The relationship between scaled selection coefficients and dN/dS

Stephanie J. Spielman * **and Claus O. Wilke** *

*Department of Integrative Biology, Center for Computational Biology and Bioinformatics, and Institute of Cellular and Molecular Biology. The University of Texas at Austin, Austin, TX 78712, USA.

Submitted to Proceedings of the National Academy of Sciences of the United States of America

The evolutionary rate ratio dN/dS, the ratio of nonsynonymous to synonymous substitution rates, is among the most commonly used metrics used to infer the strength of natural selection in a phylogenetic framework. Recently, however, mutation-selection-balance (MutSel) models ahve become a popular alternative to the dN/dS framework. MutSel models estimate site-specific amino-acid and/or codon scaled selection coefficients, which indicate the selective response to all possible mutations. However, it remains unknown how these two frameworks relate to one another, in particular whether dN/dS estimates reveal similar or distinct information from scaled selection coefficients. Therefore, we have derive a formal mathematical relationship between these two quantities. We prove that dN/dS values corresponding to MutSel scaled selection coefficients are strictly less than 1, demonstrating that MutSel models inherently cannot accomodate positive, diversifying selection. However, dN/dS can readily be greater than 1 if selection acts on synonymous changes, even though only purifying selection is occuring. Finally, we use this established relationship between models to evaluate performance of dN/dS maximum likelihood (ML) inference frameworks. We find that ML methods yield substantially biased dN/dS estimates when the fitted models do not exactly correspond to the mechanistic process that generated the data. Moreover, we show that the best-fitting model is not the model with the least bias and highest precision for parameter estimates of interest, and thus selecting models based on fit (AIC) can be counterproductive and positively misleading.

dN/dS | mutation-selection-balance models | scaled selection coefficients | Markov models of sequence evolution

Introduction

The oldest and most-widely used method to infer selection pressure in protein-coding genes calculates the evolutionary rate ratio dN/dS, which represents the ratio of nonsynonymous to synonymous substitution rates. This metric indicates how quickly a protein's constituent amino acids change, and it is commonly used to identify proteins or protein sites that experience negative selection ($dN/dS < 1$), evolve neutrally ($dN/dS \approx 1$), or experience positive, diversifying selection ($dN/dS > 1$) [1, 2, 3, 4]. Frameworks for calculating dN/dS have broadly fallen into two camps: heuristic counting methods [5, 6, 7, 8, 9] and maximum likelihood (ML) methods [10, 11, 1, 12]. The latter variety assume an explicit Markov-process model of sequence evolution to yield maximum likelihood estimates (MLEs) of the parameter ω, which represents the quantity dN/dS (although we note that other styles of these models use separate parameters for nonsynonymous and synonymous substitution rates [11, 13]). These dN/dS models have become a staple of comparative sequence analysis since their introduction in the 1990s (see ref [14] for a comprehensive review), and we will refer to them as ω models throughout this paper.

A second class of models, known as mutation-selection-balance (MutSel) models, are increasingly being viewed as a viable alternative to ω models. While ω models describe the how quickly a protein's constituent amino acids change, MutSel models assess the strength of natural selection operating on specific amino-acid or codon changes. In particular, the MutSel framework, couched firmly in population genetics theory, considers the specific selective responses to of all site-wise mutations in a protein-coding sequence [15, 16]. MutSel models yield estimates of site-wise scaled selection coefficients $S = 2N_e s$, which indicate the extent to which natural selection favors, or disfavors, particular codon or amino acid changes [15, 17, 18, 19]. Although first introduced over 15 years ago [15], MutSel models have seen little use due to their high computational expense. Recently, however, several computationally tractable model implementations have emerged [20, 21], allowing for the first time the potential for widespread adoption.

ω models have undergone rigorous development in their 20 years of existence and have advanced to high levels of sophistication. These models can accommodate a variety of evolutionary scenarios, including synonymous rate variation [11, 13], episodic [22, 23] and/or lineage-specific selection [24, 25, 26], and they can also incorporate information regarding protein structure and/or epistatic interactions [27, 28, 29, 30, 31]. This flexibility, along with accessible software implementations [32, 33, 34], make ω models a very attractive modeling choice. On the other hand, some have argued that MutSel models, given their explicit consideration of population genetics theory and attention to site-specific amino acid fitness differences, offer a more fine-grained approach to studying protein evolution than do dN/dS models [15, 18, 19, 16]. Recent phylogenetic studies have also demonstrated that evolutionary models which account for amino acid fitness values dramatically outperform other ω models, suggesting that MutSel models may more aptly represent the evolutionary process [35, 36].

Although both ω and MutSel models describe the same fundamental process of protein-coding sequence evolution along a phylogeny, it is unknown how these two modeling classes relate to one another. In particular, as these inference methods have been developed independently, it remains an open question whether or not parameter estimates from one model are comparable to those of the other model. As a consequence, although certain rhetorical arguments may be made in favor of using one method over another, there is currently no formalized, concrete rationale to guide researchers in their methodological choices. Elucidating the relationship between

Reserved for Publication Footnotes

these competing modeling frameworks will more precisely reveal under which circumstances the use of these models is justified.

Here, we formalize the relationship between ω and MutSel models by examining the extent to which their respective focal parameters, dN/dS and scaled selection coefficients, yield overlapping information about the evolutionary process. To this end, we derive a mathematical framework to calculate dN/dS values from scaled selection coefficients. We find that dN/dS values can be precisely calculated from scaled selection coefficients, and that dN/dS and the distribution of scaled selection coefficients are strongly related. Furthermore, we prove that, when synonymous mutations are neutral, dN/dS calculated from selection coefficients is necessarily less than 1. This proof demonstrates that MutSel models are inherently only able to model purifying selection, and therefore would be an inappropriate model choice if positive selection is expected. However, we also find that, when synonymous codons have different fitnesses, it is possible to recover dN/dS values above 1, even though no positive selection is occurring.

REWRITE PARAGRAPH Finally, we are able to use this robust relationship to assess the performance of ω models. If ω models are behaving as expected, we expect that they will yield the same dN/dS estimates as our calculations. We find that, in the absence of nucleotide compositional bias, dN/dS values inferred in an ML framework agree precisely with those calculated from scaled selection coefficients, meaning that MutSel and dN/dS models are in complete agreement. However, we found that, in the presence of mutational or nucleotide biases, ML inference frameworks produce systematically biased dN/dS estimates.

Results and Discussion

Theoretical model. We model sequence evolution as a continuous-time Markov process [12] under the assumptions of a fixed effective population size N_e and constant selection pressure over time. This process is governed by the 61×61 transition matrix $P(t) = e^{Qt}$, where the corresponding instantaneous rate matrix Q gives the instantaneous substitution probabilities between all 61 sense codons. We further assume that only single nucleotide changes occur instantaneously. We adopt the Halpern-Bruno [15, 17, 19, 16] MutSel modeling framework, which models the evolutionary process with explicit population genetics theory.

To being, let f_i be the fitness of codon i, and thus the selection coefficient acting on a mutation from codon i to codon j is $s_{ij} = f_j - f_i$ [37, 17]. The fixation probability for a mutation from codon i to codon j is given by

$$u_{ij} = \frac{2s_{ij}}{1 - e^{-2N_e s_{ij}}} \approx \frac{1}{N_e} \frac{2N_e s_{ij}}{1 - e^{-2N_e s_{ij}}} \qquad [1]$$

[38, 15, 17].

