Physicochemical Amino Acid Properties Better Describe Substitution Rates in Large Populations

Claudia C. Weber*[,1,2] and Simon Whelan[3]

[1]Center for Computational Genetics and Genomics, Department of Biology, Temple University, Philadelphia, PA

[2]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, United Kingdom

[3]Evolutionary Biology Center, Uppsala University, Uppsala, Sweden

*Corresponding author: E-mail: cweber@ebi.ac.uk.

Associate editor: Claus Wilke

Abstract

Substitutions between chemically distant amino acids are known to occur less frequently than those between more similar amino acids. This knowledge, however, is not reflected in most codon substitution models, which treat all nonsynonymous changes as if they were equivalent in terms of impact on the protein. A variety of methods for integrating chemical distances into models have been proposed, with a common approach being to divide substitutions into radical or conservative categories. Nevertheless, it remains unclear whether the resulting models describe sequence evolution better than their simpler counterparts.

We propose a parametric codon model that distinguishes between radical and conservative substitutions, allowing us to assess if radical substitutions are preferentially removed by selection. Applying our new model to a range of phylogenomic data, we find differentiating between radical and conservative substitutions provides significantly better fit for large populations, but see no equivalent improvement for smaller populations. Comparing codon and amino acid models using these same data shows that alignments from large populations tend to select phylogenetic models containing information about amino acid exchangeabilities, whereas the structure of the genetic code is more important for smaller populations.

Our results suggest selection against radical substitutions is, on average, more pronounced in large populations than smaller ones. The reduced observable effect of selection in smaller populations may be due to stronger genetic drift making it more challenging to detect preferences. Our results imply an important connection between the life history of a phylogenetic group and the model that best describes its evolution.

Key words: **substitution models, protein evolution, phylogenomics, effective population size, mutation–selection model.**

Introduction

Quantifying the impact of natural selection on proteins is of broad interest in evolutionary biology, providing insight into the structural and functional constraints acting on proteins and how they adapt to an organism's environment. The most widely used method for studying selection using multiple sequence alignments of protein-coding sequences is to consider the ratio of the nonsynonymous substitution rate (d*N*) to the synonymous substitution rate (d*S*), often referred to as $\omega = dN/dS$. These codon-based models assume that d*S* reflects the neutral rate of evolution and d*N* represents the rate after selection has acted. The ω measure is well characterized, and established methods allow it to vary along a sequence, through an evolutionary tree, or a combination of both (Goldman and Yang 1994; Muse and Gaut 1994; Yang 1998; Yang and Nielsen 1998, 2002). Although useful, ω is coarse grained because d*N* takes on the same value regardless of the amino acids being substituted despite widespread evidence that some amino acid substitutions are more likely to occur than others.

Early research showed that amino acid substitutions between amino acids with very different physicochemical properties occur less frequently than substitutions between more similar amino acids (Epstein 1967), leading to the introduction of a range of scales measuring physicochemical distances (Grantham 1974; Miyata et al. 1979). The original codon models used a d*N*/d*S* measure that incorporated these distances (Goldman and Yang 1994), based on the rationale that selection against more similar amino acid substitutions ought to be weaker than against more distant ones. Subsequent research, however, found that this model frequently provided a poorer fit than the simpler M0 model, which estimates d*N*/d*S* but does not capture differences in the selective pressures acting on different amino acid substitutions (Yang et al. 1998).

Other studies proposed classifying amino acid substitutions into conservative and radical substitutions, describing substitutions between similar and dissimilar amino acids, respectively. The relative amount of radical and conservative change can be described by the ratio R/C, which is interpreted as a measure of selective pressure acting on the two different

classes of nonsynonymous mutations. Several approaches have been proposed to estimate R/C by counting amino acid changes (Zhang 2000; Smith 2003; Popadin et al. 2007; Weber et al. 2014), and results suggest that radical mutations are more likely to be selected against than conservative mutations (Smith 2003). Further, organisms with a small effective population size tend to accumulate more radical substitutions than those with a larger effective population size and more efficient natural selection (Eyre-Walker et al. 2002; Smith 2003; Popadin et al. 2007), although see Figuet et al. (2016). However, count-based methods may be unreliable in the presence of mutation and composition biases (Dagan et al. 2002; Smith 2003), given their implicit incorporation of parsimony-type scoring. Attempts to incorporate R/C into a more statistically justified and robust codon substitution model have so far been limited to surveys of positive selection in genes already known to contain sites with accelerated nonsynonymous substitution rates (Sainudiin et al. 2005). Therefore, little is known about whether a model measuring R/C provides a significant improvement in model fit over M0, and to what extent statistical fit supports the notion that negative selection treats radical substitutions differently than conservative ones.

An alternative approach to describe variation in the selective pressures acting on amino acids is to model them at the amino acid—as opposed to the codon level. Here, "exchangeability" parameters, estimated from databases of sequences, capture the rate of change between pairs of amino acids and reflect a hazy combination of the genetic code and the selective pressures acting on the substitution. These models range from simple approaches that apply an averaged substitution process across the entire protein to sophisticated mixture models that have different substitution patterns for different sites (Jones et al. 1994; Whelan and Goldman 2001; Le and Gascuel 2008). They have been successfully used for a range of phylogenetic problems, with the mixture models proving valuable at resolving deep phylogenies (Lartillot et al. 2007). Modeling approaches that link the biophysical attributes of amino acid substitution models to codon models have also been developed (Yang et al. 1998; Seo and Kishino 2009) but are currently not widely used.

Recently, statistical model selection methods for choosing between codon and amino acid substitution models have been devised, allowing the state-space of the model that best describes a given sequence alignment to be determined (Seo and Kishino 2009; Whelan et al. 2015). Analyses of collections of sequence alignments show that preference for nucleotide, amino acid, or codon models is data set dependent. Moreover, the best-fitting model class is correlated with the intensity of selection as estimated by dN/dS, with more constrained alignments tending to select amino acid models and less constrained alignments selecting codon models. These results imply that the factors driving substitution rates between amino acids, and therefore differential model selection, could be driven by the selective pressures acting on biophysical properties, the structure of the genetic code, or a combination of the two (Whelan et al. 2015).

In light of the availability of vast quantities of sequence data spanning an enormous range of taxa, we return to the question of whether chemical distances predict selective preferences in protein-coding sequences. Previous analyses were restricted to a handful of mammalian sequences and therefore may not present a complete picture. To examine the relative selective pressures acting on different types of amino acid substitutions, we propose a codon model that separates nonsynonymous substitutions into conservative or radical (CoRa) categories. Our approach allows R/C to be robustly estimated in the presence of mutational and compositional bias and long branches with multiple hits. Using a mutation–selection model, we then explore how R/C will respond to variation in evolutionary variables, including effective population size (N_e), assuming that radical substitutions have larger negative effects on fitness than do conservative substitutions. We next assess whether the R/C ratios observed in empirical sequence data are consistent with radical substitutions being more disruptive and selectively unfavorable and find that this is the case for a subset of alignments overwhelmingly belonging to taxa with large effective population sizes and strong selection. Moreover, we see a preference for amino acid models over standard codon models in the same taxa, while small population size predicts a preference for codon models. Taken together, our results confirm that chemical amino acid distance is a predictor of the strength of selection and that organismal life history influences model selection.

Results

The CoRa Substitution Model

The base model for describing codon substitutions is the M0 model (Goldman and Yang 1994), which captures the relative selective pressures acting on nonsynonymous substitutions through the $\omega = dN/dS$ parameter. Here, we propose the CoRa model, a generalization of M0 that separates ω into two parameters ω_C and ω_R, which capture conservative and radical nonsynonymous substitutions, respectively (see Materials and Methods).

