On the Causes of Evolutionary Transition:Transversion Bias

Arlin Stoltzfus*[1,2] and Ryan W. Norris[3]
[1]Institute for Bioscience and Biotechnology Research, Rockville, MD
[2]Genome-scale Measurements Group, National Institute of Standards and Technology, Gaithersburg, MD
[3]Department of Evolution, Ecology and Organismal Biology, The Ohio State University
*Corresponding author: E-mail: arlin.stoltzfus@nist.gov.
Associate editor: Takashi Gojobori

Abstract

A pattern in which nucleotide transitions are favored several fold over transversions is common in molecular evolution. When this pattern occurs among amino acid replacements, explanations often invoke an effect of selection, on the grounds that transitions are more conservative in their effects on proteins. However, the underlying hypothesis of conservative transitions has never been tested directly. Here we assess support for this hypothesis using direct evidence: the fitness effects of mutations in actual proteins measured via individual or paired growth experiments. We assembled data from 8 published studies, ranging in size from 24 to 757 single-nucleotide mutations that change an amino acid. Every study has the statistical power to reveal significant effects of amino acid exchangeability, and most studies have the power to discern a binary conservative-vs-radical distinction. However, only one study suggests that transitions are significantly more conservative than transversions. In the combined set of 1,239 replacements (544 transitions, 695 transversions), the chance that a transition is more conservative than a transversion is 53 % (95 % confidence interval 50 to 56) compared with the null expectation of 50 %. We show that this effect is not large compared with that of most biochemical factors, and is not large enough to explain the several-fold bias observed in evolution. In short, the available data have the power to verify the "conservative transitions" hypothesis if true, but suggest instead that selection on proteins plays at best a minor role in the observed bias.

Key words: mutation, transition bias, replacement, substitution bias, mutation bias.

Introduction

Of the 12 types of changes from one nucleotide to another, 8 are "transversions" between a purine (A or G) and a pyrimidine (C or T), and the other 4 are "transitions." Early protein comparisons showed that related proteins often differ by transitions more than expected by chance (Fitch 1967; sources cited in Vogel 1972). By the 1980s, this "transition bias" was well known (Li et al. 1985). By the 1990s, systematists had noted effects on phylogeny inference (Wakeley 1996), and methods were revised to give more weight to transversion differences (Sinsheimer et al. 1997).

In many early works, this bias is presented as a ratio of differences, which makes the expected effect a complex function of the degree of sequence divergence. As the use of rate models became routine in comparative sequence analysis, the phenomenon of transition bias was redefined as a bias in instantaneous rates, relative to a null model of equal rates. Because every nucleotide site (e.g., a G site) may experience 1 type of transition (G→A) at rate α, and 2 types of transversion (G→C, G→T) at rate β, the aggregate rate ratio of transitions to transversions has a null expectation of $R = \alpha/(2\beta) = 0.5$. In some contexts, the ratio is expressed differently as $\kappa = \alpha/\beta = 1$. When considering amino acid changes, it is more relevant to compare the 116 possible transitions and 276 possible transversions that change a codon so as to encode a different amino acid (assuming the canonical genetic code), leading to a null expectation of

$R = 116\alpha/(276\beta) = 0.42\alpha/\beta$. Thus, the observation of roughly equal numbers of inferred transitions and transversions in classic works (Vogel and Kopun 1977), or in the extensive analysis of mammalian genes in Li (1997, table 7.2), indicates a bias of over 2-fold. Kumar (1996) estimates 2-fold to 5-fold rate biases in vertebrate mitochondrial genes (excluding 3rd positions). Other estimates may be found in the work cited by Rosenberg et al. (2003), but there is not (to our knowledge) a systematic contemporary review of this issue.

The causes of the observed bias have not been resolved. The hypothesis of a mutational cause—a transition:transversion bias in mutation—was promoted early by Vogel (1972; see also Vogel and Kopun 1977). This hypothesis was bolstered when DNA sequence comparisons revealed that a transition bias is observed in introns, pseudogenes, and other noncoding regions (Gojobori et al. 1982; Li et al. 1985), suggesting a cause that (like mutation) acts at the level of DNA, across the entire genome.

The alternative hypothesis that natural selection favors amino acid replacements via transitions is also common, and is argued on the grounds that transitions are "less severe with respect to the chemical properties of the original and mutant amino acids" (Rosenberg et al. 2003) or "tend to cause changes that conserve the chemical properties of amino acids" (Wakeley 1996), or that "the biochemical difference in the protein product tends to be greater for transversions" (Keller et al. 2007).

For purposes of evaluation, we can break down either the mutational hypothesis or the selective hypothesis into 1) a claim that there is an underlying bias (mutational or selective) favoring transitions and 2) a claim that this bias accounts for the observed evolutionary bias. For the mutational hypothesis, the existence of an underlying bias is indicated in direct studies of mutation (Schaaper and Dunn 1991; Lynch 2010; Schrider et al. 2013; Zhu et al. 2014), and by many indirect estimates based on the assumption of neutral sequence divergence (Petrov and Hartl 1999; Rosenberg et al. 2003; Zhao et al. 2004; Jiang and Zhao 2006; Morton et al. 2006), although Keller et al. (2007) report a lack of bias in grasshoppers. The bias typically is 2-fold to 4-fold over null expectations. In theory, a bias in mutation of magnitude B can cause a B-fold effect on the rate of evolution (Yampolsky and Stoltzfus 2001; McCandlish and Stoltzfus 2014). That is, the observed magnitude of mutation bias appears to be sufficient, in principle, to account for the observed evolutionary bias.

