**Abstract**

The evolutionary rate ratio $dN/dS$, representing the ratio nonsynonymous to synonymous substitution rates, is widely used to infer the strength of natural selection in protein-coding sequences, in a phylogenetic framework. Recently, mutation-selection-balance (MutSel) models have become a popular alternative to the $dN/dS$ framework. MutSel models estimate scaled selection coefficients, indicating the selective response to particular codon and/or amino-acid changes. However, it remains unknown how these two modeling frameworks relate to one another. Do $dN/dS$ estimates reveal similar or distinct information from scaled selection coefficients? To answer this question, we derive a formal mathematical relationship to calculate $dN/dS$ from scaled selection coefficients. Using this framework, we prove that MutSel models strictly produce $dN/dS \leq 1$, revealing that MutSel models inherently cannot accomodate positive, diversifying selection. However, we also find that $dN/dS$ can be greater than 1 if selection acts on synonymous changes, even though only purifying selection is occuring. In addition, we use this established relationship to investigate the behavior of maximum likelihood (ML) $dN/dS$ inference methods. In particular, we examine the extent to which ML $dN/dS$ estimates agree with $dN/dS$ as computed using selection coefficients. This approach serves as a robust and novel strategy for assessing the accuracy and utility of these modeling frameworks. We find that ML methods yield 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, based on AIC scores, is not the model with the least bias and highest precision for parameter estimates of interest, and thus selecting models based solely on fit can be counterproductive and positively misleading.

# 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 non-synonymous 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 purifying 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 continuous time Markov-process model of sequence evolution and maximum likelihood estimates (MLEs) of the parameter $\omega$, 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 $\omega$-based 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 $\omega$-based models. While $\omega$-based 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. The MutSel framework estimate site-specific 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.

$\omega$-based 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 epistatic interactions [27, 28, 29, 30, 31]. This flexibility, along with accessible software implementations [32, 33, 34], make $\omega$-based 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 [15, 18, 19, 16]. Recent phylogenetic studies have also demonstrated that evolutionary models which account for amino acid fitness values dramatically outperform $\omega$-based models, suggesting that MutSel models may more aptly represent the evolutionary process [35, 36].

Although both MutSel and $\omega$-based 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 these competing modeling frameworks will more precisely reveal under which circumstances the use of these models is justified, and will additionally reveal these models' previously unrecognized behaviors and limitations.

Here, we formalize the relationship between these two modeling frameworks 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 $dN/dS$ accurately captures the selective pressures indicated by a given distribution of scaled selection coefficients. 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.

Finally, this robust relationship provides a uniquely rigorous platform to examine the performance of $\omega$-based models. 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. [54]). 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 will serve as a novel, robust strategy to determine the accuracy and specific utility of different modeling frameworks.

Here, we adopt such a strategy to assess the inference accuracy in $\omega$-based modeling frameworks. We find that, in the absence of mutation-induced nucleotide compositional bias, $dN/dS$ values inferred in an ML framework agree precisely with those calculated from scaled selection coefficients. However, as mutational bias increases, $dN/dS$ ML inferences become increasingly biased away from their true values, even under a variety of ML model parameterizations. We also find that the best-performing ML model parameterizations are not those which exhibit the best fit to the data (measured by AIC), ultimately revealing that relying on model fit as a litmus-test for model performance is an ineffective and misleading strategy.

# Results and Discussion

## Theoretical model.

We model sequence evolution using the Halpern-Bruno MutSel modeling framework under the assumptions of a fixed effective population size $N_e$ and constant selection pressure over time [15, 17, 19, 16]. This continous-time Markov process process is governed by the $61 \times 61$ 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. To begin, let $f_i$ be the fitness of codon $i$, and let the selection coefficient acting on a mutation from codon $i$ to codon $j$ be $s_{ij} = f_j - f_i$ [37, 17]. The fixation probability for this mutation is

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

[38, 15, 17].

