Mutation bias and the predictability of evolution

Alejandro V. Cano[1,2,†], Bryan L. Gitschlag[3,†], Hana Rozhoňová[1,2],
Arlin Stoltzfus[4,5,*], David M. McCandlish[3,*], Joshua L. Payne[1,2,*]

[1]Institute of Integrative Biology, ETH Zurich, Switzerland
[2]Swiss Institute of Bioinformatics, Lausanne, Switzerland
[3]Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA
[4]Office of Data and Informatics, Material Measurement Laboratory, NIST, Gaithersburg, MD
[5]Institute for Bioscience and Biotechnology Research, Rockville, USA
† These authors contributed equally
∗ Corresponding authors (arlin@umd.edu, mccandlish@cshl.edu, joshua.payne@env.ethz.ch)

Abstract

Predicting evolutionary outcomes is an important research goal in a diversity of contexts. The focus of evolutionary forecasting is usually on adaptive processes, and efforts to improve prediction typically focus on selection. However, adaptive processes often rely on new mutations, which can be strongly influenced by predictable biases in mutation. Here we provide an overview of existing theory and evidence for such mutation-biased adaptation and consider the implications of these results for the problem of prediction, in regard to topics such as the evolution of infectious diseases, resistance to biochemical agents, as well as cancer and other kinds of somatic evolution. We argue that empirical knowledge of mutational biases is likely to improve in the near future, and that this knowledge is readily applicable to the challenges of short-term prediction.

Keywords: adaptation, mutation, prediction, theory, population genetics

1   Introduction

Predicting the dynamics and outcome of evolution is an important goal of the biological sciences, offering the potential to design better drugs, combat pathogens, and conserve endangered species (Stern and Orgogozo, 2008, 2009; Lässig et al., 2017; de Visser and Krug, 2014; de Visser et al., 2018; Fragata et al., 2019; Papp et al., 2011; Blount et al., 2018; Nosil et al., 2020; Lind, 2019; Wortel et al., 2021).

Due to the stochastic nature of the evolutionary process, forecasting offers the greatest potential over short and intermediate timescales. Our ability to make accurate forecasts depends crucially on high-quality experimental data, such as those describing the phenotypic or fitness effects of mutations. For example, over short timescales, where one may wish to predict the next beneficial mutation to arise and go to fixation, empirical knowledge of the distribution of fitness effects is key, because this provides information about the fixation probabilities of new mutations (Eyre-Walker and Keightley, 2007). At intermediate timescales, where one may wish to predict which of several possible mutational trajectories to adaptation is the most likely, empirical knowledge of the fitness effects of combinations of mutations is key, because this can be used to delineate between mutational trajectories that ascend adaptive peaks from those that fall into maladaptive valleys (de Visser and Krug, 2014). As such, the project of predicting evolution has benefited greatly from recent advances in high-throughput sequencing technologies and phenotypic assays, which ameliorate so-called "data limits" on accurate forecasting (Nosil et al., 2020). These technologies have been used to characterize the phenotypic and fitness effects of mutations in a diversity of biological systems, including regulatory elements (de Boer et al., 2019; Vaishnav et al., 2022), macromolecules (Qiu et al., 2016; Bank et al., 2016; Li and Zhang, 2018; Wu et al., 2016; Sarkisyan et al., 2016; Lite et al., 2020; Tack et al., 2021), gene regulatory circuits (Schaerli et al., 2018; Santos-Moreno et al., 2022), and metabolic pathways (Bassalo et al., 2018).

However, empirical knowledge of the phenotypic and fitness effects of mutations only takes us so far. Whereas these data provide useful information about the likelihood of mutations going to fixation, they tell us nothing about the rate with which new mutations are introduced into a population. This is an important limitation, because evolution often proceeds via the introduction of new mutations, and some types of mutations are more likely to arise than others (Yampolsky and Stoltzfus, 2001; Stoltzfus and Yampolsky, 2009). For example, studies of the rates and spectra of spontaneous mutations, such as those based on mutation accumulation experiments, have revealed a bias toward transitions (purine-to-purine or pyrimidine-to-pyrimidine changes), relative to transversions (purine-to-pyrimidine changes, or vice versa) in a wide range of species (Katju and Bergthorsson, 2019). Because such biases make some mutational steps to adaptation more likely than others, empirical knowledge of mutation bias offers the potential to improve evolutionary forecasting, both at short and intermediate timescales.

Here we address how effects of mutational biases—predictable differences in rates between different categories of mutational conversions—make evolution more predictable, focusing mostly on the case of short-term adaptation from new mutations, and setting aside some related topics such as the role of specialized mutation-generating systems (Stoltzfus, 2021, Ch. 5) and hypermutators (Couce et al., 2013). First, we review theoretical work suggesting that such biases can exert a strong influence on the outcome of evolutionary processes, including adaptive processes, that depend on new mutations. Next, we review the empirical case for an influence of mutation bias on adaptation in the lab and in nature. Finally, we discuss specific applications where empirical knowledge of mutation bias is anticipated to improve evolutionary forecasting, in regard to topics such as infectious diseases, cancer and other kinds of somatic evolution, as well as resistance to biochemical agents. We note some recurring themes: (1) the most commonly observed outcome is often

the most mutationally favorable of the adaptive options, not the most fit, (2) ordinary nucleotide mutation biases often have strong and predictable effects on the genetic changes underlying adaptation, (3) perturbing the mutation spectrum alters the distribution of such changes, and (4) the influence of mutation biases can be altered by the beneficial mutation supply and other population-genetic and environmental conditions. In general, we argue that knowledge of mutation can improve predictability in practical contexts. We conclude with comments on open questions and future prospects.

2 Theory

Under what conditions will empirical knowledge of mutation bias improve evolutionary forecasting? To address this question, we first turn to theory. The classic "Modern Synthesis" view assumes evolution from standing variation in an abundant gene pool, so that the process of evolution is formally a process of recombining and shifting frequencies of available alleles without new mutations (Beatty, 2022; Stoltzfus, 2017, 2021). In this context, adaptation happens by selectively favorable shifts in frequencies of multi-locus combinations of small-effect alleles generated by recombination (Stebbins, 1966; Simpson, 1964; Dobzhansky et al., 1977; Mayr, 1963). The role of mutation is strictly limited: recurrent mutation acts only as a weak pressure, ineffectual except when mutation rates are high and unopposed by selection (Haldane, 1927, 1933; Fisher, 1930). Therefore, in this theory, the predictability of evolution emerges from a consideration of selection: in the short-term, an evolving population ascends a fitness gradient in a multi-locus allele-frequency space; in the long-term, it approaches a local or global maximum of fitness.

A different view of the roles of mutation and selection emerged during the molecular revolution. Comparisons of protein sequences suggested that evolutionary divergence occurs by the accumulation of individual substitutions of amino acid residues, where each substitution reflects a mutation that was promoted– or at least, tolerated– by selection, which was conceptualized as a filter acting on individual mutations (Margoliash, 1963; Zuckerkandl and Pauling, 1962, 1965). This way of thinking placed the process of mutation in the more important role of offering individual variants directly for selective filtering (rather than merely filling up the gene pool to facilitate subsequent recombination). This conception of evolution as a 2-step process was formalized in "origin-fixation" models, which depict the limiting behavior of evolution when the number of new mutations introduced per generation becomes arbitrarily small (McCandlish and Stoltzfus, 2014). In an origin-fixation model, the rate of evolution is determined by the product of a rate of "introduction" or origin $N\mu$ and a probability of fixation π, i.e., $R = N\mu\pi$.

Importantly, this new way of thinking about evolution suggests an increased influence for mutation biases, because the likelihood of each possible step will depend on the likelihood of the underlying mutation. For evolution in the origin-fixation regime, mutational biases (i.e., biases in origination) and biases in fixation each have proportional effects on the course of evolution (Yampolsky and Stoltzfus, 2001), i.e., we can express a ratio of origin-fixation rates in terms of these two different types of biases:

$$\frac{R_{ij}}{R_{ik}} = \frac{\mu_{ij}N\pi_{ij}}{\mu_{ik}N\pi_{ik}} = \frac{\mu_{ij}}{\mu_{ik}} * \frac{\pi_{ij}}{\pi_{ik}}, \tag{1}$$

where R_{ij} is the rate of change from allele i to allele j, μ_{ij} is the mutation rate from allele i to allele j, π_{ij} is the chance of fixation of a new allele of type j in a population otherwise of type i, and N is the population size (see also Streisfeld and Rausher 2011). That is, the evolutionary bias between two alternative types of changes, $i \rightarrow j$ vs. $i \rightarrow k$, can be expressed as the product of a bias in origination (e.g., transition-transversion bias or GC-AT bias) and a bias in fixation (Yampolsky and Stoltzfus, 2001; Streisfeld and Rausher, 2011). This means that biases in the introduction process can influence adaptation even when mutation rates are low and selection is strong, in contrast to the classical view in which internal biases are assumed to require evolution by mutation pressure (Haldane, 1927, 1933; Fisher, 1930), which requires high rates of mutation.

The equation above reflects origin-fixation conditions, and is useful for thinking about short-term evolution, or about long-term evolution in an infinite space. What about less ideal conditions, e.g., extended adaptive walks on finite landscapes? To grasp the potential effects of mutation bias on adaptive walks, it is helpful to consider the different perspectives of points, paths, local peaks and landscapes. From a typical point on a complex landscape, multiple upward (fitness-increasing) steps are possible, and some are mutationally favorable (whereas others are not), so that the orientation of an evolving system may be biased. Any path of upward steps eventually ends at some local peak, and some paths are enriched in mutationally favorable steps (whereas others are not), so that a system evolving under a bias may favor some paths over others. From the perspective of peaks, each fitness peak is accessible by some set of upward paths, and this set of paths may differ in size, and may be more or less enriched for mutationally favorable paths, so that certain peaks may be more likely outcomes of evolution, averaging over many possible starting points. Finally, for a given landscape with many peaks, we can define all the upward paths, i.e., all the possible adaptive walks, and thus some landscapes will have more mutationally favored walks, making them more navigable.

Evolutionary simulations on complex adaptive landscapes confirm these broad expectations and provide some guidance on the size of effects (Stoltzfus, 2006; Sane et al., 2022; Tuffaha et al., 2022; Cano and Payne, 2020; Schaper and Louis, 2014). For instance, Stoltzfus (2006) modeled adaptive walks using an NK model of fitness applied to a protein encoded by a gene subject to variable GC:AT bias, finding that a several-fold bias in mutation can have a substantial impact on the amino acid composition of evolved proteins. Cano and Payne (2020) explored the effect of transition-transversion bias on the navigability of empirical landscapes for transcription-factor binding sites, finding that the landscapes are most navigable when the mutation bias matches the bias inherent in the landscape. Schaper and Louis (2014) find that RNA folds with the most sequences are more findable in adaptation.

How far do effects of mutation biases extend outside of the strict origin-fixation regime that emerges as the mutation supply μN becomes arbitrarily small? In the hypothetical case of an infinite-sites model, mutation biases are influential regardless of mutation supply (Box 1). For finite cases, the results of Yampolsky and Stoltzfus (2001) suggest that biases in the introduction process decay with mutation supply but remain influential well outside the origin-fixation regime. Subsequent work has clarified this relationship (Gomez et al., 2020; Soares et al., 2021; Cano et al., 2022). In particular, Cano et al. (2022) used simulations to study

the effect of mutation supply in a codon-based model of protein adaptation. They quantified the effect of mutation bias with a single statistic, β, which ranges from 0, indicating no influence, to 1, indicating that the spectrum of amino-acid-changing mutations has a proportional influence on the spectrum of changes fixed in adaptation. They found that $\beta \approx 1$ when the mutation supply is low ($N\mu \approx 10^{-4}$), and ultimately goes to 0 for high mutation supply, with most of the shift from 1 to 0 happening as mutation supply goes from 10^{-2} to 10^{0}.

Finally, what are the implications for predictability? As explained in Box 2, considering a single adaptive step, predictability (in the sense of repeatability) decreases with the number of possibilities, and increases with the variance in their probabilities (Lenormand et al., 2016; Bailey et al., 2017). This predictability can be partitioned further (under limiting conditions explained in Box 2) into contributions of mutation and fixation. The separate terms have the same property that, the greater the variability in the probability of fixation π, or the greater the variability in μ, the greater the contribution to repeatability. An important implication of this theoretical result is that, in designing approaches to prediction, it is important to capture as much variance as possible in elementary chances, and to treat mutation and selection comparably to avoid a skewed picture of their contributions. For instance, if 40 different beneficial mutations are possible, and we use individual fitness measurements for each s, but characterize each μ with an average rate from a model of 6 types of rates, this artificially reduces the expected contribution of mutation to repeatability, given that such simplified models capture only a minority of the variance in individual mutation rates (Hodgkinson and Eyre-Walker, 2011).

What about predictability in long-term adaptive walks? In the special case of adaptation on a fixed and finite landscape without epistasis, the evolving system will converge on a single global peak, and mutation bias will influence the trajectory and the length of the walk, but not the final destination. In any other case, mutation bias may influence the direction, length, and ultimate destination of a walk, as outlined above. Predictability has a somewhat counter-intuitive relationship to mutation bias when a system with a particular bias is on a landscape enriched for upward paths favored by that bias. In this case, as shown by Cano and Payne (2020), there is a larger set of upward paths enriched for mutationally favorable changes, and so the particular path taken in any instance of adaptation is less predictable. Predicting multi-step evolutionary trajectories may be further complicated by the evolution of mutation bias itself. In particular, adaptation under a particular mutation bias depletes the distribution of fitness-enhancing mutations for the types of mutations favored by mutation bias, so that reversing the bias temporarily enhances the occcurrence of beneficial de novo mutations, which increases the rate of adaptation (Sane et al., 2022; Tuffaha et al., 2022).