For the selective hypothesis, arguments to the effect that transitions are more conservative typically invoke a biochemical factor (or a composite such as the Grantham index) that correlates with patterns of evolutionary divergence, and is found to be more conserved by transitions than by transversions (Vogel and Kopun 1977; Zhang 2000). This form of argument suffers from a logical circularity: if mutation shapes patterns of evolutionary amino acid replacement, then biochemical factors chosen for their ability to make sense of evolutionary patterns are not independent of mutation.

Presumably no biochemical factor, nor any simple combination of factors, fully captures the effects of replacements in complex proteins operating in a complex milieu. Indeed, the use of biochemical surrogates would seem unnecessary, given the availability of more direct measurements. Systematic laboratory studies of the effects of amino acid replacements in proteins have been carried out for 25 years (Kleina and Miller 1990). Although early studies summarized by Yampolsky and Stoltzfus (2005) typically reported crude measures of biochemical or growth effects (e.g., a 2-valued scale of "−" and "+"), a number of more recent studies report a continuous measure of fitness for each mutant (Sanjuan et al. 2004; Carrasco et al. 2007; Domingo-Calap et al. 2009; Peris et al. 2010; Jacquier et al. 2013; Roscoe et al. 2013; Acevedo et al. 2014; Bloom 2014; Firnberg et al. 2014; Thyagarajan and Bloom 2014; Wu et al. 2014). Such studies provide direct evidence on the relative conservativeness of transitions and transversions that change amino acids.

Here we focus on whether direct measurements of fitness support the conservative transitions hypothesis, based on a collection of 8 studies comprising measured fitness values for 544 transitions and 695 transversions that change an amino acid. We assess the power of each study by comparing mutant fitnesses for each type of replacement (e.g., Ser to Pro) with a cross-validation predictor and with 2 existing measures of amino acid exchangeability called EX (Yampolsky and Stoltzfus 2005) and U (Tang et al. 2004). We find that for every mutation study, even the smallest, there is a significant correlation with one or more of these predictors; half of the studies show a highly significant

correlation ($P < 0.001$). More importantly, for most studies, measured fitness values correlate significantly with a conservative-vs-radical distinction based on EX or U. Specifically, a replacement designated as "conservative" has a 65% (EX) or 64% (U) chance of being more fit than a "radical" replacement.

However, the same studies typically do not show significant conservativeness of transitions. In the combined data, a transition has a 53% chance (95% confidence interval [CI] 50 to 56) of being more fit than a transversion, only slightly above the null expectation of 50%. We show that this effect is not large compared with that of most biochemical predictors, and is not large enough to explain the several-fold bias toward transition replacements observed in evolutionary studies. The mutation-bias hypothesis, though not proven, remains an obvious possibility, while the selective hypothesis would seem untenable.

Results

The literature search described in Materials and Methods (see Supplementary Material online) resulted in the eight data sets in table 1, each of which provides measures of fitness based on individual growth or paired growth (Sanjuan et al. 2004; Carrasco et al. 2007; Domingo-Calap et al. 2009; Peris et al. 2010; Jacquier et al. 2013; Rihn et al. 2013, 2015). We will refer to these 8 data sets as 8 studies, although they correspond to 7 publications, one of which (Domingo-Calap et al. 2009) reports separate mutant fitness distributions for 2 different phages.

Together these studies provide fitness data on 1,239 mutants covering 145 of the 150 possible amino acid changes that can be accomplished by single-nucleotide changes. Because measures of fitness from different studies are not scaled in the same way, we convert fitnesses to within-study quantiles, for example, the median fitness in a study is assigned a quantile of 0.5, and the fitness at the 95th percentile is assigned a quantile of 0.95. This set of mutants includes 544 transitions and 695 transversions. The ratio deviates from 1:2 because the error-prone polymerase chain reaction method used in the three largest studies produces roughly equal numbers of transitions and transversions.

Can Small Idiosyncratic Studies Detect General Amino Acid Effects?

The usefulness of these data for addressing the conservativeness of amino acid replacements in evolution might be limited for a variety of reasons including measurement error, the context-dependency of individual mutant effects—in the context of small numbers of observations from an idiosyncratic set of proteins—the use of artificial laboratory conditions, the fact that replacement mutations have effects other than the amino acid replacements (e.g., effects on mRNA stability), and the fact that fitness is not a direct or simple function of protein properties.

To assess the power of mutation studies individually and collectively, we correlate observed mutant fitness quantiles with expected values from three independent predictors: the

Table 1. Power of Eight Studies to Discern Generic Effects of Exchangeability.

Name	N	Cross-validation		EX		Tang's U	
		R^2	P value	R^2	P value	R^2	P value
Jacquier (TEM1)	757	0.021	1.0×10^{-4} **	0.085	2.9×10^{-16} **	0.053	1.6×10^{-10} **
Rihn (HIV integrase)	156	0.040	0.012*	0.069	9.0×10^{-4} **	0.074	6.1×10^{-4} **
Rihn (HIV capsid)	135	0.034	0.033*	0.056	5.9×10^{-3} **	0.17	5.6×10^{-7} **
Carrasco (TEV)	52	0.097	0.025*	0.045	0.13	0.056	0.091
Peris (f1)	51	0.12	0.014*	0.16	3.1×10^{-3} **	0.21	6.2×10^{-4} **
Domingo-Calap (Qbeta)	32	0.13	0.040*	0.10	0.073	0.17	0.018*
Sanjuan (VSV)	32	0.32	1.3×10^{-3} **	0.12	0.053	0.18	0.014*
Domingo-Calap (phiX174)	24	0.24	0.017*	0.37	1.7×10^{-3} **	0.51	9.4×10^{-5} **
Combined	1,239	0.047	6.3×10^{-7} **	0.088	9.4×10^{-21} **	0.090	3.2×10^{-16} **

NOTE.—P values for the null hypothesis of no correlation. *$P < 0.05$; **$P < 0.01$.