We further define $S_{ij} = 2N_e s_{ij}$ as the scaled selection coefficient for this mutation. We model the substitution as the product of fixation and mutation rates, μ.

Therefore, the substitution probability for codon i to codon j is

$$q_{ij} = N_e \mu_{ij} u_{ij} = \mu_{ij} \frac{S_{ij}}{1 - e^{-S_{ij}}}, \qquad [2]$$

and this expression corresponds to the instantaneous matrix element Q_{ji}. [15, 37].

Given detailed balance (reversibility), we have

$$q_{ij} P_i = q_{ji} P_j, \qquad [3]$$

where P_i is the stationary frequency of codon i.

From equations [2] and [3], we can write the ratio of substitution probabilities as

$$\frac{q_{ij}}{q_{ji}} = \frac{P_i \mu_{ij} S_{ij} (1 - e^{-S_{ji}})}{P_j \mu_{ji} S_{ji} (1 - e^{-S_{ij}})} \qquad [4]$$

Given that $S_{ij} = -S_{ji}$, we can simplify equation [4] to show that $q_{ij}/q_{ji} = e^{S_{ij}}$, and we therefore find that

$$S_{ij} = \ln \left(\frac{P_j \mu_{ji}}{P_i \mu_{ij}} \right). \qquad [5]$$

These equations establish a relationship between scaled selection coefficients and the stationary codon frequencies of the Markov model. Moreover, we note that in the specific case of symmetric mutation rates $\mu_{ij} = \mu_{ji}$, we have $S_{ij} = \ln \left(\frac{P_j}{P_i} \right)$ [37].

Mathematical relationship between scaled selection coefficients and dN/dS. Using the theory laid out in the previous subsection, we can calculate an evolutionary rate by summing over all substitution probabilities weighted by the frequency of the originating codon. Further, we can establish specific expressions for nonsynonymous and synonymous evolutionary rates, and then divide them in order to obtain a value for the evolutionary rate ratio dN/dS.

To begin, we can write the nonsynonymous rate K_{N} as

$$K_{\mathrm{N}} = N_e \sum_i \sum_{j \in \mathcal{N}_i} P_i \mu_{ij} u_{ij} \,, \qquad [6]$$

where $\mathcal{N}_i$ is the set of codons that are nonsynonymous to codon i and differ from it by one nucleotide. To normalize K_{N}, we divide it by the number of nonsynonymous sites, which we calculate according to the mutational opportunity definition of a site [10, 12] as

$$L_{\mathrm{N}} = \sum_i \sum_{j \in \mathcal{N}_i} P_i \mu_{ij} \,, \qquad [7]$$

and thus we find that

$$dN = \frac{K_{\mathrm{N}}}{L_{\mathrm{N}}} = \frac{N_e \sum_i \sum_{j \in \mathcal{N}_i} P_i \mu_{ij} u_{ij}}{\sum_i \sum_{j \in \mathcal{N}_i} P_i \mu_{ij}} \,. \qquad [8]$$

Similarly, for dS, the synonymous evolutionary rate K_{S} per synonymous site L_{S}, we have

$$dS = \frac{K_{\mathrm{S}}}{L_{\mathrm{S}}} = \frac{N_e \sum_i \sum_{j \in \mathcal{S}_i} P_i \mu_{ij} u_{ij}}{\sum_i \sum_{j \in \mathcal{S}_i} P_i \mu_{ij}} \,, \qquad [9]$$

where $\mathcal{S}_i$ is the set of codons that are synonymous to codon i and differ from it by one nucleotide substitution. The quantities K_{S} and L_{S} are defined as in Eqs. [6] and [7] but sum over $j \in \mathcal{S}_i$ instead of $j \in \mathcal{N}_i$. Moreover, we note that, if we make the dual assumptions that nucleotide mutation rates are symmetric and that all synonymous codons have equal fitness (i.e. synonymous mutations are neutral), the synonymous fixation rate $u_{ij} = 1/N_e$ [39]. Under this circumstance, the value for dS reduces to 1.

dN/dS accurately reflects selection strength. Using the theoretical framework established in equations [1] - [9], we can examine the relationship between the dN/dS values corresponding to different distributions of scaled selection coefficients. To this end, we generated 200 distinct distributions of amino acid fitness values f_a. We drew these 20 amino acid fitness values from a normal distribution $\mathcal{N}(0, \sigma^2)$, where $\sigma^2 \sim \mathcal{U}(0, 4)$. Here, σ^2 effectively represents the strength of

natural selection; higher σ^2 correspond to larger fitness difference among amino acids, prompting selection to act more strongly against nonsynonymous changes. In other words, high σ^2 values indicate strong purifying selection, while lower values indicate weaker purifying selection. We additionally note that these amino acid fitness quantities correspond to the amino acid propensity parameters estimated by currently available MutSel inference methods [20, 21].

We then converted each distribution of amino acid fitnesses to a corresponding set of codon fitnesses. For 100 of the distributions, we assumed that synonymous changes were neutral, and thus we directly assigned each codon the same fitness $f_i = f_a$. For the other 100 sets of fitnesses, we allowed synonymous codons to have different fitness values. In these circumstances, we randomly selected a preferred codon for each amino acid, and we assigned the preferred codon the fitness of $f_i = f_a + \lambda$ and all non-preferred codons the fitness $f_i = f_a - \lambda$. We drew a unique λ for each fitness distribution from $\mathcal{U}[0, 2]$. We refer to first set of codon selection coefficients as "no codon bias," and the second set as "codon bias."

Finally, using equations **[1]** - **[9]**, we computed dN/dS for each distribution of codon fitnesses. For these calculations, we also need to select mutation rates. We set the mutation rate for transitions as $\mu\kappa$, and the rate for all transversions as μ. We use the value $\mu = 10^{-6}$ for all dN/dS calculations, and we draw a unique value for κ from $\mathcal{U}[1, 6]$ for each set of codon fitnesses.

We found that dN/dS values scale excellently with the variance (σ^2) of the distribution of amino-acid scaled selection coefficients (Figure 1). As expected, dN/dS and σ^2 are strongly negatively correlated; when fitness differences among amino acids are very high, dN/dS takes on lower values, properly reflecting stronger purifying selection (Figure 1). This correlation is predictably much stronger for alignments without codon bias (Figure 1A) than for alignments with codon bias (Figure 1B). The weakened relationship for alignments with codon bias emerges from the fact that fitness differences among synonymous codons will obscure underlying amino acid fitness differences. Even so, the presence of codon bias does not remove the significant negative correlation between dN/dS and selection strength.