In sum, theory suggests that mutation bias can influence adaptation under a broad range of population genetic conditions, with the strongest signal of mutational influence appearing when the mutation supply is low. Mutation biases can influence both the outcomes of short-term adaptation, and the trajectory, length, and outcome of adaptive walks, dependent on conditions. The extent to which empirical knowledge of mutation bias will improve evolutionary forecasting depends on the extent to which natural systems evolve under conditions favorable to these effects. Because this is an empirical issue, we next turn to experimental evidence, from the lab and nature.
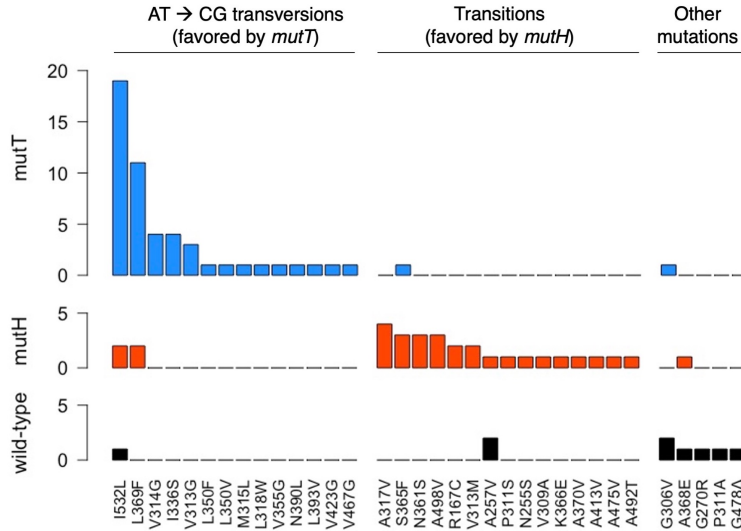
Fig. 1. The mutation spectrum shapes adaptation. Cefotaxime-resistant variants from a <u>mutT</u> parent (top row, blue) tend to have the kinds of mutations favored by <u>mutT</u> (left block of columns, $A : T > C : G$ transversions), and likewise for resistant variants from a <u>mutH</u> parent (middle row, red) which tend to have the transitions (middle block of columns) favored by <u>mutH</u>. Data for *ftsI* variants provided by Alejandro Couce.

3 Evidence

As outlined above, theory suggests that, where conditions allow, systematic biases in mutation can shape the course of adaptation via biases in the introduction process. What is the evidence that this kind of causation is real? What do we know about effect-sizes under various conditions? How well do these effects fit theoretical expectations? How broadly are such effects expected?

3.1 Causal agency

To begin, one may ask what studies establish causal agency, i.e., proving beyond any reasonable doubt that X causes Y? The gold standard is to manipulate X and show the expected effects on Y under controlled conditions. This standard is satisfied by the work of Couce et al. (2015) and Horton et al. (2021), laboratory studies with microbial systems, involving adaptation from new mutations under controlled conditions that include direct manipulation of the mutation spectrum.

Couce et al. (2015) subjected 192 replicate lines of *Escherichia coli* to increasing concentrations of the β-lactam antibiotic cefotaxime, using 3 different parental strains: wild-type, *mutH*, and *mutT*. The latter two are mutators with higher overall rates of mutation and distinctive biases toward transitions (*mutH*) or $A : T \rightarrow C : G$ transversions (*mutT*). Fig. 1 shows the resistance-conferring mutations that arose in *ftsI*, the gene in which most of these mutations are found. The resistance-conferring mutations from *mutT* isolates

(blue) tend to be $A : T \rightarrow C : G$ transversions (left block of bars), which are the type favored by *mutT*, whereas the resistance-conferring mutations that evolved in the *mutH* strain (red) tend to be the transitions (center block of bars) favored by *mutH*. That is, changing the mutation spectrum changes the spectrum of adaptive changes in a corresponding manner.

The second study, by Horton et al. (2021), was motivated by the observation that 2 different strains of *Pseudomonas fluorescens* adapt to the loss of motility in strikingly different ways. In one strain, over 95 % of the time, adaptation involved an A289C change in the *ntrB* locus, whereas in the other strain, adaptation involved mutations in diverse genes. They identified a hotspot mutation associated with synonymous sequence differences in the two strains. To test that the mutational hotspot caused the difference in adaptation, they used genetic engineering to create the hotspot in one strain, and remove it in the other — all without changing the protein sequence (because the engineered changes were synonymous). The results confirmed the mutational hypothesis. When the hotspot was removed, adaptation no longer relied on the mutation in the *ntrB* locus; and when the hotspot was engineered, adaptation no longer involved mutations in diverse genes, but rather relied on the A289C mutation.

3.2   Range of effect-sizes

Having established causal agency with studies that involve unusual conditions– some mutators and a hotspot–, let us now ask about effect-sizes when ordinary nucleotide mutation biases are involved, and particularly, let us consider whether quantitative relationships between s, μ and the frequency with which a variant evolves f are roughly what we expect from theory. Several studies are useful in this regard. We will focus here on Maclean et al. (2010), Rokyta et al. (2005), and Cano et al. (2022).

Maclean et al. (2010) tracked the emergence of resistance to Rifampicin in replicate cultures of *Pseudomonas aeruginosa*. Resistant strains typically have mutations in *rpoB*, encoding the main RNA polymerase subunit. Maclean et al. (2010) measured selection coefficients for 35 resistant variants, and mutation rates for 11 of these. The mutation rates– all for single-nucleotide substitutions– ranged 30-fold. However, the selection coefficients are very large and show a much smaller range, from 0.3 to 0.9, so that the range expected for the probability of fixation is even smaller, just 0.45 to 0.83 (using the formula of Kimura 1962). Thus, under origin-fixation conditions, we expect a 30-fold effect of mutation but only about a 2-fold effect of selection (given that clonal interference is not expected). The results shown in Fig. 2 confirm this expectation and provide some additional useful evidence (these results are also used as an example in Box 2). As shown in the left panel, the most frequent outcome is not the most fit; the top 2 most frequent outcomes fall in the middle of the fitness distribution. Meanwhile there is a strong and roughly proportional effect of mutation rate, as shown in the center panel. The right panel confirms that this effect of mutation rate is not due to confounding with selection coefficients, which are uncorrelated with the mutation rates.

In a well known study of recurrent evolution, Rokyta et al. (2005) carried out 1-step adaptation 20 times in replicate populations of bacteriophage $\phi X 174$, under conditions of adaptation from new mutations. They found that the most frequent change, repeated 6 times, was not the most fit, but rather the 4th most fit. These
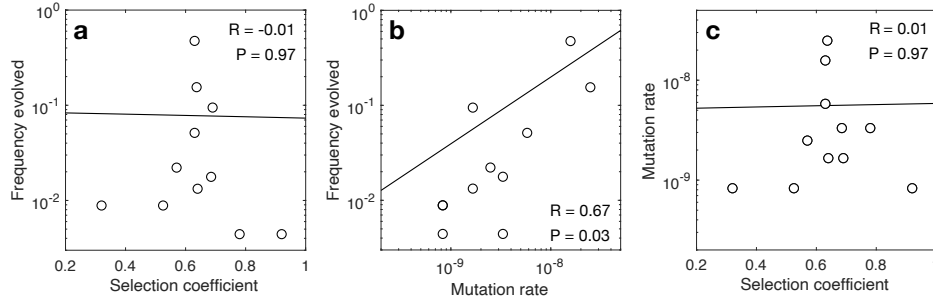
Fig. 2. Inter-relations of mutation rate, selection coefficient, and frequency evolved in a laboratory adaptation experiment (R, Pearson's correlation coefficient; P, the P-value). From 284 replicate cultures adapted to Rifampicin, Maclean et al. (2010) identified 35 different *rpoB* mutations, many emerging repeatedly, reporting their frequency of evolving (in replicate cultures), and reporting mutation rates for 11 of the variants (frequency evolved is the combined frequency in backgrounds where the mutation is possible, using data in Table 1 from Maclean et al. (2010); selection coefficients are averaged when multiple values are reported for different backgrounds). In the three panels, linear regressions were performed on the unlogged values, and the Pearson's correlation coefficients and their respective P-values are shown.

results were not in agreement with the model of Orr (2002), which assumes uniform mutation, prompting the authors to seek a mutational explanation. They found that an origin-fixation model incorporating (1) measured selection coefficients and (2) a model of nucleotide mutation rates (from comparative data) performed better in predicting outcomes than Orr's model, which assumes homogeneity in mutation rates. Thus knowledge of mutation rates improved the predictability of adaptive outcomes.

As explained in Section 2, Cano et al. (2022) developed a method to capture the influence of the mutation spectrum with a single coefficient of mutational influence β that ranges from 0 (no influence) to 1 (proportional influence). They also applied this method to three data sets of adaptive amino acid substitutions, including substitutions implicated in natural adaptation of *Mycobacterium tuberculosis* to antibiotics, as well as laboratory adaptation of *E. coli* and *Saccharomyces cerevisiae* to environmental stress, using independently curated species-specific mutation spectra that describe the relative rates of the six possible nucleotide changes within double-stranded DNA. For each species, they found that β is close to 1 and significantly different from 0, indicating a proportional influence of the mutation spectrum. Moreover, they showed this was not just an effect of transition bias, but rather of the entire distribution of rates across the six types of single-nucleotide changes. Indeed, the frequencies of the six types of nucleotide changes among adaptive substitutions are strongly correlated with the independently curated species-specific mutation spectra (Fig. 3). The authors note that the three species differ in important population genetic conditions, such as mutation supply. Whereas *M. tuberculosis* has one of the lowest mutation rates of all bacteria (Eldholm and Balloux, 2016) and is therefore likely to experience only limited clonal interference during adaptation to a new human

8

host (Liu et al., 2015; Trauner et al., 2017), *E. coli* and *S. cerevisiae* have relatively higher mutation rates (Lee et al., 2012; Zhu et al., 2014) and often experience clonal interference in laboratory evolution experiments (Lang et al., 2013; Good et al., 2017). The results of Cano et al. (2022) therefore provide empirical support for the theoretical result that mutation bias can influence adaptation across a broad range of population genetic conditions (Yampolsky and Stoltzfus, 2001; Gomez et al., 2020).
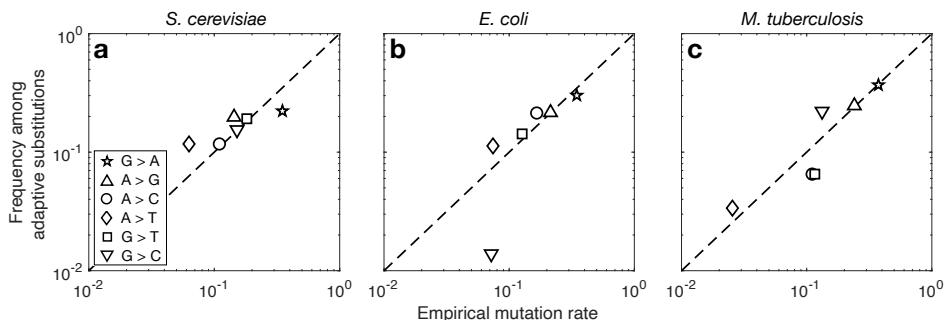

Fig. 3. The frequency of nucleotide changes among adaptive substitutions as a function of the empirical mutation rate for (a) *S. cerevisiae*, (b) *E. coli*, and (c) *M. tuberculosis*. The symbols correspond to the six different types of point mutations (inset in panel a). The dashed line shows $y = x$. Data from Cano et al. (2022).

3.3   Scope of applicability

Now, having established causal agency and the potential for large effect-sizes, let us consider the scope and generality of this kind of cause-effect relationship. How widely can we expect it to apply? An ideal way to address this question would be to carry out a meta-analysis of published studies of adaptation. We would want to include in this analysis all of the relevant work, dividing it into experimental and natural adaptation, and perhaps considering other factors such as taxonomy and population size. At present, such an analysis would be quite difficult and would cover only a very minor fraction of the literature. The difficulties may be summarized as follows. In over a century of experimental studies of adaptation, the vast majority do not include a genetic analysis. Those with a genetic analysis typically implicate loci or alleles (e.g., involved in the adaptation of quantitative traits) without identifying specific mutations. The adaptation studies that implicate specific mutations (a tiny fraction of all adaptation studies) typically do not have sufficient replicates to support powerful tests, e.g., sometimes they are a one-off case (Xie et al., 2019). In addition, most reports implicating adaptive mutations do not follow a rigorous standard for making this determination, so that mis-attributions are common (Coombes et al., 2019), a serious problem given the prior expectation that non-adaptive changes will show effects of mutation biases. Furthermore, even in cases where adaptation can be traced confidently to specific mutations, we rarely have the kind of information on mutation biases and selection coefficients that would be needed to reach the conclusion that mutational effects are consequential once selection is taken into account.

The meta-analysis strategy of Stoltzfus and McCandlish (2017), focused on transition-transversion bias among amino-acid changes, was designed to maximize the use of available data given these difficulties. Briefly, it takes advantage of the following: (1) many contemporary studies of adaptation implicate specific amino acid changes, typically caused by single-nucleotide substitutions, doing so in a rigorous way based on verifying effects with genetic comparisons or engineering, (2) for a broad range of taxa, nucleotide mutations show a bias toward transitions, typically 2- to 4-fold above null expectations (Katju and Bergthorsson, 2019; Stoltzfus and Norris, 2016), and (3) experimental studies of the fitness effects of mutations suggest that transitions and transversions that change amino acids hardly differ in their distribution of fitness effects (Stoltzfus and Norris, 2016), so that a reasonable null expectation is a simple 1:2 ratio of transitions to transversions (Stoltzfus and McCandlish, 2017).

On this basis, one may gather qualifying results and combine them, applying statistical tests for an excess of transitions relative to the 1:2 expectation. For instance, Meyer et al. (2012) carried out replicate laboratory evolution experiments with bacteriophage λ under conditions that favor changes in the J gene, the product of which helps the virus target its bacterial host. Among 241 putatively adaptive changes, the ratio of transitions to transversions was 192:49, roughly 8-fold higher than the 1:2 null expectation. The meta-analysis by Stoltzfus and McCandlish (2017) covers experimental and natural adaptation using this approach, with the added safeguard that results are restricted to recurrent amino acid changes, i.e., their data set is conditioned on parallel evolution. The experimental data set covers five different experimental systems, the largest of which are the study of Meyer et al. (2012) and the studies of Crill et al. (2000) and Bull et al. (1997) that uncovered numerous reversals and parallels in lines of $\phi X174$ propagated through successive host reversals (between *E. coli* and *Salmonella typhimurium*). Combining the data from all five studies, Stoltzfus and McCandlish (2017) find a highly significant 304:83 ratio of transitions to transversions among events of parallel adaptive amino acid changes.