EX matrix (Yampolsky and Stoltzfus 2005), the U matrix of Tang et al (2004), and a cross-validation predictor. The cross-validation predictor applied to a given target study is constructed from all other studies (i.e., excluding the target study), and is simply a matrix of mean quantiles for each type of replacement (e.g., Ala to Val).

EX and U are used on the grounds of being powerful and mutationally unbiased predictors, whereas various biochemical predictors are less powerful (as will become apparent below), and various evolution-based measures other than U (e.g., PAM, BLOSUM), though perhaps powerful, cannot be used, because they are not known to be free of the mutational effects that we wish to exclude. The EX matrix, based on a meta-analysis of early mutation studies (which reported phenotypes other than fitness), was designed specifically to serve as a mutationally unbiased measure of exchangeability in models that separate selection from mutation. In a comparative evaluation, EX was shown to be as powerful, or more powerful, than a representative sample of other predictors (Yampolsky and Stoltzfus 2005). The "universal evolutionary index" or U matrix of Tang et al (2004) is based on modeling evolution of thousands of genes, using a method designed to separate codon-level mutational effects from protein-level effects. It purports to be a measure of evolutionary acceptability that scales directly with the rate of evolution.

The results of using EX, U, and a cross-validation predictor (table 1) indicate that even small studies of mutant fitnesses have considerable power to reveal generic effects of amino acid exchangeability. For instance, for the study of 135 HIV capsid mutants by Rihn et al., there is a significant correlation between the fitness reported for a mutant and the predictor for the relevant replacement type (e.g., Val to Ala), whether the predictor is EX, U, or a cross-validation predictor based on the other studies. This shows not only that individual studies are powerful, but that there is a consistency across studies: although most effects of an amino acid replacement in a protein are very context dependent (which is why the R^2 values are small), generic effects of exchangeability are seen across sites and proteins.

Are Transition Replacements More Conservative?

The conservative transitions hypothesis proposes that transitions collectively are more conservative than transversions.

To assess how well mutant fitness studies distinguish conservative replacements from radical ones, we construct two versions of this distinction, EX_B and U_B ("B" indicates a binary distinction, as opposed to a continuous measure), simply by designating higher-exchangeability replacements as conservative and the remainder as radical.

Table 2 shows how well studies of mutant fitness distinguish conservative from radical replacements, and how well they distinguish transitions from transversions ("TiTv" column). The measure of effect-size denoted "AUC" (Area Under the Curve) is the chance that a mutant designated as conservative is more fit than a randomly chosen radical mutant. This statistic is not affected by the relative sizes of the 2 classes; its range is from 0 to 1, with a null expectation of 0.5; higher values indicate that nominally conservative changes are indeed conservative. We call this measure AUC because it has the same meaning as the area under an ROC (receiver-operating characteristic) curve for a binary classifier. That is, as pointed out by Hanley and McNeil (1982), the AUC for a binary classifier is equivalent to the chance that a randomly chosen positive instance will be ranked higher than a randomly chosen negative instance (see Materials and Methods).

Even small studies have significant power to distinguish conservative from radical substitutions based on EX_B and U_B. In the combined data set, AUC is 0.65 for EX_B and 0.64 for U_B. That is, a conservative replacement according to EX_B has a 65% chance of being more fit than a randomly drawn radical replacement.

However, the same studies typically do not distinguish transitions from transversions. Only one study shows a marginally significant result ($P = 0.019$ for Jacquier et al.). The combined results for the entire set of 1,239 replacements are shown at the bottom of table 2. For the combined data, the AUC is 0.53, with a 95 % CI of 0.50 to 0.56 (based on 400 bootstrap replicates).

One might object that this approach is framed incorrectly, in that it uses the entire distribution of mutational effects, whereas the distribution of changes fixed in evolution is obviously weighted toward more modest effects, because natural selection removes the most damaging ones whether they are transitions or transversions. If the changes actually accepted in evolution are mostly in the top 50 %, or the top 10 %, of the fitness distribution, then this is the fraction that

Table 2. The Power of Eight Studies to Detect Various Binary Distinctions.

Name	N	EX_B AUC	EX_B P value	U_B AUC	U_B P value	TiTv AUC	TiTv P value
Jacquier (TEM1)	757	0.66	2.8×10^{-14} **	0.61	4.0×10^{-8} **	0.54	0.019*
Rihn (HIV integrase)	156	0.64	2.1×10^{-3} **	0.67	7.2×10^{-5} **	0.54	0.191
Rihn (HIV capsid)	135	0.60	7.1×10^{-3} **	0.63	7.4×10^{-4} **	0.50	0.50
Carrasco (TEV)	52	0.64	0.037*	0.61	0.087	0.50	0.50
Peris (f1)	51	0.60	0.13	0.72	4.6×10^{-3} **	0.54	0.31
Domingo-Calap (Qbeta)	32	0.60	0.16	0.65	0.076	0.75	0.084
Sanjuan (VSV)	32	0.61	0.13	0.75	6.1×10^{-3} **	0.31	0.93
Domingo-Calap (phiX174)	24	0.81	5.8×10^{-3} **	0.94	1.6×10^{-4} **	0.34	0.85
Combined	1,239	0.65	6.5×10^{-18} **	0.64	1.6×10^{-12} **	0.53	0.024*

