A Novel Use of Equilibrium Frequencies in Models of Sequence Evolution

Nick Goldman and Simon Whelan

Department of Zoology, University of Cambridge, U.K.

Current mathematical models of amino acid sequence evolution are often applied in variants that match their expected amino acid frequencies to those observed in a data set under analysis. This has been achieved by setting the instantaneous rate of replacement of a residue *i* by another residue *j* proportional to the observed frequency of the resulting residue *j*. We describe a more general method that maintains the match between expected and observed frequencies but permits replacement rates to be proportional to the frequencies of both the replaced and resulting residues, raised to powers other than 1. Analysis of a database of amino acid alignments shows that the description of the evolutionary process in a majority (approximately 70% of 182 alignments) is significantly improved by use of the new method, and a variety of analyses indicate that parameter estimation with the new method is well-behaved. Improved evolutionary models increase our understanding of the process of molecular evolution and are often expected to lead to improved phylogenetic inferences, and so it seems justified to consider our new variants of existing standard models when performing evolutionary analyses of amino acid sequences. Similar methods can be used with nucleotide substitution models, but we have not found these to give corresponding significant improvements to our ability to describe the processes of nucleotide sequence evolution.

Introduction

Research into better mathematical models of molecular sequence evolution continues to be of interest for many reasons. When a complex model gives a statistically significant improvement in fit to observed data compared with a simpler model, the biological factors corresponding to the additional complexity can be taken to be important in the process of evolution for those sequences. This increases our understanding of molecular evolution, and the models may be used to investigate these biological factors in greater detail than before. An example is the improved ability to detect molecular adaptation that has come from new models of selective pressures operating on DNA, reviewed by Yang and Bielawski (2000).

We may also expect more realistic models to give more reliable inference of biological factors they share with simpler models, for example phylogenetic topologies and branch lengths. This improvement could arise from their improved ability to account for factors that simpler models neglect and whose influence on observed data might otherwise be misinterpreted. This expectation has generally been confirmed by both theoretical results from simulation studies (e.g., Kuhner and Felsenstein 1994; Zhang 1999) and empirical results (e.g., Cunningham, Zhu, and Hillis 1998; Takezaki and Gojobori 1999; Philippe and Germot 2000; Yoder and Yang 2000), although Posada and Crandall (2001) describe an interesting exception.

Much progress has recently been made with improving models of sequence evolution (reviewed, e.g., by Liò and Goldman 1998; Thorne 2000; Whelan, Liò, and Goldman 2001). This work relies on statistical tests that can compare competing models and diagnose their shortcomings (Felsenstein 1981; Goldman 1993*a*, 1993*b;* Yang, Goldman, and Friday 1994; Whelan and Goldman 1999; Goldman and Whelan 2000; Ota et al. 2000). Amino acid replacement models have been improved by the use of maximum likelihood (ML) methods to infer replacement rate parameters (Adachi and Hasegawa 1996; Yang, Nielsen, and Hasegawa 1998; Adachi et al. 2000; Whelan and Goldman 2001). Both amino acid replacement and DNA substitution models have been improved by the incorporation of amongst-site heterogeneity of evolutionary rate (Yang 1994*a,* 1996). Codon-level models can now test for effects of selection on sequence evolution and permit detection of individual sites where selection acts (Nielsen and Yang 1998; Yang et al. 2000; Anisimova, Bielawski, and Yang 2001).

The expected frequencies of nucleotides, amino acids, or codons may be matched to those observed in the data under analysis. This generally leads to significantly improved models and has become widespread. For DNA models this has often led to the use of the FEL (Felsenstein 1981) and HKY (Hasegawa, Kishino, and Yano 1985) models in preference to the JC (Jukes and Cantor 1969) and K2P (Kimura 1980) models. For amino acid models, the same has been achieved by the ‘‘+F’’ method of Cao et al. (1994). For codon models, analogous methods were first described by Muse and Gaut (1994) and Goldman and Yang (1994). All of these models achieve equality of the frequencies expected by the model and observed in the data by setting the instantaneous rates of evolutionary changes proportional to the observed frequency of the nucleotide, amino acid, or codon being changed to.

In this article, we describe a generalization of these approaches that allows evolutionary rates to be dependent on the frequency of the nucleotide, amino acid, or codon being replaced as well as on the frequency of the one that forms the replacement. This introduces just one additional parameter (or degree of freedom) to the model. We have applied the new method, denoted ‘‘+gwF,’’ to models of DNA substitution and amino acid replacement and demonstrate that the new method gives a sta-

Key words: amino acid frequencies, Markov process models, model comparison, molecular phylogenetics, nucleotide frequencies.

Address for correspondence and reprints: Nick Goldman, EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, U.K. E-mail: goldman@ebi.ac.uk.

Mol. Biol. Evol. 19(11):1821–1831. 2002

tistically significant improvement to the fit of models to data in a large proportion of the amino acid data sets in a database of protein family alignments. Our method is equally applicable to codon models, although we have not yet investigated its utility in this context.