Several subsequent studies have shown effects of transition-transversion bias. Sackman et al. (2017) extended the earlier study of Rokyta et al. (2005) by applying the same 20-replicate protocol to three additional types of phages, for a total of 80 adaptive changes. For each of the four phages, the most common variant to evolve was not the one with the largest fitness benefit. Out of $20 \times 4 = 80$ changes, the transition-transversion ratio was 74:6, a striking result. Likewise, Bertels et al. (2019) observed a strong enrichment of transitions among adaptive mutations in propagation of HIV-1 in human T-cell lines, and Katz et al. (2021) observed a bias toward transitions during adaptation of *E. coli* to long-term stationary phase.

What about adaptation in nature? The meta-analysis of Stoltzfus and McCandlish (2017) includes data from ten cases of natural adaptation traced to specific mutations, with results shown in Table 1. For example, species such as monarch butterflies (*Danaus plexippus*) evolve resistance to cardiac glycosides by changes in the sodium pump ATPα1 (Zhen et al., 2012; Aardema et al., 2012; Ujvari et al., 2015), which not only allows them to eat *Apocynaceae*, but also to sequester the toxin in their tissues, making them noxious to predators. Other cases involved adaptation to natural or anthropogenic toxins (tetrodotoxin, insecticides, benzimidazole, and the antiviral agent Ritonavir), altitude adaptation via hemoglobin changes, convergent foregut fermentation, trichromatic vision, and echolocation. Combining the data from these cases of natural

Phenotype	Taxon	Target	Ti Events Counts	Sum	Tv Events Counts	Sum
Insecticide resistance*	Insecta	Rdl, Kdr, Ace	2, 2, 5, 2, 3	14	9, 2, 4	15
Tetrodotoxin resistance	Vertebrata	Na channels	2, 6, 3	11	2, 2, 2, 3, 3	12
Glycoside resistance	Metazoa	Na^+/K^+-ATPase	4, 4, 2, 2	12	7, 2, 2, 4	15
Herbicide resistance*	Poaceae	ACCase	5, 2	7	7, 2, 4, 5	18
Altitude adaptation	Aves	β-hemoglobin	4, 13	17	2, 3, 2	7
Trichromatic vision	Vertebrata	Opsins	2, 5	7	6, 4, 2	12
Echolocation	Mammalia	Prestin	2, 2, 2	6	3, 2	5
Growth in Ritonavir*	HIV1	Protease	25, 7, 9	41	4	4
Foregut fermentation	Vertebrata	Ribonucleases	2, 4, 4	10		0
Benzimidazole resistance*	Ascomycota	β-tubulin	7	7	5, 6	11
totals				132		99

Table 1. Transition bias among natural parallelisms (Stoltzfus and McCandlish, 2017). For 10 different study systems implicating diverse taxa, the counts of parallel events are given for transitions and transversions (because this study is conditioned on parallelism, each type of change has at least 2 events). The results show a strong bias toward transitions. Note that some cases (marked by *) represent recent local adaptation of sub-populations to anthropogenic substances, while the rest refer to episodes of adaptation from the distant past.

adaptation, Stoltzfus and McCandlish (2017) uncovered a ratio of 132 transitions to 99 transversions (Table 1) — a 2.7-fold enrichment over the null.

Another example of transition-transversion bias in natural adaptation involves a very large set of resistance mutations identified clinically in the global human pathogen *M. tuberculosis*, which exhibits a strong mutation bias toward transitions (Hershberg and Petrov, 2010) and evolves resistance to antibiotics exclusively through chromosomal mutations (Gagneux, 2018). Examining two independently curated data sets, Payne et al. (2019) uncovered transition-transversion ratios of 1755:1020 and 1771:900, a 3.4-fold and 3.9-fold enrichment over the null, respectively. They also took advantage of the special case of Met-to-Ile replacements, which can occur via one transition (ATG $\rightarrow$ ATA) and two transversions (ATG $\rightarrow$ ATT and ATG $\rightarrow$ ATC). Thus a 1:2 ratio is expected under the null hypothesis in which mutation bias has no effect. Instead, they observed ratios of 88:49 and 96:39 in the two data sets, roughly in 4-fold excess of the null expectation.

What about other forms of mutation bias? In mammals and birds, mutation rates are elevated at CpG dinucleotides relative to other sequence contexts, due to the effects of cytosine methylation on DNA damage and repair (Ehrlich et al., 1990; Hess et al., 1994; Smeds et al., 2014). Genetic studies of high-altitude birds provided the first hints that this form of mutation bias may influence adaptation in nature, specifically the evolution of increased affinity of hemoglobin for oxygen, which is likely adaptive in hypoxic conditions

and preferentially occurs via missense mutations at CpG dinucleotides (Galen et al., 2015; Zhu et al., 2018). Building off these observations, Storz et al. (2019) systematically analyzed the genetic sequences of hemoglobins in 35 matched, phylogenetically independent pairs of high- and low-altitude bird populations. Among the 35 pairs, they found 22 changes in oxygen affinity plausibly linked to altitude adaptation, implicating 10 different amino acid changes in hemoglobins. Of these 10 amino acid changes, 6 involved CpG mutations, whereas only 1 CpG mutation would be expected by chance, a significant excess. Thus, altitude adaptation in natural bird populations shows a significant enrichment of mutationally-likely genetic changes, specifically mutations at CpG dinucleotides.

Taken together, the evidence summarized in this section provides robust support for a large and predictable influence of mutation biases on the changes involved in adaptation. The most common adaptive variants are often not the most fit, but the ones with the highest mutation rates. Quantitative biases in nucleotide mutation rates can have proportional effects, leading to a detailed match between the mutation spectrum and the spectrum of adaptive changes, and results from episodes of natural adaptation traced to the molecular level suggest a broad taxonomic scope.

4   Applications

Addressing ecological, agricultural and biomedical challenges often involves seeking to limit the reproduction of threatening biological agents such as microbial pathogens and parasites. Accordingly, understanding the evolutionary processes that give rise to problems of drug and pesticide resistance can lead to marked advancements in the agricultural and biomedical sciences. Extrapolating from its general utility in evolutionary modeling, here we discuss how considerations of mutation-biased evolution shows tremendous potential in addressing challenges of widespread human concern, with a particular focus on evolutionary dynamics in somatic contexts such as cancer, drug and pesticide resistance, and infectious disease.

4.1   Somatic evolution

Human somatic DNA mutates throughout adulthood in a manner that can cause disease, particularly as repeated rounds of genome replication in mitotically active cells provide opportunities for the emergence of mutant cells that have a replicative advantage, often to the detriment of the organism. Accounting for biases in mutation rates can provide improved insight into the evolutionary dynamics that occur among somatic cells (Cannataro et al., 2018). In recent years, several studies exploring the evolutionary conversion of healthy somatic cells to cancer cells have shed valuable insights on the roles played by mutation biases (Cannataro et al., 2018; Temko et al., 2018). For example, a study of APOBEC-induced mutations in head and neck squamous cell carcinoma found that the relative importance of mutations for the cancer phenotype often differed from their prevalence, with some variants occurring infrequently despite being highly favored by selection (Cannataro et al., 2019). A more comprehensive analysis featuring 7,815 cancer exomes identified dozens of highly statistically significant associations between cancer-driving mutations

and specific mutational signatures such as those associated with environmental carcinogens and mutagenic enzyme activity (Poulos et al., 2018). The vast majority of these associations include deamination by APOBEC and deficiencies in proofreading and mismatch repair during replication. Intriguingly, this study also identified a negative association between tobacco smoke and the G12D substitution in KRAS; in other words, KRAS G12D is more common among the lung cancers of non-smokers (Poulos et al., 2018). Consistent with this finding, lung cancers harboring the KRAS G12D substitution were recently associated with a lower tumor mutation burden (Gao et al., 2020; Tan et al., 2022), for which reason this mutation may serve as a negative biomarker for the success of immunotherapy. In addition to showing that the mutations most strongly favored by selection are not necessarily the most prevalent among cancer patients, these findings suggest that mutational biases facilitate a link between the source of carcinogenesis and the predicted success of a given treatment.

What other factors may alter mutation rates in a manner that predictably influences the progression and treatability of cancer? Importantly, chemotherapy itself represents a source of mutagenesis, suggesting that attempts to treat cancer may inadvertently induce adaptive changes in the cancer that complicates further treatment options. For example, although mutations at residues 12 and 13 of the cell-signaling GTPase KRAS have a higher selective advantage, the Q61H mutation is common in colorectal cancers with resistance to the anti-EGFR antibody cetuximab (Woolston et al., 2021), due to a mutational signature associated with chemotherapy. Importantly, this work suggests that mutation signatures can serve as a basis for predicting the evolution of drug resistance in cancer patients. More recent investigation has also found that depending on the cancer type, the predominant driver mutations can arise from "actionable" mutation signatures. In addition to tobacco, these drivers include mutations associated with exposure to UV light and endogenous processes associated with aging (Cannataro et al., 2022). By identifying specific causal factors underlying the likelihood of driver and drug-resistance mutations across different types of cancers, these findings provide a basis for predicting the efficacy of preventative and therapeutic strategies.

Finally, the relative over-proliferation of some blood-cell lineages is another evolutionary process at the cellular level that can lead to an elevated risk of cancer. Among common genetic variants underlying this over-proliferation phenotype—termed clonal hematopoiesis—mutation rate was recently reported to be inversely related to fitness effect (Watson et al., 2020). For example, the most commonly observed mutation of the arginine at residue 882 in DNMT3A (mutation to histidine) was estimated to have a higher mutation rate but a lower fitness effect than mutation of the same arginine to cysteine (Watson et al., 2020). Consistent with these observations, a more recent study of chronic myelogenous leukemia found that for the tyrosine kinase inhibitor imatinib, epidemiological incidences of mutations conferring drug resistance are best predicted by the likelihood of the mutations rather than by their fitness effects (Leighow et al., 2020). Together, these results highlight mutation bias as an important predictor of somatic disease risk as well as drug resistance.

4.2 Resistance to biochemical agents

The evolution of resistance to drugs and host immunity represent substantial obstacles in the fight against disease. Accordingly, by providing insights on the processes underlying adaptive evolution, accounting for the combined roles of mutation and selection can improve our ability to understand and thus predict how resistance evolves among microbial pathogens. Specifically, are some mutational trajectories toward drug resistance enriched for higher-probability mutations than others, and can this information be leveraged to fight infectious disease, in particular by tailoring treatment approaches that minimize the predicted likelihood of drug resistance? Recent work exploring large data sets of drug-resistance mutations has begun to shed light on these questions. For example, the study by Payne et al. (2019) discussed in the previous section suggests the evolution of antibiotic resistance in *M. tuberculosis* is at least partially predictable, with some mutational paths toward resistance occurring more frequently than others depending on the relative abundance of transition mutations. Moving forward, it will be greatly informative to comprehensively characterize mutational paths toward resistance in a greater diversity of infectious pathogens and across a wide panel of antibiotics. By identifying drugs or drug cocktails for which mutational paths toward resistance tend to be relatively depleted of high-probability mutations, it may be possible to employ treatment regimens that minimize the predicted likelihood of evolved resistance, enabling treatments with longer-lasting effectiveness.

Besides transition-transversion bias, how else might biases in mutation rate guide the evolution of drug resistance? The idea that the most mutationally probable changes are not necessarily the most strongly favored by selection implies the existence of potential mutations that would be highly adaptive but whose rates of occurrence are negligible. Accordingly, by altering the relative rates of mutation types, changes in the sources of mutation may promote adaptation by enhancing access to otherwise-rare beneficial mutations (Sane et al., 2022; Tuffaha et al., 2022). In one recent example, point mutations in a DNA topoisomerase gene, which is important for relieving topological stress in DNA strands, were reported to introduce mutation hotspots that result in new adaptive paths toward antibiotic resistance in *E. coli* (Bachar et al., 2020). Although the relevance to infectious pathogens remains unclear, these findings highlight a promising approach toward identifying new potential mutational paths to the evolution of antibiotic resistance. In particular, future work may be able to determine whether mutations in DNA maintenance or repair genes shift the mutation spectrum in a manner that promotes drug resistance evolution in pathogenic bacteria. Granting such insights, we anticipate the potential for bacterial genotyping as a predictor for the likelihood of evolved resistance to specific classes of antibiotic.

As with antibiotics, the widespread use of pesticides and fungicides in agriculture can select for the evolution of resistance, which has been reported in hundreds of species (Gould et al., 2018; Hawkins and Fraaije, 2021). Similar to many examples discussed in the previous sections, mutation biases have been implicated in instances of insecticide, fungicide and herbicide resistance (Table 1) (Stoltzfus and McCandlish, 2017), suggesting a broad range of potential agricultural applications for incorporating mutation bias into evolutionary forecasting. In addition to transition-transversion biases, how else might mutation biases improve predictability of resistance evolution? To address this question, we consider examples of natural

mutators. Powdery mildews are fungal plant pathogens that belong to the genera *Erysiphe* and *Blumeria* and represent a major agricultural threat (Jones et al., 2014). These taxa have undergone extensive loss of DNA mismatch repair genes throughout their evolutionary history, leading to rapid, mutation-biased genome evolution (Phillips et al., 2021). Importantly, heavy use of fungicide has been reported to accelerate the evolution of resistance in these taxa (Jones et al., 2014). Whether variation in the mismatch repair pathway predictably alters the likelihood of resistance evolution remains unclear. However, the number of mismatch repair genes lost during evolution varies greatly across taxa and correlates with nucleotide substitution and composition biases (Phillips et al., 2021), which raises the interesting possibility that the tendency for the genomic changes that facilitate fungicide resistance might also correlate with the loss of these genes. It would be fascinating to address this possibility in future work, in particular by interrogating the mutation spectra produced from targeted disruption of mismatch repair for their tendencies toward fungicide resistance.

