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

Inferring the strength of natural selection in protein-coding sequences along a phylogeny is a major objective in the field of molecular evolution. Two broad classes of Markov-process models have been developed to describe these selective pressures. The first and most widely-used variety, mechanistic codon (MC) models, estimate the evolutionary rate ratio dN/dS and have been developed to a high level of sophistication. The second class of models, known as mutation-selection-balance (MutSel) models, explicitly consider the dynamic interplay between mutation and selection, and estimate site-specific amino-acid or codon scaled selection coefficients.
By contrast, MutSel models estimate scaled selection coefficients for amino acid and/or codons [1, 2, 3, 4, 5]. Thus, while MC models describe the how quickly a protein's constituent amino acids change, MutSel models calculate the strength of natural selection operating on specific amino-acid or codon changes.
However, the extent to which these modeling frameworks relate to each other is unknown. Do dN/dS estimates yield similar or distinct information from scaled selection coefficients? To answer this question, we derive a formal mathematical relationship between these two quantities. We demonstrate that scaled selection coefficients can precisely predict dN/dS, indicating that MC and MutSel models are in complete agreement. Using this relationship, we can gain unique insight into the properties and limitations of both modeling frameworks. We find that MutSel models cannot describe positive selection and/or adaptive evolution, and are therefore only suitable when sequences evolve strictly under purifying selection. However, when synonymous mutations are not neutral, it is possible to achieve $dN/dS > 1$, even though positive selection is not occurring. Finally, we find that MC models produce systematically biased dN/dS estimates when nucleotide mutation rates are asymmetric, revealing that MC models do not properly account for nucleotide compositional bias.

mechanistic codon models | 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 ($dN/dS > 1$) [6, 7, 8, 9]. Frameworks for calculating dN/dS have broadly fallen into two camps: heuristic counting methods [10, 11, 12, 13, 14] and maximum likelihood methods [15, 16, 6, 17]. The latter variety assume an explicit Markov-process model of sequence evolution, and yield maximum likelihood estimates (MLEs) for the parameter ω, which represents the quantity dN/dS. MC models have become a staple of comparative sequence analysis since their introduction in the 1990s (see ref [18] for a comprehensive review). Throughout this paper, we will refer to these inference methods as ω models.

A second class of models, known as mutation-selection-balance (MutSel) models, are increasingly being viewed as a viable alternative to ω models. 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 [1, 4]. 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 [1, 19, 2, 3]. Although first introduced over 15 years ago [1], MutSel models have seen little use due to their high computational expense. Recently, however, several computationally tractable model implementations have emerged [20, 5], 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 [16, 21], 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 MC 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 MC models [1, 2, 3, 4]. Recent phylogenetic studies have also demonstrated that evolutionary models which account for amino acid fitness values offer dramatic improvements over MC 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 these competing modeling frameworks will more precisely re-

Reserved for Publication Footnotes

veal 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.

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 [17] 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 [1, 19, 3, 4] 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, 19]. The fixation probability for a mutation from codon i to codon j is given by

$$P_{\text{fix}}(i \to j) = \frac{2s_{ij}}{1 - e^{-2N_e s_{ij}}} \approx \frac{1}{N_e} \frac{2N_e s_{ij}}{1 - e^{-2N_e s_{ij}}} \qquad [1]$$

[38, 1, 19]. 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

$$q_{ij} = N_e \mu_{ij} P_{\text{fix}}(i \to j) = \mu_{ij} \frac{S_{ij}}{1 - e^{-S_{ij}}}, \qquad [2]$$

and this expression corresponds to the instantaneous matrix element Q_{ji}. [1, 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

$$K_{\text{N}} = N_e \sum_i \sum_{j \in \mathcal{N}_i} p_i P_{\text{fix}}(i \to j) \mu_{ij}\,, \qquad [6]$$

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 [15, 17] as

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

and thus we find that

$$dN = \frac{K_{\text{N}}}{L_{\text{N}}} = \frac{N_e \sum_i \sum_{j \in \mathcal{N}_i} p_i P_{\text{fix}}(i \to j) \mu_{ij}}{\sum_i \sum_{j \in \mathcal{N}_i} p_i \mu_{ij}}\,. \qquad [8]$$