We further define $S_{ij} = 2N_e s_{ij}$ as the scaled selection coefficient for this mutation [17]. As substitution rates are the product of fixation and mutation rates, $\mu$, this model's $Q$ matrix is populated by elements

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

for a substitution from codon $i$ to $j$ [15, 37].

Given detailed balance (reversibility), we have

$$q_{ij} P_i = q_{ji} P_j, \tag{3}$$

where $P_i$ is the stationary frequency of codon $i$.

From equations (2) and (3), we can write the ratio of substitution probabilities as

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

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

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

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

$$K_N = N_e \sum_i \sum_{j \in \mathcal{N}_i} P_i \mu_{ij} u_{ij}, \tag{6}$$

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

$$L_N = \sum_i \sum_{j \in \mathcal{N}_i} P_i \mu_{ij} , \tag{7}$$

and thus we find that

$$dN = \frac{K_N}{L_N} = \frac{N_e \sum_i \sum_{j \in \mathcal{N}_i} P_i \mu_{ij} u_{ij}}{\sum_i \sum_{j \in \mathcal{N}_i} P_i \mu_{ij}} . \tag{8}$$

Similarly, for $dS$, the synonymous evolutionary rate $K_S$ per synonymous site $L_S$, we have

$$dS = \frac{K_S}{L_S} = \frac{N_e \sum_i \sum_{j \in \mathcal{S}_i} P_i \mu_{ij} u_{ij}}{\sum_i \sum_{j \in \mathcal{S}_i} P_i \mu_{ij}} , \tag{9}$$

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, if we assume that nucleotide mutation rates are symmetric and that all synonymous codons have equal fitness (i.e. synonymous mutations are neutral), we have the synonymous fixation rate $u_{ij} = 1/N_e$ [39]. Under this circumstance, the value for $dS$ reduces to 1.

## MutSel models strictly describe purifying selection.

Using 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 scaled amino acid fitness values $F_a = 2N_e f_a$. For each fitness distribution, we drew all values $F_a$ from a normal distribution $\mathcal{N}(0, \sigma^2)$, where $\sigma^2 \sim \mathcal{U}(0, 4)$. Here, higher values for $\sigma^2$ correspond to larger fitness difference among amino acids, prompting selection to act more strongly against nonsynonymous changes. Thus, higher $\sigma^2$ values indicate strong purifying selection, while lower values indicate weaker purifying selection, and finally $\sigma^2 = 0$ indicates that all amino acids are equally fit. We note that these $F_a$ quantities correspond exactly 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 scaled 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 $\lambda$ 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 using symmetric mutation rates, where transitions occur at a rate $\mu \kappa$, and transversions at a rate $\mu$. We use the value $\mu = 10^{-6}$ for all $dN/dS$ calculations, and we draw a unique value for $\kappa$ from $\mathcal{U}[1, 6]$ for each set of codon fitnesses.

We found that $dN/dS$ values scale excellently with the variance ($\sigma^2$) of the distribution of amino-acid scaled selection coefficients (Figure 1). As expected, $dN/dS$ and $\sigma^2$ 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 fitness distribution without codon bias (Figure 1A) than for those with codon bias (Figure 1B). The weakened relationship under codon bias emerges from the fact that fitness

differences among synonymous codons 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 $\sigma^2$ 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, scaled selection coefficients strictly yield $dN/dS \leq 1$ . This proof formalizes the MutSel model's 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 $dN/dS \leq 1$ does not hold when synonymous changes are not neutral, as seen in Figure 1B. In other words, even though the underlying evolutionary model assumes a system at evolutionary equilibrium, $dN/dS$ 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. Even so, the fact that $dN/dS$ can, in theory, spuriously bear the hallmark of positive selection, even when purifying selection alone is occurring, is cautionary tale warning against naive interpretation of $dN/dS$ calues. For instance, it is theoretically possible that estimates of positive selection, as assessed by $dN/dS$, in species with high levels of codon bias driven in part by selection (e.g. *E. coli*, *Drosophila*, or certain mammalian species [42, 45, 44]) may actually signal strong codon bias, and not positive selection at all.