4.3   Infectious disease

The COVID-19 pandemic represents an exceptional case study in the importance of forecasting evolutionary trajectories among both real and potential pathogens. Since the start of the pandemic, scientists and medical professionals have sought to understand the mechanisms underlying both disease severity and viral evolution, with the goal of maximizing mitigation efforts and vaccine effectiveness. Toward this end, numerous investigations of SARS-CoV-2 genomes have identified mutation biases strongly favoring uracil content, with potential implications ranging from vaccine design to personalized therapies and the emergence of new viral variants. For example, Rice et al. (2021) recently reported that although mutation bias strongly favors U content, selection largely occurs against U content, which raises the question about how informative this mutation bias may be toward predicting adaptive changes. On the other hand, a strong C-to-U mutation bias was more recently reported to drive the diversification of CD8+ T cell epitopes and the depletion of proline residues, which has been suggested to compromise T-cell immunity among individuals carrying the human leukocyte antigen B7 serotype (Hamelin et al., 2022). Because host immunity represents a strong source of selection pressure on viral replication, the C-to-U bias may be helping to sustain high COVID-19 infection rates by facilitating immunity evasion within at least a subset of the population. Thus, in addition to the evolution of drug resistance as described above, mutation bias may shape the evolution of viral pathogens in a manner that predictably disrupts host immunity.

Given the abundance of recent changes in the spike protein (Abas et al., 2022), it may be possible to draw statistical inferences about whether, and to what extent, recent adaptations in SARS-CoV-2 are mutation-biased. This will greatly inform our ability to develop and implement accurate pandemic forecasting. In particular, the most commonly observed adaptive mutations are not necessarily the most fit. Accordingly, if the recent evolution of SARS-CoV-2 has been largely determined by amino acid changes that are mutationally favored but selectively suboptimal, then there may exist adaptive "jackpot" mutations that have yet to be sampled. In this case, a prolonged high rate of infections could be expected to enable the eventual occurrence of low-probability mutations that substantially enhance viral transmission. This scenario seems consistent

with the recurrent emergence of increasingly transmissible variants. On the other hand, high COVID-19 infection rates raise the question of whether the evolution of SARS-CoV-2 is mutation-limited, especially given the ability of new variants to spread between geographic regions and populations. In either case, the rapid accumulation of amino acid replacements provides a considerable sample of empirical data. These data could be combined with estimates of mutation rates in order to determine whether recent or future emergence of increasingly transmissible variants are driven by systematic relationships between mutation rates and fitness effects.

In addition to SARS-CoV-2, the rapid pace of adaptive evolution has made some pathogens such as HIV notorious for their ability to evade our efforts to employ treatments and vaccinations with long-term efficacy. As a retrovirus, HIV requires reverse transcriptase to infect hosts, and numerous reverse transcriptase mutations can confer resistance to reverse transcriptase inhibitors that are used to treat HIV infection. Importantly, due to a bias favoring the G-to-A transition mutation, the resistance-conferring M184I replacement in reverse transcriptase was found to occur more readily than M184V, despite the latter conferring greater replicative fitness (Keulen et al., 1997). Consistent with this finding, theoretical modeling has implicated G-to-A mutations, mediated by the APOBEC family of host deaminases, in the evolution of drug resistance in HIV (Jern et al., 2009).

How might such biases aid in the predictability of HIV evolution? Recent work suggests that instances of parallel evolution serve as a promising source of insight on this question. In particular, the relative number of independent occurrences of a given type of evolutionary change reflects its underlying probability: if one of two types of evolutionary change has a two-fold greater probability of occurring, it can be expected to occur twice as often in independent lineages. Since the chance of parallel or repeated evolution increases with greater variance in mutation rates (see Box 2), mutation biases raise the probability of particular types of evolutionary change. Consistent with these theoretical expectations, a long-term evolution experiment involving the continued passaging of HIV in human T-cells revealed numerous instances of parallel changes, characterized by a strong bias for G-to-A transitions (Bertels et al., 2019). Unfortunately, because long-term evolution is bound to involve the accumulation of both adaptive and neutral changes, such experiments pose the challenge of disentangling the roles of mutation and selection.

To overcome this difficulty, deep mutational scanning can be used to isolate the functional effects of massive numbers of individual mutations. For example, Haddox et al. (2018) used deep mutational scanning to characterize the amino acid preferences at every site in the envelope proteins from two HIV lineages. Results from such experiments can be combined with measures of mutation rates to generate pairwise estimates of rate and fitness effect for large numbers of potential mutations. Such pairwise estimates enable the prediction of likely sequence changes during evolution, since the rate of such changes are jointly proportional to both mutation rate and selection coefficient (Eq. 1). Finally, given its rapid rate of evolution, long-term evolution experiments with HIV such as the one performed by Bertels et al. (2019) provide a wealth of empirical sequence changes for testing and refining evolutionary predictions. By identifying adaptive paths involving low-probability mutations, such an approach could potentially uncover new drug and vaccine targets that minimize the likelihood of evolved resistance, leading to treatment regimens with longer-term

effectiveness.

5   Challenges

Theory and empirical evidence indicate that mutation biases can have predictable effects on the genetic changes fixed in adaptation under a broad range of population-genetic conditions. In the context of research on the predictability of evolution, the obvious application of these results is simply to absorb the science that is already well established– nucleotide substitution biases shape short-term adaptation– and apply that by leveraging available information on the mutation spectrum.

Beyond these obvious applications, what further gains would be possible with new technology or a shift in resources and attention? In this section, we suggest specific areas in which a stronger focus on effects of biases in variation might yield substantial gains, including (1) improving measurements of basic quantities, (2) expanding our attention beyond nucleotide substitution biases, and (3) assessing the predictability of mutational effects across diverse conditions and time-scales.

5.1   Expanding coverage of basic measurements

The results reported above show the value of obtaining fundamental measurements of the following three quantities, for each possible outcome: selection coefficient (s), mutation rate (μ), and frequency of evolution (f). In particular, the example of Maclean et al. (2010), as employed in Fig. 2 and Box 2, shows that such data are extraordinarily valuable, yet this case is small – just 11 variants – and we know of no other comparable data set.

More commonly, we have access to individual measures of functional effects via deep mutational scanning, but no individual mutation rates, which are instead represented by a model of average rates for classes, e.g., a model of 2 rates for transitions and transversions, or a model of 6 types of nucleotide substitutions. Yet, as explained in Box 2, prediction will always suffer when models of average rates are used. Direct and indirect evidence indicates individual rates have a large amount of variance (i.e., useful information) that simplified models of mutation rates simply do not capture, e.g., Maclean et al. (2010) find a 30-fold range in mutation rates for just 11 nucleotide substitution mutations in the same gene; Hodgkinson and Eyre-Walker (2011) use a comparative method to estimate that a triplet context model captures only about one-third of the actual variance in mutation rates.

The technical barriers to addressing this rather stunning deficit are rapidly disappearing. Until recently, methods for measuring mutation rates dated from the 1940s and were used infrequently. However, new methods for identifying and tracking mutations are now appearing rapidly, including methods based on real-time visualization (Robert et al., 2018), and methods specifically designed to measure mutation rates accurately in deep sequencing experiments (Dolan et al., 2021; Acevedo et al., 2014). We note that, if estimates of μ, s, and f are used to interrogate their relationships, it is imperative to ensure that the estimates are unbiased with respect to these relationships. For instance, some methods used to study somatic evolution,

e.g., clonal haematopoesis in Watson et al. (2020), infer both μ and s from a joint distribution of population frequency and somatic prevalence (measures obviously related to f), and this raises the question of whether they are subject to correlated errors, e.g., if under-estimation of s induces over-estimation of μ.

We look forward to a future in which quantitative scientists have access to well defined sets of fundamental measurements for diverse model systems in somatic evolution, the emergence of resistance to toxins, and the adaptive evolution of infectious agents exploiting host resources.

5.2   Exploring diverse sources of variational bias

Most approaches to analysis and modeling that incorporate rates for nucleotide substitution mutations use a simplified model, e.g., a 2-parameter model (i.e., transitions and transversions) or a 6-parameter model. However, as noted above, such models overlook most of the variance in individual rates (Hodgkinson and Eyre-Walker, 2011). This is particularly important given the common observation that adaptive outcomes are often highly enriched for a few high-rate mutations that happen to be favorable. This suggests an importance for improving models for mutation hotspots, a topic that is rarely treated (e.g., Rogozin and Pavlov 2003). One must not overlook the possibility of highly consequential correlations between mutational and functional features of genomes (Monroe et al., 2022). For instance, Sankar et al. (2016) find that, in prokaryotic genomes, transcription-replication collisions result in some very specific and large effects, including an orientation-dependent 4-fold increase in point mutations affecting promoters, mostly due to $T \rightarrow C$ mutations at position -7 relative to the start of transcription.

Although much work remains to be done in terms of basic measurements and models regarding single-nucleotide mutations, there is a far larger universe of possible mutations to explore, including multi-nucleotide mutations, compound changes affecting dispersed sites, microindels (very small insertions and deletions), the expansions and contractions of highly variable repeat loci, segmental duplications, transposable element insertions, inversions, and chromosome fission and fusion. A quantitative overview of this universe of mutations is given in Stoltzfus (2021, App. B).

Opportunities to improve prediction in this regard arise most obviously when, for some specific prediction problem, evolution commonly involves mutations other than nucleotide substitutions (and they also arise, less obviously, when such mutations are probable under some conditions but are not observed). For instance, segmental duplications occur commonly in experimental yeast adaptation (e.g., Chang et al. 2013), transposable element insertions are commonly implicated in local adaptation of bacteria in nature (e.g., Vandecraen et al. 2017), and highly variable short-tandem-repeat loci have been implicated in cases of short-term adaptation such as the domestication of dogs (Fondon and Garner, 2004). A case of particular interest are multi-nucleotide changes to codons (Schrider et al., 2011), which have been observed in studies of cancer (Wang et al., 2020), developmental disorders (Kaplanis et al., 2019), SARS-CoV2 (Borges et al., 2021; Tonkin-Hill et al., 2021), and resistance to antimicrobials, e.g., in *M. tuberculosis* (Yadon et al., 2017). Such mutations are a target of opportunity given the kind of information already available, namely (1) deep mutational scanning studies, which cover the amino acid changes that occur by double- and triple-nucleotide

changes to codons, and (2) prior information on the underlying rate for tandem double or triple mutations in eukaryotes, which appears to be (in total) about 2 or 3 orders of magnitude less than the total rate of single-nucleotide changes (Schrider et al., 2011).

Finally, we stress that the literature on natural and experimental evolution covers a variety of phenomena– under the headings of predictability, contingency (repeatability), constraints, genotype-phenotype maps, and so on– that are not usually associated with the concept of mutation bias but which are subject to the same rules of population genetics as mutational effects under a scheme of aggregation (Box 2). For instance, the genetic code is a genotype-phenotype map dictating that there is 1 single-nucleotide mutational path from Met (ATG) to Val (GTG), 2 paths from Met to Leu (TTG, CTG) and 3 paths from Met to Ile (ATT, ATC, ATA). The biases induced by this mapping are not the same thing as mutation biases in the narrow sense of biases imposed by the mechanism of mutagenesis: instead, they are induced by an asymmetric mapping of genetic changes into a phenotype space. Nevertheless, from the perspective of understanding effects of biases in the introduction of variation, a scheme of aggregation that imposes 2-fold or 3-fold biases has the same impact as a 2-fold or 3-fold mutation bias. Likewise, when the mutational target size of a trait or the mutational accessibility of an alternative phenotype is identified (Houle, 1998; Streisfeld and Rausher, 2011; Lind et al., 2015; Bailey et al., 2017; Besnard et al., 2020), this corresponds to a scheme of aggregation over elementary mutational events. For instance, the series of studies from Lind et al. (Lind et al., 2015, 2019) dissecting the emergence of the wrinkly spreader phenotype in *P. fluorescens* provides a detailed look at asymmetries in the mutational accessibility of an alternative phenotype. Analyses of genotype-phenotype maps in a wide diversity of biological systems reveal that such asymmetries are common (Ahnert, 2017). In long-term evolution, the phenotypes that have more genotypes are on the whole more "findable" (Garson et al., 2003; Schaper and Louis, 2014; Dingle et al., 2020). Understanding the extent to which these biases have predictable effects on the genetic changes fixed in adaptation is an important direction for future research, i.e., the challenge is to measure the predictive accuracy of different kinds of aggregation (and some guidance for doing so is provided in Box 2).

5.3   Considering diverse conditions and timescales

The most robust empirical and theoretical results available today focus on short-term or one-step adaptation, and the effects are best understood for the case of mutation-limited conditions, although we are beginning to get a clear sense of what happens as mutation supply increases and clonal interference becomes common in finite spaces (Yampolsky and Stoltzfus, 2001; Soares et al., 2021; Cano et al., 2022) or infinite ones (Box 1, Gomez et al. 2020). These results are relevant to many challenges in prediction, as we have argued above, e.g., antibiotic resistance. However, a challenge for future research is to expand the consideration of mutational effects to cover longer time-scales and a greater diversity of contexts, including evolution from standing variation and synthetic evolution. Attempts to predict long-term effects of mutation bias, for instance, can take advantage of limited theory currently available on how mutation bias influences multi-step trajectories to adaptation (Cano and Payne, 2020; Sane et al., 2022; Stoltzfus, 2006; Tuffaha et al., 2022).

Regarding evolution from standing genetic variation, when multiple beneficial alleles are present in numbers high enough to escape random loss, the most fit allele typically wins (Bailey et al., 2017). This appears to leave no room for effects of mutation bias, but actually it just pushes the question of origination biases into a different realm, where the primary question concerns how the distribution of standing variation is shaped by tendencies of mutation. For instance, the rate of length changes in short tandem repeat (STR) loci is so high that the vast majority of such loci will exhibit standing variation for length in a moderately sized population, and this is relevant for cases such as short-term evolution of gene expression (Gymrek et al., 2015).

