The relationship between dN/dS and scaled selection coefficients

Stephanie J. Spielman[1] and Claus O. Wilke[1]

Address:
[1]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.

*Corresponding author
Email: ??????????

Keywords: selection coefficients, dN/dS, natural selection, protein evolution, models of sequence evolution

Abstract

Two measures of the strength of selection on protein coding sequences are dN/dS and selection coefficients. Are the the same? Are the different? We don't know! But now we do. And they're the same. Also, stop using mechanistic codon models because parameterizing them is basically impossible.

Introduction

Over the years, various methods have been used to calculate the strength of natural selection acting on protein-coding sequences. Traditionally, the focus has been on estimating the evolutionary rate ratio, dN/dS, the rate of nonsynonymous to synonymous substitution rates. This metric indicates how quickly a protein's constituent amino acids change, and is widely used to identify cases of positive, diversifying selection ($dN/dS > 1$) [1–4]. Following early counting methods for estimating dN/dS (e.g. refs [5] and [6]), mechanistic codon models, which assume an explicit Markov-process model of sequence evolution (see ref. [7] for a comprehensive review), have taken a leading role as the inference method of choice since their introduction in the 1990s [1, 8, 9]. These models yield maximum likelihood estimates (MLEs) for the parameter ω, which represents the quantity dN/dS, and have seen great success in the field of molecular evolution.

A second class of models, known as mutation-selection-balance (MutSel) models, have emerged recently as a popular alternative to mechanistic codon models. The MutSel framework, couched firmly in population genetics theory, models the dynamic interplay between mutation and selection in a protein-coding sequence. MutSel models yield estimates of site-wise scaled selection coefficients, which indicate the extent to which natural selection favors, or disfavors, particular codons or amino acids at a given protein position. Although MutSel models were first introduced over 15 years ago [10], they have seen virtually no use due to their high computational expense. However, recently, several computationally tractable model implementations have emerged [11, 12], allowing for the first time the potential for widespread use.

Although both mechanistic codon models and MutSel models describe the same fundamental process of protein-coding sequence evolution along a phylogeny, it is largely unknown how these two classes of models 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. Whether dN/dS values have any correspondence with scaled selection coefficients remains an open question. Therefore, while 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.

Here, we formalize the relationship between mechanistic codon and MutSel models by examining the extent to which their focal parameters, dN/dS and scaled selection coefficients, yield overlapping information about the evolutionary process. To this end, we derive a mathematical relationship these models' primary parameters from which one can infer dN/dS values from selection coefficients alone. Using a simulation approach, we verify that dN/dS values estimated using selection coefficients alone correspond precisely to ω MLEs inferred using standard mechanistic codon models.

(???)Further, we prove that, under conditions of symmetric mutation rates, this relationship holds only under regimes of purifying selection or neutral evolution ($dN/dS \leq 1$). This proof reveals that MutSel models are inherently unable to describe accurately protein evolution under a regime of positive diversifying selection, or when $dN/dS > 1$.

Moreover, our analyses incidentally have revealed certain biases inherent in the ML ω inference approach. (...)

Methods

Sequence simulation

We simulated protein-coding sequences as a continuous-time Markov process [13] according to the MutSel model proposed by [10]. This model's instantaneous rate matrix Q is given by

$$Q_{ij} = \begin{cases} f_{ij}\mu_{ij}\kappa & \text{single nucleotide transition} \\ f_{ij}\mu_{ij} & \text{single nucleotide transversion} \\ 0 & \text{multiple nucleotide changes} \end{cases}, \quad (1)$$

where μ_{ij} is the symmetric nucleotide mutation rate and f_{ij}, the fixation probability from codon i to j, is defined as

$$f_{ij} = ln\left(\frac{\pi_j\mu_{ij}}{\pi_i\mu_{ji}}\right) \Big/ \left(1 - \frac{\pi_i\mu_{ji}}{\pi_j\mu_{ij}}\right), \quad (2)$$

where π_i is the equilibrium frequency of codon i.

We simulated protein-coding sequences along a 10-taxon phylogeny, with all branch lengths equal to 0.01, beginning with a root sequence selected using steady-state codon frequencies. We set a global nucleotide mutation rate of $\mu_{xy} = 10^{-6}$, and we drew a value for each simulation's κ from $\mathcal{U} \sim (1, 6)$, resulting in a fully symmetric mutation matrix. 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 allows us to verify our derived relationship between selection coefficients and dN/dS with a sufficiently sized data set.

Assigning scaled selection coefficients

We simulated 100 alignments under the assumption that all synonymous codons have equal fitness (e.g. no codon bias), and a second set of 100 alignments which incorporated codon bias. For both frameworks, we generated amino acid scaled selection coefficients, S_a, for each simulation, by fixing one coefficient to 0 and drawing the remaining 19 values from a normal distribution $(N) \sim (0, \sigma^2)$, where $\sigma^2 \sim U(0, 4)$. Here, σ^2 effectively represents the strength of natural selection; larger values of σ^2 will correspond to greater fitness differences among amino acids, and thus more selective pressure.

For simulations without codon bias, we assigned these amino acid selection coefficients to codons such that all synonymous codons had the same scaled selection coefficient. For simulations with codon bias, we randomly selected a preferred codon for each amino acid. We then assigned the preferred codon a selection coefficient of $S_a + b$, where $b = (2)$, and we assigned non-preferred codons selection coefficients of $S_a - b/(k-1)$, where k is the total number of synonymous codons for the given amino acid. In this way, the average selection coefficient for this amino acid remained unchanged, but one codon per amino acid was more selectively favored.

dN/dS Inference

We calculated a global dN/dS for each alignment using the mathematical framework outlined in (3)–(8) as well using standard maximum likelihood methods. Specifically, inferred dN/dS using the M0 mechanistic codon model [2], as implemented in the HyPhy batch language [14]. The M0 models uses the GY94 instantaneous rate matrix [1,8], which includes the primary parameters ω, κ, and equilibrium codon frequencies. As different κ and equilibrium codon frequency parameterizations can change ω estimates [13,15,16], we inferred ω under a variety of model parameterizations,

including three κ parameterizations (κ fixed to 1, κ as a free parameter, and κ fixed to its true value), and four codon frequency specifications (Fequal (all sense codons have an equilibrium frequency of 1/61), F3x4 codon frequencies [9], CF3x4 codon frequencies [17] and F61, or empirical codon freqencies [8]), ultimately resulting in 12 ω MLEs per simulated alignment. Note that we additionally verified that the system was evolving under a state-state process by verifying that the true, simulated codon frequencies were the same as the empirical codon frequencies calculated from the simulated alignment. All code used is freely available at **github**.