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

$$dS = \frac{K_{\text{S}}}{L_{\text{S}}} = \frac{N_e \sum_i \sum_{j \in \mathcal{S}_i} p_i P_{\text{fix}}(i \to j) \mu_{ij}}{\sum_i \sum_{j \in \mathcal{S}_i} p_i \mu_{ij}}\,, \qquad [9]$$

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. [??] 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 $P_{\text{fix}}(i \to j) = 1/N_e$ [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, **?**]/

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 **??**). As Figure 1 demonstrates, 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. Furthermore, as expected, this trend is much stronger for alignments without codon bias (Figure 1A, $r^2 = 0.83$) than for alignments with codon bias (Figure 1B, $r^2 = 0.45$). 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 **PROOF**, 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 constant selection pressure over time and therefore no positive selection is occuring, dN/dS can readily be greater than 1. Furthermore, we find that 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 has been shown to be relatively weak in most taxa [40]. For instance, experimental evidence from the yeast Hsp90 protein suggests that, while there are some fitness differences among synonymous codons, these differences are exceedingly minimal compared to fitness differences among amino acids [41, 42]. Moreover, our implementation of codon bias explicitly assumed that selection alone, and not mutation, was the sole source of codon bias. This implementation might not be entirely biologically realistic, as both mutational and selective forces likely contribute to codon bias in real genomes [43, 44, 40, 45, 46]. Even so, it is 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 [44, 47, 46], may not be true cases of positive selection, but rather signals of strong codon bias.

Application of above. A useful outcome of the relationship we present is that we can use it to benchmark inference methods.... **introductory sentence/paragraph of some variety, here are some thoughts.** An important application of the robust relationship we have established between dN/dS and selection coefficients is that we can it it to examine how the different modeling frameworks broadly relate to one another. For instance, this relationship might serve as a benchmark to determine if dN/dS and MutSel inference methods yield comparable results. Typically, we assess model performance by simulating data under simple circumstances which adhere to the model's underlying assumptions. While this is a valuable, it's not all we can do. For instance, that Holder study was able to identify unknown limitations in phylogenetic methods.

Therefore, we simulated alignments as a continuous-time Markov process [17] according to the Halpern-Bruno model [1] to examine how our dN/dS derivation performs relative to standard dN/dS inference methods. We simulated 200 alignments using the scaled selection coefficients generated in the previous subsection, and for each alignment, inferred dN/dS using the M0 model [15, 7], as implemented in the HyPhy batch language [32]. This model uses the GY94 instantaneous rate matrix, which includes the primary parameters κ, or the transition/transversion bias, equilibrium codon frequencies π, and finally ω, which represents the dN/dS rate ratio [15, 6]. Throughout the remaining text, we refer to dN/dS as inferred using maximum likelihood by its parameter ω, and to dN/dS computed using equations [**1**] - [**9**] simply as dN/dS.

The relationship between dN/dS and ω measurements is shown in Figure 2A-B. It is clear that dN/dS values computed from scaled selection coefficient agree nearly perfectly with ω, and the presence of codon bias does not influence this robust relationship. 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 demonstrate that the dN/dS quantity is fully contained within MutSel model parameters, and that both the and MutSel models yield consistent information.

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 models are still able to perform really well.

Now that we confirm that dN/dS inference frameworks play nicely with MutSel models, we can test the accuracy of ω models using more realistic parameters. To this end, we made use of realistic amino acid fitness and nucleotide mutation rate parameters. In particular, we used influenza nucleoprotein (NP) site-specific amino acid preference values, given by ref. [35]. These data 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 [48], or polio virus [49]. 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 the corresponding mutation rates to simulate and infer dN/dS.

When inferring ω, we used five different codon frequency model parameterizations. These parameterizations included F_equal, which assigns frequencies of 1/61 to all sense codons, [17] and the common frequency estimators F61 [15], F3x4 [16], and CF3x4 [50]. The F61 estimator approximates these model parameters using an alignment's empirical codon frequencies, while the F3x4 and CF3x4 estimators approximate codon frequency parameters using positional nucleotide frequencies. Additionally, we inferred ω using a fifth frequency parameterization which consisted of the codon frequencies that would arise strictly from mutation, in the absence of natural selection. We term this parameterization "Ftrue," as its values are precisely those intended for these model parameters [15, 16, 14, 17].

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 codon frequency parameterizations (full regression plots are shown in Figure S3). Figure 3A displays the bias, or systematic deviation from a 1:1 relationship, between dN/dS and ω, and Figure 3B displays r^2 values between dN/dS and ω. Note that a bias of 0 would indicate that .....