Importantly, Figure 1A demonstrates that, in the limiting case when σ^2 approaches 0, and thus all codons have virtually the same fitness, dN/dS converges to 1. More precisely, the largest dN/dS value recovered for alignments without codon bias was 0.997, and this alignment featured a $\sigma^2 = 0.08$. This result properly reflects the case of neutral evolution. In fact, in Appendix 1, we prove that, when synonymous changes are neutral and mutation is symmetric (i.e. $\mu_{xy} = \mu_{yx}$), dN/dS is necessarily always less than or equal to 1. We have also proven that, when synonymous changes are neutral and mutation rates are symmetric, dN/dS as calculated from scaled selection coefficients will always be less than 1. This proof formalizes the MutSel model underlying assumption that selection pressure is constant over the phylogeny and confirms that MutSel models are inherently unable to describe positive, diversifying selection. Although this proof assumes symmetric nucleotide mutation rates, we do not expect that deviations from this assumption will have dramatic effects on dN/dS estimates. Therefore, we conclude that the MutSel framework is an inappropriate model when positive selection is expected, as the model may yield spurious and misleading results.

However, this restriction of $dN/dS < 1$ does not hold when synonymous changes are not neutral, as seen in Figure 1B. In other words, even though the underlying evolutionary model assumes evolutionary equilibrium, i.e. selection pressures remain constant over time and hence there is no positive, diversifying selection, dN/dS can readily be greater than 1. Indeed, it is theoretically possible to achieve arbitrarily high dN/dS values when there are fitness differences among synonymous codons; in the most extreme case of codon bias, where only a single codon per amino acid is selectively tolerated, the number of synonymous sites $L_S = 0$, and thus the value for dN/dS approaches infinity. Given that the MutSel model framework assumes an overarching regime of purifying selection, this finding might seem paradoxical. However, the logical argument that $dN/dS > 1$ represents positive, diversifying selection assumes that the rate of synonymous change may be used as a neutral benchmark, an assumption which codon bias clearly violates. Thus, in theory, what is classically termed positive selection can result simply from strong synonymous fitness differences.

We acknowledge that it is unlikely that this result will have a strong influence in real analyses, as selection on synonymous codons is likely relatively weak in most taxa [40]. For instance, experimental evidence from the yeast Hsp90 protein has shown that any fitness differences among synonymous codons are exceedingly small compared to fitness differences among amino acids. This implementation might not be entirely biologically realistic, as both mutational and selective forces likely contribute to codon bias in real genomes [41, 42, 40, 43, 44]. Even so, the fact that dN/dS can theoretically bear the hallmark of positive selection, even when purifying selection alone is occurring, is an important insight. It is, therefore, possible that estimates of positive selection in species with high levels of codon bias driven in part by selection, such as bacterial, *Drosophila*, or certain mammalian species [42, 45, 44], may not be true cases of positive selection, but rather signals of strong codon bias.

dN/dS calculated from ssc's as a model benchmark. The relationship we have established between dN/dS and scaled selection coefficients offers a unique opportunity to assess the robustness of dN/dS inference methods. It is conventional practice in model development to benchmark models against data simulated according to the model itself. While this strategy is crucial for testing whether correct model implementation, it is inherently incapable of discerning limitations and properties of the inference framework under more general conditions. Moreover, it cannot confirm that the underlying model accurately represents the evolutionary process. Therefore, we apply a novel benchmarking approach which uses the theoretical relationship among modeling frameworks to assess the accuracy and specific utility of those models. This approach, outlined in Figure 2A, entails comparing dN/dS values calculated from selection coefficients to those inferred by an ω-based model, in order to benchmark the model's accuracy. NOTE: this approach is not possible without our relationship, because we simltaneously know the true underlying value without having data that conforms exactly to the model. Confounding effects are therefore removed.

Using the selection coefficients and mutation rates used in the previous subsection, we simulated alignments using standard methods [12] according the Halpern-Bruno MutSel model [15]. We then inferred dN/dS for each alignment using the M0 model [10, 2], as implemented in the HyPhy batch language [32]. The M0 model uses the GY94 instantaneous rate matrix, which includes parameters for transition/transversion bias (κ), equilibrium codon frequencies (π), and finally the dN/dS rate ratio (ω) (see equation **[13]**). Throughout the remaining text, we refer to dN/dS as inferred by M0 as ω, and to dN/dS computed using equations **[1]** - **[9]** simply as dN/dS.

We found that dN/dS values agree nearly perfectly with ω values (Figure 2B). This agreement was neither influenced by the presence of codon bias, nor by nucleotide compostitional bias; indeed, simulated alignments featured a wide range (0.21-0.89) of GC contents. Additionally, in Figure 2C, we demonstrate that ω converges to the true dN/dS value as the size of the data set, represented by simulated alignment length, increases. These results unequivocally show that the dN/dS quantity is fully contained within MutSel model parameters, and importantly that ω-based model inference methods behave exactly as expected, yielding precise dN/dS estimates. This finding has important implications for modeling choices; although the MutSel framework might model the sequence evolution in a way that more mechanistically matches the evolutionary process, ω-based models do not dramatically suffer from any modeling limitations.

dN/dS inference with realistic data. Having confirmed that ω-based and MutSel models **agree** under general conditions, we sought to test the accuracy of ω-based models using more realistic data. To this end, we made use of realistic amino acid fitness and nucleotide mutation rate parameters to construct a series of experimental evolutionary models. In particular, we used influenza nucleoprotein (NP) site-specific amino acid preference values, given by ref. [35], which consisted of experimentally-determined fitness values for each individual amino acid across all sites in NP, yielding 498 distinct amino acid propensity distributions. We combined these experimental fitness parameters with three sets of experimentally determined mutation rates, either for NP [35], yeast [46], or polio virus [47]. While each of these mutation matrices is asymmetric, they feature differing degrees of asymmetry, with NP mutation rates being the most symmetric and polio mutation rates the most asymmetric. More precisely, in the absence of amino-acid level selection, the GC contents that the NP, yeast, and polio mutation rates would generate are 0.518, 0.336, and 0.192, respectively. Finally, we built a unique experimentally-informed evolutionary model for all combinations of amino acid fitness distributions and mutation rates using the approach outlined in refs. [35, 36]. We calculated stationary codon frequencies for each experimental model, and used these values in combination with their corresponding mutation rates to calculate dN/dS, and simulate alignments from which ω may be inferred.

In order to properly accomodate the underlying mutational asymmetry in these alignments, a proper M0 parameterization must be chosen. Two broad frameworks have been proposed to account for such nucleotide biases; first, GY-style models employ target codon frequencies π_i [10], whereas MG-style models employ target nucleotide frequencies π_n [11]. For example, the instantaneous rate matrix element giving the substitution probability from codon AAA to AAG would contain the target codon frequency π_{AAG} under the former framework, and the target nucleotide frequency π_{G} under the latter framework. Previous works have shown that MG-style and GY-style models can yield different ω estimates [13, 48] and thus we inferred ω according to a variety of frequency parameterizations. Frequency parameterizations for GY-style models included the frequency estimators F61 [10], F3x4 [10], CF3x4 [49], and finally F1x4 [11]. For MG-style models, we considered both a parameterization in which four global nucleotide frequency parameters were used [11], and a parameterization which employed twelve nucleotide frequency parameters to allow for different frequencies at each codon position (MG3) [13]. We term the former framework MG1, and the latter MG3.

Figure 3 shows the resulting relationships between dN/dS and ω MLEs for each set of mutation rates (NP, yeast and polio), across M0 model frequency parameterizations (full regression plots are shown in Figure S1). Figure 3A displays the estimator bias, or the systematic discrepency between the true dN/dS value and the ω MLEs, and Figure 3B displays r^2 values between dN/dS and ω. Note that the precise bias and r^2 values are shown in Tables S1 and S2, respectively.