The CoRa model adds one additional degree of freedom compared with M0, allowing a significance test of the relative fit of the two models by comparing the standard likelihood ratio test statistic to a χ^2_1 distribution. In this case, M0 is the simpler model and is nested within CoRa under the condition $\omega_C = \omega_R$, which places conservative and radical mutations under the same amount of selective pressure. This restriction is absent in CoRa, so a measure of R/C describes the relative probability of fixing a radical versus a conservative mutation.

Factors Affecting R/C Estimates

In the Materials and Methods section, we describe two broad types of method for inferring R/C. The first relies on the formulation of the CoRa model and directly estimates the relative instantaneous substitution rates of $dR/dS = \omega_R$ and $dC/dS = \omega_C$. Here, dR and dC refer to radical and conservative nonsynonymous substitutions, respectively. The second uses observed sequences to count the relative frequency of radical to conservative substitutions, K_r/K_c, either 1) by using

a parsimony approach to count the observable differences between sequences to obtain $\frac{K_r}{K_c}(t)$, which misses cases where more than one substitution has occurred at a site or 2) through stochastic mapping (Minin and Suchard 2008) to obtain $\frac{K_r}{K_c}(\mathbf{Q})$, which can count all the substitutions that occurred. Using an analytic approach (see Materials and Methods), we examine how evolutionary forces can affect these different R/C measures, demonstrating the robustness of the $dR/dC = \omega_R/\omega_C$ measure and the difficulties of working with K_r/K_c measures. All of the results presented here depend both on the classification of radical and conservative amino acid substitutions used by the CoRa model, and the structure of the genetic code, which determines the relative rates of different nonsynonymous and synonymous substitutions.

The Effects of Genomic Variation on R/C Measures

We start by assessing how K_r/K_c measures are affected by variation in the parameters of the M0 substitution model, which has equal rates of substitution between conservative and radical amino acids. Figure 1*a–c* shows the effect of variation in ω, GC-content and transition–transversion (κ) bias on the different K_r/K_c measures. The first observation for all figures is that K_r/K_c is not equal to one even when there is no difference between conservative and radical substitution rates. The deviation from one matches previous observations (Eyre-Walker et al. 2002; Weber et al. 2014) and arises because radical substitutions are more likely to occur by chance due to the structure of the genetic code and the relative frequencies of the codons. Inferring the relative strength of selection acting on radical and conservative substitutions is therefore difficult when only raw K_r/K_c is considered, although changes in K_r/K_c might still be informative of a shift in those values (see below). This is not a problem for the dR/dC estimate since it is inferred directly from the substitution model, which accounts for the genetic code.

In figure 1a, we see that the ratio $\frac{K_r}{K_c}(\mathbf{Q})$ (solid black line) is not affected by changes in ω. This stability is expected since under the M0 model $\omega_R = \omega_C = \omega$. Both ω_R and ω_C scale in direct proportion to ω, so one may expect that changes to ω have little effect on the K_r/K_c measure. The alternative $\frac{K_r}{K_c}(t)$ measure counts only the observable changes occurring on a branch of length t and is affected by variation in ω because its values are affected by multiple hits. For short distances where few substitutions are missed $\frac{K_r}{K_c}(t)$ closely resembles $\frac{K_r}{K_c}(\mathbf{Q})$. However, when t becomes larger there is an increasingly strong response in $\frac{K_r}{K_c}(t)$ to the value of ω that can be attributed to the undercounting of more common nonsynonymous substitutions, which becomes more severe as ω increases. Hence, $\frac{K_r}{K_c}(t)$ is prone to inflation for long branches. Meanwhile, variation in ω has no effect on the $dR/dC = 1$ estimate (not shown).

The effect of %GC-content and transition–transversion rate on R/C is shown in figure 1b and c, respectively. For simplicity, the effect of %GC is assessed by changing the GC-content of each codon position (GC1, GC2, and GC3) in an identical manner to obtain a given %GC value. This approach is simpler than what occurs in nature, where GC3 is more variable than GC1 or GC2 but provides an adequate insight into the overall effect of %GC variation. Given that the CoRa model accounts explicitly for variation in %GC and κ, estimates of dR/dC are unaffected. In contrast, %GC variation has a substantial and nonmonotonic effect on both $\frac{K_r}{K_c}(\mathbf{Q})$ and $\frac{K_r}{K_c}(t)$. The effect of κ is comparatively smaller, but there is still a negative nonlinear correlation between the value of κ and $\frac{K_r}{K_c}(\mathbf{Q})$ and $\frac{K_r}{K_c}(t)$. The length of time t tends to inflate $\frac{K_r}{K_c}(t)$ estimates in a manner dependent on both GC and κ, with the magnitude of the increase reflecting the specific %GC or κ value examined.

Finally, figure 1d shows how the different K_r/K_c measures respond to variation in the intended target variable dR/dC. There is a clear linear relationship between dR/dC and K_r/K_c measures, with the slope of the line being determined by a complex function of the parameters in the M0 model under the conditions considered in the preceding panels, the genetic code, and $\mathbf{M}_{rad}$. This observation is encouraging since it suggests that although K_r/K_c might not be informative about the absolute selective pressures acting on conservative versus radical substitutions, relative changes in K_r/K_c might correspond to clear increases or decreases in those selective pressures. When comparing genomic sequences with the same %GC and κ, this relationship applies, but it is unlikely to hold for general comparisons where the values of these parameters vary.

To provide an idea of how compositional variation might affect K_r/K_c, we examine two sets of M0 parameters: 1) %GC $= 0.5$ and $\kappa = 2.5$ where the slope of the line is 1.77 and 2) %GC $= 0.65$ $\kappa = 2.5$ where the slope of the line increases to 1.97 (not shown in figure). The differing slope makes comparisons between regions with different properties problematic since it is difficult to disentangle changes in K_r/K_c resulting from changes in selection and those resulting from differences in sequence properties. From these results, we conclude that both K_r/K_c measures are flawed and difficult to interpret, and dR/dC hence provides the most natural and accurate estimates of the relative selective pressures acting on conservative and radical changes in proteins. We, therefore, consider only dR/dC in the following sections.

Effective Population Size

The relative rates of radical to conservative substitution have been suggested to be associated with or even be suitable to predict N_e (Nabholz et al. 2013). To explore how variation in N_e relates to changes in the rates of nonsynonymous substitution, we consider a simple mutation–selection model (see Materials and Methods).