Methods

Existing Amino Acid Replacement Models

All models discussed in this article assume that amino acid sites in an alignment evolve independently and according to the same stationary, homogeneous, and reversible Markov process (Swofford et al. 1996; Liò and Goldman 1998). The probability of amino acid i being replaced by amino acid j after time t is denoted $P_{ij}(t)$, with i and j each representing one of the 20 different amino acids. These probabilities can be written as a 20×20 matrix that may be calculated as $P(t) = \exp(tQ)$, where Q is a matrix with off-diagonal elements q_{ij} being the instantaneous rates of change of amino acid i to amino acid j and diagonal elements q_{ii} being fixed so that the row sums of Q equal 0. The rate matrix Q may also be parameterized $q_{ij} \equiv \pi_j s_{ij}$ for all $i \neq j$. This identifies frequency parameters π_j that describe the equilibrium frequencies of the amino acids and parameters s_{ij} (sometimes called exchangeabilities: see, e.g., Whelan and Goldman 2001; Whelan, Liò, and Goldman 2001) that form a symmetric matrix (i.e., $s_{ij} = s_{ji}$). The matrix Q is scaled so that $-\Sigma_i\ \pi_i q_{ii} = 1$, causing evolutionary times to be measured in expected replacements per site; for clarity, we omit this scaling from our equations below. Estimation of Q from large sequence databases has led to various standard models of amino acid replacement. Examples derived from globular protein sequences include the Dayhoff (Dayhoff, Schwartz, and Orcutt 1978), JTT (Jones, Taylor, and Thornton 1992) and WAG (Whelan and Goldman 2001) models. We denote estimates from databases with a tilde; for example, $\tilde{Q}$; $\tilde{q}_{ij} \equiv \tilde{\pi}_j \tilde{s}_{ij}$.

Typical applications of these models to phylogenetic inference from amino acid sequences combine the database estimates $\tilde{Q}$ with information from the observed sequences using the +F method first described by Cao et al. (1994). This assumes that in the parameterization $\tilde{q}_{ij} = \tilde{\pi}_j \tilde{s}_{ij}$ the exchangeabilities $\tilde{s}_{ij}$ are fundamental properties of the amino acids (i, j). In contrast, the frequencies $\tilde{\pi}_j$ represent properties that depend on factors such as mutation bias, selective pressure (the relative advantage/disadvantage of different amino acids) and codon usage, which are assumed to differ from one protein sequence data set to another. As a consequence, the +F method proposes that during the analysis of a specific data set the fundamental properties described by the $\tilde{s}_{ij}$ derived from a database analysis remain unaltered, but that the $\tilde{\pi}_j$, describing evolutionary pressures acting on the sequences, be replaced with a set of frequencies more appropriate to the data being analyzed. Denoting these frequencies π_j, this gives a rate matrix defined by the off-diagonal elements

$$q_{ij} = \pi_j \tilde{s}_{ij} \tag{1}$$

that uses the values $\tilde{s}_{ij}$ derived from a database yet ensures that the expected amino acid frequencies under the model match those of the data being analyzed. If the $\tilde{s}_{ij}$ used are those of the JTT model, for example, then the model defined by equation (1) would be denoted JTT+F (i.e., JTT base model with +F method). The π_j are usually estimated by simply counting the residues of all the sequences under analysis to find the observed proportions of each amino acid, although they may also be estimated by ML (see also Whelan and Goldman 1999).

A New Generalization of Amino Acid Replacement Models

The parameterization in equation (1) and the +F method assume that the rate of change of amino acid i to amino acid j is proportional to π_j, the equilibrium frequency of j in the sequences under study, and to the exchangeability $\tilde{s}_{ij}$. There seems no reason why this parameterization, and its implicit assumption of the fundamental nature of the exchangeabilities $\tilde{s}_{ij}$, should be the only one considered. Other formulations exist that combine information derived from database analyses (embodied in estimates $\tilde{Q}$) with observed amino acid frequencies in a manner that equates the expected and observed frequencies: for example, making the rate of change of i to j inversely proportional to π_i. In the former case, a low frequency of amino acid l (small π_l) is explained by there being only low rates of change to (or fixation of) l ($q_{il} \propto \pi_l$), perhaps reflecting a selective disadvantage of l. In the latter case it is explained by there being rapid mutation away from l ($q_{lj} \propto 1/\pi_l$), perhaps representing a high mutability of l. Both these factors seem biologically plausible.