NOTE.—P values for a one-sided test where the alternative is that nominally conservative replacements, or transitions, are more fit. *$P < 0.05$; **$P < 0.01$.
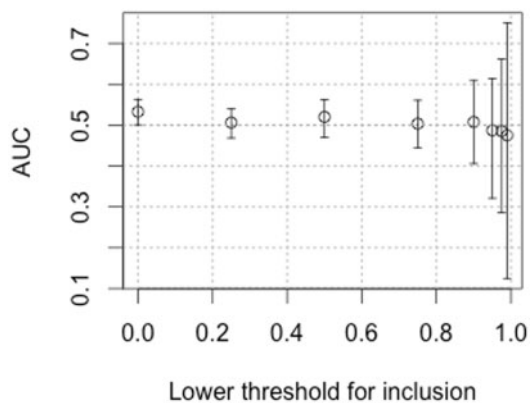

FIG. 1. The conservativeness of transitions when the distribution of mutant effects is truncated at the low end. The advantage of transitions (AUC) is shown as a function of threshold quantile for left-truncated data, for example, the AUC value for $x = 0.25$ is computed without the bottom 25 % of the distribution. Under the conservative transitions hypothesis, one might expect that, even if there is no advantage over the entire distribution, an advantage will appear at the high end. In fact, this is not observed. As mentioned in the text, AUC = 0.53 for the complete set of data, corresponding to a truncation threshold of 0, that is, no truncation. As the threshold increases, AUC decreases (rather than increases), although the differences are insignificant.

should be examined most closely to test the conservative transition hypothesis.

The effect of testing for a transition:transversion effect at successively higher thresholds of fitness is shown in figure 1. In fact, the AUC does not go up if we filter out the low end, but stays close to 0.5.

Another way to explore the upper end of the fitness distributions is to consider studies of mutational effects that focus on beneficial mutations (Ferris et al. 2007; MacLean et al. 2010; Miller et al. 2011; Schenk et al. 2012). These studies are small, with only 15 to 38 mutants. For the combined set of 111 beneficial mutants shown in table 3, the AUC for the conservative transitions hypothesis is 0.40 (95 % CI 0.28 to 0.51), suggesting that perhaps beneficial transitions are not more, but less fit than beneficial transversions. We note in passing that beneficial mutations seem less predictable in their effects than random mutations. That is, we would expect from table 2 that a set of 111 random mutants

would show predictable effects of exchangeability, because most smaller studies show significant effects.

What Is the Expected Evolutionary Effect Size?

As mentioned above, the conservative transitions hypothesis has 2 parts, a claim that transitions are conservative, and a claim that this conservativeness accounts for an evolutionary pattern. The present set of studies suggests that transitions are more conservative, but only slightly. How important could an effect of this size be?

One way to ask this question is to compare the transition:transversion distinction with various biochemical distinctions. Any quantitative property of an amino acid can be used to create a conservative-vs-radical distinction: for example, for a measure of the polarity of each amino acid, the conservative changes will be the ones with the least change in polarity. The AAindex database (Kawashima and Kanehisa 2000) has data on nearly 250 biochemical factors (see Materials and Methods). The results in figure 2 indicate that biochemical predictors typically are 1) considerably more powerful than the transition:transversion distinction and 2) considerably less powerful than EX and U.

Yet, one might object that natural selection has the ability to amplify small differences into major effects. Perhaps a difference with an effect size of AUC = 0.53 might translate into a several-fold bias in terms of evolutionary acceptance.

How do these two relate to each other? The U matrix illustrates this relationship, because values of U scale with evolutionary rates, and U_B has a known power as a binary predictor, namely AUC = 0.64. The ratio of U values for

Table 3. Relative Advantage of Transitions in Four Studies of Beneficial Mutations.

Name	N	EX_B AUC	EX_B P value	U_B AUC	U_B P value	TiTv AUC	TiTv P value
Schenk (TEM1)	38	0.46	0.63	0.40	0.84	0.49	0.53
MacLean (RpoB)	31	0.45	0.68	0.47	0.62	0.35	0.93
Miller (ID11)	27	0.58	0.30	0.62	0.18	0.25	0.96
Ferris (phi6)	15	0.69	0.18	0.64	0.20	0.39	0.76
Combined	111	0.52	0.50	0.50	0.57	0.40	0.95

NOTE.—P values for a one-sided test where the alternative is that nominally conservative replacements, or transitions, are more fit. *$P < 0.05$; **$P < 0.01$.

conservative replacements relative to radical ones is 2.7. That is, conservative replacements as defined by U_B are 2.7-fold more likely to be accepted in evolution than radical ones.

This pair of values, AUC = 0.64 and evolutionary bias = 2.7, represents one point in the relationship between evolutionary acceptability and classification power for mutant fitness effects. There is another point where AUC = 0.5 (no power) and evolutionary bias = 1 (no effect). We can fill in the relationship further by randomizing U_B, as shown in figure 3. The results show that, when about 75 % of the values are randomized, U_B has an AUC of 0.53, equal to that of the transition:transversion distinction. This corresponds to an evolutionary bias of 1.3. The CI of AUC from 0.50 to 0.56 for the transition:transversion distinction corresponds to the interval of 1.0 to 1.6 in evolutionary bias. That is, the expected evolutionary effect of the transition:transversion bias is a 1.3-fold bias, with a CI of 1.0 (no effect) to 1.6. This makes it unlikely that selection plays the major role in causing the evolutionary transition:transversion bias, which typically is several-fold favoring transitions.

Discussion

Based on a collection of eight studies that report fitnesses for replacement mutations, we have assessed the prospects for the hypothesis that the conservativeness of replacements via transition accounts for their increased frequency in evolution. Even small studies reveal predictable patterns of amino acid exchangeability, and most have sufficient power to distinguish a binary conservative-vs-radical distinction. However, the same studies typically do not show significant conservativeness of transitions. Overall, the chance of a transition mutation being more fit than a transversion is 53 % (95 % CI 50 to 56). This effect size is not large compared with that of

most biochemical predictors, and is not large enough to explain the several-fold bias toward transition replacements observed in evolutionary studies.