Figure 2a shows how ω responds to variation in N_e for various levels of negative selective pressure, s, where s_c and s_r represent selection on conservative and radical changes, respectively. The solid line represents the case where $s = s_c = s_r = -0.0020$ and demonstrates how ω changes from $\sim$1.0 when N_e is small to $\sim$0.0 as N_e becomes very large. Under the log transform of N_e, the different values of s transpose this curve on the x axis, with the thin lines showing

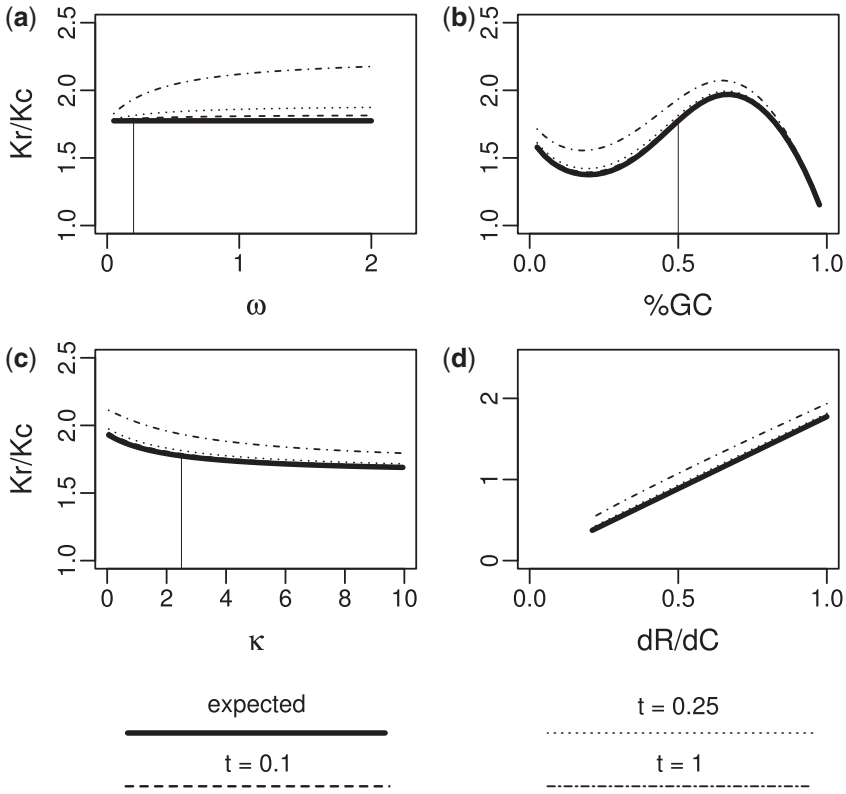

Fig. 1. Factors affecting estimates of K_r/K_c obtained through counting (as opposed to dR/dC, which is obtained from estimated instantaneous substitution rates). Panels (a)-(c) show the impact of ω, %GC and κ, while (d) shows how Kr/Kc responds to dR/dC. Solid lines represent the expected value of K_r/K_c when every substitution can be observed, whereas the thin dashed lines represent expected values of K_r/K_c observable between a pair of sequences after differing amounts of time: wide-dash for $t = 0.1$; dotted for $t = 0.25$; and dot-dash for $t = 1.0$. In all panels but (*d*), $\omega_R = \omega_C = \omega$. The vertical lines indicate the values of ω, %GC and κ held constant across the other panels.

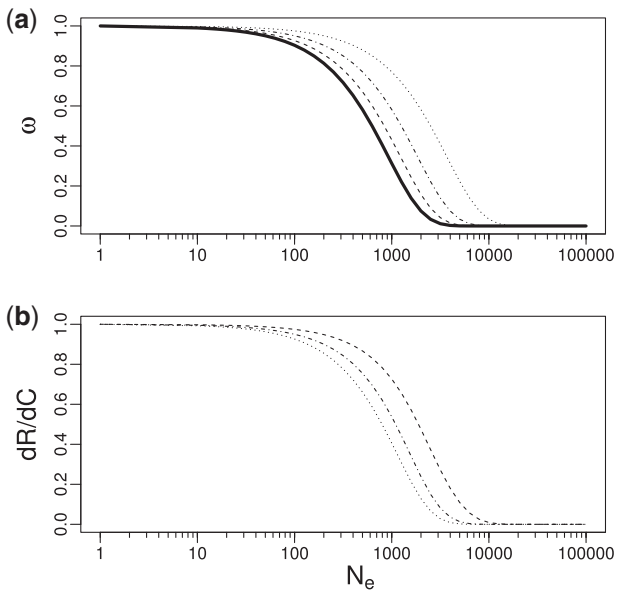

Fig. 2. The effect of N_e on (*a*) ω estimates, and (*b*) dR/dC estimates under the mutation–selection model. See text for more details. Note that the selective coefficients represented here are arbitrarily chosen examples intended for illustration.

$s_c = \{-0.0015, -0.0010, -0.0005\}$ from left to right. Note that this leads to a diminishing effect of N_e on ω as purifying selection weakens. For example, to obtain a decrease in ω from 0.75 to 0.25 under $s = -0.0020$ (thick line) requires an increase in N_e of around 900, whereas the same reduction in ω for $s = -0.0010$ (dash-dot line) requires an increase in N_e of around 1,800.

When considering R/C we are interested in the ratio of differential selective pressures affecting different types of non-synonymous change, with the assumption that conservative substitutions are more likely to be fixed than radical substitutions, due to stronger selection acting against the latter (selective coefficients: $s_c, s_r < 0$; $s_c > s_r$; i.e., s_c is less negative than s_r). Figure 2b shows how dR/dC responds to variation in N_e according to equation (9) (see Materials and Methods), with each thin dotted line corresponding to different ratios of $s_r/s_c = 4$ (dotted; left), 2 (dot-dash; mid), and 4/3 (dashed; right). These selective coefficients will produce a range of observed dR/dC values depending on N_e. For sufficiently small populations $dR/dC \to 1$ because selection is not strong enough to differentiate between radical and conservative mutations. Meanwhile, for sufficiently large populations $dR/dC \to 0$ as radical substitutions are far less likely to fix than conservative substitutions. (Note that as N_e gets large enough neither radical nor conservative substitutions are likely, but if one did occur it would likely be conservative under this model.)

We find that different ratios of selective coefficients for conservative and radical mutations do indeed respond differently to variation in N_e, so for a fixed population size

comparisons of d*R*/d*C* are informative of the selective pressures acting during evolution. For instance when comparing two genes from the same set of taxa, a lower value of d*R*/d*C* in one gene might be directly indicative of relatively stronger selection acting on radical mutations for that gene. Figure 2b demonstrates that these differences are difficult to interpret directly (note the nonlinear spacing between d*R*/d*C* values representing different ratios of selective coefficients for a given N_e). Nevertheless, relative rankings would be informative.

Genomic Analyses of Amino Acid Substitution Patterns

The previous section outlines how we expect d*R*/d*C* to respond to different values of N_e and the range of values it might take. Next, we investigate whether results from real data match these expectations by examining a range of phylogenomic data sets. It is not possible to obtain accurate estimates of N_e directly, so we separate our data a priori into two broad categories of high and low N_e based on other authors' observations. Mammals, birds, and vertebrates are assumed to have relatively low N_e relative to yeast, arthropods, and insects, which are assumed to have relatively high N_e.

Effective Population Size Predicts Relative Rates of Conservative and Radical Substitutions

Comparisons between the CoRa and M0 models provide a direct measure of statistical support for the notion that radical and conservative substitutions are subject to different selective pressures. Figure 3 shows the distributions of the likelihood ratio test statistic, $2\delta = 2[\ln(X;\hat{\theta}_{\text{CoRa}}) - \ln(X;\hat{\theta}_{\text{M0}})]$, for our low and high N_e genomic data sets. Here, $\hat{\theta}$ refers to the parameter estimates from the model and X to the data. We find that the distributions for 2δ differ markedly between the two sets, with the majority of high N_e alignments falling above the 95% significance cutoff ($\chi^2_1 = 3.84$), whereas a large fraction of the low N_e alignments are below the threshold. That is, CoRa provides a better fit than M0 for groups with high N_e, demonstrating significant support for $\omega_R \neq \omega_C$ but often gives no significant improvement in fit for low N_e, where we find little support for differences between the rates of radical and conservative substitutions.