We generalize the formulations given by Cao et al. (1994) and above by proposing the parameterization

$$q_{ij} = \frac{\pi_j^g}{\pi_i^f} c_{ij} \tag{2}$$

with the c_{ij} representing as-yet unknown proportionality constants. We impose two constraints on f, g and the c_{ij}. First, to use the information derived from database analyses, we require that if the observed frequencies π_j match the $\tilde{\pi}_j$ from the database analysis then the rate matrices should also match (i.e., $q_{ij} = \tilde{q}_{ij}$). From equation (2), this is achieved when $c_{ij} = \tilde{\pi}_i^f \tilde{\pi}_j^{-g} \tilde{q}_{ij}$. Second, to ensure that this model has the required equilibrium distribution, q_{ij}/π_j must be symmetric in i and j. Noting the previous expression for c_{ij} and the fact that the $\tilde{q}_{ij}/\tilde{\pi}_j$ are symmetric, this is achieved only if $g = 1 - f$. This leads to:

$$q_{ij} = \left(\frac{\pi_j}{\tilde{\pi}_j}\right)^{1-f} \left(\frac{\tilde{\pi}_i}{\pi_i}\right)^f \tilde{q}_{ij}. \tag{3}$$

Because database estimates $\tilde{Q}$ are often reported in terms of their frequencies and exchangeabilities $\tilde{\pi}_j$ and $\tilde{s}_{ij}$, the following equivalent formulation may be useful:

$$q_{ij} = \frac{\pi_j^{1-f}}{\tilde{\pi}_j^{-f}} \left(\frac{\tilde{\pi}_i}{\pi_i}\right)^f \tilde{s}_{ij}. \tag{4}$$

We propose the suffix "+gwF" (*g*eneralized *w*eighted *F*requencies) to indicate use of the model of equation (3). The +gwF formulation requires the estimation of amino acid frequencies for the observed data (π_j), as do +F models, and it includes one extra free parameter (f) that may generally take any value (but see below) and may be estimated by ML.

Occasionally, sequence alignments may be observed in which one or more amino acids do not appear; that is, $\pi_z = 0$ for some amino acid z. (We assume that this is never the case for the $\tilde{\pi}_j$ because any reasonable database is expected to contain each amino acid at least once.) This causes no problem when $f < 0$, but equation (3) generates infinite rates of replacement of z (q_{zj}) when $f \geq 0$. Note, however, that the proportion of all replacements occurring that are from i to z is given by $\pi_i q_{iz} = (\pi_i \pi_z)^{1-f} \tilde{\pi}_i^f \tilde{\pi}_z^{f-1} \tilde{q}_{iz}$. This equals 0 when $0 \leq f \leq 1$: amino acid z never arises, and so the infinite values of q_{zj} suggested by equation (3) can be replaced with arbitrary finite values without altering the probabilities $P(t)$ necessary for likelihood calculations. If $f > 1$, amino acid z both arises and is replaced at an infinite rate: the new model is then meaningless in the context of sequence evolution. In effect, f must be ≤ 1 when some observed amino acid frequencies equal zero.

When $f = 0$, equation (4) reduces to equation (1) and the +F method results. From equation (3) above, the dependency of the model on the observed data is then of the form $q_{ij} \propto \pi_j/\tilde{\pi}_j$ and data sets for which the estimate of f (denoted $\hat{f}$) is near 0 may be considered to have replacements rates largely determined by factors dependent on the resulting residue j (e.g., mutation bias or selection acting on j). If $f = 1$ we have $q_{ij} \propto \tilde{\pi}_i/\pi_i$, and data sets with $\hat{f}$ near 1 may have replacement rates largely determined by factors dependent on the replaced residue i (e.g., its mutability). Other values of $\hat{f}$ indicate the relative strength of these two types of factors. Intermediate values indicate a more even combination of both of these effects (e.g., $f = 1/2$ leads to $q_{ij} \propto (\pi_j/\tilde{\pi}_j)^{1/2}(\tilde{\pi}_i/\pi_i)^{1/2}$). More extreme values indicate stronger tendencies for factors dependent on the resulting ($\hat{f} < 0$) or replaced ($\hat{f} > 1$) residue to determine replacements. The dependence of equation (3) on the terms $(\pi_j/\tilde{\pi}_j)^{1-f}$ and $(\tilde{\pi}_i/\pi_i)^f$ means that the most extreme values of f lead to replacements being dominated by interchanges between the amino acids that are commonest (when $f \ll 0$) or rarest (when $f \gg 1$) in the data set under analysis relative to the database values. But we have observed no data sets where such very extreme estimates $\hat{f}$ were found (see *Results and Discussion,* below).

If the sets of observed and database amino acid frequencies are equal (i.e., $\pi_j = \tilde{\pi}_j$ for all j) then the +gwF model has no dependency on the parameter f (see eq. 3). Although this is improbable, there will be an effect of decreased dependency on f when the π_j are close to the $\tilde{\pi}_j$. We discuss some consequences of this effect below.

Application of Generalized Amino Acid Replacement Models

Application of the new models of amino acid replacement is via standard ML methods, as described by Swofford et al. (1996). No closed form solution of $P(t) = \exp(tQ)$ is generally available for equation (3), and so spectral decomposition must be used (Liò and Goldman 1998). The amino acid frequencies $\tilde{\pi}_j$ and the $\tilde{Q}$ or exchangeability parameters $\tilde{s}_{ij}$ for the standard Dayhoff, JTT, and WAG models are widely available. We estimate the amino acid frequencies π_j for each data set analyzed by counting the residues of the sequences. The parameter f of +gwF models is estimated by ML simultaneously with any other free parameters (branch lengths, etc.).