Overall, Figure 3 demonstrates that ω estimates are systematically biased downwards as mutation rates become increasingly asymmetric. This trend exists across all 5 codon frequency parameterizations; ω inferences on alignments with NP mutation rates have the least amount of bias, followed by yeast and finally polio mutation rates, and a parallel trend exists for the correlation strengths. Strikingly, even though Ftrue uses the exact values intended for these parameters, it does not strongly outperform the other frequency parameterizations. Instead, Ftrue features the same general trend of decreasing accuracy with increasing mutational asymmetry. Surprisingly, the F_equal parameterization performs similarly, if not better than, the other frequency estimators examined here. While it correlates worse for polio, it additionally has the least amount of bias. F_equal features the least amount of bias, and has very high correlations 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 relative strength of the r^2 values for the latter three estimators is misleading given their relatively high biases. In fact, these increased biases demonstrate that these estimators systematically underestimate dN/dS. Taken together, these results imply that the codon frequency parameters in MC models cannot adequately account for nucleotide compositional bias as generated by mutation.

ω inferences are least biased from Fequal, although the lower correlations for this one indicates more noise.

There is a noteworthy exception to the overall trend of ω underestimation; ω MLEs for NP mutation rates, when computed using F61, were actually overestimates of dN/dS. We attribute this result to the fact that the NP mutation rates were only minimally asymmetric, and in the absence of natural selection, these mutation rates would produce nearly even nucleotide frequencies. Importantly, when nucleotide mutation rates are symmetric, steady-state codon frequencies are controlled only by selection, as there is no opportunity to generate compositional bias through mutation [37]. Therefore, because the F61 estimator directly uses empirical codon frequencies, the resulting M0 codon frequency parameters actually contain information about the strength of natural selection. The end result is that selection pressures which should be strictly incorporated in the ω parameter are inadvertently contained within the codon frequency parameters. Thus, the model infers selection to be weaker than it actually is, producing elevated ω MLEs. Although the ω overestimation was relatively small in this particular case, this example highlights that it is crucial to properly parameterize MC models. If these models are incorrectly parameterized, the ω parameter will no longer accurately represent the dN/dS evolutionary rate ratio, but will instead be a meaningless quantity.

We are confident that this bias definitively resulted from asymmetric mutation rates, and not from uneven nucleotide compositions. Indeed, our alignments simulated with symmetric mutation rates featured a wide array of GC-contents, ranging from 0.22-0.82. The underlying mutational symmetry means that unequal nucleotide compositions arose strictly from natural selection favoring particular amino acids, and MC models inferred dN/dS perfectly in these circumstances.

Future Directions

We will conclude with insights gained from our study and recommendations for using this inference frameworks going forward.

1. mutsel and omega models basically agree. anything a mutsel model can do, an omega model can. however, this is not the case in the reverse. positive selection, which we expect to accompany adaptation vs equilibrium evolution, which corresponds to a static fitness landscape. therefore, there can be no adaptive evolution in MutSel models, so stop looking for it. This is mirrored by our finding that dN/dS ¡ 1, under the assumptions of symmetry and neutral changes are synonymous. let's make the distinction between beneficial changes and positive selection, please. you can move to an amino acid of higher fitness at steady state - this does not imply positive, diversifying selection, and treating them as such is misleading. semantics are really important. 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. [53]). 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.

Methods

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.

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. [53]). 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.

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

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

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 [17] according to the Halpern-Bruno MutSel model [1]. In simplified form, this model's instantaneous rate matrix Q is given by

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

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 [48], or polio virus [49] 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 [7], 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 [15, 6]. 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 [15, 16, 14, 17]. Therefore, when inferring ω for alignments simulated with symmetric mutation rates, used the F_equal frequency parameterization, which assigns equal values of 1/61 for all π_i [17]. F_equal gives the exact codon frequencies expected under symmetric mutation, in the absence of selection. Alternatively, when inferring ω for alignments simulated with experimental mutation rates, we used 5 different sets of equilibrium frequency parameterizations. First, we inferred ω by specifying codon frequencies which would arise strictly from mutational processes in the absence of natural selection. We computed these codon frequency values using the same approach as we did in calculating the true steady-state codon frequencies, except instead of using the experimental amino acid preference data, we assigned all amino acids the same preference value of 0.05. This strategy eliminated amino-acid level selection and allowed mutation rates alone to determine equilibrium codon frequencies. We term this frequency parameterization "Ftrue." Finally, we used the common frequency estimators F61 [15], F3x4 [16], and CF3x4 [50]. 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. Finally, we used ω with the F_equal parameterization.