There are two possible explanations for this difference in model fit. First, there might be no significant difference in the selective constraints acting on conservative (s_c) and radical (s_r) substitutions in the vertebrate data we examined. This explanation is implausible since there is no reason to believe that the set of proteins and their functions and levels of constraint differ between the high and low N_e sets, as both categories include genome-scale data and are not restricted to curated (potentially biased) phylogenetic markers. Second, the rates of fixation of conservative and radical substitutions in low N_e populations might be similar for much of the proteome resulting in d*R*/d*C* values close to 1, whereas high N_e populations have more noticeable differences in d*R* and d*C*. This second explanation is consistent with predictions from

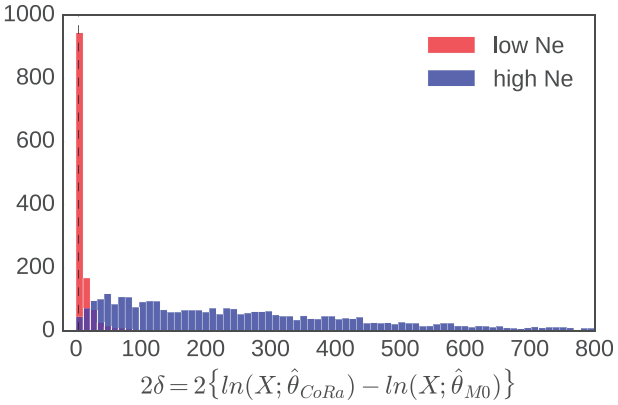

Fig. 3. The histogram shows the distribution of the 2δ statistic for CoRa compared with M0 for alignments from groups with small populations (red) compared with large populations (blue). The dashed vertical line indicates the threshold above which the statistic is significant (3.84). Significant likelihood ratio tests indicate a preference for CoRa over M0. CoRa describes alignments from large populations significantly better than M0. These results are not explained by alignment length or tree depth (see supplementary figs. S1 and S3, Supplementary Material online).

figure 2, with the ratio of selective coefficients describing the sequences resulting in a curve that is shifted to the right relative to the examples shown. The fact that the whole distribution is shifted also confirms that our observations are not driven by any individual alignment.

To further explore this hypothesis, we examine figure 4, which plots $\hat{\omega}_R$ from the CoRa model against $\hat{\omega}_C$. For the high N_e data sets, we find the dashed blue lowess line falls below $x = y$ suggesting $\hat{\omega}_R$ is systematically lower than $\hat{\omega}_C$. Combined with the likelihood ratio test results, this supports the hypothesis that radical substitutions are, on average, subject to stronger negative selective pressure, consistent with the expectation that they are more often deleterious. We also observe that the high N_e data sets tend to have lower estimates for both ω_C and ω_R. In contrast, for the low N_e genomic data, the values of $\hat{\omega}_R$ and $\hat{\omega}_C$ tend to be much closer to each other, with the lowess line closer to $x = y$, demonstrating that selection is less efficient at differentiating between radical and conservative substitutions. These observations are all consistent with our theoretical expectations and with N_e leading to differences in the inferred patterns of amino acid substitution.

Effective Population Size Drives Model State-Space Selection

An alternative approach to incorporating the selective pressures acting on amino acid changes with different effects on biophysical properties is to describe the substitution process at the amino acid sequence level. Performing phylogenetic model selection across state-spaces using the approach of Whelan et al. (2015) allows us to directly compare M0-based codon models, which do not account for variation in selection acting on amino acid substitutions, to amino acid models, which include empirically derived exchangeability parameters capturing variation in rates of substitution

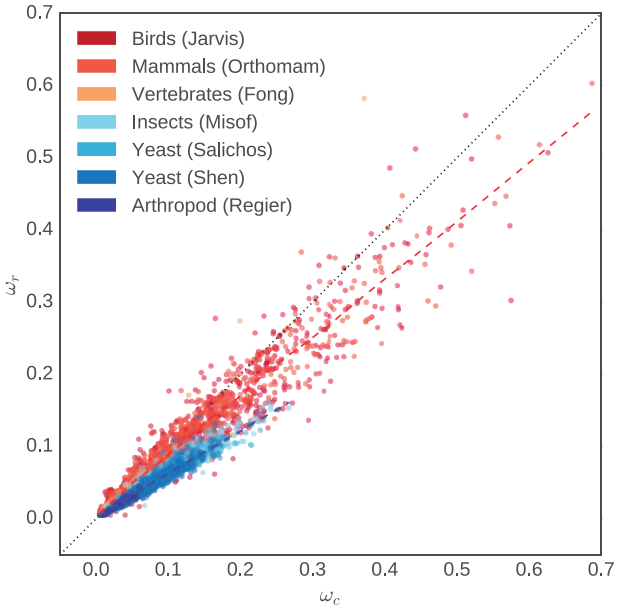

FIG. 4. Phylogenomic data sets from taxa with different population sizes differ in terms of the relationship between $\hat{\omega}_C$ and $\hat{\omega}_R$. The blue and red points represent low N_e and high N_e phylogenomic data sets, respectively. Each point represents an individual gene and its shade indicates the phylogenomic data set it is drawn from. Although $\hat{\omega}_C > \hat{\omega}_R$ for the majority of sequences from large N_e groups (insects, arthropods, and yeast), vertebrate sequences show a mixed pattern and overall higher rates of radical substitution. The dotted black line indicates $x = y$, that is, $\hat{\omega}_C = \hat{\omega}_R$. The blue lowess line represents putatively high N_e groups, whereas red represents low N_e groups.

between amino acids. Following our observation that N_e affects radical and conservative substitution rates, we might expect high and low N_e populations to systematically prefer different types of model. We used the ModelOMatic program to infer both the best-fitting model and the state-space of the preferred model.

Table 1 shows the overall preferred state-space for each of our phylogenomic data sets. Our inferences are consistent with both our theoretical observations and our CoRa-based analyses: high N_e data sets overwhelmingly select amino acid models, which account for biophysical differences between amino acids. In contrast, low N_e data sets select codon models, which do not (see table 1). Hence, models that incorporate information about amino acid properties, either in the form of binary classification via $\mathbf{M}_{rad}$ or a matrix describing the probability of exchange, provide a better fit for sequences from large populations with more effective selection. Here, the varying probabilities of exchange between amino acid pairs described by empirical models are analogous to assigning different selective coefficients to different classes of substitution in the CoRa model (fig. 2*a*).

To further explore these differences in preferred state-space, we examine the difference in Akaike's information criterion (AIC) between the best-fitting codon model and the best-fitting amino acid model for each alignment (Posada and Buckley 2004). Figure 5 shows the low N_e data sets (red) universally and strongly select codon models. In contrast, high N_e data sets (blue) have a lower level of support for selecting amino acid models, which nevertheless encompasses the vast majority of alignments. The relatively weaker support for amino acid models may reflect their empirical derivation, their lower parameterization, and their grouping codons together when describing amino acids. That is, they discard information about the evolution of individual sequences that codon models are able to capture. They might, therefore, be less well tuned to the fine-grained process, but their ability to describe average fitness differences between amino acids overcomes that limitation.

We have also considered the possibility that the amount of divergence affects our ability to detect preferences for different types of amino acid substitutions. We find that the tree lengths (sum of branch lengths) from the low N_e data sets tend to be substantially shorter than those from the high N_e data sets, but this difference does not seem adequate to account for the differences in the substitution process between the two groups. When examining genes with similar tree lengths across data sets the selection of CoRa versus M0 or amino acid versus codon model appears to be primarily predicted by whether the gene has evolved under high or low N_e. See supplementary material, Supplementary Material online, for more information.