Our main aim here is to assess the utility of +gwF models, measured by any improvement to the statistical fit of models to observed data. We compare the +F and +gwF versions of the existing JTT and WAG models, each considered both without and with the addition of a discrete (four category) Γ distribution (Yang 1994*a*, 1996; denoted "+Γ") to model heterogeneity of evolutionary rates amongst amino acid sites. This leads to a total of eight analyses for each aligned sequence family studied, with the most important comparisons being the following four:

A: JTT+F versus JTT+gwF
B: JTT+F+Γ versus JTT+gwF+Γ
C: WAG+F versus WAG+gwF
D: WAG+F+Γ versus WAG+gwF+Γ.

Likelihood ratio tests (LRTs) can be performed to discover whether the additional parameter makes a statistically significant improvement to the fit of the model to observed data for these comparisons (Yang, Goldman, and Friday 1994; Huelsenbeck and Rannala 1997). As noted above, equation (1) is recovered from equation (4) when $f = 0$. This indicates that +F models are nested within the new +gwF models and differ by one degree of freedom. The test statistic of the difference in maximized log-likelihoods is compared with a $\chi_1^2/2$ distribution (Yang, Goldman, and Friday 1994; Whelan and Goldman 1999) and a value in excess of the 95%-point of this distribution ($\chi^2_{1,95\%}/2 = 1.92$) indicates the rejection of the simpler (+F) model in favor of the more complex (+gwF) model.

An asymptotically equivalent test of the significance of the +gwF model is to estimate a confidence interval (CI) for the parameter f and see if that CI excludes 0, the value of f at which the +gwF model reduces to the +F form. The required CI can be estimated by the curvature method (Kendall and Stuart 1979, pp. 45–49; Yang, Goldman, and Friday 1995).

A New Generalization of Nucleotide Substitution Models

For models of nucleotide substitution, standard parameters $\tilde{q}_{ij}$ or $\tilde{\pi}_j$ and $\tilde{s}_{ij}$ have not been derived from large databases. Instead, $\tilde{s}_{ij}$ have often been taken to equal 1 for all substitutions, or else taken to equal κ (an unknown parameter, often to be estimated from observed data) when the substitution i to j is a transition (A ↔ G, C ↔ T) and 1 otherwise (transversions A, G ↔ C, T). These exchangeability parameters have been used either without consideration of observed π_j (effectively setting $\pi_j = 1/4$ for all j), as in the models JC (Jukes and Cantor 1969) and K2P (Kimura 1980), or with the π_j estimated from the observed sequences, as in the models FEL (Felsenstein 1981) and HKY (Hasegawa, Kishino, and Yano 1985). Following the notation used for amino acid models, we note that the nucleotide models FEL and HKY might equally be denoted JC+F and K2P+F, respectively.

Development of +gwF models for nucleotide substitution follows the same arguments as for amino acid replacement models. In the absence of nucleotide frequency values from databases, we set $\tilde{\pi}_j = 1/4$ for all j. Because this is a constant and all rates q_{ij} are ultimately scaled so that the mean rate of evolution is 1, for clarity we can omit the $\tilde{\pi}_j$ and the scaling factor from our equations. Using $\tilde{s}_{ij} = 1$ for all nucleotides i and j we get a generalized version of the JC model, denoted JC+gwF, defined by

$$q_{ij} = \frac{\pi_j^{1-f}}{\pi_i^f}. \tag{5}$$

This reduces to the FEL (JC+F) model when $f = 0$. Alternatively, we can derive a generalized version of the K2P model, denoted K2P+gwF, defined by

$$q_{ij} = \begin{cases} \kappa \dfrac{\pi_j^{1-f}}{\pi_i^f} & \text{when the change } i \text{ to } j \text{ is a transition} \\ \dfrac{\pi_j^{1-f}}{\pi_i^f} & \text{when the change } i \text{ to } j \text{ is a transversion.} \end{cases} \tag{6}$$

This reduces to the HKY (K2P+F) model when $f = 0$. The case of $\pi_z = 0$ is ignored because it is unlikely to arise for realistic data sets. Application of these models is again by standard ML methods, with the new parameter f estimated by ML simultaneously with branch lengths and other parameters of the substitution model (e.g., κ in eq. 6).

Utility of the new models in terms of any improved fit to observed data may be assessed via the following comparisons:

FEL (JC+F) versus JC+gwF
FEL+Γ (JC+F+Γ) versus JC+gwF+Γ
HKY (K2P+F) versus K2P+gwF versus REV
HKY+Γ (K2P+F+Γ) versus K2P+gwF+Γ versus REV+Γ.