ACKNOWLEDGMENTS. This work was supported by the army and by NIH.

1. Halpern AL, Bruno WJ (1998) Evolutionary distances for protein-coding sequences: modeling site-specific residue frequencies. Mol Biol Evol 15:910–917.
2. 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.
3. 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.
4. 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).
5. 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.
6. 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.
7. Yang ZH, Nielsen R, Goldman N, Pedersen AMK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449.
8. Kosakovsky Pond S, Frost SD (2005) Not so different after all: A comparison of methods for detecting amino acid sites under selection. Mol Biol Evol 22:1208–1222.
9. 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.
10. 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.
11. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426.
12. Pamilo P, Bianchi NO (1993) Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. Mol Biol Evol 10:271–281.
13. Ina Y (1995) New methods for estimating the numbers of synonymous and nonsynonymous substitutions. Journal of Molecular Evolution 40:190–226.
14. Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17:32–42.
15. Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725–736.
16. 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.
17. Yang Z (2006) Computational Molecular Evolution (Oxford University Press).
18. Anisimova M, Kosiol C (2009) Investigating protein-coding sequence evolution with probabilistic codon substitution models. Mol Biol Evol 26:255–271.
19. 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.
20. Rodrigue N, Lartillot N (2014) Site-heterogeneous mutation-selection models within the PhyloBayes-MPI package. Bioinformatics :1020–1021.

21. Kosakovsky Pond S, Muse SV (2005) Site-to-site variation of synonymous substitution rates. Mol Biol Evol 22:2375–2385.
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. Hietpas RT, Jensen JD, Bolon DNA (2011) Experimental illumination of a fitness landscape. Proc Natl Acad Sci USA 108:7896–7901.
42. Hietpas RT, Bank C, Jensen JD, Bolon DNA (2013) Shifting fitness landscapes in response to altered environments. Evolution 67:3512–3522.
43. Blumer M (1991) The selection-mutation-drift theory of synonymous codon usage. Genetics 129:897–907.
44. Duret L (2002) Evolution of synonymous codon usage in metazoans. Curr Opin Genet Dev 12:640–649.
45. 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.
46. Plotkin JB, Kudla G (2011) Synonymous but not the same: the causes and consequences of codon bias. Nature Rev Genet 12:32–42.
47. Chamary JV, Parmley JL, Hurst LD (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. Nature Rev Genet 7:98–108.
48. Zhu YO, Siegal ML, Hall DW, Petrov DA (2014) Precise estimates of mutation rate and spectrum in yeast. Proc Natl Acad Sci USA :FORTHCOMING.
49. Acevedo A, Brodsky L, Andino R (2014) Mutational and fitness landscapes of an RNA virus revealed through population sequencing. Nature 505:686 – 690.
50. Kosakovsky Pond SL, Delport W, Muse SV, Scheffler K (2010) Correcting the bias of empirical frequency parameter estimators in codon models. PLoS One 5:e11230.
51. Duret L, Galtier N (2005) Biased gene conversion and the evolution of mammalian genomic landscapes. Annu Rev Genomics Hum Genet 10:285–311.
52. Webster M, Hurst L (2012) Direct and indirect consequences of meiotic recombination: implications for genome evolution. Trends Genet 28:101–109.
53. 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

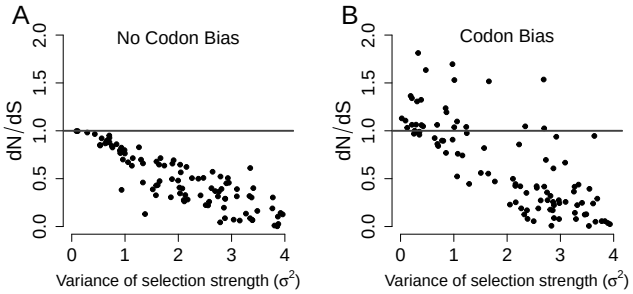

Fig. 1. dN/dS decreases in proportion to amino-acid level selection strength. dN/dS is plotted against the σ^2 of the simulated distribution of amino-acid scaled selection coefficients. Higher values of σ^2 indicate larger fitness differences among amino acids, whereas the limiting value of $\sigma^2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