Two distinct trends emerge from Figure 3. First, increasing asymmetry in the mutational process induces substantial and statistically significant bias in M0 ω estimates. Most often, the M0 model underestimates ω relative to the true dN/dS value. Indeed, for all frequency parameterizations, ω estimates are most accurate under NP mutation rates, and both accuracy and precision tend to decrease as asymmetry progresses from yeast to polio mutation rates. Second, frequency parameterizations which more closely match the mechanistic process that generated the data (MG1, MG3) strongly outperform all other frequency estimators. **how to explain f1x4 vs f3x4?**. Moreover, it is clear that the MG1 parameterization yields the overall best performance of all frequency estimators considered, and in particular it features by far the least amount of bias for the highly asymmetric polio mutation rates.

Strikingly, when we examined model fit, using AIC scores [50], for the different frequency parameterization, we found that the F61 parameterization was unequivocally the best performing model, on average, for all datasets (Table 1). This result dramatically juxtaposes the substantial inaccuracy and imprecision that F61 typically yields. Indeed, as Figure 3 shows, F61 has the most estimator bias for NP datasets as well as the least precision for both NP and polio datasets. This result shows reveals that selecting models based strictly on model fit can be counterproductive, as model fit is clearly at odds with model performance. Previous studies which have examined how the choice of frequency estimators influenced model fit have found that F61 offers improvements over F3x4 [10], and that MG3 was largely comparable to F3x4 [13]. However, the unique approach of our study, in which the alignments are simulated according to the distinct MutSel model, reveals that these differences in model fit are not informative, but rather misleading.

Conclusions

We will conclude with insights gained from our study and recommendations for using ω-based and MutSel modeling frameworks going forward. We have shown that dN/dS be accurately calculated from selection coefficients, revealing that ω-based and MutSel models yield consistent and overlapping information about the strength natural selection. Importantly, our proof that $dN/dS \leq 1$, when calculated from selection coefficients and when synonymous mutations are neutral, indicates that the use of MutSel models is only justified under conditions of strictly purifying selection, or neutral evolution. This restriction is in part indicated by the basic MutSel model assumption of constant selection pressures over time, or in other words a static fitness landscape [15, 28, **?**, 16].

Thus, if the aim is to identify positive selection, only ω-based models, of the two frameworks examined here, are justified. However, there are still some theoretical issues with the approach that these models take, namely in their use of equilibrium codon frequency parameters. We expect that, if positive selection is occurring, there is necessarily a shift in

the underlying fitness landscape causing different amino acids to become preferred.

ω-based models are an apt model choice for examining positive, diversifying selection. However, if one desires site-specific point estimates of dN/dS, ω-based models

2. if and when you use omega models, you absolutely must parameterize them properly, otherwise dnds is a meaningless quantity. this seems difficult to do. there is an internal tension in omega models, wrt to equilibrium codon frequencies. If codon frequencies are assumed to be at equilibrium, how can we ever properly account for adaptive processes, such as positive selection? moreover, if we screw up any model parameter, dnds may be wrong. suggests that point estimates for dnds from these models are not ideal, see that one plos response. precise point estimates can, however, be calculated from a mutsel model, provided purifying selection. but if detecting positive selection is your goal, seriously do NOT use mutsel models. future work should investigate modeling frameworks which fully account for nonequilibrium evolutionary processes.

We additionally emphasize that improper model parameterizations lead to spurious ω MLEs which do not accurately represent dN/dS. If other model parameters (κ and equilibrium codon frequencies) are specified incorrectly, or inadvertently contain information about amino-acid level natural selection, the resulting ω MLE will not represent the true dN/dS evolutionary rate ratio. Only by ensuring that ω is the only model parameter which contains information about natural selection will it assuredly represent dN/dS.

Taken together, these results strongly suggest that the MC model's codon frequency parameters are ill-suited to accomodate compositional biases which result from forces other than amino-acid level selection. We therefore suggest that future work investigate the utility of novel parameters for MC models which better account for asymmetry in the mutational process.

In sum, we have garnered several important insights into the behavior of MC and MutSel models, as well as the dN/dS metric. These results were only made possible through establishing a formal mathematical relationship between distinct modeling frameworks. We believe that the approach presented in this paper represents a promising future avenue for methodological benchmarking. Typically, researchers assess the performance of a given inference framework through simulations which adhere to the underlying model's assumptions (with a notable exception of ref. [51]). While this strategy is critical for testing whether a model implementation behaves as expected, it is innately incapable of assessing the limitations and properties of the inference framework under more general conditions, and it cannot confirm that the underlying model accurately represents the evolutionary process. Therefore, we suggest an alternate approach to benchmark inference methods: assessing the extent to which distinct models agree may serve as a novel, robust strategy to determine the accuracy and specific utility of different modeling frameworks. As we have shown here, this approach has great potential to reveal previously unrecognized model properties or biases and will help ensure robust model development going forward.

We additionally emphasize that improper model parameterizations lead to spurious ω MLEs which do not accurately represent dN/dS. If other model parameters (κ and equilibrium codon frequencies) are specified incorrectly, or inadvertently contain information about amino-acid level natural selection, the resulting ω MLE will not represent the true dN/dS evolutionary rate ratio. Only by ensuring that ω is the only model parameter which contains information about natural selection will it assuredly represent dN/dS.

Methods

Simulation of scaled selection coefficients We first examined the relationship between dN/dS and scaled selection coefficients by simulating 200 distributions of amino acid scaled fitness values, $F_a = 2Nf_a$, from a normal distribution $\mathcal{N}(0, \sigma^2)$, where a unique σ^2 was drawn from $\mathcal{U}(0, 4)$ for each fitness distribution. We converted these amino acid fitnesses to codon fitnesses, F_i. For 100 of the fitness distributions, we directly assigned all codons within a given amino acid family the fitness f_a, giving all synonymous codons the same fitness. For the other 100 fitness distributions, we assigned synonymous codons different fitnesses by randomly selected a preferred codon for each amino acid. This preferred codon was assigned the fitness of $F_i = F_a + \lambda$, and all non-preferred codons were given the fitness $F_i = F_a - \lambda$. We drew a unique λ for each fitness distribution from $\mathcal{U}[0, 2]$. We then computed stationary codon frequencies as

$$p_i = \frac{e^{F_i}}{\sum e^{F_k}}, \qquad [10]$$

where the sum in the denominator runs over all 61 sense codons [37]. Equation **[10]** gives the analytically precise stationary frequencies for a MutSel model, under the assumption of symmetric nucleotide mutation rates, i.e. where $\mu_{xy} = \mu_{yx}$ [37]. For each resulting set of stationary codon frequencies, we used equations **[6]** - **[9]** to compute a dN/dS value. For these calculations, we set the mutation rate for transitions as $\mu\kappa$, and the rate for all transversions as μ. We used the value $\mu = 10^{-6}$ for all dN/dS calculations, and we drew a unique value for κ from $\mathcal{U}[1, 6]$ for each set of codon frequencies.