Discussion

In this study, we examine how the physicochemical properties of amino acids might improve the ability of phylogenetic models to describe protein evolution. We predict a direct link between rates of amino acid substitution and effective population size (N_e) mediated by variation in the efficacy of natural selection. One aspect of amino acid substitution patterns that has recently received renewed attention is whether more chemically distant changes are less frequently fixed (Nabholz et al. 2013; Weber et al. 2014; Figuet et al. 2016). We show that fully investigating this relationship requires a parametric substitution model since simplified counting methods are susceptible to a range of established biases, such as variation in GC-content and transition–transversion bias (Dagan et al. 2002; Smith 2003), which can make the results challenging to interpret. Additionally, we find that long branch lengths and high proportions of nonsynonymous substitutions can exacerbate these issues. We therefore propose the CoRa codon model that allows different types of amino acid substitution to be treated separately.

We test our theoretical predictions by comparing the relative statistical fit of phylogenetic substitution models across two broad bins of population sizes. We find that phylogenetic data from low N_e populations provide little information to differentiate between broad classes of amino acid change, and simple models in the codon state-space tend to fit better than more general models in the amino acid state-space. In contrast, for high N_e populations, we observe distinct differences between rates of amino acid substitution and complex models in the amino acid state-space fit substantially better than simple codon-based models. In the latter case, we also confirm that radical substitutions are more strongly selected against than conservative substitutions, consistent with their

Table 1. Model Selection Results by Alignment Data Set.

Group	No. of Species	State-Space of Preferred Model	CoRa Selected and $\omega_R < \omega_C$	ω	Pop. Size
Mammals (Douzery et al. 2014)	43	Codon (100%)	202/473 (43%)	0.1152	Small
Birds (Jarvis et al. 2015)	48	Codon (100%)	276/725 (38%)	0.1027	Small
Vertebrates (Fong et al. 2012)	129	Codon (100%)	17/66 (26%)	0.0213	Small
Yeast (Salichos and Rokas 2013)	23	AA (97%)	665/669 (>99%)	0.0615	Large
Yeast (Shen et al. 2016)	96	AA (97%)	1,214/1,217 (>99%)	0.0518	Large
Arthropods (Regier et al. 2010)	75	AA (98%)	67/68 (99 %)	0.0164	Large
Insects (Misof et al. 2014)	103	AA (96%)	1,380/1,389 (>99%)	0.0533	Large

NOTE.—CoRa is selected when the likelihood ratio test statistic 2δ is significant.

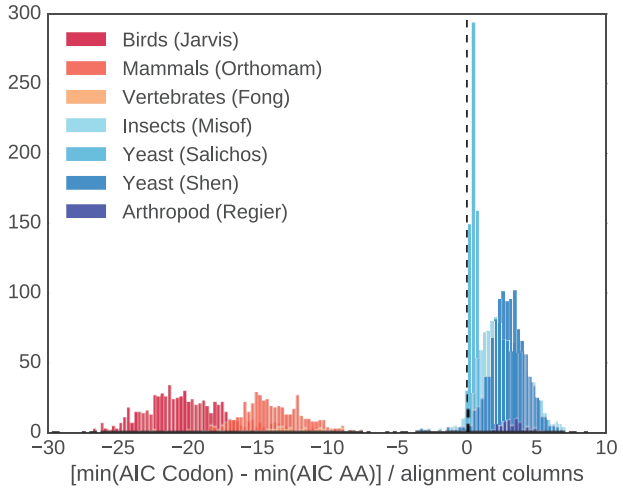

FIG. 5. Density plot showing ΔAIC for the best-fitting AA model versus the best-fitting codon model. Blue shades indicate high N_e, whereas red shades low N_e. Each data point represents one alignment.

being more deleterious. These results hold across several different phylogenetic data sets and, to the best of our knowledge, are not accounted for by other factors, such as sequence divergence, differences in protein function between data sets or artifacts caused by low-quality sequence data.

Given that one might naively expect constraints on structure to be better captured when amino acid preferences are considered, it may appear counterintuitive that some sequences are better described by simpler models. One explanation is that the structure of the genetic code leads to an inherent bias toward less disruptive changes compared with what would be observed under a random alternative code (Haig and Hurst 1991; Freeland and Hurst 1998). The genetic code alone is unlikely to explain, however, the preference for simple models in low N_e populations. It is more likely that we lack either the statistical power or the model resolution to detect that selective differences between broadly binned amino acids for low N_e populations. More phylogenetic diversity or longer sequences might help with statistical power, and methods that allow across-site variation in the nonsynonymous rates might also be helpful.

It might also be the case that the amino acid partition considered here does not reflect the optimal model, and that another partition would pick up sufficient signal from sequences that select M0 in our comparisons. However, determining how said partition would look is a nontrivial problem, given the number of possible classifications that would have to be considered. Although it has been reported that models that offer improvements over M0 on individual alignments are readily found (Delport et al. 2010), it is unclear how this translates to large databases of sequences. It is also not straightforward to interpret partitions from randomly assigned categories in a biologically meaningful manner. Hence, an exhaustive search for the optimal model is beyond the scope of this work. In addition, the existence of a better classification that performs well across all population sizes would be in line with the observation that finer resolution may be required to detect amino acid preferences in some data sets.

The very clear support for codon over amino acid models observed for low N_e alignments may also reflect the drawbacks of not considering information about the codon-level processes that influenced the evolution of the sequences (Seo and Kishino 2008). The observation that a subset of the same sequence alignments that reject amino acid models nonetheless select CoRa over M0 aligns with the idea that accounting for the structure of the genetic code does offer advantages. Overall, our results are in accord with the notion that selective constraints on amino acid changes are, on average, greater than the relative constraints on radical changes compared with conservative ones (Smith 2003) and that this is more readily detectable in high N_e populations.

That the rate of molecular evolution correlates with population size as a function of the strength of selection is not surprising and has previously been widely discussed (Woolfit 2009; Bromham 2011; Hua and Bromham 2017). Here, we confirm that this observation also extends to the ratio of radical to conservative substitutions, R/C, in accord with theoretical expectations. Although this relationship has previously been hinted at, we can now rule out the possibility that it is merely an artifact of the effect of compositional bias on count-based estimates of R/C (see discussion in Weber et al. [2014]). In addition, we find that life history does not just affect rates of evolution but also more broadly predicts which models are suitable for describing them.

Our results may help explain many previous observations in the literature and provide insight into possible causes for the differences in relative fit of existing phylogenetic models. There is a wide variety of available amino acid substitution models, trained on data ranging from a handful of mammalian mitochondrial sequences (Yang et al. 1998) to vast groups of species covering the entire tree of life (Le and Gascuel

2008). The difference between these models is usually discussed in terms of the varying protein contents they describe (Keane et al. 2006) or the parameters they infer, but our results suggest that other factors such as differences in N_e affecting their training data could also play an important role. Our results suggest, for instance, that amino acid substitution models trained on low N_e mammalian data might be more reflective of the genetic code whereas models trained across the tree of life, which consider sequences from extremely high N_e species, could capture more subtle preferences driven by physicochemical differences between amino acids.

The effect of N_e on substitution rates may also explain the limited improvement in fit observed in early attempts to incorporate physicochemical distances between amino acids into codon models (Yang et al. 1998). The sequences examined in these studies came from mammals, which have small populations and would therefore likely require a substantial amount of phylogenetic information to discriminate between different categories of amino acids.