Note that the general time reversible model for DNA substitution (REV; Yang 1994*b*) is more complex than other existing models and subsumes the +gwF models described here. It is of interest to see whether this extra complexity is justified relative to the K2P+gwF model. All the above comparisons are between nested models and can be assessed using LRTs with test statistics based on a χ^2_1 distribution (all FEL vs. JC+gwF and HKY vs. K2P+gwF comparisons, with or without +Γ) or a χ^2_3 distribution (all K2P+gwF vs. REV comparisons).

Data Used for Investigating New Models

To investigate the performance of the new +gwF models of amino acid replacement, we analyzed multiple sequence alignments from the BRKALN database (D. Jones, personal communication). Our working version of this database contains 182 aligned protein families, totaling 3,905 sequences. These data have previously been used to estimate the WAG model of amino acid replacement (Whelan and Goldman 2001) and amino acid replacement models specific to different protein secondary structures (Thorne, Goldman, and Jones 1996; Goldman, Thorne, and Jones 1998). For each alignment in our database only one, previously established, topology was considered; see Whelan and Goldman (2001) for details of how these topologies were derived. Although there are likely to be some suboptimal topologies amongst those used, we expect that on the whole the topologies are optimal or near optimal and no systematic errors or biassing of results will occur (see also Whelan and Goldman 2001).

Investigation of the performance of +gwF models of nucleotide substitution used two data sets: the well-known 895-bp alignment of mtDNA sequences of five primates (Brown et al. 1982), and an alignment of 2000 bp from the *gag* and *pol* genes of six isolates of HIV-1 (Goldman, Anderson, and Rodrigo 2000).

Results and Discussion

Generalized Amino Acid Replacement Models

Likelihoods and Fit of Models to Data

Figure 1 shows the log-likelihood surface for an alignment of catalase sequences, one of the families of our database, under the WAG+gwF+Γ model (log-likelihood maximized over branch lengths for the single topology studied, as a function of parameters f and α of the +gwF and +Γ components, respectively). The log-likelihood surfaces for all analyses performed were generally similarly well-behaved (e.g., smooth, no strong correlation between parameters, easily identified maxima) and posed few difficulties for finding ML values and parameter estimates. Some care had to be used when parameter estimates were near to their boundaries (e.g., $\alpha \to \infty$, or $f \approx 1$ when some $\pi_z = 0$). An animated version of figure 1 can be viewed at http://www.ebi.ac.uk/goldman/publs/gwF.html, as can an animation illustrating how the instantaneous rate matrix Q and the log-likelihood vary with f.

Table 1, rows A–D, summarize the results of four model comparisons (described above) designed to show the utility of the +gwF method and applied to each of the 182 alignments in our database. In each comparison, it is evident that the +gwF models have an improved fit to the observed data for the majority of the alignments, with statistically significant improvements in approximately 70% of cases.

In a typical analysis it is likely that the choice of evolutionary model (e.g., JTT or WAG, +F or +gwF, no rate heterogeneity or +Γ) will be made with reference to the data, and not predetermined as with the comparisons above. For this reason, we also considered the

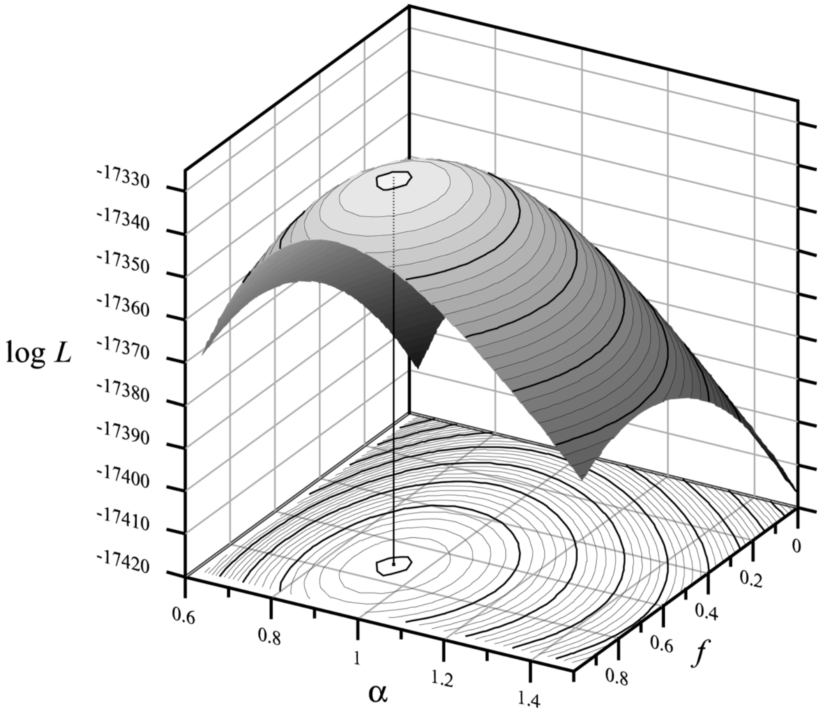

FIG. 1.—Typical log-likelihood surface. The surface shows the log-likelihood value maximized over branch lengths and as a function of parameters f and α for the WAG+gwF+Γ model applied to an alignment of 29 catalase sequences. The global maximum, log $L = -17329.68$ occurring when $f = 0.72$, $\alpha = 0.93$, is indicated and the contours and maximum point are also projected onto the (f, α)-plane.