Results

Mathematical relationship between selection coefficients and omega

We describe here how to calculate dN/dS from the parameters of a MutSel model. Under the assumption that the mutational process is symmetric, e.g. $\mu_{xy} = \mu_{yx}$ for all nucleotide pairs xy, we can write the steady-state frequency of codon i as

$$\pi_i = \frac{e^{S_i}}{\sum_k e^{S_k}}, \tag{3}$$

where the sum in the denominator runs over all 61 sense codons [18]. Here, S_i is the scaled selection coefficient for codon i; larger S_i values correspond to higher frequencies of codon i.

The fixation probability for a mutation from codon i to codon j is [10, 18]

$$f_{ij} = \frac{1 - (\pi_i/\pi_j)^{1/N_e}}{1 - \pi_i/\pi_j} \approx \frac{1}{N_e} \frac{\ln \pi_j - \ln \pi_i}{1 - \pi_i/\pi_j}\,, \tag{4}$$

where N_e is the effective population size. Using this framework, we can calculate an evolutionary rate by summing over all substitution probabilities weighted by the frequency of the originating codon. Further, we can establish specific expressions for nonsynonymous and synonymous evolutionary rates, and then divide them in order to obtain a value for the evolutionary rate ratio dN/dS.

To begin, we can write the nonsynonymous rate K_N as

$$K_\mathrm{N} = N_e \sum_i \sum_{j \in \mathcal{N}_i} \pi_i f_{ij} \mu_{ij}\,, \tag{5}$$

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 [8, 13] as

$$L_\mathrm{N} = \sum_i \sum_{j \in \mathcal{N}_i} \pi_i \mu_{ij}\,, \tag{6}$$

and thus we find that

$$dN = \frac{K_\mathrm{N}}{L_\mathrm{N}} = \frac{N_e \sum_i \sum_{j \in \mathcal{N}_i} \pi_i f_{ij} \mu_{ij}}{\sum_i \sum_{j \in \mathcal{N}_i} \pi_i \mu_{ij}}\,. \tag{7}$$

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

$$dS = \frac{K_\mathrm{S}}{L_\mathrm{S}} = \frac{N_e \sum_i \sum_{j \in \mathcal{S}_i} \pi_i f_{ij} \mu_{ij}}{\sum_i \sum_{j \in \mathcal{S}_i} \pi_i \mu_{ij}}\,, \tag{8}$$

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. (5) and (6) but summing over $j \in \mathcal{S}_i$ instead of $j \in \mathcal{N}_i$.

Equations (3)–(8) establish a connection between the scaled selection coefficients and the evolutionary rate ratio dN/dS. Moreover, we note that, if we assume that all synonymous codons have equal fitness (e.g. synonymous mutations are neutral), the synonymous fixation rate $f_{ij} = 1/N_e$ [19]. Under this circumstance, the value for dS reduces to 1.

dN/dS can be accurately predicted from scaled selection coefficients

To validate our derived relationship between scaled selection coefficients and dN/dS, we simulated protein-coding sequences along a 10-taxon phylogeny according to a mutation-selection model framework [10, 18]. We simulated 100 alignments which gave equal fitness values to synonymous codons, and 100 alignments which incorporated codon bias (see Method for details). All simulations assumed a symmetric nucleotide rate matrix. For each alignment, we calculated dN/dS using equations (3)–(8) as well as using the M0 mechanistic codon model [1], as implemented in the HyPhy batch language [14].

The relationship between dN/dS calculations is shown in Figure 1A (for simulations with no codon bias) and Figure 1B (for simulations with codon bias). It is clear that dN/dS values derived using selection coefficients agree nearly perfectly with those inferred using standard maximum likelihood methods, and fitness differences among synonymous codons do not influence this robust relationship. Additionally, in Figure 1C, we demonstrate convergence of dN/dS estimates as the size of the data set, represented by simulated alignment length,increases. Taken together, these results demonstrate that MutSel model parameters fully encapsulate information regarding dN/dS, and that the results from MutSel and mechanistic codon models are in complete agreement.

Moreover, the strength of selection pressure scales fairly well with dN/dS. Figure 2 displays the relationship between dN/dS and the standard deviation, σ^2, of the distribution of amino acid selection coefficients. Higher values of σ^2 indicate higher fitness differences among amino acids, thus leading to stronger selection pressure acting on nonsynonymous substitutions. Figure 2 demonstrates that when fitness differences among amino acids are very high, dN/dS takes on lower values, properly reflecting stronger purifying selection. As expected, this trend is more robust for alignments without codon bias (Figure 2A, $r^2 = 0.83$) than for alignments with codon bias (Figure 2B, $r^2 = 0.70$). This difference emerges from the fact that fitness differences among synonymous codons will obscure the underlying amino acid fitness differences.

Importantly, Figure 2A shows that, in the limiting case when σ^2 approaches 0, and thus amino acids have virtually the same fitness values, dN/dS approaches the value of 1. Indeed, the largest dN/dS value for alignments without codon bias is 0.9996. This result properly reflects the case of neutral evolution, in which amino acids . In fact, in **SI proof**, we prove that, when synonymous codons have equal fitness values, dN/dS is necessarily always less than or equal to 1. This restriction does not, however, hold in the face of codon bias, which can readily yield dN/dS values greater than 1 (Figures 1B and 2B), even though the protein sequence is evolving under equilibrium conditions. We discuss the implications of these findings in depth in *Discussion*.