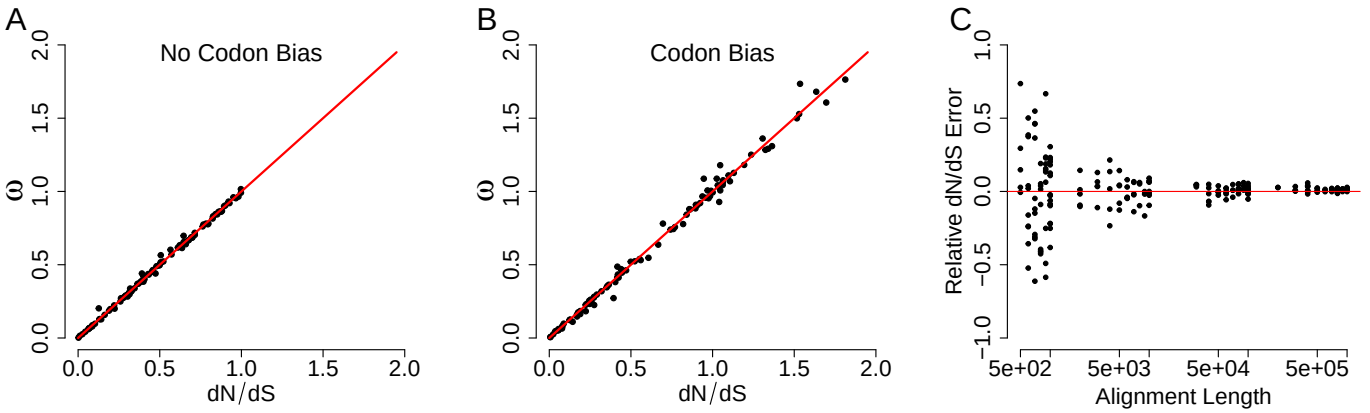

Fig. 2. Regressions between dN/dS values as calculated from scaled selection coefficients and as inferred using the M0 mechanistic codon model. Each point corresponds to a single simulated alignment. All ω values shown here were inferred by parameterizing the M0 model with κ fixed to its true, simulated value as well as the Fequal codon frequency specification [17]. 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.

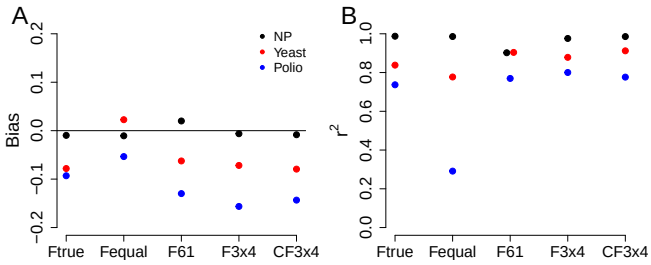

Fig. 3. Bias and correlation between dN/dS and ω MLEs across M0 codon frequency parameterizations, for each set of nucleotide mutation rates. (A) Bias, or systematic deviation from a 1:1 relationship. (B) r^2 values between dN/dS and ω. All bias and r^2 values are highly statistically significant, with all $p < 10^{-12}$.

Supplementary Information

Table S1. Estimator bias, representing the mean difference between the expected dN/dS value and estimated ω. representing the systematic deviation from a 1:1 relationship, between ω MLEs and dN/dS, 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 $p < 10^{-12}$.

	M0 model codon frequency parameterization				
Mutation rate	Ftrue	Fequal	F61	F3x4	CF3x4
NP	-0.00971	-0.0107	0.02	-0.0063	-0.0084
Yeast	-0.0782	0.0226	-0.0625	-0.0718	-0.0794
Polio	-0.0931	-0.0534	-0.130	-0.156	-0.143

Table S2. NYP r^2 values between ω MLEs and dN/dS, for all mutation rates and M0 codon frequency parameterizations examined. All r^2 values are statistically significant, with all $p < 10^{-15}$.

