984) High guanine plus cytosine content in the third letter of codons of an extreme thermophile. DNA sequence of the isopropylmalate dehydrogenase of *Thermus thermophilus*. *J Biol Chem* 259:2956–2960.
13. Naya H, Romero H, Zavala A, Alvarez B, Musto H (2002) Aerobiosis increases the genomic guanine plus cytosine content (GC%) in prokaryotes. *J Mol Evol* 55:260–264.
14. Singer CE, Ames BN (1970) Sunlight ultraviolet and bacterial DNA base ratios. *Science* 170:822–825.
15. Crow JF (2010) Wright and Fisher on inbreeding and random drift. *Genetics* 184: 609–611.
16. McCutcheon JP, Moran NA (2011) Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 10:13–26.
17. Betancourt AJ, Welch JJ, Charlesworth B (2009) Reduced effectiveness of selection caused by a lack of recombination. *Curr Biol* 19:655–660.
18. Gillespie JH (2000) Genetic drift in an infinite population. The pseudohitchhiking model. *Genetics* 155:909–919.
19. Neher RA (2013) Genetic draft, selective interference, and population genetics of rapid adaptation. *Annu Rev Ecol Evol Syst* 44:195–215.
20. Douillard FP, et al. (2016) Polymorphisms, chromosomal rearrangements, and mutator phenotype development during experimental evolution of *Lactobacillus rhamnosus* GG. *Appl Environ Microbiol* 82:3783–3792.
21. Sniegowski PD, Gerrish PJ, Lenski RE (1997) Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* 387:703–705.
22. Notley-McRobb L, Pinto R, Seeto S, Ferenci T (2002) Regulation of mutY and nature of mutator mutations in *Escherichia coli* populations under nutrient limitation. *J Bacteriol* 184:739–745.
23. Denamur E, et al. (2002) High frequency of mutator strains among human uropathogenic *Escherichia coli* isolates. *J Bacteriol* 184:605–609.

24. Couce A, Alonso-Rodriguez N, Costas C, Oliver A, Blázquez J (2016) Intrapopulation variability in mutator prevalence among urinary tract infection isolates of *Escherichia coli*. *Clin Microbiol Infect* 22:566.e1–566.e7.
25. Chopra I, O'Neill AJ, Miller K (2003) The role of mutators in the emergence of antibiotic-resistant bacteria. *Drug Resist Updat* 6:137–145.
26. Jolivet-Gougeon A, et al. (2011) Bacterial hypermutation: Clinical implications. *J Med Microbiol* 60:563–573.
27. Eisen JA, Hanawalt PC (1999) A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat Res* 435:171–213.
28. Denamur E, et al. (2000) Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell* 103:711–721.
29. Rocha EPC, Touchon M, Feil EJ (2006) Similar compositional biases are caused by very different mutational effects. *Genome Res* 16:1537–1547.
30. Fall S, et al. (2007) Horizontal gene transfer regulation in bacteria as a "spandrel" of DNA repair mechanisms. *PLoS One* 2:e1055.
31. Brown EW, LeClerc JE, Li B, Payne WL, Cebula TA (2001) Phylogenetic evidence for horizontal transfer of *mutS* alleles among naturally occurring *Escherichia coli* strains. *J Bacteriol* 183:1631–1644.
32. Elena SF, Whittam TS, Winkworth CL, Riley MA, Lenski RE (2005) Genomic divergence of *Escherichia coli* strains: Evidence for horizontal transfer and variation in mutation rates. *Int Microbiol* 8:271–278.
33. Nilsson AI, et al. (2005) Bacterial genome size reduction by experimental evolution. *Proc Natl Acad Sci USA* 102:12112–12116.
34. Lind PA, Andersson DI (2008) Whole-genome mutational biases in bacteria. *Proc Natl Acad Sci USA* 105:17878–17883.
35. 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 USA* 109:E2774–E2783.
36. 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:E5990–E5999.
37. Wiser MJ, Ribeck N, Lenski RE (2013) Long-term dynamics of adaptation in asexual populations. *Science* 342:1364–1367.
38. Lenski RE, et al. (2015) Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with *Escherichia coli*. *Proc Biol Sci* 282:20152292.
39. Tenaillon O, et al. (2016) Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* 536:165–170.
40. Wielgoss S, et al. (2013) Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc Natl Acad Sci USA* 110:222–227.
41. Morcos F, et al. (2011) Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc Natl Acad Sci USA* 108:E1293–E1301.
42. Coulondre C, Miller JH, Farabaugh PJ, Gilbert W (1978) Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature* 274:775–780.