Influence of mechanistic codon model parameterizations

The mechanistic codon dN/dS estimates (ω MLEs) reported in the previous subsection were obtained by fixing the κ parameter in the GY94 rate matrix to its true simulated value and by using the Fequal codon frequency parameterization. However, as different model parameterizations are

known to influence the resulting ω MLE [13,15,16], we inferred ω according to a total of 12 distinct model parameterizations which varied the κ and equilibrium codon frequency specifications. In particular, we examined the effects of fixing κ to its true value, fixing κ to 1, or allowing κ to be a free parameter of the model. We also examined four equilibrium codon frequency specifications: Fequal [13, 15], which assigns an equal frequency of 1/61 to all sense codons, F3x4 [9], CF3x4 [17] and finally F61 [**?**, 8, 15], which uses the empirical codon frequencies.

Table 1 shows how the ω values estimated according to these different ML parameterizations relate to the dN/dS values as calculated using equations (3)–(8), specifically for the simulations without codon bias (Table S1 gives the equivalent results for simulations with codon bias). Results in Table 1 yield several important insights into the behavior of mechanistic codon models. First, these results emphasize the importance of including the κ parameter in mechanistic codon models, as fixing $\kappa = 1$ (effectively removing this parameters) reduces the accuracy of dN/dS estimation. Second, it is clear that these models only estimate dN/dS accurately under the Fequal frequency parameterization. Third, the F3x4 and CF3x4 frequency estimators perform nearly identically to one another, but neither yield yield ω MLEs which correlate to the true dN/dS values. However, when codon bias is present, F3x4 and CF3x4 instead produce ω MLEs which moderately moderate correlate with dN/dS. As F3x4 and CF3x4 yielded comparable performances, we continue our discussion with just F3x4. Fourth, under the F61 parameterization, ω MLEs negatively correlate with dN/dS. In other words, as the codon frequency parameterization were more and more tailored to the given data set, ω MLEs decreased in accuracy. This relationship can be readily seen in Figure 3, which displays regressions between dN/dS and ω MLEs for Fequal, F3x4, and F61 parameterizations; while Fequal leads to perfectly correlating dN/dS estimates, F3x4 and F61 parameterizations frequently produce dramatically elevated ω estimates. Similar regressions for all κ and codon frequency specifications can be found in Figure S1 for alignments without codon bias, and in Figure S2 for alignments with codon bias.

Finally, we examined whether strength of natural selection, as represented by the σ^2 for the distribution of amino acid selection coefficients, influenced the accuracy in ω MLEs. For each dataset (no codon bias and codon bias alignments), we built a mixed-effects linear model using the R package "nlme" [20]. Each model featured the absolute value of the error between dN/dS and ω MLE values as the response, σ^2 and frequency parameterization as fixed effects, and simulation replicate as a random effect. Graphical results for this model are shown in Figure **??** for alignments without codon bias, and in Figure **??** for alignments with codon bias. As F3x4 and CF3x4 performed identically, the figure show only results from the more commonly used F3x4 frequency estimator. For alignments without codon bias (Figure **??**), we detected a significant interaction effect between the codon frequency parameterization and σ^2 ($p < 1^{-10}$), such that the increase in dN/dS error was much more dramatic for F3x4 and F61 results than for Fequal results. Importantly, for all frequency parameterizations, dN/dS error was minimal when σ^2 was near 0, indicating neutral evolution. This finding is also reflected in Figure 3, in which F3x4 and F61 appear to estimate dN/dS much more accurately as dN/dS approaches 1. This result likely stems from the fact that empirical codon frequencies for these alignments were quite similar to Fequal, as fitness differences among codons were minimal. However, as σ^2 increased, the dN/dS error for Fequal remained very low, whereas error for F3x4 and F61 increased dramatically.

Not yet sure where to place this bit - However, it is interesting to note that mechanistic codon models yield similarly accurate κ MLEs under both Fequal and F61 frequency parameterizations, but the F3x4 and CF3x4 estimators infer κ more poorly. Regression plots displaying the relationship between true κ values and κ MLEs are shown in Figures S3 and S4, for simulations without and with codon bias, respectively.

Discussion

The oldest and most-widely used method to infer selection pressure in protein-coding genes calculates calculates the ratio of non-synonymous (dN) to synonymous (dS) substitution rates dN/dS to identify sites that experience negative selection ($dN/dS < 1$), sites that evolve neutrally ($dN/dS \approx 1$), and sites that experience positive diversifying selection ($dN/dS > 1$). By contrast, MutSel models estimate scaled selection coefficients, either for individual amino acids, [**?**, 10, 12, 21, 22], for codons [23], or for both. Thus, while mechanistic codon models describe the how quickly a protein's constituent amino acids change, MutSel models calculate the strength of natural selection operating on the specific amino-acid changes.

Until now, however, it has been an open question how these two modeling frameworks relate to one another. Some have argued that MutSel models, given their firm grounding in population genetics theory and attention to site-specific amino acid fitness differences, offer a more fine-grained approach to studying protein evolution than do mechanistic codon models [10, 21]. Recent phylogenetic studies have also demonstrated that evolutionary models which explicitly consider amino acid fitness values offer dramatic improvements over other models, including mechanistic codon models, suggesting that MutSel models may more aptly represent the process of coding-sequence evolution [24, 25].

Here, we have derived a formal mathematical relationship between the quantities dN/dS and scaled codon selection coefficients, the primary parameters of mechanistic codon and MutSel models, respectively. Through a simulation approach, we find that these two models are in full agreement, and that the value for dN/dS can be precisely calculated from selection coefficients alone. Furthermore, this mathematical equivalency between selection coefficients and dN/dS values is robust to fitness differences among synonymous codons. However, it is important to note that our implementation of codon bias explicitly assumed that frequency differences among synonymous codons resulted from fitness differences alone. In other words, the sole source of codon bias in our simulations was selection, not mutation. This implementation might not be entirely biologically realistic, as both mutational and selective forces likely contribute to codon bias in real genomes [26–29]. However, the key finding that we present is that fitness differences among synonymous codons do not affect the robust mathematical equivalency between scaled selection coefficients and dN/dS.