## Relationship provides a novel benchmarking approach.

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 a given model has been correctly implementation, it inherently cannot discern 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 $\omega$-based model, in order to benchmark the model's accuracy. In this way, we overcome a common pitfall of standard benchmarking strategies in which model performance is assessed using data simulated according to the model itself.

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]. Throughout the remaining text, we refer to $dN/dS$ inferred using ML as $\omega$, and to $dN/dS$ computed using equations (1) - (9) simply as $dN/dS$.

We found that $dN/dS$ values agree nearly perfectly with $\omega$ MLEs (Figure 2B). This agreement was neither influenced by the presence of codon bias, nor by variable nucleotide compostition; indeed, simulated alignments featured a wide range (0.21-0.89) of GC contents. Additionally, in Figure 2C, we demonstrate that $\omega$ 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 $\omega$-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, $\omega$-based models do not dramatically suffer from any modeling limitations.

## Best performing models do not display best model fit.

We next sought to test the accuracy of $\omega$-based models using more realistic parameter values. To this end, we determined distinct codon fitness distributions using 498 unique distributions of experimentally-derived, site-specific amino acid fitnesses for H3N2 influenza nucleoprotein (NP), given by ref. [35]. We combined each of these fitness distributions with three sets of experimentally-determined mutation rates, either for NP [35], yeast [46], or polio virus [47] in order to determine $498 \times 3 = 1494$ distinct distributions of steady-state codon frequencies (see *Methods* section for details). While all three mutation matrices are asymmetric, each features a differing degree of mutational bias; 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. For each resulting set of stationary codon frequencies in combination with its respective set of mutation rates, we calculated $dN/dS$ and simulated alignments from which we inferred $\omega$.

$\omega$-based models account for the nucleotide mutational bias by incorporating either target codon [10] or target nucleotide [11] frequencies; these frameworks are known, respectively, as GY-style and MG-style models. For example, the instantaneous rate matrix element giving the substitution probability from codon AAA to AAG would contain the target codon frequency $\pi_{\text{AAG}}$ in GY-style models, but the target nucleotide frequency $\pi_{\text{G}}$ in MG-style models. Previous works have shown that MG-style and GY-style models can yield different $\omega$ estimates [13, 48] and thus we inferred $\omega$ 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 [13]. We term the former framework MG1, and the latter MG3. Note that MG1 corresponds to the original MG-style model [11], whereas MG3 corresponds to the so-called MG94xHKY84 model [13]. (In Appendix 2 we show how the MG-style models can be written in a GY-style framework.)

Figure 3 shows the resulting relationships between $dN/dS$ and $\omega$ MLEs for each set of mutation rates (NP, yeast and polio), across 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 $\omega$ MLEs, and Figure 3B displays $r^2$ values between $dN/dS$ and $\omega$. The precise bias and $r^2$ values are given 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 $\omega$ estimates. Most often, the model underestimates $\omega$ relative to the true $dN/dS$ value. Indeed, for all frequency parameterizations, $\omega$ 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 and MG3) generally outperform all other frequency estimators, with a notable exception of F1x4. In particular, MG1 clearly performs the best 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 parameterizations, 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 frequently yielded. 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. Therefore, evaluating model performance based strictly on model fit can be counterproductive, as model fit is clearly at odds with model performance, and rather misleading. In sum, going forward, we highly recommend that researchers employ MG-style matrix frameworks in their $dN/dS$ inferences.

## Conclusions

We have shown that $dN/dS$ be accurately calculated from selection coefficients, revealing that $\omega$-based and MutSel models yield consistent and overlapping information regarding the strength of natural selection. Importantly, our proof that $dN/dS \leq 1$ (assuming symmetric mutation rates and that synonymous changes 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, 18, 16]. Thus, if the aim is to identify positive selection, only $\omega$-based models, of the two frameworks examined here, are justified.