Finally, mutation bias may improve evolutionary forecasting in synthetic evolution (Simon et al., 2019), such as in laboratory evolution experiments with genomically recoded organisms (Pines et al., 2017; Moratorio et al., 2017; Calles et al., 2019; Drienovska and Roelfes, 2020) or engineered mutagenesis mechanisms (Ravikumar et al., 2018). For instance, a directed evolution technique called Orthorep uses an orthogonal DNA polymerase to introduce mutations to plasmid-borne target genes (Ravikumar et al., 2018). The polymerase's mutation spectrum is heavily biased toward $G : C \rightarrow A : T$ transversions (Ravikumar et al., 2014), which may influence evolutionary outcomes, such as the primary and promiscuous functions of enzymes (Rix et al., 2020). More broadly, synthetic evolution provides a useful testbed for mutation-biased adaptation theory, as the mutation spectrum can be manipulated under controlled laboratory conditions, and evolutionary outcomes can be quantified with DNA sequencing and phenotypic assays.

6 Conclusion

We have presented theoretical arguments and empirical evidence that mutation bias can have predictable effects on the genetic changes fixed in adaptation. In studies of adaptation in diverse contexts, where fitness effects have been measured, it is regularly observed that the most common outcome is not the most fit: instead it is often a beneficial variant with an extreme rate of mutational origin. More generally, the spectrum of changes observed in adaptation reflects the mutation spectrum. Sometimes this effect can be quite strong, even proportional. The study of mutation-biased adaptation has achieved some degree of quantitative and theoretical sophistication, although much remains to be determined about factors such as the influence of population-genetic conditions, and the scope of applicability in natural adaptation.

On this basis, we can make a strong argument that knowledge of mutation bias can be leveraged to improve evolutionary forecasting. We have highlighted applications where we think this approach may prove particularly valuable, in relation to somatic evolution, resistance to toxins, and the adaptation of infectious agents to make use of host resources. Our hope is that this review will serve as a useful source of guidance for those implementing approaches to prediction, and that the information contained in it will be quickly eclipsed by more diverse, general, and precise results.

7 Acknowledgments

Box 1   Mutational biases in models with finite and infinite sites

Consider the case where there are m possible beneficial mutations, and the i-th beneficial mutation has selection coefficient s_i and mutation rate μ_i. We assume that the s_i's are drawn independently from the same distribution of mutational effects on fitness. For a population of size N, under what circumstances do the μ_i's influence which of these m mutations will be the first to reach fixation?

To gain some intuition for this question it is helpful to consider two limiting situations:

1. In the limit of very large mutation supply, in particular if $N\mu_i \gg 1$ for all μ_i, $i = 1, \ldots, m$, all possible beneficial mutations are introduced into the population in each generation. These mutations will all compete with each other and, assuming that mutation rates are small relative to selection coefficients, ultimately the fittest of the m mutations will go to fixation. To see the consequences of this result, consider single-nucleotide substitutions, which can be classified as either transitions or transversions. Because there are twice as many transversions than transitions, the fittest variant, that is the one that will eventually go to fixation, has two times the probability of being a transversion than being a transition. Thus, under these conditions, the expected transition:transversion ratio among adaptive substitutions would be 1:2, regardless of whether transitions arise at a higher rate.

2. Now let us consider the limit where all the $N\mu_i$ are small (specifically, assume $\sum_i N\mu_i s_i \ll \min_i s_i / \log N$, so that the first beneficial mutation to become established in the population will typically have sufficient time to reach fixation before the next beneficial mutation is able to become established, (Gerrish and Lenski, 1998)). In this setting, the probability that a mutation will be the first to go to fixation is proportional to both its mutation rate and its selection coefficient (Eq. 1), so that all other things being equal, we expect that classes of mutation with high mutation rates, such as transitions, will be over-represented amongst fixed mutations. For instance, if the mutation rate for individual transitions is κ times the mutation rate for individual transversions ($\mu_i/\mu_j = \kappa$ if i is a transition and j is a transversion), then we expect a transition:transversion ratio of κ:2 among fixed mutations.

Thus, broadly speaking, mutational biases will tend to have a stronger influence on molecular adaptation when the beneficial mutation supply is low, because in this regime the first beneficial mutation that becomes established in the population is likely to go to fixation rather than the fittest possible mutation, and the waiting time until a mutation becomes established is inversely proportional to its mutation rate.

Another common class of models are the infinite sites models, where we assume that each new mutation has a selection coefficient that is drawn independently from some distribution of fitness effects. In this class of models, if different types of mutations share the same distribution of fitness effects, then the relative proportions of different types of mutations among fixed mutations always varies in a manner directly proportional to the mutational bias. For example, if we consider transitions and transversions, for each selection coefficient s, the transitions:transversion ratio among mutations with that selection coefficient is

κ:2. Thus no matter how competition between co-segregating mutations alters the distribution of fixed selection coefficients relative to the overall distribution of fitness effects, the expected transition:transversion ratio among fixed mutations will always be κ:2. The results of this intuition are shown graphically using evolutionary simulations in Fig. B1.1. Even though we see the effects of competition between multiple adaptive mutations as a departures from the origin-fixation expectation that sets in at a total mutation supply $2N\mu \approx 1$ (Fig. B1.1a), the ratios of fixed mutations are proportional to the introduction rates regardless of the value of the mutation supply (Fig. B1.1b). In the more general case of infinite sites models for mutational types that do not share the same distribution of fitness effects, the strict proportionality with mutation rates need not hold, however a similar intuition applies in that we can consider the relative proportion of each mutational type for each possible selection coefficient s, and then the relative frequencies of fixed mutations can be determined by averaging these proportions over the characteristic distribution of selection coefficients (Neher, 2013) fixed during adaptation.
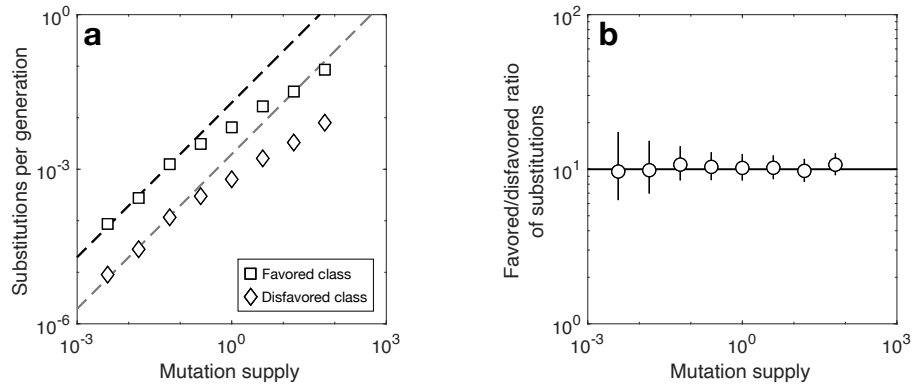

Fig. B1.1. Simulation of the infinite sites model. Two classes of mutations were simulated, one occurring with ten times the frequency of the other. The fitness effects of the two classes of mutations have the same exponential distribution and they combine additively. (a) The number of fixed mutations per generation. The dashed lines show the expectation under the origin-fixation regime (i.e., no clonal interference) for the favored class and the disfavored class in black and gray, respectively. Note the departure of the observed number of fixed mutations from the origin-fixation expectation as $2N\mu$ approaches 1, as the fixation of new mutations is slowed down by clonal interference. (b) The ratio of the favored and disfavored class among the fixed mutations. Vertical lines represent 95 % CIs estimated using the Poisson distribution. Analogous results for one-step adaptation are shown in Fig. S1.

The simulation code is available in a GitHub repository at https://github.com/alejvcano/infiniteSites.

Box 2   Quantifying contributions of mutation and selection to repeatability

How do mutation and selection contribute to the predictability of evolution? Can we partition their effects? One sense of predictability is repeatability, the chance that the outcome of evolution will match what we have seen before. Let us define repeatability as the chance of parallel evolution between a pair of trials. If we have n elementary outcomes, each happening with some probability p_i, then repeatability P_{para} is the sum of squares of p_i:

$$P_{\text{para}} = \sum_{i=1}^{n} p_i^2. \tag{B2.1}$$

This is equivalent to the measure known as the Simpson's index $S(\boldsymbol{p})$, where by $\boldsymbol{p}$ we denote the vector of p_i's (analogously, we use bold symbols to denote vectors of variables in the following). Simpson's index has a simple relationship to the variance $\text{Var}(\boldsymbol{p})$ or coefficient of variation $c(\boldsymbol{p})$ of the distribution of elementary probabilities:

$$P_{\text{para}} = S(\boldsymbol{p}) = n\,\text{Var}(\boldsymbol{p}) + 1/n = \frac{c(\boldsymbol{p})^2 + 1}{n}. \tag{B2.2}$$

Under uniformity, $\text{Var}(\boldsymbol{p}) = c(\boldsymbol{p}) = 0$, and $P_{\text{para}} = 1/n$. The greater the variance in probabilities, the greater the chance of parallelism. Any factor that increases variance, increases parallelism. Likewise, any approach to prediction that ignores variance, e.g., by aggregating outcomes into classes whose members are assigned the average behavior of the class, will underestimate parallelism.

One of the ways to quantify the effect of heterogeneity is to compute an effective number of options with the same chance of recurrence under uniformity, equal to the inverse of the probability of parallelism, i.e., $n_e = 1/P_{\text{para}}$, comparing that to n. If one state has $p = 1$ and the others have $p = 0$, then repeatability is 1, and $n_e = 1$. If all states are equally likely, then $p_i = 1/n$ for all i, $n_e = n$ and the repeatability is $\sum(1/n)^2 = 1/n$. From the 20 replicates of Rokyta et al. (2005), the counts ranked by selection coefficient are $\boldsymbol{k} = [1, 5, 3, 6, 1, 1, 1, 1, 1]$, thus repeatability is $\sum_i (k_i/20)^2 = 0.19$, and $n_e = 5.2$, i.e., the effect of heterogeneity in p_i's is like reducing the choices from 9 to 5.2. Note this calculation ignores sampling error by treating the observed frequency f_i as the true probability p_i.

How could we partition repeatability to effects of mutation and selection? This is possible for the special case of origin-fixation dynamics (McCandlish and Stoltzfus, 2014). For event i, an origin-fixation process specifies a rate $N\mu_i\pi_i$, with π_i being the fixation probability of the event, so that its application in the current context means that $p_i \propto \mu_i\pi_i$. Then the chance of parallelism is given by

$$P_{\text{para}} = \frac{(c(\boldsymbol{\mu})^2 + 1)(c(\boldsymbol{\pi})^2 + 1)}{n}, \tag{B2.3}$$

if we can assume that covariance of $\boldsymbol{\mu}$ and $\boldsymbol{\pi}$, as well as the covariance of the squares of μ_i's and squares of π_i's is 0. In general, the fixation probability $\pi(\cdot)$ is a function of s, the selection coefficient, and N, population size (Kimura, 1962). However, under strong selection weak mutation conditions $\pi(s, N) \approx 2s$

(Haldane, 1927), and this leads directly to the result of Chevin et al. (2010):

$$P_{\text{para}} \approx \frac{(c(\boldsymbol{\mu})^2 + 1)(c(\boldsymbol{s})^2 + 1)}{n}. \tag{B2.4}$$

Consider some applications of Eqn B2.3. For the results of Rokyta et al. (2005), $c(\boldsymbol{\pi})^2 + 1 = 1.086$ for the probabilities of fixation computed from the reported selection coefficients, and $c(\boldsymbol{\mu})^2 + 1 = 1.33$ for mutation rates (given the model described by the authors), so mutational heterogeneity contributes slightly more. For the 11 variants from Maclean et al. (2010) with known mutation rates, $c(\boldsymbol{\mu})^2 + 1 = 7.93$ and $c(\boldsymbol{\pi})^2 + 1 = 1.05$, so mutational heterogeneity is contributing much more. This disparity reflects a 30-fold variability in mutation rates, but only about 2-fold range in probabilities of fixation, given that all the resistant variants have large selection coefficients (because $\pi(s, N) \approx 2s$ does not apply for large s, we must use Eqn B2.3 instead of B2.4). Note that this framework does not fully partition the effects of mutation and selection, as these factors are conflated in determining n, which in practice reflects the number of mutations that are both sufficiently beneficial and sufficiently mutationally likely to have an appreciable chance of being detected.

Finally, it is of interest to consider repeatability when outcomes are aggregated. In the case of predicting phenotypes, a genotype-phenotype map is used to assign a phenotype to each of the n elementary outcomes. Alternatively, the categories could be defined by genes (Bailey et al., 2017, 2018), pathways, or GO categories, as per Tenaillon et al. (2012). Suppose that the n elementary outcomes are assigned fully to mutually exclusive groups of size $m_1, m_2, ...m_\ell$ by a function f such that $f(i) = j$ when outcome i is in group j, and suppose further that $f(\cdot)$ assigns outcomes randomly. Then (as given in the mathematical appendix), the expected probability P_{gpara} that two outcomes of evolution are in the same group is

$$\mathbb{E}P_{\text{gpara}} = \sum_{j=1}^{\ell} \left(\sum_{i|f(i)=j} p_i^2 \right) = \frac{n}{n-1} \left(S(\boldsymbol{p}) + S(\boldsymbol{g}) - S(\boldsymbol{p})S(\boldsymbol{g}) - \frac{1}{n} \right) \approx S(\boldsymbol{p}) + S(\boldsymbol{g}) - S(\boldsymbol{p})S(\boldsymbol{g}), \tag{B2.5}$$

where $S(\boldsymbol{p})$ is the Simpson's index over n elementary outcomes and $S(\boldsymbol{g}) = \sum_{j=1}^{\ell}(m_j/n)^2$ is the Simpson's index of the partition into ℓ groups (and where the approximation is valid for large n). The effect of aggregation is always to increase parallelism. Note that Eqn B2.5 is symmetric in $S(\boldsymbol{p})$ and $S(\boldsymbol{g})$, so the effect of the distribution of the elementary events' probabilities and the distribution of the categories' probabilities on the probability of parallelism is the same. The utility of this formula is that the prediction success of a concrete scheme of aggregation (e.g., GO categories) can be compared to the baseline expectation for a random aggregation with the same distribution of category sizes.