Our results rest on the key assumptions that the protein sequence is evolving under steady-state, or equilibrium, conditions, and the nucleotide mutation rates are symmetric (e.g. $\mu_{xy} = \mu_{yx}$). The first assumption recapitulates the population genetics theory behind MutSel models, which assume that selection coefficients remain constant over the phylogeny, and therefore the protein is evolving along a static fitness landscape [10,21,22]. While the latter assumption of symmetric mutation rates may not be biologically realistic **cite papers which show asym mu**, it allowed us to investigate the mathematical relationship between dN/dS and selection coefficients under broad conditions. In particular, we have proven that, when synonymous codons have equal fitness and mutation rates are symmetric, dN/dS will always be less than 1. However, when synonymous codons were allowed to have different selection coefficients, dN/dS can easily be greater than 1, and indeed frequently was (Figures 1B and 2B). In fact, when synonymous codons have different fitnesses, it is possible to have arbitrarily high dN/dS values; in the most extreme case of codon bias, in which 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. We additionally expect that an asymmetric mutation rate matrix could yield $dN/dS > 1$.

In other words, even if the protein is evolving along to a static fitness landscape, it is possible to that the sequence will feature dN/dS values greater than 1. This finding seems paradoxical to classic interpretations of $dN/dS > 1$. Typically, these values are viewed as hallmarks of positive

or diversifying selection, which is assumed to occur when the protein sequence experiences strong selective pressure to change its constituent amino acids. Positive selection, therefore, necessarily implies that the protein does not evolve at equilibrium, but rather a shift in selective constraint caused new amino acids to be favored. However, one must also recognize that the contention that $dN/dS > 1$ represents positive selection assumes that synonymous substitutions are selectively neutral, which is likely not the case when codon bias exists. Thus, what is classically termed positive selection can result simply from strong synonymous fitness differences. Therefore, it is entirely possible that estimates of positive selection as inferred via dN/dS estimates in species with high levels of codon bias, such as many bacterial, *Drosophila*, or certain mammalian species [27–30], may not be true cases of positive selection, but rather simply signals of strong codon bias.

Incidentally, our study recovered that mechanistic codon models can produce strongly biased inferences when parameters are incorrectly specified. In particular, ω MLE values only corresponded to the true dN/dS value when the equilibrium codon frequency parameters were specified as equal (e.g. each codon had an equilibrium frequency of 1/61). Alternatively, the common approaches of using F61 (empirical) frequencies [8]) or frequency estimators such as F3x4 [9] and CF3x4 [17] always yielded incorrect and highly elevated ω MLEs. We explain this phenomenon by recognizing that the rationale for including codon frequency parameters in mechanistic codon models is to account for unequal nucleotide frequencies specifically caused by mutational and not selective forces [13,15]. The proper values for these parameters, then, should be the codon frequencies which would exist *in the absence of natural selection*. This approach is the only way to ensure that ω is the sole model parameter which contains information about natural selection. Otherwise, the ω parameter will no longer represent the true dN/dS evolutionary rate ratio.

Moreover, frequency estimators such as F3x4 and CF3x4, which use positional nucleotide frequencies to calculate codon frequencies, must make the implicit assumption that observed unequal base frequencies result from biased mutation rates. This assumption, however, may not be fully justified. Indeed, our simulated alignments featured a wide array of nucleotide compositions, with GC-contents ranging from 0.22-0.79. Given that we simulated sequences according to a symmetric mutation matrix, all compositional biases in our data sets resulted entirely from natural selection favoring particular codons, not by any bias towards unequal base frequencies. Therefore, the proper equilibrium frequency parameterization for our alignments was indeed equal codon frequencies, which would be expected in the absence of natural selection and when mutation rates are symmetric.

These results emphasize that it is crucial to parameterize mechanistic codon models properly. Indeed, their primary parameter ω will only truly represent dN/dS when all other parameters are properly specified. If codon frequencies are not properly specified, which we suspect is the case in most analyses, then the ω MLEs are virtually meaningless and do not represent selective pressure. Therefore, we contend that there is hardly ever a justification to specify empirical codon frequencies, also known as the F61 frequency estimator [8], as natural selection has clearly produced the observed frequencies. Unfortunately, the F61 frequencies are the default parameterization in the widely-used PAML software's codeml implementation [31], so we strongly recommend that users take great care when using this package. In addition, the only robust way to ensure that codon frequencies are properly specified is through experimentally calculating mutation rates. Luckily, this data already exists for a variety of taxa, including **citations for papers uncovering mutation rates**. We recommend that, if experimental data is absent, users err on the side of caution and specify equal codon frequencies to reduce the possibility of false positives.

Finally, we contend that this methods presented in this paper reveal 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. However,

this strategy can only confirm that inference methods are behaving as expected; it cannot confirm that the underlying model accurately represents the evolutionary process. Instead, we suggest an alternate approach to benchmark inference methods, and indeed evolutionary models: assessing the extent to which distinct models agree may serve as a novel, robust strategy to determine the accuracy of different modeling frameworks.

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 SD (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] Anisimova M, Kosiol C (2009) Investigating protein-coding sequence evolution with probabilistic codon substitution models. Mol Biol Evol 26: 255–271.

[8] Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11: 725–736.

[9] 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.

[10] Halpern AL, Bruno WJ (1998) Evolutionary distances for protein-coding sequences: modeling site-specific residue frequencies. Mol Biol Evol 15: 910–917.

[11] Rodrigue N, Lartillot N (2014) Site-heterogeneous mutation-selection models within the PhyloBayes-MPI package. Bioinformatics : 1020–1021.

[12] 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.

[13] Yang Z (2006) Computational Molecular Evolution. Oxford University Press.

[14] Kosakovsky Pond SL, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenetics. Bioinformatics 21: 676–679.

[15] Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17: 32–42.

[16] Zhang Z, Yu J (2006) Evaluation of six methods for estimating synonymous and nonsynonymous substitution rates. Geno Prot Bioinfo 4: 173–181.

[17] 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.

[18] Sella G, Hirsh AE (2005) The application of statistical physics to evolutionary biology. Proc Natl Acad Sci USA 102: 9541–9546.

[19] Crow JF, Kimura M (1970) An Introduction to Population Genetics Theory. California: Burgess Pub. Co.

[20] Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2014) nlme: Linear and Nonlinear Mixed Effects Models. URL `http://CRAN.R-project.org/package=nlme`. R package version 3.1-117.

[21] 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.

[22] 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.

[23] 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.

[24] Bloom JD (2014) An experimentally determined evolutionary model dramatically improves phylogenetic fit. Mol Biol Evol : To appear.