However, we also find that, when selection acts on synonymous changes, $dN/dS$ can readily be greater than 1, even though the protein sequence is strictly evolving under purifying selection. This seemingly paradoxical finding actually reflects an assumption violation; the assertion that $dN/dS > 1$ necessarily corresponds to positive, diversifying selection requires that synonymous changes are neutral, which clearly does not hold if codon bias is present. This result builds on previous studies which have shown that purifying selection can also yield $dN/dS > 1$ if sequences considered contain segregating polymorphisms rather than strictly fixed differences [51, 52, 53]. Thus, it is becoming increasingly clear that the $dN/dS = 1$ neutral threshold typically used to distinguish purifying and positive selection is highly sensitive to violations in model assumptions.

Finally, we emphasize the utility of examining the relationships among distinct modeling frameworks. Not only can such an approach uncover certain mathematical properties of model parameters, but it also provides a robust framework for assessing the performance of various inference method. In particular, our approach allowed us to identify biases in $dN/dS$ inference frameworks which were not reflected in model fit assessments. Indeed, the best-performing model frameworks were not the models with the best fit to the data, ultimately revealing that relying strictly on

model performance is a counterproductive strategy in model selection. As we have shown here, this approach has great potential to reveal previously unrecognized model properties or biases that quantities of model fit are unable to realize, and will help ensure robust model development going forward.

# 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 $\sigma^2$ was drawn from $\mathcal{U}(0, 4)$ for each fitness distribution. We converted these amino acid fitnesses to codon fitnesses, $F_i$, as follows. For 100 of the fitness distributions, we directly assigned all codons within a given amino acid family the fitness $F_a$, and thus all synonymous codons had 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 $\lambda$ for each fitness distribution from $\mathcal{U}[0, 2]$. We then computed stationary codon frequencies as

$$p_i = \frac{e^{F_i}}{\sum e^{F_k}}, \tag{10}$$

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]. We used equations (6) - (9) to compute $dN/dS$ value for each resulting set of stationary codon frequencies. For these calculations, we set the mutation rate for transitions as $\mu\kappa$, and the rate for all transversions as $\mu$. We used the value $\mu = 10^{-6}$ for all $dN/dS$ calculations, and we drew a unique value for $\kappa$ 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

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

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 (Figure 4), 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 rate matrix $Q$ for each experimental model is populated by

$$q_{ij} = \begin{cases} \frac{F_j}{F_i}\mu_{ij} & \text{single nucleotide change, where } F_j \geq F_i \\ \mu_{ij} & \text{single nucleotide change, where } F_j < F_i \\ 0 & \text{multiple nucleotide changes} \end{cases}, \tag{12}$$

for a substitution from codon $i$ to codon $j$, 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 constituent 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.

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,

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

where $\kappa$ is the transition-transversion bias, $\pi_j$ is the equilibrium frequency of the target codon $j$, and $\omega$ represents $dN/dS$ [10, 1]. Importantly, this model's $\pi$ 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]. Thus, when inferring $\omega$ on datasets which used symmetric mutation rates, we used assigned the value 1/61 to all parameters $\pi_i$, as all codons are equally probable in the absence of mutational bias.

Alternatively, when inferring $\omega$ for alignments simulated with experimental fitness and mutation rates, we used several different frequency parameterizations, including GY-style [10] (target codon frequency) and MG-style [11] (target nucleotide frequency) parameterizations.**FILL IN DESCRIPTION OF FREQUENCIES.**

# 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 $dN = K_N/L_N \leq 1$. 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 $\mu$ 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

$$N_e\mu\big(P_i u_{ij} + P_j u_{ji}\big) = \frac{2P_i P_j(\log(P_i) - \log(P_j))}{P_i - P_j} \,. \tag{14}$$

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

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

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

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

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

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

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

$$\frac{\partial F(x, y)}{\partial x} = \frac{-2y^2 \sum\limits_{n=3}^{\infty} \frac{1}{n}\left(1 - \frac{x}{y}\right)^n}{(x - y)^2} \,, \tag{18}$$

which is clearly negative.