The finding that the conservativeness of transitions is a rather weak effect increases the prospects for the alternative mutational explanation, in which the rate at which new alleles are introduced by transition mutations is several-fold higher than for transversions, and this bias predisposes evolutionary change to happen via transitions (for a general explanation, see Stoltzfus and Yampolsky 2009).

Although this idea may be familiar, it relates to a rather substantial and unresolved issue in evolutionary genetics, which is the extent to which evolution in nature happens in the "gene pool" regime supposed by the architects of the Modern Synthesis, in the kind of mutation-driven regime supposed by early mutationists and later molecular evolutionists, or something in between (McCandlish and Stoltzfus 2014). The idea of mutation and selection as opposing forces suggests that mutation bias will be influential only when selection is absent, thus hypotheses that invoke mutation bias are often interpreted as neutral models (as noted by Yampolsky and Stoltzfus 2001). Presumably, this is why researchers have pursued selective explanations for transition:transversion bias among amino acid changes, even while accepting a mutational explanation for noncoding changes in the same genes: the proteins are assumed to be "under selection" and thus not susceptible to mutation bias. However, this way of depicting mutation and selection as opposing forces is only justified under the special conditions of the gene pool regime. Outside of this regime, mutation and selection can both contribute to orientation or direction in evolution (Yampolsky and Stoltzfus 2001; McCandlish and Stoltzfus 2014).
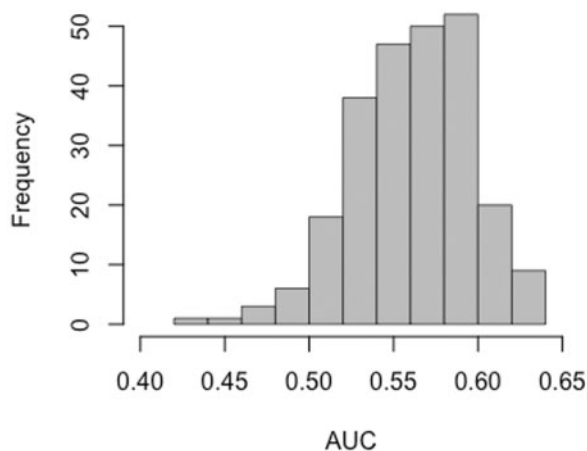
Fig. 2. The power of conservative–radical distinctions based on biochemical factors. The 245 biochemical factors from the AAIndex database were used to construct 245 conservative-vs-radical distinctions, which were then applied to the prediction of mutant fitnesses from mutation scanning experiments. The AUC is the chance that a nominally conservative mutant has a higher fitness than a randomly chosen radical mutant. The range of AUC is thus from 0 to 1, with a null expectation of 0.5 for a random predictor. Most predictors (84 %) are more powerful than the transition:transversion distinction (AUC = 0.53), and all are less powerful than EX_B or U_B (AUC = 0.65 or 0.64, respectively).
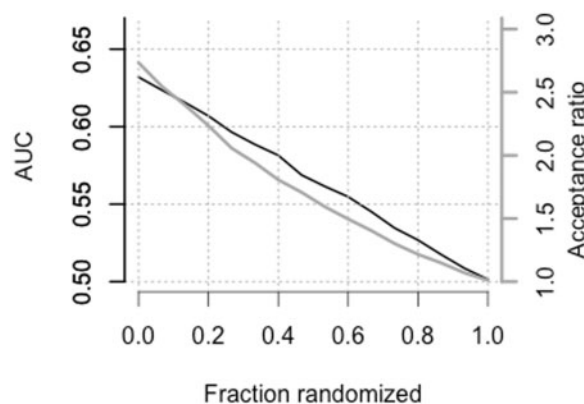
Fig. 3. Relationship between the power to predict mutant fitnesses and evolutionary effect size. AUC (black line scaled to left axis) and evolutionary acceptance ratio (gray line scaled to right axis) are shown for increasingly randomized versions of U_B. For the unrandomized U_B, the power in predicting mutational effects is AUC = 0.64, and this corresponds to an evolutionary acceptance ratio of 2.7 for conservative versus radical replacements. To estimate the evolutionary acceptance ratio for more modest values of AUC, we can weaken U_B by randomly reassigning conservative or radical labels to an increasingly large fraction of replacement types (200 replicates at each level of randomization). The AUC of 0.53 is reached at about 75 % randomization, where the evolutionary effect size is 1.3.

The results presented here also prompt the question of how it came to be so widely supposed that transitions are conservative. In a survey of the literature, we found that, when the alleged conservativeness of transitions is attributed to a source, the source is often Zhang (2000), or early works such as Fitch (1967), Grantham (1974), or Vogel and Kopun (1977). Grantham (1974) does not address this issue explicitly, but a genetic code-based calculation shows that the mean Grantham distance for transition-mediated replacements is lower than that for transversions, for example, as indicated in table 2 of Xia et al (1998). The study by Vogel and Kopun is often cited as evidence for the conservative transitions hypothesis, because they present a calculation that, for three different biochemical measures, suggests that transitions are more conservative.