following more realistic and commonly used model selection scenarios:

E: choose between JTT+F and JTT+F+Γ (LRT based on a $\overline{\chi}_1^2$ mixture distribution; Goldman and Whelan 2000, table 2); compare the better model with the +gwF version of the same model (LRT based on a χ_1^2 distribution);

F: choose between WAG+F and WAG+F+Γ; compare the better model with the +gwF version of the same model (LRTs as [E]);

G: choose between JTT+F and WAG+F, and between JTT+F+Γ and WAG+F+Γ (in each case, simply choose the model with higher likelihood because there is no difference in complexity); then choose the better of the preferred models (test based on a $\overline{\chi}_1^2$ mixture distribution; we neglect the possibility that the competing models may not be nested at this stage); finally, compare the resulting model with the +gwF version of the same model (χ_1^2 LRT).

Scenario G probably most resembles a typical phylogenetic analysis. Summaries of the results of applying scenarios E–G to each of our 182 alignments are given in table 1. We still see the +gwF models giving a significant improvement over their +F versions in the final

Table 1
Summary of Numbers of Cases for Which +gwF Models of Amino Acid Replacement are Significantly Better than +F Models Under Different Testing Scenarios

	MODELS COMPARED		SIGNIFICANT IMPROVEMENTS UNDER +GWF, BASED ON			
			LRT		CI	No. in Common[b]
ROW	Base Model	Compared with	No. of Cases	%[a]	No. of Cases	
A . . .	JTT+F	JTT+gwF	125	68.7	126	125
B . . .	JTT+F+Γ	JTT+gwF+Γ	128	70.3	130	127
C . . .	WAG+F	WAG+gwF	126	69.2	126	126
D . . .	WAG+F+Γ	WAG+gwF+Γ	126	69.2	130	126
E . . .	Better of JTT+F, JTT+F+Γ	Base model +gwF	126	69.2	—[c]	—[c]
F. . . .	Better of WAG+F, WAG+F+Γ	Base model +gwF	126	69.2	—[c]	—[c]
G . . .	Best model without +gwF	Base model +gwF	127	69.8	129	127

NOTE.—LRT: likelihood ratio test results; CI: confidence interval test results.

[a] % Calculated out of 182 aligned amino acid sequence families.

[b] Cases in common between LRT and CI tests.

[c] Comparison not performed.

stage of each scenario in approximately 70% of cases. Of the 182 alignments studied, 116 show the +gwF model giving a significant improvement in the fit of evolutionary model to observed data in every one of the seven LRT comparisons A–G.

When the CI procedure is followed for assessing the significance of comparisons A–D and G, the results correlate very well with those already described (see table 1). The +gwF model is significantly better than +F for almost the same sets of families when assessed by the LRT or CI methods. All the results of LRTs and CI tests indicate that assessment of the improvement of +gwF models over +F models is not strongly affected by other components (e.g., base model JTT or WAG; no rate heterogeneity or $+\Gamma$) of the models used.

We also note that in comparison G, 142 out of the 182 alignments (78.0%) lead to a model based on the WAG standard model being preferred, reinforcing evidence suggesting that WAG is a useful alternative to JTT (Whelan and Goldman 2001).

Figure 2 shows the relationship between the improvements in log-likelihood ($\Delta\log L$) and $\hat{f}$ for comparison B of table 1, and the distribution of $\hat{f}$ over 182 sequence alignments under the JTT+gwF+Γ model. This plot is typical of those obtained for any of the comparisons and models described in this article. Recall that when $\hat{f} = 0$, +gwF reduces to +F and $\Delta\log L = 0$ necessarily. In figure 2 the distinction is made between cases when the +gwF model is statistically preferred to +F according to the LRT, and cases where it is not (i.e., whether or not $\Delta\log L > 1.92$). Over the 182 alignments studied, and for whichever +gwF model is used, we find that the statistically significant inferred values of $\hat{f}$ are in the main distributed between 0.25 and 1.5, with a peak around 0.6 (e.g., under the best +gwF model as selected by scenario G above, the distribution of values of $\hat{f}$ significantly different from 0 has mean = 0.69, median = 0.64, standard deviation (SD) 0.42, interquartile range = 0.26).

Although a number of values of $\hat{f}$ shown in figure 2 are <0, notice that only one such is statistically significant ($\hat{f} = -2.36$ with $\Delta\log L = 2.09$)—this is the only statistically significant $\hat{f} < 0$ under any model studied in this article, according to a LRT. No values of $\hat{f} < 0$ are significant according to the CI test.