## Appendix 2

**I seem to be completely incapable of getting this section to sound anything but confusing, and I'm taking a break after staring at it for 2 hours.** GY-style matrices may be expressed in the framework of the general-time reversible (GTR) model, in which the instantaneous matrix $Q$ can be obtained from a $61 \times 61$ symmetric substitution rate matrix and a $1 \times 61$ vector containing the equilibrium codon frequencies, which correspond to the stationary distribution of the Markov chain. On the other hand, MG-style rate matrices explicitly consider nucleotide,

not codon, frequencies, and thus do not clearly follow this paradigm. We now describe how the MG-style matrix can be rewritten in a form that follows the GTR framework.

To begin, we note that mutation is a nucleotide-level, and not a codon-level process, and the mutation process can be represented by a $4 \times 4$ symmetric matrix of mutation rates $\mu_{xy}$ and a vector of stationary nucleotide frequencies $\pi_y$. If the target is nucleotide $y$, the instantaneous substutition rate matrix $Q$ must incorporate the factor $\mu_{xy}\pi_y$ to properly reflect the mechanism of mutation, even in a codon framework. Substitution can additionally be described by a symmetric $61 \times 61$ matrix. We further note that the MG-style matrix yields the stationary codon frequency $\pi_i = \pi_{i_1}\pi_{i_2}\pi_{i_3}C$ for a given codon $i$, where $C = 1 - \Pi_{\text{stop}}$ and $\Pi_{\text{stop}} = \pi_T\pi_A\pi_G + \pi_T\pi_G\pi_A + \pi_T\pi_A\pi_A$ [11].

We can therefore write the mutation rate $\mu_{xy}\pi_y$ instead as to $\mu_{xy}/(\pi_m\pi_n) \times C$, where $\pi_m$ and $\pi_n$ are the nucleotides which do not change in a given instantaneous codon substitution. This allows us to rewrite the rate instantaneous matrix as

$$
Q_{ij} = \begin{cases}
\dfrac{\mu_{ij}C}{\pi_{m_{ij}}\pi_{n_{ij}}}\pi_j & \text{synonymous change} \\[2em]
\dfrac{\mu_{ij}C}{\pi_{m_{ij}}\pi_{n_{ij}}}\omega\pi_j & \text{nonsynonymous change} \\[2em]
0 & \text{multiple nucleotide changes}
\end{cases}
\tag{19}
$$

for a substitution from codon $i$ to codon $j$, where $\pi_{x_{ij}}$ is the frequency of the nucleotide $x$ which does not change during the codon $i$ to $j$ substitution. This matrix indeed does indeed conform to the GTR framework, and is equivalent to the standard MG-style rate matrix [11] in the special case of the HKY85 mutation model.

# References

[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] Rocha E, *et al.* (2006) Comparisons of *dN/dS* are time dependent for closely related bacterial genomes. *J Theor Biol* 239:226 – 235.

[52] Kryazhimskiy S, Plotkin JB (2008) The population genetics of *dN/dS*. *PLoS Genet* 4:e1000304.

[53] Mugal CF, Wolf JBW, Kaj I (2014) Why time matters: Codon evolution and the temporal dynamics of *dN/dS*. *Mol Biol Evol* 31:212–231.