These prior studies are inconclusive for two general reasons. The first is that none reports an effect size sufficient to account for the evolutionary bias. Indeed, Vogel and Kopun themselves favored a mutational explanation for the evolutionary bias on the grounds that the effect size for conservativeness of transitions seemed to be too small (see hypothesis 3 on p. 179). Zhang's (2000) analysis of three possible conservative:radical distinctions finds that the distinction based on Miyata et al (1979) yields the largest evolutionary effect size, which is a 2-fold effect, that is, radical replacements are roughly half as likely to accumulate, relative to null expectations. However, although the effect of conservativeness is 2-fold, the link reported between transitions and conservativeness is weak. According to Zhang (2000), the chance that a transition is conservative by Miyata's measure is 35 %, compared with 33 % for transversions, a proportional difference of only 6 % (i.e., 2/33 = 0.06). Miyata-conservativeness may be a 2-fold evolutionary effect, but if transitions are only 6 % more Miyata-conservative than transversions, the overall bias will be far less than 2-fold.

Second, none of these works escapes the kind of logical circularity pointed out by Di Giulio (2001), see also Yampolsky and Stoltzfus (2005), in which a measure of evolutionary tendencies is invoked to argue for effects of selection rather than mutation, ignoring the possibility that the pattern of evolution is itself influenced by mutational effects. This is an indirect (and thus presumably unintended) form of the Panglossian fallacy, that is, it is formally a fallacy of arguing that transitions are better simply because they happen more often, without inquiring into why they happen more often.

The circularity is not avoided by invoking biochemical factors. The popular composite indices of "biochemical" distance constructed by Grantham (1974) and Miyata et al (1979) are based on choosing biochemical factors that fit well with observed evolutionary patterns from earlier protein comparisons. Likewise, all three biochemical measures used by Vogel and Kopun (1977) are based on fitting to protein comparisons. The problem with this approach is suggested in figure 4, which shows the conservativeness of transitions for biochemical indices in the AAindex database (Kawashima and Kanehisa 2000). About 3/5 make transitions seem conservative, and the other 2/5 make them seem radical.

As figure 2 indicates, this is not because biochemical indices are generally poor predictors of exchangeability. Instead, among many moderately powerful predictors, there are ones that make transitions seem favorable, and others that make transversions seem favorable. Thus, converting evolutionary patterns into biochemical descriptors before reapplying them to the analysis of evolutionary patterns does not allow one to escape a logical circularity: some biochemical factors can be invoked to rationalize the conservativeness of transitions, whereas others can be invoked to rationalize the conservativeness of transversions.

Materials and Methods

Identification of Studies and Data Sets for Inclusion

An initial core set of studies (Sanjuan et al. 2004; Carrasco et al. 2007; Roscoe et al. 2013) was expanded by including other works cited by these studies. Then this set was expanded further by open-ended searches based on keywords or by tracking citations. In general, no text-based search does a good job of recovering mutation scanning studies of the desired type. Narrow searches (e.g., "distribution of mutational effects") implicate only a fraction of true positives and did little to expand the core set of studies; broad searches (e.g., "mutation" plus "fitness") implicate so many false positives that they are impractical and were abandoned. Most relevant studies cite the pioneering work of Sanjuan et al. (2004) or the seminal review by Eyre-Walker and Keightley (2007). Candidate studies identified in this manner were screened for appropriateness, ultimately resulting in the eight studies listed in table 1. The search covered literature
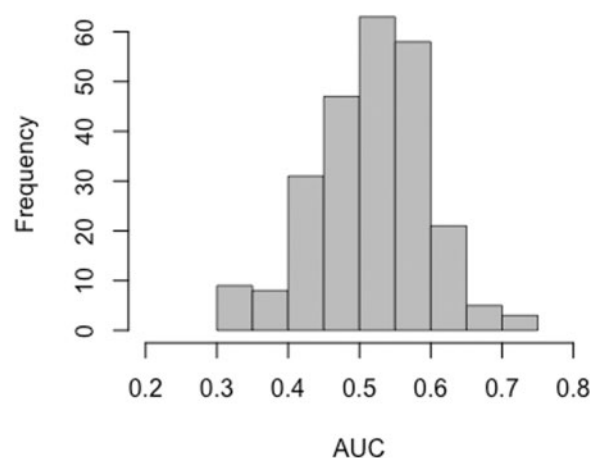

Fig. 4. The advantage of transitions implied by various biochemical factors. The 245 biochemical factors from AAindex were used to compute a pairwise similarity measure for amino acids indicating their biochemical similarity, then these measures were used to assess whether transitions are more conservative than transversions. AUC is the chance that a replacement due to a transition has a higher similarity score than a randomly chosen transversion (where the random sampling of transitions and transversions is based on the pool of actual mutants from the eight studies). The resulting distribution indicates that transitions are more conservative according to about 3/5 of biochemical factors (AUC > 0.5), and less conservative according to the other 2/5 of factors (AUC < 0.5).

published through December 2014 and does not include more recent studies.

We restricted our attention to studies with 1) a size of at least 20 replacement mutants (or, for beneficial studies, at least 10); 2) measures of growth (fitness) rather than biochemical activities; and 3) a random or arbitrary set of mutants. Most excluded studies of mutational effects have only a few mutants, or they report effects on binding or activity (but not on fitness), or they are focused on achieving particular outcomes rather than exploring a random set of variants, or they use deep sequencing to identify and quantify mutants, an approach that introduces uncontrolled nucleotide biases (supplementary material, section 2).

Processing and Management of Mutation Data

Starting from raw data tables supplied by authors (either directly, or via published supplements), all further processing and analysis steps were encoded in scripts. For each study used here, there is an R-Markdown (Rmd) file that (when executed in an appropriate environment, such as RStudio) describes and executes the steps (e.g., cleaning, recoding, sequence integration) to convert input data into a standard tabular form in which there is a single row describing each mutant and its effects. The figures and tables in this article are generated by further Rmd scripts that operate on the standardized input data.