[25] Bloom JD (2014) An experimentally informed evolutionary model improves phylogenetic fit to divergent lactamase homologs. Mol Biol Evol 31: 1956-1978.

[26] Blumer M (1991) The selection-mutation-drift theory of synonymous codon usage 129: 897–907.

[27] Duret L (2002) Evolution of synonymous codon usage in metazoans. Curr Opin Genet Dev 12: 640–649.

[28] Hershberg R, Petrov D (2008) Selection on codon bias. Annu Rev Genet 42.

[29] Plotkin JB, Kudla G (2011) Synonymous but not the same: the causes and consequences of codon bias. Nature Rev Genet 12: 32–42.

[30] Chamary JV, Parmley JL, Hurst LD (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. Nature Rev Genet 7: 98–108.

[31] Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24: 1586–1591.

Table 1: Effect of ML parameterizations on inference.

Codon frequencies	κ parameterization	dN/dS correlation	dN/dS error	κ correlation	κ error
Fequal	True	1	0.015		
Fequal	Free	0.998	0.042	0.829	0.151
Fequal	1	0.986	0.198		
F3x4	True	0.019	8.277		
F3x4	Free	-0.013	8.383	0.477	0.269
F3x4	1	0.002	5.825		
CF3x4	True	0.015	8.252		
CF3x4	Free	-0.015	8.352	0.476	0.276
CF3x4	1	-0.008	5.782		
F61	True	-0.522	56.495		
F61	Free	-0.557	46.346	0.828	0.231
F61	1	-0.522	38.668		

Codon frequency specifications were either set as equal (1/61 per codon), calculated from the F3x4 estimator [9], calculated from the CF3x4 estimator [17], or set equal to the simulated alignment's empirical frequencies. κ was specified as either a fixed value, its true simulated value or 1, or as a free parameter of the model. Correlations given are between the ML ω estimate and our derived ω values. Error refers to the mean absolute error between these two ω estimates. Similar values for κ are shown for those inferences where κ was a free parameter of the model. Note that all is significant.

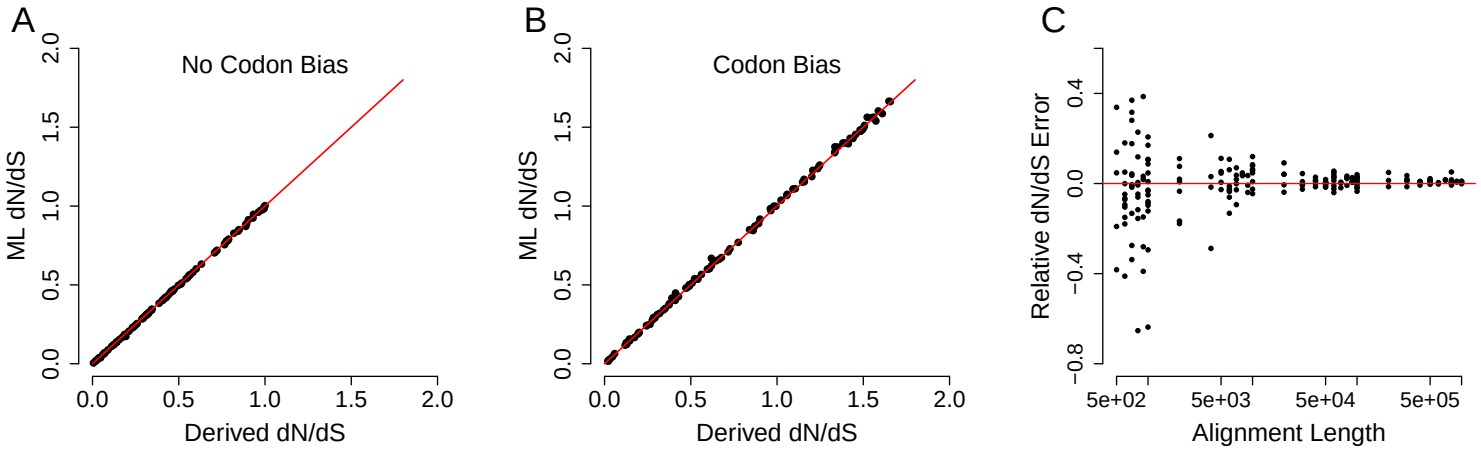

Figure 1: Relationship works exceedingly well. Left panel shows 100 points, each of which corresponds to single simulation. Note that here the ml inference is shown for equal codon frequency specs and kappa fixed to true value (a similar plot for free kappa is shown in suppfigs, but results are qualitatively identical.) Right panels shows convergence of omega values as data set size (represented as simulated alignment length) increases. The y-axis indicates relative error of the ML dN/dS estimates, and the x-axis indicates sequence length on a log-scale. As the sequence length, or the data set size, increases, the two dN/dS estimates converge to the same value.

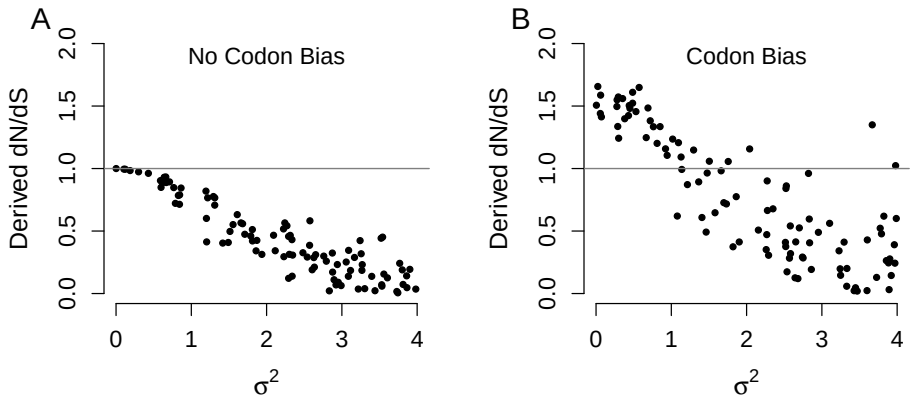

Figure 2: ssc's (selection strength) scales well with dnds but the strength of the relationship diminishes with codon bias as synonymous now have fitness differences, so dnds is less of a reliable indicator of selection strength.