As noted above, if the observed π_j values are close to the database values $\tilde{\pi}_j$ then the dependency of +gwF models on f is reduced. Measuring the similarity of the π_j and $\tilde{\pi}_j$ by the G statistic of the LRT test of the goodness-of-fit of the π_j to the $\tilde{\pi}_j$ (Sokal and Rohlf 1994, p. 690; $G = 0$ when $\pi_j = \tilde{\pi}_j$ for all j, and increasing values of G indicate decreasing similarity), we find no correlations between G and estimates $\hat{f}$ but negative correlations between G and the widths of CIs for f (results not shown). This demonstrates that similarity of the π_j and $\tilde{\pi}_j$ has no biassing effect on estimates of f, but reduces the information available about f (e.g., making it harder to reject the null hypothesis $f = 0$).

Behavior of +gwF Models

The statistical tests described above indicate that +gwF models of amino acid sequence evolution give significant improvements in many cases. It is also interesting to see if the parameter f is ‘‘well-behaved.’’ If so, this is a further indication that the new model is accurately reflecting a biological feature of sequence evolution and not simply improving the fit of model to data by the addition of a meaningless parameter to a poor model. We performed various analyses to investigate this.

Figure 3*A* shows the correlation of estimates $\hat{f}$ derived from the 182 alignments in our database under the WAG+gwF and WAG+gwF+Γ models and distinguishes the cases where neither, one, or both of these models give a significant improvement over their respective +F versions. The estimates $\hat{f}$ are generally approximately the same under the two models, and the correlation is high (for the 121 alignments for which WAG+gwF and WAG+gwF+Γ are significantly better than WAG+F and WAG+F+Γ, respectively, Spearman's rank correlation coefficient $\rho = 0.87$, $P < 0.001$). Figure 3*B* illustrates the same effect in the comparison of the JTT+gwF and WAG+gwF models. Here, the estimates $\hat{f}$ are less closely related ($\rho = 0.67$, $P < 0.001$, for the 119 cases in which JTT+gwF and WAG+gwF are significantly better than JTT+F and WAG+F, respectively) and it appears that choice of base model has a greater effect on inferred values $\hat{f}$ than does inclusion of $+\Gamma$. Nevertheless, estimates of f are clearly not strongly affected by either the base model or other components of the model used.

The interaction of the +gwF and other components of evolutionary models can also be assessed by looking at the relative improvements in likelihood when one component (e.g., +gwF) is added either with or without another component (e.g., $+\Gamma$) already present. If the components are capturing information on entirely independent factors in sequence evolution, the changes in likelihood due to the incorporation of each component will be independent. We investigated the increase in likelihood caused by the replacement of +F by +gwF and the incorporation of $+\Gamma$, each either with or without the other already being applied. For example, figure 4*A* plots the relationship between $\Delta\log L$(WAG±gwF) (defined as $\log L$(WAG+gwF) − $\log L$(WAG+F)) and $\Delta\log L$(WAG±gwF+Γ) (defined as $\log L$(WAG+gwF+Γ) − $\log L$(WAG+F+Γ)) for the 182 sequence alignments in our database. This shows how the use of $+\Gamma$ affects the increase in likelihood afforded by the use of +gwF when using the WAG base model. We see there is a good correlation between the log-likelihood differences ($\rho = 0.98$, $P < 0.001$), indicating that in general +gwF and $+\Gamma$ are responding independently to different evolutionary signals in the data. For the larger improvements, note that the log-likelihood improvement gained from using +gwF tends to be greater in the absence of $+\Gamma$ than with it (i.e., points fall below the 45° line). This nonindependence may be because without $+\Gamma$, +gwF may also be able to adapt the evolutionary model to represent some of the effects of rate heterogeneity, whereas with $+\Gamma$ implemented these effects are accounted for elsewhere and are not picked up by +gwF.