Alignment simulations. We simulated protein-codong sequences as a continuous-time Markov process using standard methods [12] according to the Halpern-Bruno MutSel model [15]. In simplified form, this model's instantaneous rate matrix Q is given by

$$Q_{ji} = \begin{cases} \mu_{ij} \frac{S_{ij}}{1 - 1/S_{ij}} & \text{single nucleotide change} \\ 0 & \text{multiple nucleotide changes} \end{cases}, \qquad [11]$$

for a mutation from codon i to j, where is the mutation rate, p_i is the stationary frequency for codon i, and the scaled selection coefficient S_{ij} is defined in equation **[5]**. All alignments presented here were simulated along a 4-taxon phylogeny, beginning with a root sequence selected from stationary codon frequencies. Unless otherwise stated, all simulated alignments contained 500,000 codon positions. A single evolutionary model was applied to all positions in the simulated sequences. While this lack of site-wise heterogeneity is unrealistic for real sequence evolution, it allowed us to verify our derived relationship between scaled selection coefficients and dN/dS with a sufficiently sized data set.

Computation of stationary frequencies for experimental data sets. We used experimentally-determined site-specific amino acid fitness parameters (F_a) for influenza nucleoprotein (NP), from Bloom 2014 [35], in combination with experimental nucleotide mutation rates for either NP [35], yeast [46], or polio virus [47] to derive realistic distributions of stationary codon frequencies. Bloom 2014 reported 498 distinct site-wise amino acid preference distributions for NP [35]. We combined these 498 amino acid preference sets with each of the three mutation rate matrices sets to construct a total of $498 \times 3 = 1494$ unique experimental evolutionary Markov models, using the approach in refs. [35, 36]. The instantaneous matrix for these

experimental models is given by

$$Q_{ji} = \begin{cases} \frac{F_j}{F_i}\mu_{ij} & \text{single nucleotide change, where } F_j \geq F_i \\ \mu_{ij} & \text{single nucleotide change, where } F_j < F_i \\ 0 & \text{multiple nucleotide changes} \end{cases}, \quad [12]$$

where F_i is the fitness of codon i [35, 36]. We calculated F_i values by simply assigning a given amino acid's experimental fitness F_a to each of its consistituent codons; thus, all synonymous changes are neutral. We determined the stationary codon frequencies for each resulting experimental model from the matrix's eigenvector corresponding to the eigenvalue 0. Finally, we simulated alignments for each set of stationary frequencies and corresponding mutation rates according to equation [**11**].

Maximum likelihood inference of dN/dS. We inferred ω for all simulated alignments using the M0 model For the 200 alignments simulated with symmetric mutation rates, we inferred dN/dS using the M0 model [2], as implemented in the HyPhy batch language [32]. The M0 model uses the GY94 instantaneous rate matrix,

$$Q_{ji} = \begin{cases} \pi_j & \text{synonymous transversion} \\ \kappa\pi_j & \text{synonymous transition} \\ \omega\pi_j & \text{nonsynonymous transversion} \\ \omega\kappa\pi_j & \text{nonsynonymous transition} \\ 0 & \text{multiple nucleotide changes} \end{cases}, \quad [13]$$

where κ is the transition-transversion bias, π_j is the equilibrium frequency of the target codon j, and ω represents dN/dS [10, 1]. Importantly, this model's π parameters are intended to represent those codon frequencies which would exist in absence of selection pressure, but those which would result from mutation alone [10, 11, 9, 12]. **need transition here** Therefore, as all codons are equally probable when nucleotide mutation is symmetric, we assigned values of 1/61 to all codon frequencies π for these inferences.

REWRITE THIS PARAGRAPH Alternatively, when inferring ω for alignments simulated with experimental fitness and mutation rates, we used 5 different sets of equilibrium frequency parameterizations. we used the common frequency estimators F61 [10], F3x4 [11], and CF3x4 [**?**]. As typical analyses consider model frequency parameters as protein-wide (not site-specific) parameters, we computed these parameter values by pooling, for each set of mutation rates, all 498 steady-state codon frequencies to derive average codon frequencies. This approach yielded a set of global equilibrium frequencies for each set of mutation rates, and we calculated the F61, F3x4, and CF3x4 frequencies from these distributions.

Appendix 1

Here, we prove that $dN/dS \leq 1$ when calculated from scaled selection coefficients. We assume that nucleotide mutation rates are symmetric ($\mu_{xy} = \mu_{yx}$) and that synonymous codons have the same fitness (synonymous changes are neutral). As described in the main text of this paper, these assumptions yield $dS = 1$, and hence we have to show that $dN = K_{\text{N}}/L_{\text{N}} \leq 1$. To this end, we note that the summation series defining dN can be rearranged such that substitution probability from codon i to j is always added to the substitution probability from codon j to i. It can be shown, for each of these pairs, that $dN \leq 1$, and hence $dN/dS \leq 1$.

For this proof, we consider the pair of nonsynonymous codons i and j, where $P_i \leq P_j$ that $P_j > 0$ (P_i represents the stationary frequency of codon i). Moreover, as we assume $\mu_{ij} = \mu_{ji}$, we will simply write μ for each of these quantities throughout the following. As follows from equation [**6**], the sum of the probability weights of evolving from codon i to j and from codon j to i is

$$N_e\mu\big(P_i u_{ij} + P_j u_{ji}\big) = \frac{2P_iP_j(\log(P_i) - \log(P_j))}{P_i - P_j}. \quad [14]$$

This quantity represents the K_{N} (numerator) calculation for dN. To prove $dN \leq 1$, we must show that this quantity is less than or equal to $P_i + P_j$, which represents the L_{N} (denominator) in the dN calculation. To this end, we introduce the function

$$F(x,y) = x + y - \frac{2xy[\log(x) - \log(y)]}{x - y}, \quad [15]$$

and we will now show that $F(x,y) \geq 0$ for $x \leq y$ and $y \geq 0$. It is straightforward to show this for $x = y$. For $x < y$, we show that the first derivative of equation [**15**] is negative throughout $x \in (0,y)$, which proves that the function monotonically decreases, and thus $F(x,y) > 0$, in this interval. We calculate the first derivative as

$$\frac{\partial F(x,y)}{\partial x} = \frac{\big[(x-3y)(x-y) - 2y^2(\log x - \log y)\big]}{(x-y)^2}. \quad [16]$$

We now replace the expression $\log x - \log y$ by its Taylor expansion, yielding

$$\frac{\partial F(x,y)}{\partial x} = \frac{\left[(x-3y)(x-y) - 2y^2\left(\sum_{n=1}^{\infty} \frac{1}{n}(1-x/y)^n\right)\right]}{(x-y)^2}. \quad [17]$$

We further note that the first two terms of the Taylor series equal $(x-3y)(x-y)$, and thus expression [**17**] simplifies to

$$\frac{\partial F(x,y)}{\partial x} = \frac{-2y^2\sum_{n=3}^{\infty}\left(1-\frac{x}{y}\right)^n}{(x-y)^2}, \quad [18]$$

which is clearly negative.

Appendix 2

We consider mutation as a nucleotide-level, and not a codon-level process. We can generally describe the mutation process by a 4×4 symmetric matrix of mutation rates $\mu_x y$ times a vector of equilibrium nucleotide frequencies π_x, such that the probability of a mutation from nucleotide x to y is $\mu_{xy}\pi_y$. Thus, we desire a framework to incoroporate this mutational information into a codon model of sequence evolution.