[54] 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.
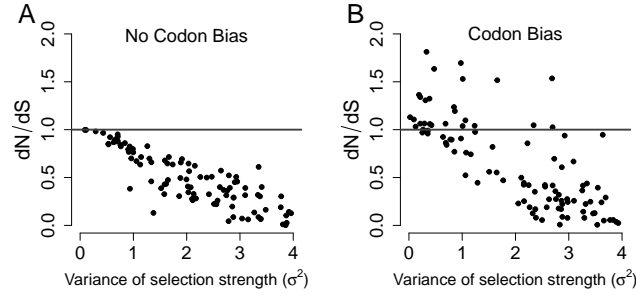
Figure 1: $dN/dS$ decreases in proportion to amino-acid level selection strength. $dN/dS$ is plotted against the $\sigma^2$ of the simulated distribution of amino-acid scaled selection coefficients. Higher values of $\sigma^2$ indicate larger fitness differences among amino acids, whereas the limiting value of $\sigma^2 = 0$ indicates 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.
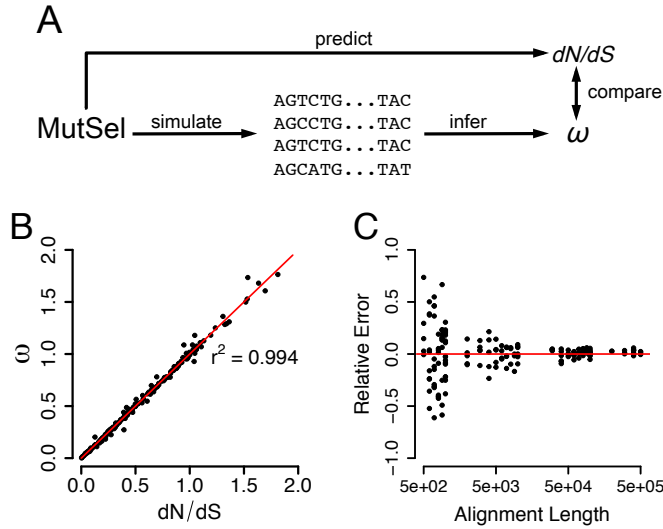


Figure 2: Combined modeling approach to assess performance of $dN/dS$ inference frameworks. (A) Protein-coding alignments are simulated in the MutSel modeling framework. $dN/dS$ can then be calculated from scaled selection coefficients as well as through a ML inference framework. Comparing resulting quantities reveals the accuracy in the chosen $dN/dS$ inference framework. (B) Regression between $dN/dS$ values as calculated from scaled selection coefficients and as inferred using the M0 model [10, 1, 2]. Each point corresponds to a single simulated alignment, and the red line is the $x = y$ line. (C) Convergence of $\omega$ 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 is the $y = 0$ line, indicating no error.
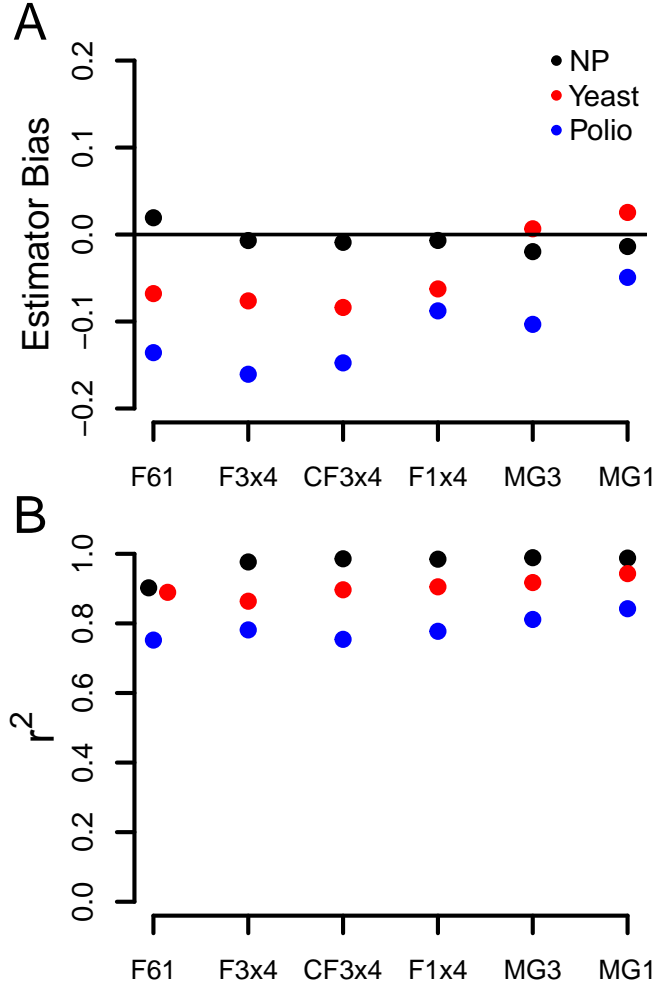
Figure 3: (A) Estimator bias and (B) $r^2$ values between $dN/dS$ and $\omega$ MLEs across model frequency parameterizations, for each set of nucleotide mutation rates. Negative biases indicate $\omega$ estimates that are, on average, lower than $dN/dS$. For all frequency parameterizations, bias generally increases as mutation rates become increasingly asymmetric. Even so, MG-style models tend to yield far less biased $\omega$ estimates than do GY-style models.