	M0 model codon frequency parameterization				
Mutation rate	Ftrue	Fequal	F61	F3x4	CF3x4
NP	0.988	0.986	0.902	0.976	0.986
Yeast	0.938	0.777	0.904	0.878	0.912
Polio	0.737	0.291	0.770	0.800	0.776

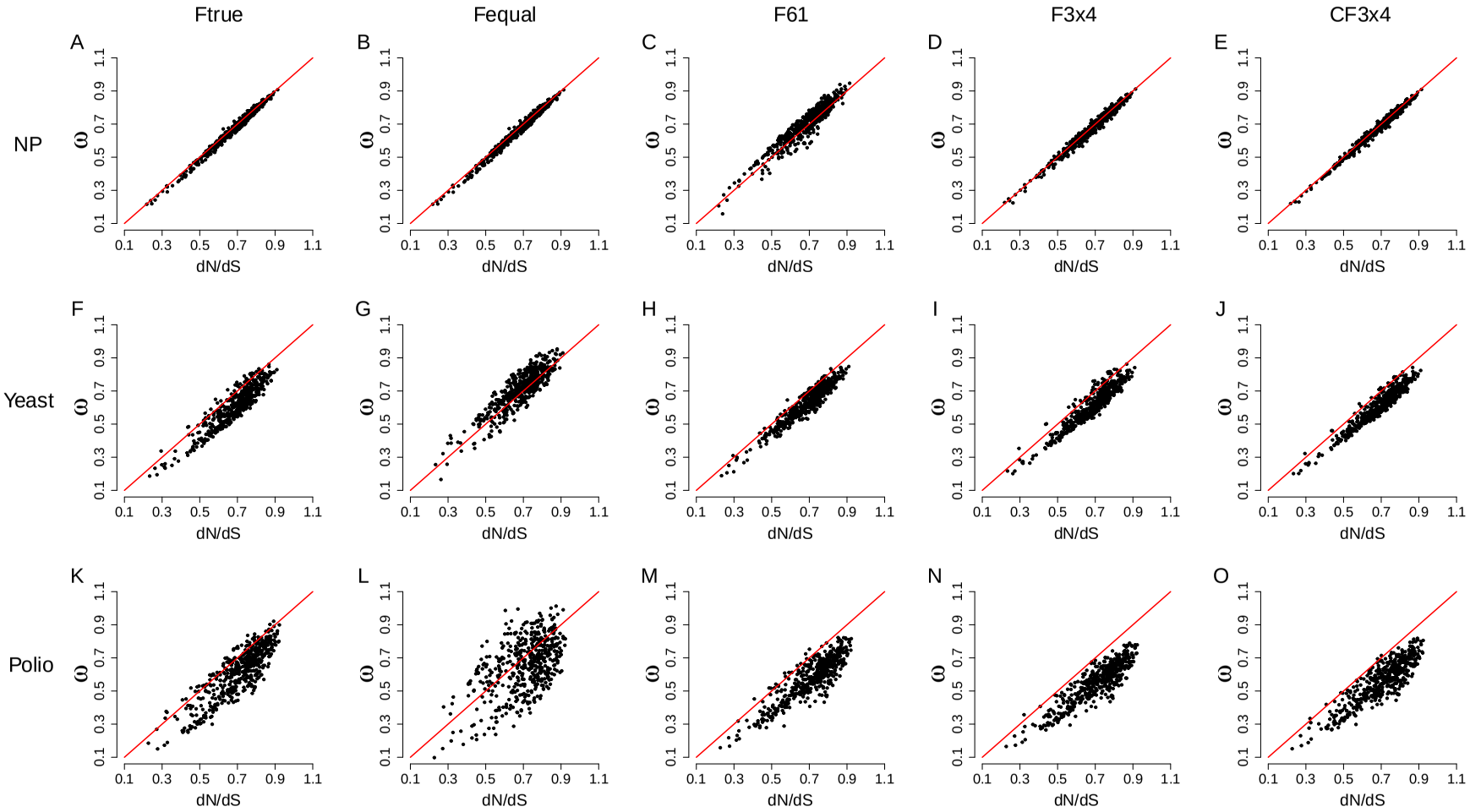

Figure S3. Regression plots for ω MLEs versus dN/dS values computed from scaled selection coefficients, for each set of nucleotide mutation rates and all M0 codon frequency parameterizations. Each point represents a single simulated alignment, and the red lines correspond to $x = y$. (A) Simulations which assigned equal fitness values to synonymous codons. (B) Simulations which allowed different fitness values among synonymous codons.