To derive the Fnuc model parameterization, we define n_{ij} as the target nucleotide in a substitution from codon i to codon j, and π_x as the equilibrium frequency of nucleotide x. We begin with a modified version of the MG94 [11] instantaneous rate matrix

$$Q_{ji} = \begin{cases} \pi_{n_{ij}} & \text{synonymous transversion} \\ \kappa\pi_{n_{ij}} & \text{synonymous transition} \\ \omega\pi_{n_{ij}} & \text{nonsynonymous transversion} \\ \omega\kappa\pi_{n_{ij}} & \text{nonsynonymous transition} \\ 0 & \text{multiple nucleotide changes} \end{cases}. \quad [19]$$

The primary difference between equation [**19**] and the original MG94 rate matrix is that we consider a single parameter, ω for the nonsynonymous/synonymous rate ratio, whereas the original formulation used parameters β and α to describe separately nonsynonymous and synonymous rates, respectively.

As Muse and Gaut [11] noted, the stationary frequency for a given codon i is

$$P_i = \frac{\pi_{i_1}\pi_{i_2}\pi_{i_3}}{1-\Pi_{\text{stop}}}\,, \qquad [20]$$

where where i_1, i_2, and i_3 are the three nucleotides of codon i, and $\Pi_{\text{stop}} = \pi_{\text{T}}\pi_{\text{A}}\pi_{\text{G}} + \pi_{\text{T}}\pi_{\text{G}}\pi_{\text{A}} + \pi_{\text{T}}\pi_{\text{A}}\pi_{\text{A}}$. The expression in equation [**20**] is better known as the frequency estimator F1x4. We emphasize that F1x4 considers global, not positional, nucleotide frequencies; in other words, 4 parameters, not 12, describe nucleotide equilibrium frequency values. Typically, these F1x4 stationary codon frequencies are subsequently used in the rate matrix, rather than the target nucleotide frequencies. Instead, we propose a simple model reformulation, which we term Fnuc, to actually incorporate the target nucleotide frequencies, as was originally intended. These target nucleotide frequencies are given by

$$\pi_x = \frac{P_i(1-\Pi_{\text{stop}})}{\pi_y\pi_z}, \qquad [21]$$

where P_i is the F1x4 frequency for codon i and is comprised of the nucleotides x, y and z. It is easy to show that, if applied to all 61 codons, 4 unique nucleotide frequency values will result, which can be readily incorporated into the matrix given in equation [**19**] to reveal the Fnuc parameterization.

ACKNOWLEDGMENTS. This work was supported by the army and by NIH. Computational resources were provided by CCBB.

1. Nielsen R, Yang Z (1998) Likelihood models for detecting positive selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929–936.
2. Yang ZH, Nielsen R, Goldman N, Pedersen AMK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449.
3. Kosakovsky Pond S, Frost S (2005) Not so different after all: A comparison of methods for detecting amino acid sites under selection. Mol Biol Evol 22:1208–1222.
4. Huelsenbeck JP, Jain S, Frost SWD, Kosakovsky Pond SL (2006) A Dirichlet process model for detecting positive selection in protein-coding DNA sequences. Proc Natl Acad Sci USA 103:6263–6268.
5. Li WH, Wu CI, Luo CC (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution consider the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174.
6. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426.
7. Pamilo P, Bianchi NO (1993) Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. Mol Biol Evol 10:271–281.
8. Ina Y (1995) New methods for estimating the numbers of synonymous and nonsynonymous substitutions. Journal of Molecular Evolution 40:190–226.
9. Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17:32–42.
10. Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725–736.
11. Muse SV, Gaut BS (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol Biol Evol 11:715–724.
12. Yang Z (2006) Computational Molecular Evolution (Oxford University Press).
13. Kosakovsky Pond S, Muse S (2005) Site-to-site variation of synonymous substitution rates. Mol Biol Evol 22:2375–2385.
14. Anisimova M, Kosiol C (2009) Investigating protein-coding sequence evolution with probabilistic codon substitution models. Mol Biol Evol 26:255–271.
15. Halpern AL, Bruno WJ (1998) Evolutionary distances for protein-coding sequences: modeling site-specific residue frequencies. Mol Biol Evol 15:910–917.
16. Thorne JL, Lartillot N, Rodrigue N, Choi SC (2012) Codon models as vehicles for reconciling population genetics with inter-specific data. In Cannarozzi G, Schneider A, eds., Codon evolution: mechanisms and models (Oxford University Press, New York).
17. Yang Z, Nielsen R (2008) Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. Mol Biol Evol 25:568–579.
18. Rodrigue N, Philippe H, Lartillot N (2010) Mutation-selection models of coding sequence evolution with site-heterogeneous amino acid fitness profiles. Proc Natl Acad Sci USA 107:4629–4634.
19. Tamuri AU, dos Reis M, Goldstein RA (2012) Estimating the distribution of selection coefficients from phylogenetic data using sitewise mutation-selection models. Genetics 190:1101–1115.
20. Rodrigue N, Lartillot N (2014) Site-heterogeneous mutation-selection models within the PhyloBayes-MPI package. Bioinformatics :1020–1021.
21. Tamuri AU, Goldman N, dos Reis M (2014) A penalized-likelihood method to estimate the distribution of selection coefficients from phylogenetic data. Genetics 197:257–271.
22. Kosakovsky Pond S, et al. (2011) A random effects branch-site model for detecting episodic diversifying selection. Mol Biol Evol 28:3033–3043.
23. Murrell B, et al. (2012) Detecting individual sites subject to episodic diversifying selection. PloS Genet 8:e1002764.
24. Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19:908–917.
25. Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol 22:2472–2479.
26. Kosakovsky Pond S, Frost S (2005) A genetic algorithm approach to detecting lineage-specific variation in selection pressure. Mol Biol Evol 22:478–485.
27. Robinson DM, Jones DT, Kishino H, Goldman N, Thorne JL (2003) Protein evolution with dependence among codons due to tertiary structure. Mol Biol Evol 20:1692–1704.
28. Thorne J, Choi S, Yu J, Higgs P, Kishino H (2007) Population genetics without intraspecific data. Mol Biol Evol 24:1667–1677.
29. Rodrigue N, Kleinman C, Phillipe H, Lartillot N (2000) Computational methods for evaluating phylogenetic models of codong sequence evolution with dependence between codons. Mol Biol Evol 26:1663–1676.
30. Scherrer MP, Meyer AG, Wilke CO (2012) Modeling coding-sequence evolution within the context of residue solvent accessibility. BMC Evol Biol 12:179.
31. Meyer AG, Wilke CO (2012) Integrating sequence variation and protein structure to identify sites under selection. Mol Biol Evol 30:36–44.
32. Kosakovsky Pond SL, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenetics. Bioinformatics 21:676–679.
33. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24:1586–1591.
34. Delport W, Poon A, Frost S, Pond S (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26:2455–2457.
35. Bloom JD (2014) An experimentally determined evolutionary model dramatically improves phylogenetic fit. Mol Biol Evol :To appear.
36. Bloom JD (2014) An experimentally informed evolutionary model improves phylogenetic fit to divergent lactamase homologs. Mol Biol Evol 31:1956–1978.
37. Sella G, Hirsh AE (2005) The application of statistical physics to evolutionary biology. Proc Natl Acad Sci USA 102:9541–9546.
38. Kimura M (1962) On the probability of fixation of mutant genes in a population. Genetics 4:713–719.
39. Crow JF, Kimura M (1970) An Introduction to Population Genetics Theory (Burgess Pub. Co., California).
40. Hershberg R, Petrov D (2008) Selection on codon bias. Annu Rev Genet 42.
41. Blumer M (1991) The selection-mutation-drift theory of synonymous codon usage. Genetics 129:897–907.
42. Duret L (2002) Evolution of synonymous codon usage in metazoans. Curr Opin Genet Dev 12:640–649.
43. Chen SL, Lee W, Hottes AK, Shapiro L, McAdams HH (2009) Codon usage between genomes is constrained by genome-wide mutational processes. Proc Natl Acad Sci USA 101:3480–3485.
44. Plotkin JB, Kudla G (2011) Synonymous but not the same: the causes and consequences of codon bias. Nature Rev Genet 12:32–42.
45. Chamary JV, Parmley JL, Hurst LD (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. Nature Rev Genet 7:98–108.
46. Zhu YO, Siegal ML, Hall DW, Petrov DA (2014) Precise estimates of mutation rate and spectrum in yeast. Proc Natl Acad Sci USA :FORTHCOMING.
47. Acevedo A, Brodsky L, Andino R (2014) Mutational and fitness landscapes of an RNA virus revealed through population sequencing. Nature 505:686 – 690.
48. Yap LHES V B, Huttley G (2010) Estimates of the effect of natural selection on protein-coding content. Mol Biol Evol 27:726 – 734.
49. Kosakovsky Pond S, Delport W, Muse S, Scheffler K (2010) Correcting the bias of empirical frequency parameter estimators in codon models. PLoS One 5:e11230.
50. Akaike H (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control 19:6:716–723.
51. Holder M, Zwickl D, Dessimoz C (2008) Evaluating the robustness of phylogenetic methods to among-site variability in substitution processes. Phil Trans R Soc B 363:4013–4021.