# Figures and Tables

Table 1: Mean $\Delta$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 |
| MG1 | -13325.5 | -10042.0 | -5147.6 |
| F1x4 | -13524.5 | -13658.5 | -15468.3 |
| MG3 | -14401.3 | -12851.6 | -8624.9 |
| F3x4 | -14807.2 | -17385.3 | -19384.6 |

Based on AIC scores, the F61 parameterization strongly outperforms all other model parameterizations for all mutation rates, even so the F61 framework yields neither the most accurate nor the most precise parameter estimate. Note that the order of frequency models shown in the table corresponds to the model ranking for NP, and the ranking differs somewhat for yeast and polio datasets. AIC is computed as $AIC = 2(k - \ln(L))$, where $k$ is the number of free parameters of the model, and $\ln(L)$ is the log-likelihood. As codon and/or nucleotide frequency parameters are directly estimated from the data, all models have 3 free parameters ($\omega$, $\kappa$, and a global branch length scaling parameter).
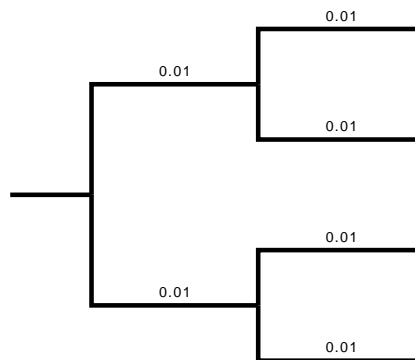
Figure 4: Phylogeny used for all simulated alignments.

# Supplementary Information

Table S1. Estimator bias between $\omega$ MLEs and the expected, true $dN/dS$ values, for all mutation rates and model frequency parameterizations examined. Negative bias values indicate that $\omega$ 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 MG3, where $P = 5.4 \times 10^{-5}$.

| Mutation rate | MG1 | F1x4 | MG3 | 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 $\omega$ MLEs and the expected, true $dN/dS$ values, for all mutation rates and model frequency parameterizations examined. All values shown are statistically significant, with all $P < 2 \times 10^{-16}$ .