43. Hall JD, Mount DW (1981) Mechanisms of DNA replication and mutagenesis in ultraviolet-irradiated bacteria and mammalian cells. *Prog Nucleic Acid Res Mol Biol* 25:53–126.
44. Bailey SF, Hinz A, Kassen R (2014) Adaptive synonymous mutations in an experimentally evolved *Pseudomonas fluorescens* population. *Nat Commun* 5:4076.
45. Pál C, Papp B, Hurst LD (2001) Highly expressed genes in yeast evolve slowly. *Genetics* 158:927–931.
46. Feugeas J-P, et al. (2016) Links between transcription, environmental adaptation and gene variability in *Escherichia coli*: Correlations between gene expression and gene variability reflect growth efficiencies. *Mol Biol Evol* 33:2515–2529.
47. Drummond DA, Wilke CO (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134:341–352.
48. Sharp PM, Li WH (1987) The codon Adaptation Index–A measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 15:1281–1295.
49. Goodman AL, et al. (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* 6:279–289.
50. Grantham R (1974) Amino acid difference formula to help explain protein evolution. *Science* 185:862–864.
51. Jacquier H, et al. (2013) Capturing the mutational landscape of the beta-lactamase TEM-1. *Proc Natl Acad Sci USA* 110:13067–13072.
52. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4:1073–1081.
53. Adzhubei IA, et al. (2010) A method and server for predicting damaging missense mutations. *Nat Methods* 7:248–249.
54. Figliuzzi M, Jacquier H, Schug A, Tenaillon O, Weigt M (2016) Coevolutionary landscape inference and the context-dependence of mutations in beta-lactamase TEM-1. *Mol Biol Evol* 33:268–280.
55. Clermont O, et al. (2008) Evidence for a human-specific *Escherichia coli* clone. *Environ Microbiol* 10:1000–1006.
56. Studier FW, Daegelen P, Lenski RE, Maslov S, Kim JF (2009) Understanding the differences between genome sequences of *Escherichia coli B* strains REL606 and BL21 (DE3) and comparison of the *E. coli B* and *K-12* genomes. *J Mol Biol* 394:653–680.
57. Dixit PD, Pang TY, Studier FW, Maslov S (2015) Recombinant transfer in the basic genome of *Escherichia coli*. *Proc Natl Acad Sci USA* 112:9070–9075.
58. Barrick JE, Lenski RE (2013) Genome dynamics during experimental evolution. *Nat Rev Genet* 14:827–839.
59. Cooper VS, Lenski RE (2000) The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* 407:736–739.
60. Leiby N, Marx CJ (2014) Metabolic erosion primarily through mutation accumulation, and not tradeoffs, drives limited evolution of substrate specificity in *Escherichia coli*. *PLoS Biol* 12:e1001789.
61. Maddamsetti R, Lenski RE, Barrick JE (2015) Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with *Escherichia coli*. *Genetics* 200:619–631.
62. Arjan JA, et al. (1999) Diminishing returns from mutation supply rate in asexual populations. *Science* 283:404–406.
63. Desai MM, Fisher DS, Murray AW (2007) The speed of evolution and maintenance of variation in asexual populations. *Curr Biol* 17:385–394.
64. Schiffels S, Szöllosi GJ, Mustonen V, Lässig M (2011) Emergent neutrality in adaptive asexual evolution. *Genetics* 189:1361–1375.
65. Woods RJ, et al. (2011) Second-order selection for evolvability in a large *Escherichia coli* population. *Science* 331:1433–1436.
66. Barroso-Batista J, et al. (2014) The first steps of adaptation of *Escherichia coli* to the gut are dominated by soft sweeps. *PLoS Genet* 10:e1004182.
67. Lang GI, et al. (2013) Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* 500:571–574.
68. Lee M-C, Marx CJ (2013) Synchronous waves of failed soft sweeps in the laboratory: Remarkably rampant clonal interference of alleles at a single locus. *Genetics* 193:943–952.
69. Levy SF, et al. (2015) Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* 519:181–186.
70. Harrison E, Laine A-L, Hietala M, Brockhurst MA (2013) Rapidly fluctuating environments constrain coevolutionary arms races by impeding selective sweeps. *Proc Biol Sci* 280:20130937.
71. Hallsson LR, Björklund M (2012) Selection in a fluctuating environment leads to decreased genetic variation and facilitates the evolution of phenotypic plasticity. *J Evol Biol* 25:1275–1290.