Figures and Tables

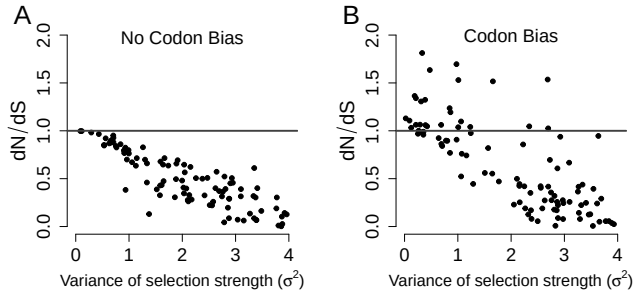

Fig. 1. dN/dS decreases in proportion to amino-acid level selection strength. dN/dS is plotted against the σ^2 of the simulated distribution of amino-acid scaled selection coefficients. Higher values of σ^2 indicate larger fitness differences among amino acids, whereas the limiting value of $\sigma^2 = 0$ means that all amino acids have the same fitness. (A) Synonymous codons have equal fitness values ($r^2 = 0.83$). (B) Synonymous codons have different fitness values ($r^2 = 0.45$). Importantly, (B), but not (A) shows dN/dS values greater than 1, in spite of the steady-state evolutionary process.

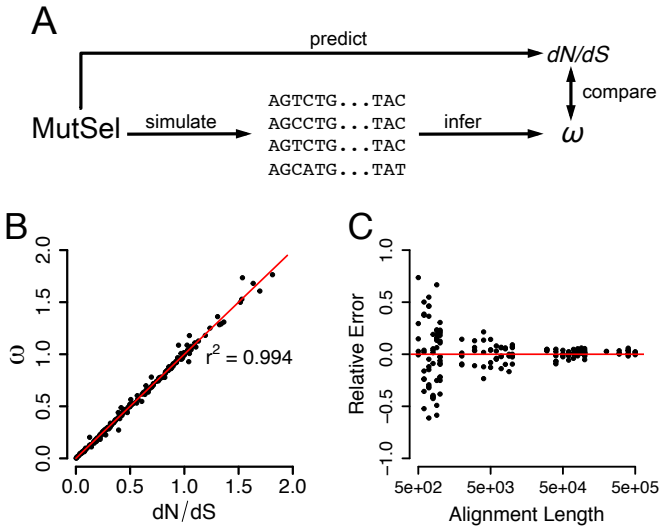

Fig. 2. Regressions between dN/dS values as calculated from scaled selection coefficients and as inferred using the M0 mechanistic codon model. Each point corresponds to a single simulated alignment. All ω values shown here were inferred by parameterizing the M0 model with κ fixed to its true, simulated value as well as the Fequal codon frequency specification [12]. The red line in panels (A-B) is the $x = y$ line. (A) Synonymous codons have equal fitness ($r^2 = 0.997$). (B) Synonymous codons have different fitness values ($r^2 = 0.992$). (C) Convergence of ω MLEs to the true dN/dS value. The y-axis indicates the relative error of the maximum likelihood dN/dS estimate, and the x-axis indicates the number of positions in the simulated alignment. As the number of positions, and hence the size of the data set, increases, the maximum likelihood estimates converge to the dN/dS values calculated using equations [**??**]-[**9**]. The red line in panel (C) is the line $y = 0$, indicating no error.