Other Data Sources

Values of U are from Tang et al. (2004), and values of EX are from published supplements. Biochemical indices from the AAIndex database were accessed via the Interpol package (Heider 2012) and custom R code.

Note that, although AAindex lists 533 biochemical indices, less than half are pure biochemical indices. The others are based on some method of counting occurrences in naturally evolved proteins, for example, frequency with which an amino is found in a helix. Because the distribution of an amino acid in an evolving set of natural proteins will depend on the distributions of its closest mutational neighbors, such measures are not mutationally unbiased. They were removed using a custom list of name exclusion patterns ("[fF]requenc," "[pP]reference," "[cC]ompositi," "[pP]ropensit," "[dD]istribution," "[iI]nformation," "[wW]eights," "[oO]ccurrence," "[Pp]roportion," "probability," "mutability," "Geisow," "Janin"), resulting in a set of 245 indices.

Tests of Power and Effect

The results presented here rely mainly on standard statistical procedures. When P values are reported in table 1 for a linear predictor, this is from the t-test in the built-in linear model (lm) function in R. When P values are reported for binary predictors in tables 2 and 3, this is based on the Wilcoxon–Mann–Whitney test as implemented in the "wilcox.test" function of the R "stats" package, using a one-sided test. When summary P values are reported at the bottom of a table, these are based on Stouffer's method of combining

P values. When CIs are given on an AUC value, this is based on resampling using 400 bootstrap replicates.

The only unfamiliar methods involve the use of binary predictors. To convert a biochemical index C to a binary distinction, we first convert it to a pairwise similarity by the formula $S_{ij} = 1 - \text{abs}(C_i - C_j)/\text{max}$, where max is the maximum absolute difference. Converting a continuous measure of similarity into a binary measure is a simple matter of assigning all values above a particular quantile to the conservative class, and the rest to the radical class. To ensure that a constructed predictor is comparable with the transition:transversion distinction, the threshold is chosen so that the conservative class is the same size as the transition class in the data to be tested.

As explained above, we can define a measure of effect size with intuitive properties that we designate as AUC, based on an application of ROC analysis that may not be obvious. In ROC analysis of a binary classifier, each instance has a binary state (e.g., disease vs. nondisease), and the classifier makes a ranking of instances and predicts the binary state based on a threshold. The ROC curve plots the true-positive rate against the false-negative rate as the threshold varies, and the area under this curve is equivalent to the chance that a randomly chosen positive instance is ranked higher than a randomly chosen negative one (Hanley and McNeil 1982). If we treat the mutant fitness study as the classifier that supplies a ranking for each mutant, and the conservative-radical distinction as the binary state of a mutant, then the AUC is the chance that a mutant of a nominally conservative type has a higher fitness than a randomly chosen mutant of a nominally radical type. The relationship of AUC to the Wilcoxon–Mann–Whitney test is explained by Hanley and MacNeil (1982). Calculating AUC from the test statistic is an algebraic conversion based on the formula AUC = (pairs − WMW_statistic(x, y))/pairs, where x and y are vectors representing the two samples, and pairs = length(x) × length(y). This formula applies specifically to wilcox.test in the R stats package (some other implementations define the test statistic in a different way).

Note that converting fitnesses to within-study quantiles allows us to compare studies, and allows us to combine data for across-study tests. The use of quantiles rather than absolute fitnesses does not have any effect on a within-study AUC or Wilcoxon–Mann–Whitney test, which is a nonparametric statistic based on ranks.

Supplementary Material

Supplementary material and figure S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

or endorsement by the National Institute of Standards and Technology.

References

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

Bloom JD. 2014. An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Mol Biol Evol.* 31:1956–1978.

Carrasco P, Daros JA, Agudelo-Romero P, Elena SF. 2007. A real-time RT-PCR assay for quantifying the fitness of tobacco etch virus in competition experiments. *J Virol Methods.* 139:181–188.

Di Giulio M. 2001. The origin of the genetic code cannot be studied using measurements based on the PAM matrix because this matrix reflects the code itself, making any such analyses tautologous. *J Theor Biol.* 208:141–144.

Domingo-Calap P, Cuevas JM, Sanjuan R. 2009. The fitness effects of random mutations in single-stranded DNA and RNA bacteriophages. *PLoS Genet.* 5:e1000742.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet.* 8:610–618.

Ferris MT, Joyce P, Burch CL. 2007. High frequency of mutations that expand the host range of an RNA virus. *Genetics* 176:1013–1022.

Firnberg E, Labonte JW, Gray JJ, Ostermeier M. 2014. A comprehensive, high-resolution map of a gene's fitness landscape. *Mol Biol Evol.* 31:1581–1592.

Fitch WM. 1967. Evidence suggesting a non-random character to nucleotide replacements in naturally occurring mutations. *J Mol Biol.* 26:499–507.

Gojobori T, Li WH, Graur D. 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. *J Mol Evol.* 17:245–250.

Grantham R. 1974. Amino acid difference formula to help explain protein evolution. *Science* 185:862–864.

Hanley JA, McNeil BJ. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29–36.

Heider D. 2012. Interpolation of amino acid sequences, version 1.3.1. Essen (Germany): University of Duisburg-Essen.

Jacquier H, Birgy A, Le Nagard H, Mechulam Y, Schmitt E, Glodt J, Bercot B, Petit E, Poulain J, Barnaud G, et al. 2013. Capturing the mutational landscape of the beta-lactamase TEM-1. *Proc Natl Acad Sci U S A.* 110:13067–13072.

Jiang C, Zhao Z. 2006. Mutational spectrum in the recent human genome inferred by single nucleotide polymorphisms. *Genomics* 88:527–534.