Variation in N_e across the tree of life and its differential impact on radical and conservative substitutions might also help explain other widely discussed observations in evolutionary biology. One example is the pervasive temporal heterogeneity in substitution rates throughout the tree of life. This heterogeneity is known to cause significant problems for phylogenetic inference (Tuffley and Steel 1998; Galtier 2001; Huelsenbeck 2002; Wang et al. 2007; Whelan 2008; Whelan et al. 2011), although its exact causes remain nebulous. Our results suggest that variation in N_e across the tree might be an important causal factor for this temporal heterogeneity. For instance, a change from a low to high N_e along a branch might lead to a switch from "on" to "off" in a covarion process (Tuffley and Steel 1998). Intriguingly, our results suggest this covarion-type switch could be an extreme example and there might be more subtle, but pervasive, changes in amino acid substitution patterns across the tree, whereby radical changes are more or less efficiently purged by selection depending on N_e. In addition to being a clear model violation, there might appear to be convergent evolution in taxa across the tree with high N_e simply because there are fewer amino acids that are accepted by selection at individual sites on these lineages, even when there are multiple amino acids that are roughly functionally equivalent. In such cases, methods that look for convergent shifts in amino acid profiles rather than shifts toward a single amino acid (Rey et al. 2018) may be more robust.

At the very least, our argument suggests one should be cautious when building trees from groups of taxa with very different life histories since even the most sophisticated amino acid substitution models, such as CAT and PMSF (Lartillot and Philippe 2004; Blanquart and Lartillot 2008; Si Quang et al. 2008; Wang et al. 2018) only account for spatial heterogeneity along the sequence and cannot capture variation in substitution patterns through the tree.

Finally, our findings also suggest that one path toward capturing and understanding protein evolution is to begin explicitly incorporating N_e into phylogenetic models. One particularly promising route would be the implementation of mutation–selection models (Halpern and Bruno 1998; Yang and Nielsen 2008; Tamuri et al. 2012; Thorne et al. 2012). These MutSel models attempt to disentangle the effects of mutation and selection, allowing the product $N_e \times s$ to be estimated for different types of substitutions from phylogenetic data. The models might then be extended to estimate relative values of N_e for branches or clades. Although the results we present in this manuscript allow us to establish that radical substitutions are preferentially selected against, a MutSel model might also allow us to obtain a more detailed picture of what constitutes a "radical" substitution in a specific site or protein. An alternative approach might be to use MCMC to estimate temporal mixture models with different substitution models at different sites on different branches. Either way, unlocking the relationship between N_e and substitution rate may help resolve some of the most difficult questions in phylogenetics.

Materials and Methods

Models

Existing Codon Models

Standard phylogenetic methods for analyzing codon sequences are typically based around the M0 model derived from Goldman and Yang (1994). This substitution model is defined through the off-diagonal elements of the instantaneous rate matrix, where $\omega = \mathrm{d}N/\mathrm{d}S$, as follows:

$$Q^{\mathrm{M0}}_{i,j} = \pi_j \begin{cases} 0 & i \text{ and } j \text{ differ by more than a single nucleotide} \\ 1 & i \text{ and } j \text{ differ by a single synonymous transversion} \\ \kappa & i \text{ and } j \text{ differ by a single synonymous transition} \\ \omega & i \text{ and } j \text{ differ by a single nonsynonymous transversion} \\ \omega\kappa & i \text{ and } j \text{ differ by a single nonsynonymous transition.} \end{cases} \quad (1)$$

Following standard approaches, a substitution model can be used to produce transition matrices describing the probability of change between different codons over time t, such that $P^{\mathrm{M0}}(t) = e^{\mathbf{Q}t}$. The diagonal elements of the instantaneous matrix are set to the negative row sums to ensure the models produce a valid set of probability matrices for different time points. This model assumes that all amino acid mutations are equally likely to be fixed, which ignores differences in the physicochemical properties of the amino acids.

The CoRa Codon Model Describing Radical and Conservative Amino Acid Change

The frequency at which conservative and radical amino acid substitutions occur provides some insight into the evolutionary forces acting on proteins (Zhang 2000; Smith 2003; Sainudiin et al. 2005), but there is limited discussion regarding how different estimates of these values relate to the well-

studied parameters of the M0 model. Here, we present the CoRa codon model of amino acid substitution, which divides $\omega = dN/dS$ from M0 into the two subcategories $\omega_C = dC/dS$ and $\omega_R = dR/dS$ in the following instantaneous rate matrix:

$$Q_{i,j}^{\text{CoRa}} = \pi_j \begin{cases} 0 & i \text{ and } j \text{ differ by more than a single nucleotide} \\ 1 & i \text{ and } j \text{ differ by a single synonymous transversion} \\ \kappa & i \text{ and } j \text{ differ by a single synonymous transition} \\ \omega_C & i \text{ and } j \text{ differ by a single nonsynonymous conservative transversion} \\ \omega_C\kappa & i \text{ and } j \text{ differ by a single nonsynonymous conservative transition} \\ \omega_R & i \text{ and } j \text{ differ by a single nonsynonymous radical transversion} \\ \omega_R\kappa & i \text{ and } j \text{ differ by a single nonsynonymous radical transition.} \end{cases} \quad (2)$$

In order to fully resolve this matrix, we need to define the radical matrix $\mathbf{M}_{\text{rad}}$, which is a 20×20 binary matrix that labels each amino acid substitution $i \rightarrow j$ as a 0 or 1, representing a conservative and a radical amino acid substitution, respectively. Here, we distinguish between polar and nonpolar amino acids and amino acids with large and small volumes, where Y, W, H, K, R, E, Q, T, D, N, S, and C are polar and L, I, F, M, Y, W, H, K, R, E, and Q are large. Any substitution involving a change in category for polarity or volume is considered radical, whereas substitutions altering neither property's category are considered conservative (Sainudiin et al. 2005). In a preliminary study, we examined other conservative and radical classifications based on chemical properties, but the above definition tended to provide better fitting models.

Under this model, we can calculate $dR/dC = \omega_R/\omega_C$, which is a R/C measure describing the relative rates of conservative and radical substitutions. Under the constraint $\omega_R = \omega_C$, conservative and radical changes cannot be distinguished by selection and CoRa reduces to the simpler model M0. This nesting means that standard likelihood ratio tests with 1 degree of freedom may be used to assess the fit of CoRa relative to M0. When CoRa provides no significant improvement on M0 there is no evidence for differences in the rate of conservative and radical substitutions. With a significant improvement in fit, then $dR/dC < 1$ represents our expectation that radical changes, which are likely to disrupt the protein structure, occur at a lower instantaneous rate than conservative changes, which are more likely to maintain the current structure. A significant improvement in fit coupled with values of $dR/dC > 1$ represents the case where radical substitutions occur significantly more frequently than conservative substitutions.