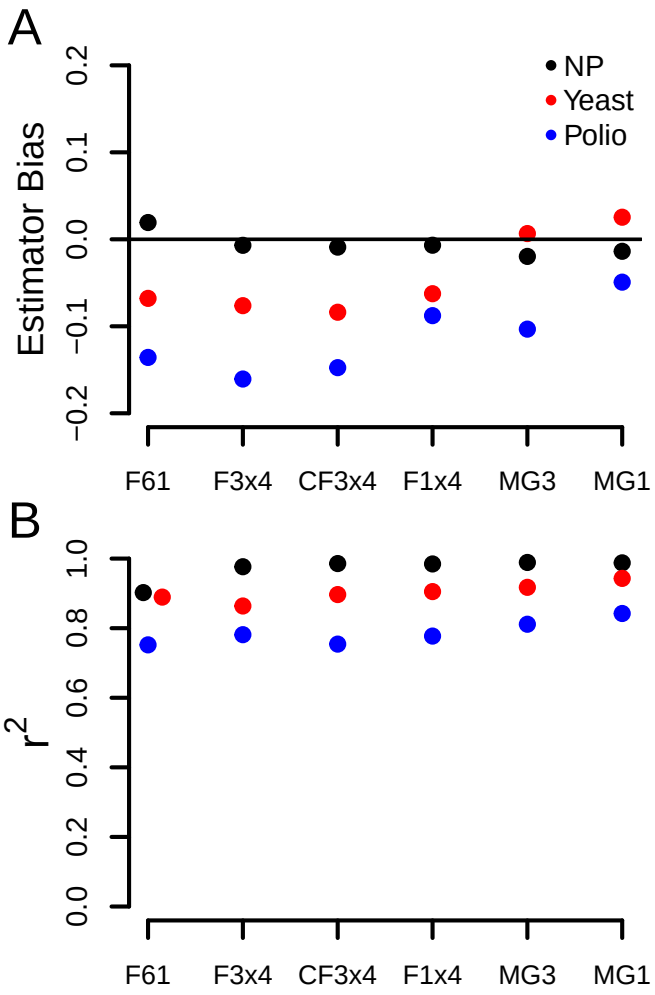

Fig. 3. (A) Estimator bias and (B) r^2 values between dN/dS and ω MLEs across M0 codon frequency parameterizations, for each set of nucleotide mutation rates. Note that negative biases indicate ω values that are, on average, lower than dN/dS. All bias and r^2 values are highly statistically significant, with all $p < 10^{-12}$. In this figure, we see that ω-based models tend to systematically underestimate dN/dS, across all codon frequency parameterizations. Generally, F_equal features the least amount of bias, and has very high r^2 values for both NP and yeast mutation rates. Although Fequal yields lower r^2 values for polio mutation rates than do F61, F3x4, and CF3x4, the latter three estimators also have relatively high biases, demonstrating that they systematically underestimate dN/dS. That Ftrue, which assigns codon frequencies to those which would exist in the absence of amino-acid level selection, also underestimates dN/dS implies that codon frequency parameters are ill-suited to accomodate mutation-induced nucleotide compositional bias.

Table 1.

Table 1. Mean △AIC for datasets simulated with NP, Yeast, or Polio mutation rates.

Frequencies	NP	Yeast	Polio
F61	0	0	0
CF3x4	-9627.5	-7951.8	-7975.9
Fnuc1	-13325.5	-10042.0	-5147.6
F1x4	-13524.5	-13658.5	-15468.3
Fnuc3	-14401.3	-12851.6	-8624.9
F3x4	-14807.2	-17385.3	-19384.6

Note that all models have 3 free parameters. All values in table are the mean dAIC with F61 as the reference, as in best model fit. The ranking shown in table is the order for NP, but note that yeast and polio is different. Even so, F61 provides the best model fit for all mutation rates, although it certainly does not infer the most accurate ω values.

Supplementary Information

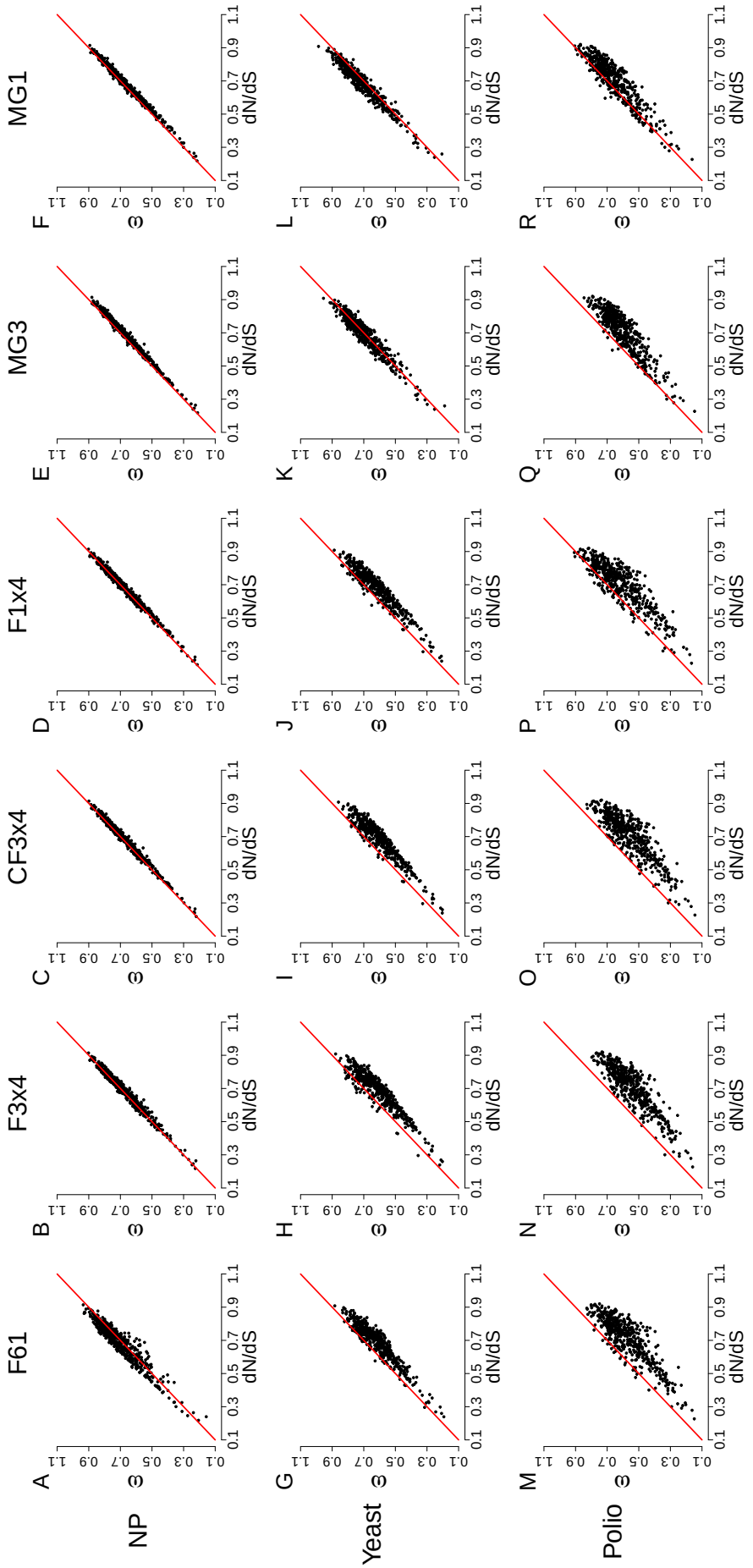

Figure S1. omega vs dn/ds for simulations using experimental parameters.

Table S1. Estimator bias between ω MLEs and the expected, true dN/dS values, for all mutation rates and M0 codon frequency parameterizations examined. Negative bias values indicate that ω MLEs are, on average, lower than dN/dS. All biases are statistically significant, with all $P < 2 \times 10^{-16}$ except for the estimator bias associated with Yeast mutation rates for Fnuc3, where $P = 5.4 \times 10^{-5}$.

Mutation rate	Fnuc1	F1x4	Fnuc3	CF3x4	F3x4	F61
NP	-0.014	-0.02	-0.007	-0.009	-0.007	0.019
Yeast	0.025	0.007	-0.063	-0.084	-0.076	-0.068
Polio	-0.049	-0.103	-0.088	-0.148	-0.161	-0.136

Table S2. r^2 between ω MLEs and the expected, true dN/dS values, for all mutation rates and M0 codon frequency parameterizations examined. All values shown are statistically significant, with all $P < 2 \times 10^{-16}$.

Mutation rate	Fnuc1	F1x4	Fnuc3	CF3x4	F3x4	F61
NP	0.988	0.989	0.985	0.986	0.977	0.902
Yeast	0.943	0.917	0.905	0.897	0.864	0.889
Polio	0.842	0.811	0.777	0.754	0.781	0.752