Kawashima S, Kanehisa M. 2000. AAindex: amino acid index database. *Nucleic Acids Res.* 28:374.

Keller I, Bensasson D, Nichols RA. 2007. Transition-transversion bias is not universal: a counter example from grasshopper pseudogenes. *PLoS Genet.* 3:e22.

Kleina LG, Miller JH. 1990. Genetic studies of the lac repressor. XIII. Extensive amino acid replacements generated by the use of natural and synthetic nonsense suppressors. *J Mol Biol.* 212:295–318.

Kumar S. 1996. Patterns of nucleotide substitution in mitochondrial protein coding genes of vertebrates. *Genetics* 143:537–548.

Li WH. 1997. Molecular evolution. Sunderland (MA): Sinauer.

Li WH, Luo CC, Wu CI. 1985. Evolution of DNA sequences. In: MacIntyre RJ, editor. Molecular evolutionary genetics. New York: Plenum.

Lynch M. 2010. Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci U S A.* 107:961–968.

MacLean RC, Perron GG, Gardner A. 2010. Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for rifampicin resistance in *Pseudomonas aeruginosa*. *Genetics* 186:1345–1354.

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

Miller CR, Joyce P, Wichman HA. 2011. Mutational effects and population dynamics during viral adaptation challenge current models. *Genetics* 187:185–202.

Miyata T, Miyazawa S, Yasunaga T. 1979. Two types of amino acid substitutions in protein evolution. *J Mol Evol.* 12:219–236.

Morton BR, Bi IV, McMullen MD, Gaut BS. 2006. Variation in mutation dynamics across the maize genome as a function of regional and flanking base composition. *Genetics* 172:569–577.

Peris JB, Davis P, Cuevas JM, Nebot MR, Sanjuan R. 2010. Distribution of fitness effects caused by single-nucleotide substitutions in bacteriophage f1. *Genetics* 185:603–609.

Petrov DA, Hartl DL. 1999. Patterns of nucleotide substitution in *Drosophila* and mammalian genomes. *Proc Natl Acad Sci U S A.* 96:1475–1479.

Rihn SJ, Hughes J, Wilson SJ, Bieniasz PD. 2015. Uneven genetic robustness of HIV-1 integrase. *J Virol.* 89:552–567.

Rihn SJ, Wilson SJ, Loman NJ, Alim M, Bakker SE, Bhella D, Gifford RJ, Rixon FJ, Bieniasz PD. 2013. Extreme genetic fragility of the HIV-1 capsid. *PLoS Pathog.* 9:e1003461.

Roscoe BP, Thayer KM, Zeldovich KB, Fushman D, Bolon DN. 2013. Analyses of the effects of all ubiquitin point mutants on yeast growth rate. *J Mol Biol.* 425:1363–1377.

Rosenberg MS, Subramanian S, Kumar S. 2003. Patterns of transitional mutation biases within and among mammalian genomes. *Mol Biol Evol.* 20:988–993.

Sanjuan R, Moya A, Elena SF. 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc Natl Acad Sci U S A.* 101:8396–8401.

Schaaper RM, Dunn RL. 1991. Spontaneous mutation in the *Escherichia coli* lacI gene. *Genetics* 129:317–326.

Schenk MF, Szendro IG, Krug J, de Visser JA. 2012. Quantifying the adaptive potential of an antibiotic resistance enzyme. *PLoS Genet.* 8:e1002783.

Schrider DR, Houle D, Lynch M, Hahn MW. 2013. Rates and genomic consequences of spontaneous mutational events in *Drosophila melanogaster*. *Genetics* 194:937–954.

Sinsheimer JS, Lake JA, Little RJ. 1997. Inference for phylogenies under a hybrid parsimony method: evolutionary-symmetric transversion parsimony. *Biometrics* 53:23–38.

Stoltzfus A, Yampolsky LY. 2009. Climbing mount probable: mutation as a cause of nonrandomness in evolution. *J Hered.* 100:637–647.

Tang H, Wyckoff GJ, Lu J, Wu CI. 2004. A universal evolutionary index for amino acid changes. *Mol Biol Evol.* 21:1548–1556.

Thyagarajan B, Bloom JD. 2014. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *Elife* 3.

Vogel F. 1972. Non-randomness of base replacement in point mutation. *J Mol Evol.* 1:334–367.

Vogel F, Kopun M. 1977. Higher frequencies of transitions among point mutations. *J Mol Evol.* 9:159–180.

Wakeley J. 1996. The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. *Trends Ecol Evol.* 11:158–162.

Wu NC, Young AP, Al-Mawsawi LQ, Olson CA, Feng J, Qi H, Chen SH, Lu IH, Lin CY, Chin RG, et al. 2014. High-throughput profiling of influenza A virus hemagglutinin gene at single-nucleotide resolution. *Sci Rep.* 4:4942.

Xia X. 1998. The rate heterogeneity of nonsynonymous substitutions in mammalian mitochondrial genes. *Mol Biol Evol.* 15:336–344.

Yampolsky LY, Stoltzfus A. 2001. Bias in the introduction of variation as an orienting factor in evolution. *Evol Dev.* 3:73–83.

Yampolsky LY, Stoltzfus A. 2005. The exchangeability of amino acids in proteins. *Genetics* 170:1459–1472.

Zhang J. 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *J Mol Evol.* 50:56–68.

Zhao Z, Li H, Wu X, Zhong Y, Zhang K, Zhang YP, Boerwinkle E, Fu YX. 2004. Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evol Biol.* 4:21.

Zhu YO, Siegal ML, Hall DW, Petrov DA. 2014. Precise estimates of mutation rate and spectrum in yeast. *Proc Natl Acad Sci U S A.* 111:E2310–E2318.