Counting-Based K_r/K_c Measures and Their Relationship to the CoRa Model

The definition of dR and dC described above has a clear relationship to the commonly used M0 model and the well-defined ω parameter. Other studies have developed alternative R/C measures. Here, we therefore characterize some of these measures and show that dR/dC provides the most convenient and interpretable value. The two alternatives we examine involve different approaches to counting conservative (K_c) and radical (K_r) nonsynonymous substitutions, and using K_r/K_c as a measure of the relative selective pressure acting on conservative and radical substitutions. The first involves tallying the total number of substitutions that occur per unit time. Here, all cases of "multiple hits" where several substitutions occur at a site and can be captured through stochastic mapping (Nielsen and Huelsenbeck 2002; Minin and Suchard 2008) or by counting the number of conservative and radical changes over a very short period of time. Using the CoRa model, the expected number of conservative and radical substitutions per unit time can be computed as

$$K_r(\mathbf{Q}) = \sum_i \sum_j 1_{\text{Radical}}(i \rightarrow j)\pi_i Q_{i,j}, \quad (3)$$

$$K_c(\mathbf{Q}) = \sum_i \sum_j 1_{\text{Conservative}}(i \rightarrow j)\pi_i Q_{i,j}, \quad (4)$$

where 1_{Radical} and $1_{\text{Conservative}}$ are indicator functions that return 1 when $i \rightarrow j$ are radical and conservative nonsynonymous substitutions, respectively, and 0 otherwise. Expected values calculated using this approach will be referred to as $\frac{K_r}{K_c}(\mathbf{Q})$ to indicate they have been derived from the instantaneous rate matrix $\mathbf{Q}$. The second approach to estimating K_r/K_c is to calculate the relative numbers of conservative and radical substitutions after a given evolutionary time, t. This approach is essentially the same as above but replaces $\mathbf{Q}$ with the probability matrix $\mathbf{P}(t)$, and fails to correct for cases where more than one substitution occurs at a site. For a given value of t, these quantities can be calculated as

$$K_r(t) = \sum_i \sum_j 1_{\text{Radical}}(i \rightarrow j)\pi_i P(t)_{i,j}, \quad (5)$$

$$K_c(t) = \sum_i \sum_j 1_{\text{Conservative}}(i \rightarrow j)\pi_i P(t)_{i,j}. \quad (6)$$

For both of these cases, the K_r and K_c measures are comparable to those obtained from amino acid sequences when counting the number of radical and conservative substitutions. Estimates made using this approach will be referred to as $\frac{K_r}{K_c}(t)$.

Defining dR/dC in Terms of N_e

Several studies have proposed that radical and conservative substitutions might predict N_e, as radical substitutions are expected to be removed more effectively from large populations (Eyre-Walker et al. 2002). Following the formulation of mutation–selection models, where the substitution rate (Q_{ij})

is a product of the mutation rate from i to j and the probability of fixation of j relative to wildtype i, the CoRa model describes the relative fixation probabilities through the ω_C and ω_R parameters. Using the weak mutation model of Golding and Felsenstein (Golding and Felsenstein 1990) to define these parameters in terms of the selective pressures acting on conservative (s_c) and radical changes (s_r) for diploids we have

$$\omega_C = \frac{2N_e(1 - e^{2s_c})}{1 - e^{4N_e s_c}} \approx \frac{4N_e s_c}{1 - e^{4N_e s_c}}, \quad (7)$$

$$\omega_R = \frac{2N_e(1 - e^{2s_r})}{1 - e^{4N_e s_r}} \approx \frac{4N_e s_r}{1 - e^{4N_e s_r}}. \quad (8)$$

Under this approach, both conservative and radical changes will be assumed to be selectively disadvantageous, which allows us to express dR/dC in terms of these selective pressures and N_e:

$$\frac{dR}{dC} = \frac{\omega_R}{\omega_C} \approx \frac{4N_e s_r}{1 - e^{4N_e s_r}} \times \frac{1 - e^{4N_e s_c}}{4N_e s_c} \approx \frac{s_r}{s_c} \times \frac{1 - e^{2N_e s_c}}{1 - e^{2N_e s_r}}. \quad (9)$$

Note that this expression only allows us to examine how dR/dC varies according to N_e for relative values of s_c and s_r since they either only occur as a fraction or in terms with N_e. (The effect of N_e on the values of ω_C and ω_R can also be computed in this manner, with Yang and Nielsen providing details (Yang and Nielsen 2008).) Our approach allows us to analytically examine how dN/dS might respond for different values of N_e, s_c, and s_r.

Analysis of Phylogenomic Data Sets

We take previously published phylogenomic data sets each consisting of a set of alignments from between 23 and 129 OTUs from birds (Jarvis et al. 2014, 2015), mammals (Douzery et al. 2014), insects (Misof et al. 2014), arthropods (Regier et al. 2010), and yeast (Salichos and Rokas 2013; Shen et al. 2016) (see https://doi.org/10.5061/dryad.157d7, http://doi.org/10.6084/m9.figshare.3370675.v1; last accessed January 2019), with the range of taxonomic groups providing comparisons between a range of effective population sizes (Charlesworth 2009; Skelly et al. 2009) and tree depths.

We performed basic quality control, including removing all alignments with internal stops; CTG codons were masked out in the Candida clade due to a nonstandard genetic code. To obtain codon alignments for the data of Salichos and Rokas (2013), amino acid alignments were mapped back to nucleotide sequences using PAL2NAL (Suyama et al. 2006). The sequence alignments in each data set were analyzed independently. The optimal state-space (purine or pyrimidine, also referred to as RY, nucleotide, codon or amino acid) for the phylogenetic analysis was determined using ModelOMatic with default settings (Whelan et al. 2015).

To ensure AIC values from ModelOMatic are comparable between alignments of different lengths, we scaled them by the number of alignment columns present after filtering sparse columns. Purpose-written software was used to estimate parameters from M0 and the CoRa model using maximum likelihood. BioNJ trees for each multiple sequence alignment were used as a starting point to fit the models (Gascuel 1997). The CoRa model is available in the Bio++ libraries (Guéguen et al. 2013) (https://github.com/BioPP/) and can be run by specifying the model=CodonAAClustFreq in bppml, with the default partition and frequencies=F3X4 corresponding to the settings used in this work. The results presented in this work can be reproduced using bppml (for more details and configuration see https://github.com/claudia-c-weber/CoRa).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We would like to thank Joe Bielawski for the discussion that prompted the early stages of this work, Nick Goldman for suggesting the control in supplementary figure S3, Supplementary Material online, and Laurent Guéguen for making the CoRa model available in Bio++. We also thank two reviewers for their thoughtful comments on the manuscript.

References

Blanquart S, Lartillot N. 2008. A site-and time-heterogeneous model of amino acid replacement. *Mol Biol Evol*. 25(5):842–858.

Bromham L. 2011. The genome as a life-history character: why rate of molecular evolution varies between mammal species. *Philos Trans R Soc Lond B Biol Sci*. 366(1577):2503–2513.

Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. *Nat Rev Genet*. 10(3):195–205.

Dagan T, Talmor Y, Graur D. 2002. Ratios of radical to conservative amino acid replacement are affected by mutational and compositional factors and may not be indicative of positive Darwinian selection. *Mol Biol Evol*. 19(7):1022–1025.

Delport W, Scheffler K, Gravenor MB, Muse SV, Pond SK. 2010. Benchmarking multi-rate codon models. *PLoS One* 5(7):e11587.

Douzery EJP, Scornavacca C, Romiguier J, Belkhir K, Galtier N, Delsuc F, Ranwez V. 2014. OrthoMaM v8: a database of orthologous exons and coding sequences for comparative genomics in mammals. *Mol Biol Evol*. 31(7):1923–1928.

Epstein CJ. 1967. Non-randomness of amino-acid changes in the evolution of homologous proteins. *Nature* 215(5099):355–359.

Eyre-Walker A, Keightley PD, Smith NGC, Gaffney D. 2002. Quantifying the slightly deleterious mutation model of molecular evolution. *Mol Biol Evol*. 19(12):2142–2149.