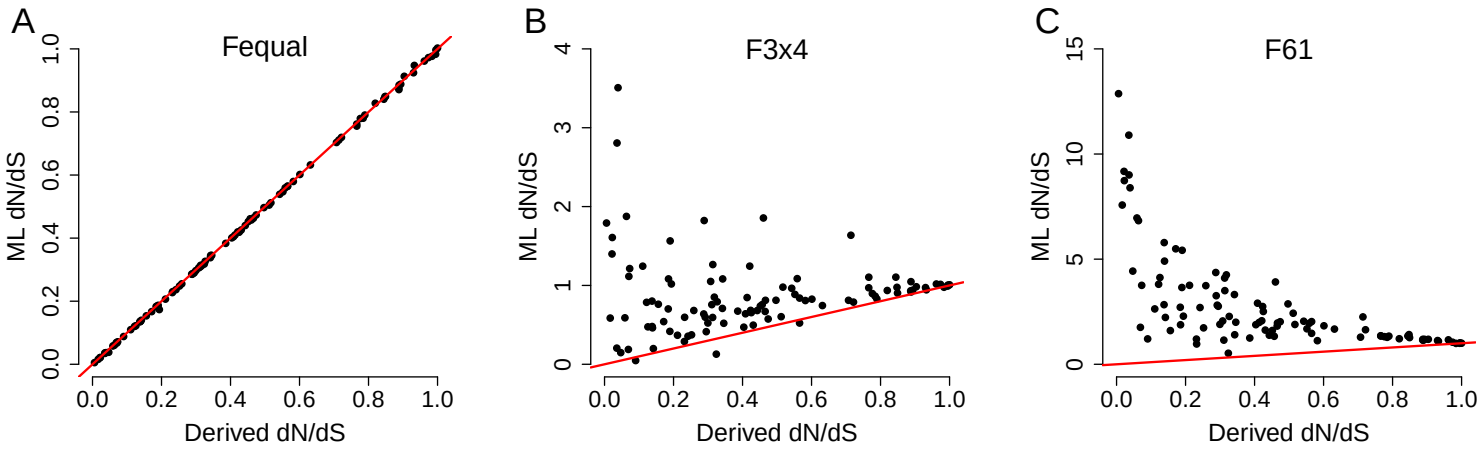

Figure 3: Issues with frequency specifications abound. In each plot, red line indicates 1:1 agreement, so note the y-axis differences. Relationship between omega values only really exists when equal codon frequencies are specified. When f3x4 or true freqs used, there is the potential to end up with dramatically inflated values. cf3x4 not shown because its results are statistically the same as f3x4.

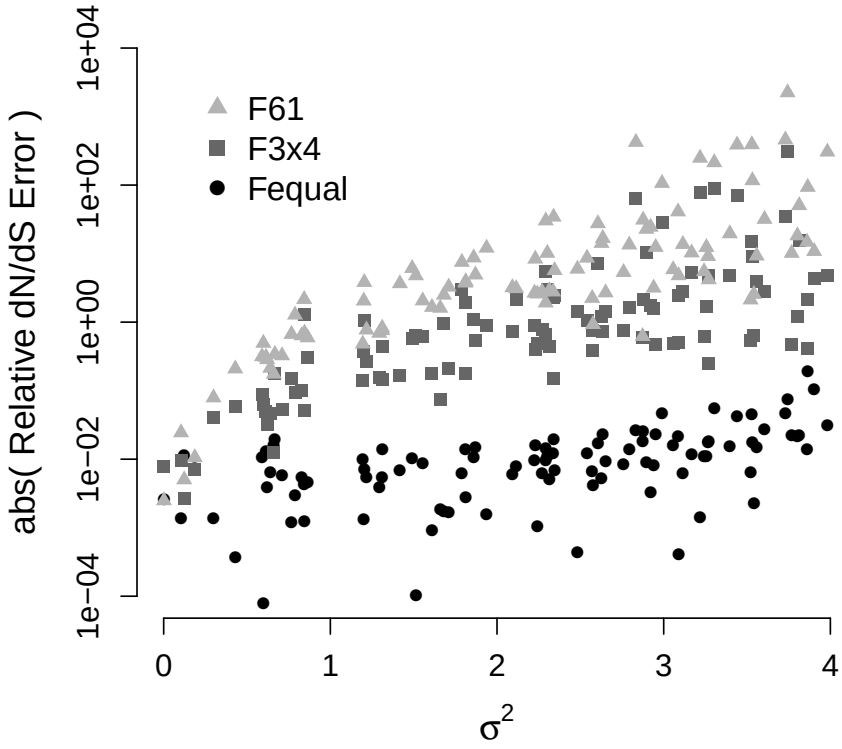

Figure 4: Interaction detected in no codon bias but none in the codon bias dataset. Xaxis is stddev, Yaxis is logscale absolute value of relative error.

Supplementary Figures and Tables

Codon frequencies	κ parameterization	dN/dS correlation	dN/dS error	κ correlation	κ error
Fequal	True	1	0.02		
Fequal	Free	0.971	0.105	0.699	0.262
Fequal	1	0.974	0.263		
F3x4	True	0.646	2.134		
F3x4	Free	0.552	2.342	0.212	0.774
F3x4	1	0.597	1.847		
CF3x4	True	0.642	2.105		
CF3x4	Free	0.553	2.316	0.212	0.765
CF3x4	1	0.581	1.845		
F61	True	-0.591	19.067		
F61	Free	-0.281	27.714	0.315	0.927
F61	1	-0.56	12.272		

Results from runs with codon bias. Codon frequency specifications were either set as equal (1/61 per codon), calculated from the F3x4 estimator [9], calculated from the CF3x4 estimator [17], or set equal to the simulated alignment's empirical frequencies. κ was specified as either a fixed value, its true simulated value or 1, or as a free parameter of the model. Correlations given are between the ML ω estimate and our derived ω values. Error refers to the mean absolute error between these two ω estimates. Similar values for κ are shown for those inferences where κ was a free parameter of the model. Note that all is significant.