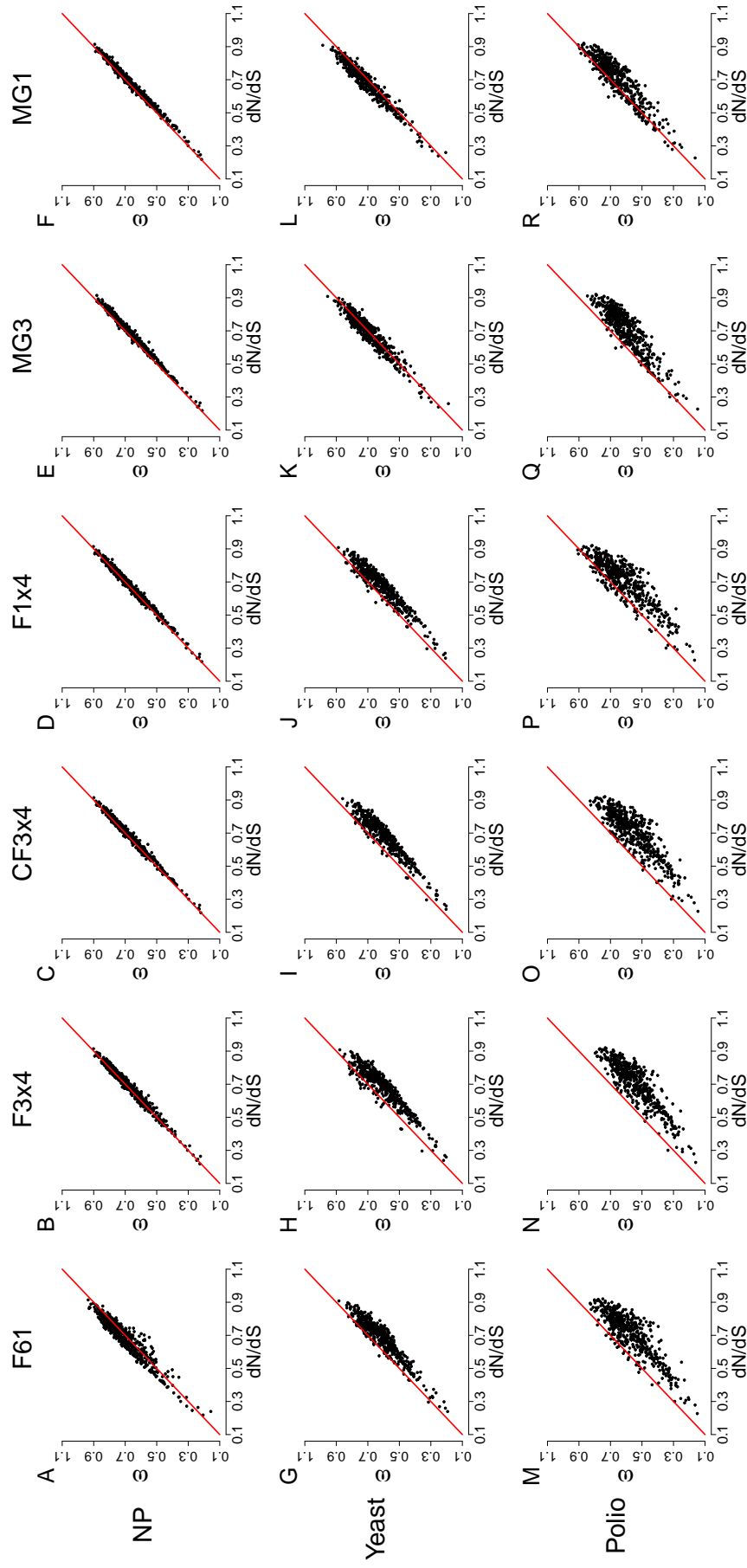| Mutation rate | MG1 | F1x4 | MG3 | CF3x4 | F3x4 | F61 |
|---|---|---|---|---|---|---|
| NP | 0.988 | 0.989 | 0.985 | 0.986 | 0.977 | 0.902 |
| Yeast | 0.943 | 0.917 | 0.905 | 0.897 | 0.864 | 0.889 |
| Polio | 0.842 | 0.811 | 0.777 | 0.754 | 0.781 | 0.752 |

Figure S1. Regressions for inferred $\omega$ estimates and $dN/dS$ values, as calculated from scaled selection coefficients, for datasets simulated using experimental fitnesses and mutation rates. Each point represents an alignment, and red lines are the $x = y$ line.