Mathematical Appendix

Here we derive the result given in Eqn B2.5.

Suppose that there are n possible outcomes occurring with probabilities $p_1, p_2, \ldots, p_n$. These outcomes are partitioned into ℓ groups of sizes $m_1, m_2, \ldots, m_\ell$ based on some function f such that $f(i) = j$ when

outcome i is in group j. Having fixed the assignment function $f(\cdot)$, we are interested in the probability P_{gpara} that two independent outcomes of evolution belong to the same group or category:

$$P_{\text{gpara}} = \sum_{j=1}^{\ell} \left(\sum_{i|f(i)=j} p_i \right)^2. \tag{A.1}$$

For purposes of establishing a baseline expectation, let us suppose that outcomes are assigned to categories randomly. We will now compute the expectation of P_{gpara} over all possible random assignments of events to categories, $\mathbb{E}P_{\text{gpara}}$.

By the linearity of expectations, $\mathbb{E}P_{\text{gpara}}$ is equal to

$$\mathbb{E}P_{\text{gpara}} = \mathbb{E}\left[\sum_{j=1}^{\ell} \left(\sum_{i|f(i)=j} p_i \right)^2 \right] = \sum_{j=1}^{\ell} \mathbb{E}\left(\sum_{i|f(i)=j} p_i \right)^2 = \sum_{j=1}^{\ell} \mathbb{E}X_j^2 \tag{A.2}$$

where X_j is a random variable defined as the sum of the probabilities of events assigned by a given instance of the assignment function f to category j. Each term of Eqn A.2 can then be computed using the standard formula for variance as

$$\mathbb{E}X_j^2 = \text{Var}\, X_j + (\mathbb{E}X_j)^2. \tag{A.3}$$

$\text{Var}\, X_j$ is given by the formula for the variance of the sum of m_j values drawn without replacement from a finite population of size n:

$$\text{Var}\, X_j = \frac{m_j(n - m_j)}{n - 1} \left(\frac{\sum_{i=1}^{n} p_i^2}{n} - \left(\frac{1}{n} \right)^2 \right) \tag{A.4}$$

where the term in parentheses is the population variance, and $\mathbb{E}X_j = m_j/n$. Rearranging, we get

$$\mathbb{E}X_j^2 = \frac{m_j}{n} \frac{n - m_j}{n - 1} \left(\sum_{i=1}^{n} p_i^2 \right) + \frac{m_j}{n} \frac{m_j - 1}{n - 1}. \tag{A.5}$$

Then

$$\mathbb{E}P_{\text{gpara}} = \sum_{j=1}^{\ell} \mathbb{E}X_j^2 = \sum_{j=1}^{\ell} \left(\frac{m_j}{n} \frac{n - m_j}{n - 1} \left(\sum_{i=1}^{n} p_i^2 \right) + \frac{m_j}{n} \frac{m_j - 1}{n - 1} \right), \tag{A.6}$$

which can be rewritten as

$$\mathbb{E}P_{\text{gpara}} = \frac{n}{n - 1} \left(S(p) + S(g) - S(p)S(g) - \frac{1}{n} \right) \approx S(p) + S(g) - S(p)S(g), \tag{A.7}$$

where $S(g) = \sum_{j=1}^{\ell} (m_j/n)^2$ is the Simpson's index of the partition into ℓ groups and $S(p) = \sum_{i=1}^{n} p_i^2$ is the Simpson's index over n outcomes (and where the approximation is valid for large n).
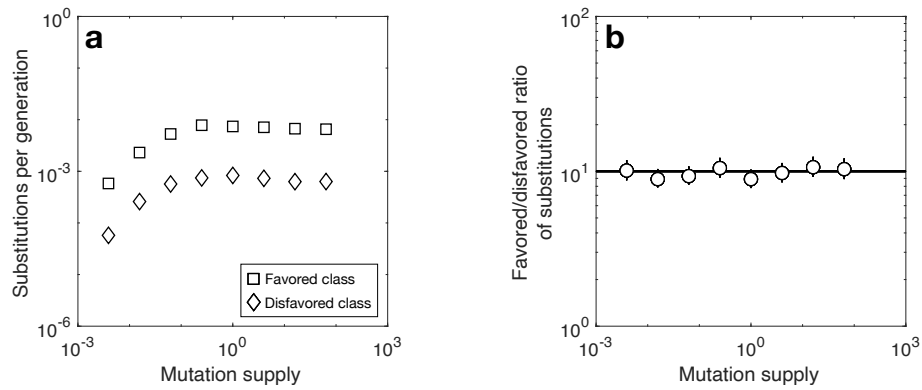
Fig. S1. Simulation of the infinite sites model in the case of one-step adaptation. Two classes of mutations were simulated, one occurring with ten times the frequency of the other. The fitness effects of the two classes of mutations have the same exponential distribution and they combine additively. We ran each simulation until the first fixation event, and repeat this process 1,000 times to record each type class of beneficial mutation that went to fixation, as well as the total number of generations. (a) The number of fixed mutations per generation. (b) The ratio of the favored and disfavored class among the fixed mutations. Vertical lines represent 95 % CIs estimated using the Poisson distribution.

The simulation code is available in a GitHub repository at https://github.com/alejvcano/infiniteSites.

References

M. L. Aardema, Y. Zhen, and P. Andolfatto. The evolution of cardenolide-resistant forms of na(+),k(+)-atpase in danainae butterflies. *Mol. Ecol.*, 21:340–9, 2012. doi: 10.1111/j.1365-294X.2011.05379.x.

A. H. Abas, S. Marfuah, R. Idroes, D. Kusumawaty, Fatimawali, M. N. Park, A. Siyadatpanah, F. A. Alhumaydhi, S. Mahmud, T. E. Tallei, T. B. Emran, and B. Kim. Can the SARS-CoV-2 omicron variant confer natural immunity against COVID-19? *Molecules*, 27:2221, 2022. doi: 10.3390/molecules27072221.

A. Acevedo, L. Brodsky, and R. Andino. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature*, 505:686–690, 2014. doi: 10.1038/nature12861.

S. Ahnert. Structural properties of genotype-phenotype maps. *J. Roy. Soc. Interface*, 14:20170275, 2017. doi: 10.1098/rsif.2017.0275.

A. Bachar, E. Itzhaki, S. Gleizer, M. Shamshoom, R. Milo, and N. Antonovsky. Point mutations in topoisomerase I alter the mutation spectrum in E. coli and impact the emergence of drug resistance genotypes. *Nucl. Acids Res.*, 48:761–769, 2020. doi: 10.1093/nar/gkz1100.

S. F. Bailey, F. Blanquart, T. Bataillon, and R. Kassen. What drives parallel evolution?: How population size and mutational variation contribute to repeated evolution. *Bioessays*, 39:1–9, 2017. doi: 10.1002/bies.201600176.

S. F. Bailey, Q. Guo, and T. Bataillon. Identifying drivers of parallel evolution: a regression model approach. *Genome Biol. Evol.*, 10:2801–2812, 2018. doi: 10.1093/gbe/evy210.

C. Bank, S. Matuszewski, R. Hietpas, and J. Jeffrey. On the (un)predictability of a large intragenic fitness landscape. *PNAS*, 113:14085–14090, 2016. doi: 10.1073/pnas.1612676113.

M. Bassalo, A. Garst, A. Choudhury, W. Grau, E. Oh, E. Spindler, T. Lipscomb, and R. Gill. Deep scanning lysine metabolism in Escherichia coli. *Mol. Syst. Biol.*, 14:e8371, 2018.

J. Beatty. The synthesis and the two scenarios. *Evolution*, 76:6–14, 2022. doi: 10.1111/evo.14423.

F. Bertels, C. Leemann, K. J. Metzner, and R. R. Regoes. Parallel evolution of HIV-1 in a long-term experiment. *Mol. Biol. Evol.*, 36:2400–2414, 2019. doi: 10.1093/molbev/msz155.

F. Besnard, J. Picao-Osorio, C. Dubois, and M. A. Felix. A broad mutational target explains a fast rate of phenotypic evolution. *Elife*, 9, 2020. doi: 10.7554/eLife.54928.

Z. Blount, R. Lenski, and J. Loso. Contingency and determinism in evolution: Replaying life's tape. *Science*, 362:eaam5979, 2018. doi: 10.1126/science.aam5979.

V. Borges, M. Alves, M. Amicone, J. Isidro, L. Zé-Zé, S. Vaz Duarte, L. Vieira, R. Guiomar, J. Gomes, and I. Gordo. Mutation rate of sars-cov-2 and emergence of mutators during experimental evolution. *Evol. Med. Public Health*, page 142–155, 2021. doi: 10.1101/2021.05.19.444774.

J. J. Bull, M. R. Badgett, H. A. Wichman, J. P. Huelsenbeck, D. M. Hillis, A. Gulati, C. Ho, and I. J. Molineux. Exceptional convergent evolution in a virus. *Genetics*, 147:1497–507, 1997. doi: 10.1093/genetics/147.4.1497.

J. Calles, I. Justice, D. Brinkley, A. Garcia, and D. Endy. Fail-safe genetic codes designed to intrinsically contain engineered organisms. *Nucl. Acids Res.*, 47:10439–10451, 2019. doi: 10.1093/nar/gkz745.

V. L. Cannataro, S. G. Gaffney, and J. P. Townsend. Effect sizes of somatic mutations in cancer. *J. Natl. Cancer Inst.*, 110:1171–1177, 2018. doi: 10.1093/jnci/djy168.

V. L. Cannataro, S. G. Gaffney, T. Sasaki, N. Issaeva, N. K. S. Grewal, J. R. Grandis, W. G. Yarbrough, B. Burtness, K. S. Anderson, and J. P. Townsend. APOBEC-induced mutations and their cancer effect size in head and neck squamous cell carcinoma. *Oncogene*, 38:3475–3487, 2019. doi: 10.1038/s41388-018-0657-6.

V. L. Cannataro, J. D. Mandell, and J. P. Townsend. Attribution of cancer origins to endogenous, exogenous, and preventable mutational processes. *Mol. Biol. Evol.*, 39, 2022. doi: 10.1093/molbev/msac084.

A. V. Cano and J. L. Payne. Mutation bias interacts with composition bias to influence adaptive evolution. *PLoS Computat. Biol.*, 16:1–26, 2020. doi: 10.1371/journal.pcbi.1008296.

A. V. Cano, H. Rozhonova, A. Stoltzfus, D. M. McCandlish, and J. L. Payne. Mutation bias shapes the spectrum of adaptive substitutions. *PNAS*, 119:e2119720119, 2022. doi: 10.1073/pnas.2119720119.

S. L. Chang, H. Y. Lai, S. Y. Tung, and J. Y. Leu. Dynamic large-scale chromosomal rearrangements fuel rapid adaptation in yeast populations. *PLoS Genet.*, 9:e1003232, 2013. doi: 10.1371/journal.pgen.1003232.

L. M. Chevin, G. Martin, and T. Lenormand. Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evolution*, 64:3213–3321, 2010. doi: 10.1111/j.1558-5646.2010.01058.x.

D. Coombes, J. W. B. Moir, A. M. Poole, T. F. Cooper, and R. C. J. Dobson. The fitness challenge of studying molecular adaptation. *Biochem. Soc. Trans.*, 47:1533–1542, 2019. doi: 10.1042/BST20180626.

A. Couce, J. Guelfo, and J. Blázquez. Mutational spectrum drives the rise of mutator bacteria. *PLoS Genet.*, 9:e1003167, 2013. doi: 10.1371/journal.pgen.1003167.

A. Couce, A. Rodríguez-Rojas, and J. Blazquez. Bypass of genetic constraints during mutator evolution to antibiotic resistance. *Proc. R. Soc. B*, 282:20142698, 2015. doi: 10.1098/rspb.2014.2698.

W. D. Crill, H. A. Wichman, and J. J. Bull. Evolutionary reversals during viral adaptation to alternating hosts. *Genetics*, 154:27–37, 2000. doi: 10.1093/genetics/154.1.27.

C. de Boer, E. Vaishnav, R. Sadeh, E. Abeyta, N. Friedman, and A. Regev. Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nat. Biotechnol.*, 38:56–65, 2019. doi: 10.1038/s41587-019-0315-8.

J. A. de Visser and J. Krug. Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.*, 15:480–490, 2014. doi: 10.1038/nrg3744.

J. A. de Visser, S. F. Elena, I. Fragata, and S. Matuszewski. The utility of fitness landscapes and big data for predicting evolution. *Heredity*, 121:401–405, 2018. doi: 10.1038/s41437-018-0128-4.

K. Dingle, F. Ghaddar, P. Šulc, and A. A. Louis. Phenotype bias determines how RNA structures occupy the morphospace of all possible shapes. *Mol. Biol. Evol.*, 39:msab280, 2020. doi: 10.1093/molbev/msab280.

T. Dobzhansky, F. Ayala, G. Stebbins, and J. Valentine. *Evolution*. W.H. Freeman, 1977.

P. T. Dolan, S. Taguwa, M. A. Rangel, A. Acevedo, T. Hagai, R. Andino, and J. Frydman. Principles of dengue virus evolvability derived from genotype-fitness maps in human and mosquito cells. *eLife*, 10: e61921, 2021. doi: 10.7554/eLife.61921.

I. Drienovska and G. Roelfes. Expanding the enzyme universe with genetically encoded unnatural amino acids. *Nat. Catal.*, 3:193–202, 2020. doi: 10.1038/s41929-019-0410-8.

M. Ehrlich, X.-Y. Zhang, and N. M. Inamdar. Spontaneous deamination of cytosine and 5-methylcytosine residues in DNA and replacement of 5-methylcytosine residues with cytosine residues. *Mutat. Res.*, 238: 277–286, 1990. doi: https://doi.org/10.1016/0165-1110(90)90019-8.

V. Eldholm and F. Balloux. Antimicrobial resistance in *Mycobacterium tuberculosis*: The odd one out. *Trends Microbiol.*, 24:637–648, 04 2016. doi: 10.1016/j.tim.2016.03.007.

A. Eyre-Walker and P. Keightley. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.*, 8: 610–618, 2007. doi: 10.1038/nrg2146.