Figuet E, Nabholz B, Bonneau M, Carrio EM, Nadachowska-Brzyska K, Ellegren H, Galtier N. 2016. Life history traits, protein evolution, and the nearly neutral theory in amniotes. *Mol Biol Evol*. 33(6):1517–1527.

Fong JJ, Brown JM, Fujita MK, Boussau B. 2012. A phylogenomic approach to vertebrate phylogeny supports a turtle-archosaur affinity and a possible paraphyletic lissamphibia. *PLoS One* 7(11):e48990.

Freeland SJ, Hurst LD. 1998. The genetic code is one in a million. *J Mol Evol*. 47(3):238–248.

Galtier N. 2001. Maximum-likelihood phylogenetic analysis under a covarion-like model. *Mol Biol Evol*. 18(5):866–873.

Gascuel O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol*. 14(7):685–695.

Golding B, Felsenstein J. 1990. A maximum likelihood approach to the detection of selection from a phylogeny. *J Mol Evol*. 31(6):511–523.

Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol*. 11(5):725–736.
Grantham R. 1974. Amino acid difference formula to help explain protein evolution. *Science* 185(4154):862–864.
Guéguen L, Gaillard S, Boussau B, Gouy M, Groussin M, Rochette NC, Bigot T, Fournier D, Pouyet F, Cahais V, et al. 2013. Bio++: efficient extensible libraries and tools for computational molecular evolution. *Mol Biol Evol*. 30(8):1745–1750.
Haig D, Hurst LD. 1991. A quantitative measure of error minimization in the genetic code. *J Mol Evol*. 33(5):412–417.
Halpern AL, Bruno WJ. 1998. Evolutionary distances for protein-coding sequences: modeling site-specific residue frequencies. *Mol Biol Evol*. 15(7):910–917.
Hua X, Bromham L. 2017. Darwinism for the genomic age: connecting mutation to diversification. *Front Genet*. 8(12):12.
Huelsenbeck JP. 2002. Testing a covariotide model of DNA substitution. *Mol Biol Evol*. 19(5):698–707.
Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346(6215):1320–1331.
Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, et al. 2015. Phylogenomic analyses data of the avian phylogenomics project. *GigaScience* 4(1):4.
Jones D, Taylor W, Thornton J. 1994. A mutation data matrix for transmembrane proteins. *FEBS Lett*. 339(3):269–275.
Keane TM, Creevey CJ, Pentony MM, Naughton TJ, Mclnerney JO. 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol*. 6(1):29.
Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol*. 7(Suppl 1):S4.
Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol*. 21(6):1095–1109.
Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol Biol Evol*. 25(7):1307–1320.
Minin VN, Suchard MA. 2008. Fast, accurate and simulation-free stochastic mapping. *Philos Trans R Soc Lond B Biol Sci*. 363(1512):3985–3995.
Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346(6210):763–767.
Miyata T, Miyazawa S, Yasunaga T. 1979. Two types of amino acid substitutions in protein evolution. *J Mol Evol*. 12(3):219–236.
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(5):715–724.
Nabholz B, Uwimana N, Lartillot N. 2013. Reconstructing the phylogenetic history of long-term effective population size and life-history traits using patterns of amino acid replacement in mitochondrial genomes of mammals and birds. *Genome Biol Evol*. 5(7):1273–1290.
Nielsen R, Huelsenbeck J. 2002. Mapping mutations on phylogenies. *Syst Biol*. 51(5):729–739.
Popadin KY, Polishchuk LV, Mamirova L, Knorre D, Gunbin K. 2007. Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. *Proc Natl Acad Sci U S A*. 104(33):13390–13395.
Posada D, Buckley TR. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst Biol*. 53(5):793–808.
Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, Martin JW, Cunningham CW. 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* 463(7284):1079–1083.
Rey C, Guéguen L, Sémon M, Boussau B. Forthcoming 2018. Accurate detection of convergent amino-acid evolution with PCOC. *Mol Biol Evol*. 35(9): 2296–2306.
Sainudiin R, Wong WSW, Yogeeswaran K, Nasrallah JB, Yang Z, Nielsen R. 2005. Detecting site-specific physicochemical selective pressures: applications to the Class I HLA of the human major histocompatibility complex and the SRK of the plant sporophytic self-incompatibility system. *J Mol Evol*. 60(3):315–326.
Salichos L, Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* 497(7449):327–331.
Seo T-K, Kishino H. 2008. Synonymous substitutions substantially improve evolutionary inference from highly diverged proteins. *Syst Biol*. 57(3):367–377.
Seo T-K, Kishino H. 2009. Statistical comparison of nucleotide, amino acid, and codon substitution models for evolutionary analysis of protein-coding sequences. *Syst Biol*. 58(2):199–210.
Shen X-X, Zhou X, Kominek J, Kurtzman CP, Hittinger CT, Rokas A. 2016. Reconstructing the backbone of the saccharomycotina yeast phylogeny using genome-scale data. *G3 (Bethesda)* 6(12):3927–3939.
Si Quang L, Gascuel O, Lartillot N. 2008. Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* 24(20):2317–2323.
Skelly DA, Ronald J, Connelly CF, Akey JM. 2009. Population genomics of intron splicing in 38 *Saccharomyces cerevisiae* genome sequences. *Genome Biol Evol*. 1:466–478.
Smith NGC. 2003. Are radical and conservative substitution rates useful statistics in molecular evolution? *J Mol Evol*. 57(4):467–478.
Suyama M, Torrents D, Bork P. 2006. Pal2nal: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 34(Web Server):W609–W612.
Tamuri AU, dos Reis M, Goldstein RA. 2012. Estimating the distribution of selection coefficients from phylogenetic data using sitewise mutation-selection models. *Genetics* 190(3):1101–1115.
Thorne JL, Lartillot N, Rodrigue N, Choi SC. 2012. Codon models as a vehicle for reconciling population genetics with inter-specific sequence data. New York: Oxford University Press. p. 97–110.
Tuffley C, Steel M. 1998. Modeling the covarion hypothesis of nucleotide substitution. *Math Biosci*. 147(1):63–91.
Wang H-C, Minh BQ, Susko E, Roger AJ. 2018. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst Biol*. 67(2):216–235.
Wang H-C, Spencer M, Susko E, Roger AJ. 2007. Testing for covarion-like evolution in protein sequences. *Mol Biol Evol*. 24(1):294–305.
Weber CC, Nabholz B, Romiguier J, Ellegren H. 2014. Kr/Kc but not dN/dS correlates positively with body mass in birds, raising implications for inferring lineage-specific selection. *Genome Biol*. 15(12):542.
Whelan S. 2008. Spatial and temporal heterogeneity in nucleotide sequence evolution. *Mol Biol Evol*. 25(8):1683–1694.
Whelan S, Allen JE, Blackburne BP, Talavera D. 2015. ModelOMatic: fast and automated model selection between RY, nucleotide, amino acid, and codon substitution models. *Syst Biol*. 64(1):42–55.
Whelan S, Blackburne BP, Spencer M. 2011. Phylogenetic substitution models for detecting heterotachy during plastid evolution. *Mol Biol Evol*. 28(1):449–458.
Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*. 18(5):691–699.
Woolfit M. 2009. Effective population size and the rate and pattern of nucleotide substitutions. *Biol Lett*. 5(3):417–420.
Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol*. 15(5):568–573.
Yang Z, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol*. 46(4):409–418.

Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol*. 19(6):908–917.

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(3):568–579.

Yang Z, Nielsen R, Hasegawa M. 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. *Mol Biol Evol*. 15(12):1600–1611.

Zhang J. 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *J Mol Evol*. 50(1):56–68.