72. Guttman DS, Dykhuizen DE (1994) Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science* 266:1380–1383.
73. Tenaillon O, Toupance B, Le Nagard H, Taddei F, Godelle B (1999) Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. *Genetics* 152:485–493.
74. Shaver AC, Sniegowski PD (2003) Spontaneously arising *mutL* mutators in evolving *Escherichia coli* populations are the result of changes in repeat length. *J Bacteriol* 185:6076–6082.
75. Abel S, Abel zur Wiesch P, Davis BM, Waldor MK (2015) Analysis of bottlenecks in experimental models of infection. *PLoS Pathog* 11:e1004823.
76. Moxon ER, Murphy PA (1978) *Haemophilus influenzae* bacteremia and meningitis resulting from survival of a single organism. *Proc Natl Acad Sci USA* 75:1534–1536.
77. Lieberman TD, et al. (2011) Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat Genet* 43:1275–1280.
78. Partensky F, Garczarek L (2010) Prochlorococcus: Advantages and limits of minimalism. *Annu Rev Mar Sci* 2:305–331.
79. Marais GAB, Calteau A, Tenaillon O (2008) Mutation rate and genome reduction in endosymbiotic and free-living bacteria. *Genetica* 134:205–210.
80. Osburne MS, Holmbeck BM, Coe A, Chisholm SW (2011) The spontaneous mutation frequencies of *Prochlorococcus* strains are commensurate with those of other bacteria. *Environ Microbiol Rep* 3:744–749.
81. Koonin EV, Dolja VV, Krupovic M (2015) Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology* 479-480:2–25.
82. Podlaha O, Riester M, De S, Michor F (2012) Evolution of the cancer genome. *Trends Genet* 28:155–163.
83. Tenaillon O, Le Nagard H, Godelle B, Taddei F (2000) Mutators and sex in bacteria: Conflict between adaptive strategies. *Proc Natl Acad Sci USA* 97:10465–10470.
84. Buskirk SW, Peace RE, Lang GI (2017) Hitchhiking and epistasis give rise to cohort dynamics in adapting populations. *Proc Natl Acad Sci USA* 114:8330–8335.
85. Parekh S, Vinci VA, Strobel RJ (2000) Improvement of microbial strains and fermentation processes. *Appl Microbiol Biotechnol* 54:287–301.
86. Klein-Marcuschamer D, Stephanopoulos G (2008) Assessing the potential of mutational strategies to elicit new phenotypes in industrial strains. *Proc Natl Acad Sci USA* 105:2319–2324.
87. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461.
88. Fijalkowska IJ, Jonczyk P, Tkaczyk MM, Bialoskorska M, Schaaper RM (1998) Unequal fidelity of leading strand and lagging strand DNA replication on the *Escherichia coli* chromosome. *Proc Natl Acad Sci USA* 95:10020–10025.
89. Finn RD, et al. (2014) Pfam: The protein families database. *Nucleic Acids Res* 42: D222–D230.
90. R Development Core Team R Development Core Team (2013) *R: A Language and Environment for Statistical Computing* (R Found Stat Comput, Vienna), Version 3.1.1. Available at www.R-project.org. Accessed July 10, 2014.
91. Finn RD, et al. (2016) The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res* 44:D279–D285.
92. Ekeberg M, Lövkvist C, Lan Y, Weigt M, Aurell E (2013) Improved contact prediction in proteins: Using pseudolikelihoods to infer Potts models. *Phys Rev E Stat Nonlin Soft Matter Phys* 87:012707.
93. Hopf TA, et al. (2017) Mutation effects predicted from sequence co-variation. *Nat Biotechnol* 35:128–135.
94. Morgan RD, Bhatia TK, Lovasco L, Davis TB (2008) MmeI: A minimal Type II restriction-modification system that only modifies one DNA strand for host protection. *Nucleic Acids Res* 36:6558–6570.
95. Gerdes SY, et al. (2003) Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J Bacteriol* 185:5673–5684.
96. Liberati NT, et al. (2006) An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc Natl Acad Sci USA* 103:2833–2838.
97. Andersson DI, Hughes D (2009) Gene amplification and adaptive evolution in bacteria. *Annu Rev Genet* 43:167–195.
98. Yamamoto N, et al. (2009) Update on the Keio collection of *Escherichia coli* single-gene deletion mutants. *Mol Syst Biol* 5:335.