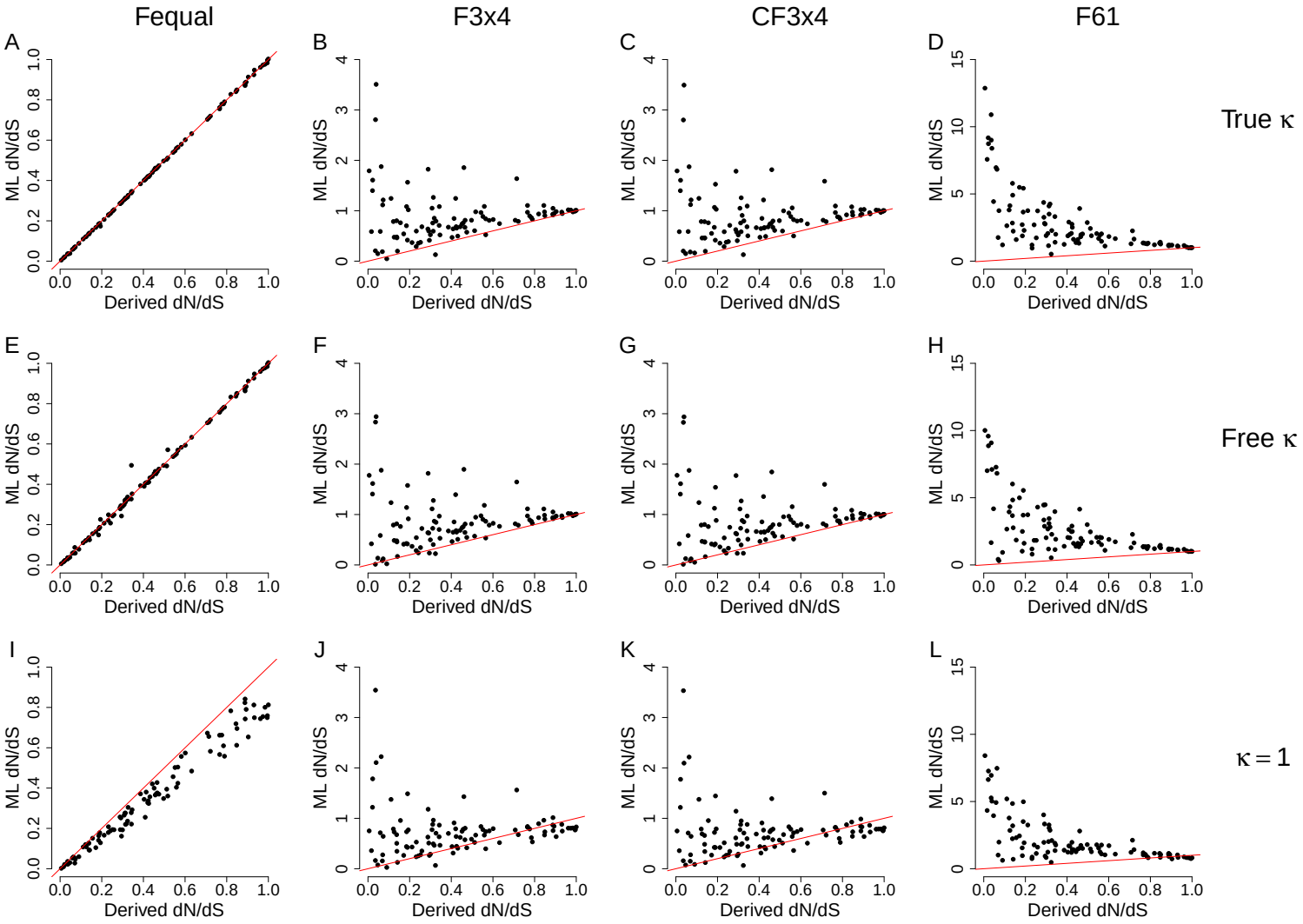

Fig. S1 Omega regression for all ML parameterizations, without codon bias.