R. Fisher. *The Genetical Theory of Natural Selection*. Oxford University Press, London, 1930.

J. W. Fondon and H. R. Garner. Molecular origins of rapid and continuous morphological evolution. *PNAS*, 101:18058–18063, 2004. doi: 10.1073/pnas.0408118101.

I. Fragata, A. Blanckaert, M. A. Dias Louro, D. A. Liberles, and C. Bank. Evolution in the light of fitness landscape theory. *Trends Ecol. Evol.*, 34:69–82, 2019. doi: 10.1016/j.tree.2018.10.009.

S. Gagneux. Ecology and evolution of *Mycobacterium tuberculosis*. *Nat. Rev. Microbiol.*, 16:202–213, 2018. doi: 10.1038/nrmicro.2018.8.

S. C. Galen, C. Natarajan, H. Moriyama, R. E. Weber, A. Fago, P. M. Benham, A. N. Chavez, Z. A. Cheviron, J. F. Storz, and C. C. Witt. Contribution of a mutational hot spot to hemoglobin adaptation in high-altitude andean house wrens. *PNAS*, 112:13958–13963, 2015. doi: 10.1073/pnas.1507300112.

G. Gao, W. Liao, Q. Ma, B. Zhang, Y. Chen, and Y. Wang. KRAS G12D mutation predicts lower TMB and drives immune suppression in lung adenocarcinoma. *Lung Cancer*, 149:41–45, 2020. doi: 10.1016/j.lungcan.2020.09.004.

J. Garson, L. Wang, and S. Sarkar. How development may direct evolution. *Biology and Philosophy*, 18(2): 353–370, 2003.

P. J. Gerrish and R. E. Lenski. The fate of competing beneficial mutations in an asexual population. *Genetica*, 102:127–144, 1998. doi: 10.1023/A:1017067816551.

K. Gomez, J. Bertram, and J. Masel. Mutation bias can shape adaptation in large asexual populations experiencing clonal interference. *Proc. Roy. Soc. B*, 287:20201503, 2020. doi: 10.1098/rspb.2020.1503.

B. Good, M. Mcdonald, J. Barrick, R. Lenski, and M. Desai. The dynamics of molecular evolution over 60,000 generations. *Nature*, 551:45–50, 2017. doi: 10.1038/nature24287.

F. Gould, Z. S. Brown, and J. Kuzma. Wicked evolution: Can we address the sociobiological dilemma of pesticide resistance? *Science*, 360:728–732, 2018. doi: 10.1126/science.aar3780.

M. Gymrek, T. Willems, A. Guilmatre, H. Zeng, B. Markus, S. Georgiev, M. Daly, A. Price, J. Pritchard, A. Sharp, and Y. Erlich. Abundant contribution of short tandem repeats to gene expression variation in humans. *Nat. Genet.*, 48:22–29, 2015. doi: 10.1038/ng.3461.

H. K. Haddox, A. S. Dingens, S. K. Hilton, J. Overbaugh, and J. D. Bloom. Mapping mutational effects along the evolutionary landscape of HIV envelope. *eLife*, 7, 2018. doi: 10.7554/eLife.34420.

J. Haldane. A mathematical theory of natural and artificial selection. v. selection and mutation. *Proc. Cam. Phil. Soc.*, 26:220–230, 1927. doi: 10.1017/S0305004100015644.

J. B. Haldane. The part played by recurrent mutation in evolution. *Am. Nat.*, 67:5–19, 1933.

D. J. Hamelin, D. Fournelle, J. C. Grenier, J. Schockaert, K. A. Kovalchik, P. Kubiniok, F. Mostefai, J. D. Duquette, F. Saab, I. Sirois, M. A. Smith, S. Pattijn, H. Soudeyns, H. Decaluwe, J. Hussin, and E. Caron. The mutational landscape of SARS-CoV-2 variants diversifies T cell targets in an HLA-supertype-dependent manner. *Cell Syst.*, 13:143–157, 2022. doi: 10.1016/j.cels.2021.09.013.

N. J. Hawkins and B. A. Fraaije. Contrasting levels of genetic predictability in the evolution of resistance to major classes of fungicides. *Mol. Ecol.*, 30(21):5318–5327, 2021.

R. Hershberg and D. A. Petrov. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet.*, 2010. doi: 10.1371/journal.pgen.1001115.

S. T. Hess, J. D. Blake, and R. Blake. Wide variations in neighbor-dependent substitution rates. *J. Mol. Biol.*, 236:1022–1033, 1994. doi: https://doi.org/10.1016/0022-2836(94)90009-4.

A. Hodgkinson and A. Eyre-Walker. Variation in the mutation rate across mammalian genomes. *Nat. Rev. Genet.*, 12:756–766, 2011. doi: 10.1038/nrg3098.

J. S. Horton, L. M. Flanagan, R. W. Jackson, N. K. Priest, and T. B. Taylor. A mutational hotspot that determines highly repeatable evolution can be built and broken by silent genetic changes. *Nat. Comm.*, 12:6092, 2021. doi: 10.1038/s41467-021-26286-9.

D. Houle. How should we explain variation in the genetic variance of traits? *Genetica*, 102:241–253, 1998. doi: 10.1023/A:1017034925212.

P. Jern, R. A. Russell, V. K. Pathak, and J. M. Coffin. Likely role of APOBEC3G-mediated G-to-A mutations in HIV-1 evolution and drug resistance. *PLoS Pathog.*, 5:e1000367, 2009. doi: 10.1371/journal.ppat.1000367.

L. Jones, S. Riaz, A. Morales-Cruz, K. C. Amrine, B. McGuire, W. D. Gubler, M. A. Walker, and D. Cantu. Adaptive genomic structural variation in the grape powdery mildew pathogen, Erysiphe necator. *BMC Genomics*, 15:1081, 2014. doi: 10.1186/1471-2164-15-1081.

J. Kaplanis, N. Akawi, G. Gallone, J. McRae, E. Prigmore, C. Wright, D. Fitzpatrick, H. Firth, J. Barrett, and M. Hurles. Exome-wide assessment of the functional impact and pathogenicity of multinucleotide mutations. *Genome Res.*, 29:1047–1056, 2019. doi: 10.1101/gr.239756.118.

V. Katju and U. Bergthorsson. Old trade, new tricks: Insights into the spontaneous mutation process from the partnering of classical mutation accumulation experiments with high-throughput genomic approaches. *Genome Biol. Evol.*, 11:136–165, 2019. doi: 10.1093/gbe/evy252.

S. Katz, S. Avrani, M. Yavneh, H. S., J. Gross, and R. Hershberg. Dynamics of adaptation during three years of evolution under long-term stationary phase. *Mol. Biol. Evol.*, 38:2778–2790, 2021. doi: https://doi.org/10.1093/molbev/msab067.

W. Keulen, N. K. Back, A. van Wijk, C. A. Boucher, and B. Berkhout. Initial appearance of the 184Ile variant in lamivudine-treated patients is caused by the mutational bias of human immunodeficiency virus type 1 reverse transcriptase. *J. Virol.*, 71:3346–3350, 1997. doi: 10.1128/JVI.71.4.3346-3350.1997.

M. Kimura. On the probability of fixation of mutant genes in a population. *Genetics*, 47:713–719, 1962. doi: 10.1093/genetics/47.6.713.

G. I. Lang, D. P. Rice, M. J. Hickman, E. Sodergren, G. M. Weinstock, D. Botstein, and M. M. Desai. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature*, 500: 571–574, 2013. doi: 10.1038/nature12344.

H. Lee, E. Popodi, H. Tang, and P. L. Foster. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *PNAS*, 2012. doi: 10.1073/pnas.1210309109.

S. Leighow, C. Liu, H. Inam, B. Zhao, and J. Pritchard. Multi-scale predictions of drug resistance epidemiology identify design principles for rational drug design. *Cell Rep.*, 30:3951–3963, 2020. doi: 10.1016/j.celrep.2020.02.108.

T. Lenormand, L. M. Chevin, and T. Bataillon. Parallel evolution: what does it (not) tell us and why is it (still) interesting? In G. Ramsey and C. Pence, editors, *Chance in Evolution*, pages 196–220. University of Chicago Press, Chicago, 2016.

C. Li and J. Zhang. The fitness landscapes of a tRNA gene. *Science*, 352:1025–1032, 2018. doi: 10.1126/science.aae0568.

P. Lind. Repeatability and predictability in experimental evolution. In P. Pontarotti, editor, *Evolution, Origin of Life, Concepts and Methods*, pages 57–83. Springer, Cham, Switzerland, 2019.

P. A. Lind, A. D. Farr, and P. B. Rainey. Experimental evolution reveals hidden diversity in evolutionary pathways. *eLife*, 4, 2015. doi: 10.7554/eLife.07074.

P. A. Lind, E. Libby, J. Herzog, and P. B. Rainey. Predicting mutational routes to new adaptive phenotypes. *eLife*, 8:e38822, 2019. doi: 10.7554/eLife.38822.

T.-L. Lite, R. Grant, I. Nocedal, M. Littlehale, M. Guo, and M. Laub. Uncovering the basis of protein-protein interaction specificity with a combinatorially complete library. *eLife*, 9:e60924, 2020. doi: 10.7554/eLife.60924.

Q. Liu, L. Via, T. Luo, L. Liang, X. Liu, S. Wu, Q. Shen, W. Wei, X. Ruan, X. Yuan, G. Zhang, C. Barry, and Q. Gao. Within patient microevolution of *Mycobacterium tuberculosis* correlates with heterogeneous responses to treatment. *Sci. Rep.*, 5:17507, 2015. doi: 10.1038/srep17507.

M. Lässig, V. Mustonen, and A. M. Walczak. Predicting evolution. *Nat. Ecol. Evol.*, 1:77, 2017. doi: 10.1038/s41559-017-0077.

C. Maclean, G. Perron, and A. Gardner. Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for Rifampicin resistance in *Pseudomonas aeruginosa*. *Genetics*, 186:1345–54, 2010. doi: 10.1534/genetics.110.123083.

E. Margoliash. Primary structure and evolution of Cytochrome C. *PNAS*, 50:672–9, 1963. doi: 10.1073/pnas.50.4.672.

E. Mayr. *Animal Species and Evolution*. Harvard University Press, Cambridge, Massachusetts, 1963.

D. McCandlish and A. Stoltzfus. Modeling evolution using the probability of fixation: history and implications. *Q Rev. Biol.*, 89:225–252, 2014. doi: 10.1086/677571.

J. R. Meyer, D. T. Dobias, J. S. Weitz, J. E. Barrick, R. T. Quick, and R. E. Lenski. Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science*, 335:428–32, 2012. doi: 10.1126/science.1214449.

J. G. Monroe, T. Srikant, P. Carbonell-Bejerano, C. Becker, M. Lensink, M. Exposito-Alonso, M. Klein, J. Hildebrandt, M. Neumann, D. Kliebenstein, M. L. Weng, E. Imbert, J. Agren, M. T. Rutter, C. B. Fenster, and D. Weigel. Mutation bias reflects natural selection in arabidopsis thaliana. *Nature*, 602: 101–105, 2022. doi: 10.1038/s41586-021-04269-6.

G. Moratorio, R. Henningsson, C. Barbezange, L. Carrau, A. Bordería, H. Blanc, S. Beaucourt, E. Poirier, T. Vallet, J. Boussier, B. Mounce, M. Fontes, and M. Vignuzzi. Attenuation of RNA viruses by redirecting their evolution in sequence space. *Nat. Microbiol.*, 2:17088, 2017. doi: 10.1038/nmicrobiol.2017.88.

R. A. Neher. Genetic draft, selective interference, and population genetics of rapid adaptation. *Annu. Rev. Ecol. Evol. Syst.*, 44:195–215, 2013. doi: 10.1146/annurev-ecolsys-110512-135920.

P. Nosil, S. Flaxman, J. Feder, and Z. Gompert. Increasing our ability to predict contemporary evolution. *Nat. Comm.*, 11:5592, 2020. doi: doi.org/10.1038/s41467-020-19437-x.

H. A. Orr. The population genetics of adaptation: the adaptation of DNA sequences. *Evolution*, 56: 1317–1330, 2002. doi: 10.1111/j.0014-3820.2002.tb01446.x.

B. Papp, R. Notebaart, and C. Pál. Systems-biology approaches for predicting genomic evolution. *Nat. Rev. Genet.*, 12:591–602, 2011. doi: 10.1038/nrg3033.

J. L. Payne, F. Menardo, A. Trauner, S. Borrell, S. M. Gygli, C. Loiseau, S. Gagneux, and A. R. Hall. Transition bias influences the evolution of antibiotic resistance in *Mycobacterium tuberculosis*. *PLoS Biol.*, 17(5), 2019. doi: 10.1371/journal.pbio.3000265.

M. A. Phillips, J. L. Steenwyk, X. X. Shen, and A. Rokas. Examination of gene loss in the DNA mismatch repair pathway and its mutational consequences in a fungal phylum. *Genome Biol. Evol.*, 13(10), 2021. doi: 10.1093/gbe/evab219.

G. Pines, J. Winkler, P. A., and R. Gill. Refactoring the genetic code for increased evolvability. *mBio*, 8: e10654–17, 2017. doi: 10.1128/mBio.01654-17.

R. Poulos, Y. Wong, R.Ryan, H. Pang, and J. Wong. Analysis of 7,815 cancer exomes reveals associations between mutational processes and somatic driver mutations. *PLoS Genet.*, 14:e1007779, 2018. doi: 10.1371/journal.pgen.1007779.

C. Qiu, O. Erinne, J. Dave, P. Cui, H. Jin, N. Muthukrishnan, L. Tang, S. Babu, K. Lam, V. Vandeventer, R. Strohner, J. Van den Brulle, S. Sze, and C. Kaplan. High-resolution phenotypic landscape of the RNA polymerase II trigger loop. *PLoS Genet.*, 12:e1006321, 2016. doi: 10.1371/journal.pgen.1006321.

A. Ravikumar, A. Arrieta, and L. C.C. An orthogonal DNA replication system in yeast. *Nat. Chem. Biol.*, 10:175–177, 2014. doi: 10.1038/nchembio.1439.

A. Ravikumar, A. G.A., M. Obadi, A. Javanpour, and L. C.C. Scalable, continuous evolution of genes at mutation rates above genomic error thresholds. *Cell*, 175:1946–1957, 2018. doi: 10.1016/j.cell.2018.10. 021.

A. M. Rice, A. Castillo Morales, A. T. Ho, C. Mordstein, S. Muhlhausen, S. Watson, L. Cano, B. Young, G. Kudla, and L. D. Hurst. Evidence for strong mutation bias toward, and selection against, U content in SARS-CoV-2: Implications for vaccine design. *Mol. Biol. Evol.*, 38:67–83, 2021. doi: 10.1093/molbev/ msaa188.

G. Rix, E. Watkins-Dulaney, P. Almhjell, C. Boville, F. Arnold, and C. Liu. Scalable, continuous evolution for the generation of diverse enzyme variants encompassing promiscuous activities. *Nat. Comm.*, 11: 5644, 2020. doi: 10.1038/s41467-020-19539-6.

L. Robert, J. Ollion, J. Robert, X. Song, I. Matic, and M. Elez. Mutation dynamics and fitness effects followed in single cells. *Science*, 359:1283–1286, 2018. doi: 10.1126/science.aan0797.

I. B. Rogozin and Y. I. Pavlov. Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutat. Res.*, 544:65–85, 2003. doi: 10.1016/s1383-5742(03)00032-2.

D. Rokyta, P. Joyce, S. Caudle, and H. Wichman. An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. *Nat. Genet.*, 37:441–444, 2005. doi: 10.1038/ng1535.

A. M. Sackman, L. W. McGee, A. J. Morrison, J. Pierce, J. Anisman, H. Hamilton, S. Sanderbeck, C. Newman, and D. R. Rokyta. Mutation-driven parallel evolution during viral adaptation. *Mol. Biol. Evol.*, 34: 3243–3253, 2017. doi: 10.1093/molbev/msx257.

M. Sane, G. D. Diwan, B. A. Bhat, L. M. Wahl, and D. Agashe. Shifts in mutation spectra enhance access to beneficial mutations. *bioRxiv*, 2022. doi: 10.1101/2020.09.05.284158.

T. Sankar, B. Wastuwidyaningtyas, Y. Dong, S. Lewis, and J. Wang. The nature of mutations induced by replication-transcription collisions. *Nature*, 535:178–181, 2016. doi: 10.1038/nature18316.

K. Santos-Moreno, E. Tasiudi, H. Kusumawardhani, J. Stelling, and Y. Schaerli. Synthetic genotype networks. *bioRxiv*, 2022. doi: 10.1101/2022.09.01.506159.

K. Sarkisyan, B. D.A., V. Meer, D. Usmanova, S. Mishin, G. Sharonov, D. Ivankov, N. Bozhanova, M. Baranov, O. Soylemez, N. Bogatyreva, P. Vlasov, E. Egorov, M. Logacheva, S. Kondrashov, D. Chudakov, E. Putintseva, I. Mamedov, D. Tawfik, K. Lukyanov, and F. Kondrashov. Local fitness landscape of the green fluorescent protein. *Nature*, 533:397–401, 2016. doi: 10.1038/nature17995.

Y. Schaerli, A. Jimenez, J. Duarte, L. Mihajlovic, J. Renggli, M. Isalan, J. Sharpe, and A. Wagner. Synthetic circuits reveal how mechanisms of gene regulatory networks constrain evolution. *Mol. Syst. Biol.*, 14: e8102, 2018. doi: 10.15252/msb.20178102.

S. Schaper and A. A. Louis. The arrival of the frequent: how bias in genotype-phenotype maps can steer populations to local optima. *PLoS One*, 9:e86635, 2014. doi: 10.1371/journal.pone.0086635.

D. Schrider, J. Hourmozdi, and M. Hahn. Pervasive multinucleotide mutational events in eukaryotes. *Curr. Biol.*, 21:1051–1054, 2011. doi: 10.1016/j.cub.2011.05.013.

A. Simon, S. Doelsnitz, and A. Ellington. Synthetic evolution. *Nat. Biotechnol.*, 37:730–746, 2019. doi: 10.1038/s41587-019-0157-4.

G. G. Simpson. Organisms and molecules in evolution. *Science*, 146:1535–1538, 1964. doi: 10.1126/science.146.3651.1535.

L. Smeds, A. Qvarnström, and H. Ellegren. Direct estimate of the rate of germline mutation in a bird. *Genome Res.*, 26:1211–1218, 2014. doi: 10.1101/gr.204669.116.

A. A. Soares, L. Wardil, L. B. Klaczko, and R. Dickman. Hidden role of mutations in the evolutionary process. *Phys. Rev. E*, 104:044413, 2021. doi: 10.1103/PhysRevE.104.044413.

G. Stebbins. *Processes of Organic Evolution*. Prentice Hall, Englewood Cliffs, NJ, 1966.

D. L. Stern and V. Orgogozo. The loci of evolution: how predictable is genetic evolution? *Evolution*, 62: 2155–2177, 2008. doi: 10.1111/j.1558-5646.2008.00450.x.

D. L. Stern and V. Orgogozo. Is genetic evolution predictable? *Science*, 323:746–751, 2009. doi: 10.1126/science.1158997.

A. Stoltzfus. Mutation-biased adaptation in a protein NK model. *Mol. Biol. Evol.*, 23:1852–1862, 2006. doi: 10.1093/molbev/msl064.

A. Stoltzfus. Why we don't want another "synthesis". *Biol. Direct*, 12:23, 2017. doi: 10.1186/ s13062-017-0194-1.

A. Stoltzfus. *Mutation, Randomness and Evolution*. Oxford, 2021.

A. Stoltzfus and D. M. McCandlish. Mutational biases influence parallel adaptation. *Mol. Biol. Evol.*, 34: 2163–2172, 2017. doi: 10.1093/molbev/msx180.

A. Stoltzfus and R. W. Norris. On the causes of evolutionary transition:transversion bias. *Mol. Biol. Evol.*, 33:595–602, 2016. doi: 10.1093/molbev/msv274.

A. Stoltzfus and L. Y. Yampolsky. Climbing mount probable: mutation as a cause of nonrandomness in evolution. *J Hered.*, 100:637–647, 2009. doi: 10.1093/jhered/esp048.

J. F. Storz, C. Natarajan, A. V. Signore, C. C. Witt, D. M. McCandlish, and A. Stoltzfus. The role of mutation bias in adaptive molecular evolution: insights from convergent changes in protein function. *Phil. Trans. Roy. Soc. B*, 374:20180238, 2019. doi: 10.1098/rstb.2018.0238.

M. A. Streisfeld and M. D. Rausher. Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution. *Evolution*, 65:629–42, 2011. doi: 10.1111/j.1558-5646.2010.01165.x.

D. Tack, P. Tonner, A. Pressman, N. Olson, S. Levy, E. Romantseva, N. Alperovich, O. Vasilyeva, and D. Ross. The genotype-phenotype landscape of an allosteric protein. *Mol. Syst. Biol.*, 17:e10179, 2021. doi: 10.15252/msb.202010179.

C. Tan, J. D. Mandell, K. Dasari, V. L. Cannataro, J. A. Alfaro-Murillo, and J. P. Townsend. Heavy mutagenesis by tobacco leads to lung adenocarcinoma tumors with KRAS G12 mutations other than G12D, leading KRAS G12D tumors-on average-to exhibit a lower mutation burden. *Lung Cancer*, 166: 265–269, 2022. doi: 10.1016/j.lungcan.2021.10.008.

D. Temko, I. Tomlinson, S. Severini, B. Schuster-Bockler, and T. Graham. The effects of mutational processes and selection on driver mutations across cancer types. *Nat. Comm.*, 9:1857, 2018. doi: 10.1038/s41467-018-04208-6.

O. Tenaillon, A. Rodríguez-Verdugo, R. L. Gaut, P. McDonald, A. F. Bennett, A. D. Long, and B. S. Gaut. The molecular diversity of adaptive convergence. *Science*, 2012. doi: 10.1126/science.1212986.

G. Tonkin-Hill, I. Martincorena, R. Amato, A. Lawson, M. Gerstrung, I. Johnston, D. Jackson, N. Park, S. Lensing, M. Quail, S. Gonçalves, C. Ariani, M. Chapman, W. Hamilton, L. Meredith, G. Hall, A. Jahun, Y. Chaudhry, M. Hosmillo, and D. Kwiatowski. Patterns of within-host genetic diversity in sars-cov-2. *eLife*, 10:e66857, 2021. doi: 10.7554/eLife.66857.

A. Trauner, Q. Liu, L. Via, X. Liu, X. Ruan, L. Liang, H. Shi, Y. Chen, Z. Wang, R. Lian, W. Zhang, W. Wang, J. Gao, G. Sun, D. Brites, K. England, G. Zhang, S. Gagneux, C. Barry, and Q. Gao. The within-host population dynamics of *Mycobacterium tuberculosis* vary with treatment efficacy. *Genome Biol.*, 18:71, 2017. doi: 10.1186/s13059-017-1196-0.

M. Tuffaha, S. Varakunan, D. Castellano, R. Gutenkunst, and L. Wahl. Shifts in mutation bias promote mutators by altering the distribution of fitness effects. *bioRxiv*, 2022. doi: 10.1101/2022.09.27.509708.

B. Ujvari, N. R. Casewell, K. Sunagar, K. Arbuckle, W. Wuster, N. Lo, D. O'Meally, C. Beckmann, G. F. King, E. Deplazes, and T. Madsen. Widespread convergence in toxin resistance by predictable molecular evolution. *PNAS*, 112:11911–11916, 2015. doi: 10.1073/pnas.1511706112.

E. Vaishnav, C. de Boer, J. Molinet, M. Yassour, L. Fan, X. Adiconis, D. Thompson, J. Levin, F. Cubillos, and A. Regev. The evolution, evolvability, and engineering of gene regulatory DNA. *Nature*, 603:455–463, 2022. doi: 10.1038/s41586-022-04506-6.

J. Vandecraen, M. Chandler, A. Aertsen, and R. Van Houdt. The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit. Rev. Microbiol.*, 43:709–730, 2017. doi: 10.1080/1040841X. 2017.1303661.

Q. Wang, E. Pierce-Hoffman, B. Cummings, J. Alföldi, L. Francioli, L. Gauthier, A. Hill, A. O'Donnell-Luria, K. Karczewski, and D. Macarthur. Landscape of multi-nucleotide variants in 125,748 human exomes and 15,708 genomes. *Nat. Comm.*, 11:2539, 2020. doi: 10.1038/s41467-019-12438-5.

C. J. Watson, A. L. Papula, G. Y. P. Poon, W. H. Wong, A. L. Young, T. E. Druley, D. S. Fisher, and J. R. Blundell. The evolutionary dynamics and fitness landscape of clonal hematopoiesis. *Science*, 367: 1449–1454, 2020. doi: 10.1126/science.aay9333.

A. Woolston, L. J. Barber, B. Griffiths, O. Pich, N. Lopez-Bigas, N. Matthews, S. Rao, D. Watkins, I. Chau, N. Starling, D. Cunningham, and M. Gerlinger. Mutational signatures impact the evolution of anti-EGFR antibody resistance in colorectal cancer. *Nat. Ecol. Evol.*, 5:1024–1032, 2021. doi: 10.1038/s41559-021-01470-8.

M. Wortel, D. Agashe, S. Bailey, C. Bank, K. Bisschop, T. Blankers, J. Cairns, E. S. Colizzi, D. Cusseddu, M. Desai, M. Egas, J. Ellers, A. Groot, D. Hackel, M. Johnson, K. Kraaijeveld, J. Krug, L. Laan, M. Laessig, and P. Pennings. The why, what and how of predicting evolution across biology: from disease to biotechnology to biodiversity. *ecoevoarxiv*, 2021. doi: 10.32942/osf.io/4u3mg.

N. Wu, C. Dai, C. Olson, J. Lloyd-Smith, and R. Sun. Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife*, 5:e16965, 2016. doi: 10.7554/eLife.16965.

K. Xie, G. Wang, A. Thompson, J. Wucherpfennig, T. Reimchen, A. MacColl, D. Schluter, M. Bell, K. Vasquez, and D. Kingsley. DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science*, 363:81–84, 2019. doi: 10.1126/science.aan1425.

A. Yadon, K. Maharaj, J. Adamson, Y.-P. Lai, J. Sacchettini, T. Ioerger, E. Rubin, and A. Pym. A comprehensive characterization of pnca polymorphisms that confer resistance to pyrazinamide. *Nat. Comm.*, 8: 588, 2017. doi: 10.1038/s41467-017-00721-2.

L. Y. Yampolsky and A. Stoltzfus. Bias in the introduction of variation as an orienting factor in evolution. *Evol. Dev.*, 3:73–83, 2001. doi: https://doi.org/10.1046/j.1525-142x.2001.003002073.x.

Y. Zhen, M. L. Aardema, E. M. Medina, M. Schumer, and P. Andolfatto. Parallel molecular evolution in an herbivore community. *Science*, 337:1634–1637, 2012. doi: 10.1126/science.1226630.

X. Zhu, Y. Guan, A. V. Signore, C. Natarajan, S. G. DuBay, Y. Cheng, N. Han, G. Song, Y. Qu, H. Moriyama, F. G. Hoffmann, A. Fago, F. Lei, and J. F. Storz. Divergent and parallel routes of biochemical adaptation in high-altitude passerine birds from the qinghai-tibet plateau. *PNAS*, 115:1865–1870, 2018. doi: 10.1073/pnas.1720487115.

Y. O. Zhu, M. L. Siegal, D. W. Hall, and D. A. Petrov. Precise estimates of mutation rate and spectrum in yeast. *PNAS*, 2014. doi: 10.1073/pnas.1323011111.

E. Zuckerkandl and L. Pauling. *Molecular Disease, Evolution, and Genetic Heterogeneity*, pages 189–225. Academic Press, New York, 1962.

E. Zuckerkandl and L. Pauling. *Evolutionary Divergence and Convergence in Proteins*. Academic Press, New York, 1965.