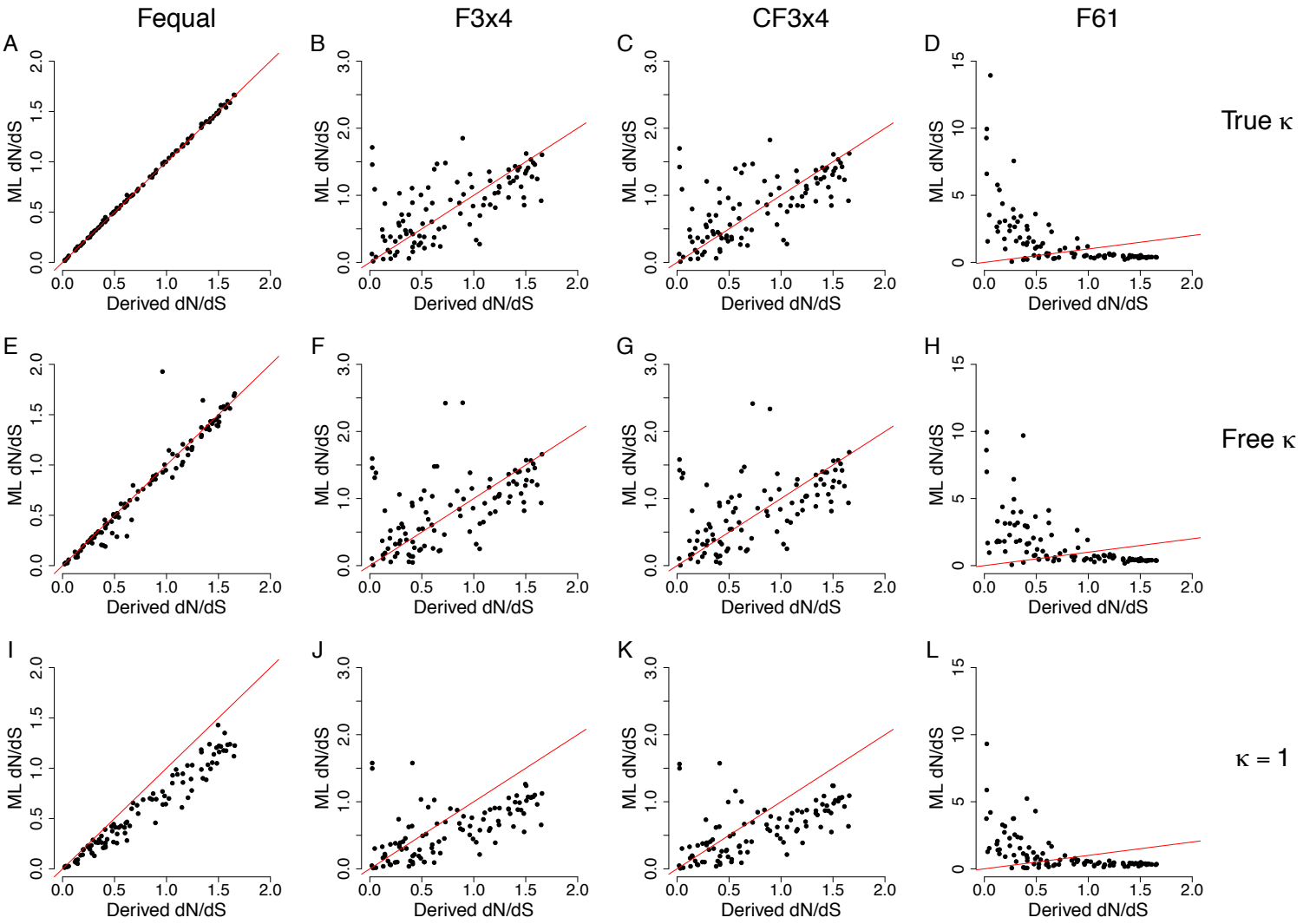

Fig. S2 Omega regression for all ML parameterizations, with codon bias.

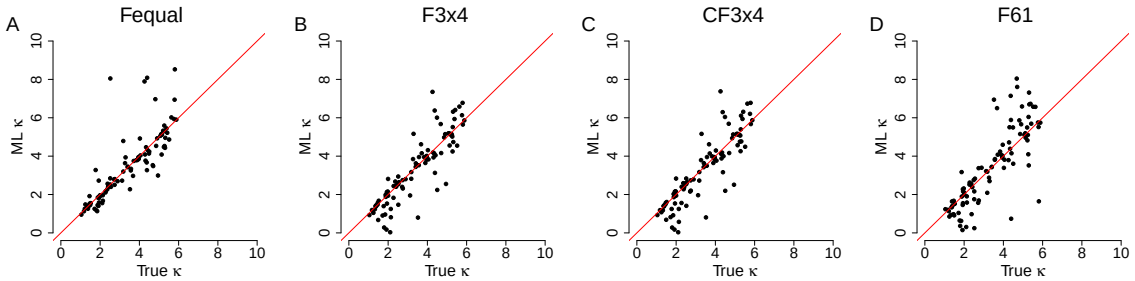

Fig. S3 Kappa regression for all ML freqspec parameterizations where kappa is a free parameter, without codon bias.

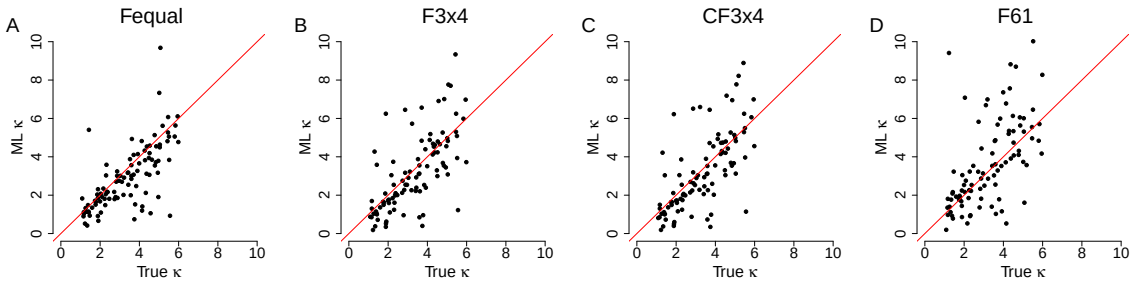

Fig. S4 Kappa regression for all ML freqspec parameterizations where kappa is a free parameter, without codon bias.

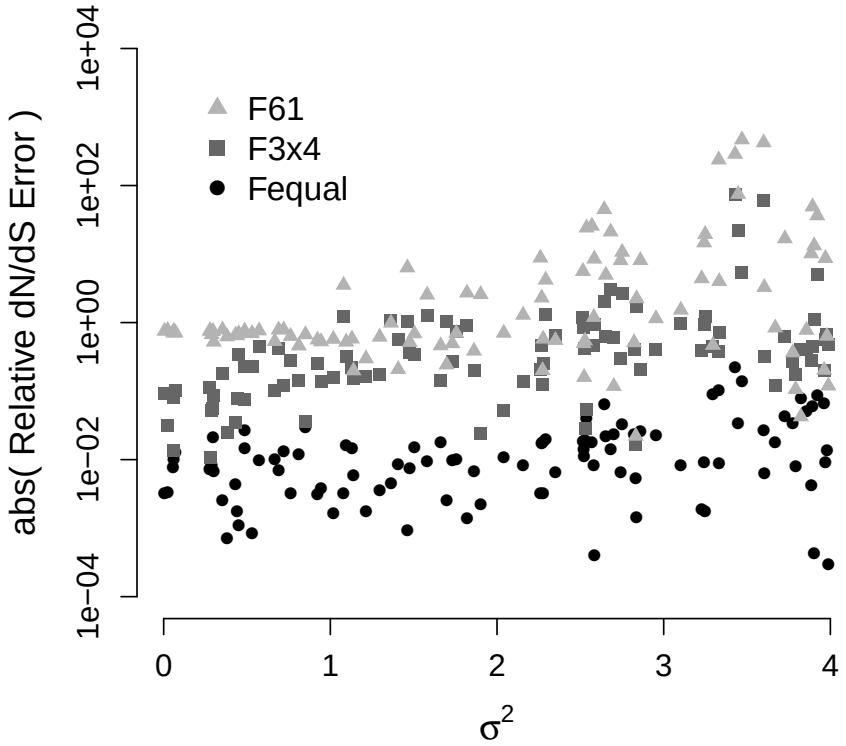

Fig. S5 codon bias fspec, sel strength, dnds error.