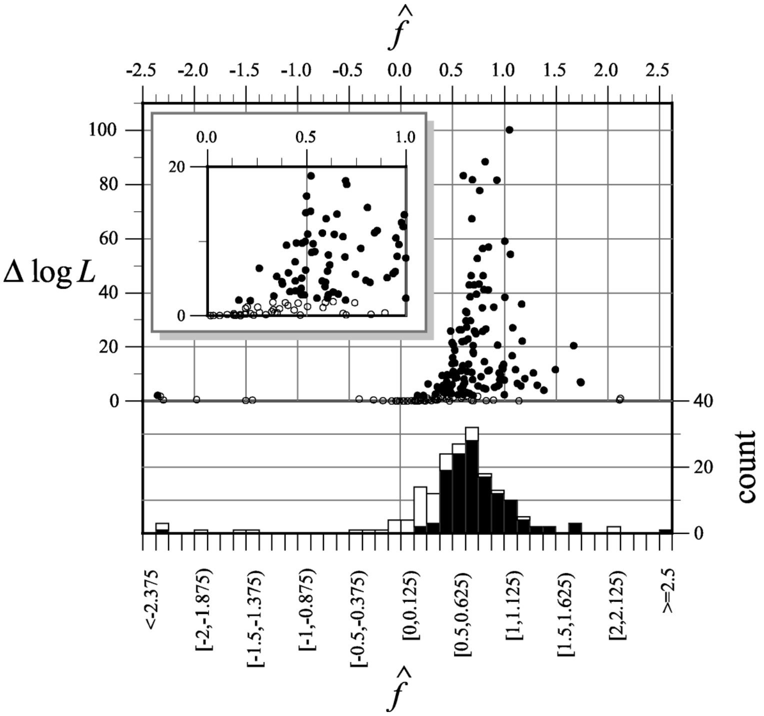

FIG. 2.—Log-likelihood improvements ($\Delta\log L$) and the distribution of optimal values $\hat{f}$ for the JTT+gwF+Γ model (table 1, row B). The upper part of the graph shows how $\Delta\log L$ and $\hat{f}$ are related for each of the 182 alignments in our database. Solid points are used for cases where the JTT+gwF+Γ model is significantly better than JTT+F+Γ according to the LRT; open points are used for cases where it is not. The inset is an expanded version of the most densely populated area of the plot. The lower part of the graph shows the distribution of estimates $\hat{f}$, with significant cases plotted in black and nonsignificant cases in white. The distribution for the significant cases has mean = 0.83, median = 0.69, SD = 1.30, interquartile range = 0.39.

Conversely, figure 4*B* plots the relationship between $\Delta\log L$(WAG+F±Γ) (defined as log *L*(WAG+F+Γ) − log *L*(WAG+F)) and $\Delta\log L$(WAG+gwF±Γ) (defined as log *L*(WAG+gwF+Γ) − log *L*(WAG+gwF)). This shows how the use of +gwF affects the increase in likelihood afforded by the use of +Γ when using the WAG base model. Note again the very high correlation ($\rho = 1.00$, $P < 0.001$) and adherence to the 45° line, confirming that these two model components are largely independent.

Choice of base model has a greater effect on the improvement in likelihood afforded by the +gwF method. This is illustrated by comparing $\Delta\log L$(JTT±gwF) and $\Delta\log L$(WAG±gwF), (fig. 4*C*) and $\Delta\log L$(JTT±gwF +Γ) and $\Delta\log L$(WAG±gwF+Γ) (fig. 4*D*). Note that the correlations, though lower, are still highly significant (both have $\rho = 0.93$, $P < 0.001$). These plots also show a tendency for points to fall below the 45° line, indicating that greater improvements in likelihoods tend to be obtained with the JTT base model than with WAG. This may be because of the superiority of WAG over JTT for the majority of alignments in our database (see above and Whelan and Goldman 2000), leading to a greater scope for +gwF to deliver improvements in likelihoods under JTT than under WAG.

We also investigated the effect the use of +gwF has on estimates of the parameter α of the Γ distribution that describes heterogeneity of evolutionary rates amongst sequence sites. Figure 5*A* illustrates this with a comparison of values $\hat{\alpha}$ estimated under the WAG+F+Γ and WAG+gwF+Γ models. These are shown for the 133 alignments in our database for which the WAG+gwF+Γ model is significantly better than the WAG+gwF model—in other words, alignments for which there is reasonable evidence that +Γ is contributing to an improved fit of evolutionary model to observed data. We see a very high correlation of estimates $\hat{\alpha}$ ($\rho = 0.97$, $P < 0.001$), with the estimates being almost equal under the two models for most data sets. This also confirms that the +gwF and +Γ components of evolutionary models are broadly independent, with no awkward interactions between them suggesting any weaknesses in their joint use.

Heterogeneity of Evolutionary Rate Within Amino Acid Sequences

Many studies have used a Γ distribution to study heterogeneity of evolutionary rates amongst the sites of DNA sequences (see Yang 1996). When these models are used for the evolutionary analysis of protein-coding DNA sequences it is generally assumed that rate heterogeneity is due to the genetic code affecting the conservativeness of different nucleotide substitutions at different codon positions and selection for protein function. Γ distributions, now modeling simply selection for function, are also often found to improve the statistical fit of evolutionary models to amino acid sequence data. We know of no compilations of estimates of α obtained in amino acid sequence studies, and so it is of interest to look at the distribution of values we obtain from our database. This is shown in figure 5*B*, which is derived from the 133 families in our database for which the optimum model defined by scenario G above includes a Γ distribution. We note an approximately bell-shaped distribution, with its upper tail somewhat elongated, and find in general that the estimates of α are higher than are often found for DNA sequence alignments (see Yang 1996 for a compilation of values estimated from DNA). This indicates lower levels of rate heterogeneity among amino acid positions of protein sequences than among the sites of nucleotide sequences, as would be expected from the biological interpretations above.

Effects on Branch Length Estimates

Our primary interest so far has been to see if the new +gwF method is able to improve the fit of evolutionary model to observed data. To investigate the potential of +gwF for improving phylogenetic inferences, we studied its effect on inferred branch lengths. For almost all families in our database, branch length estimates were only very slightly affected by the addition of the +gwF method to the JTT or WAG base model (results not shown). The implications of this finding are discussed below.

Generalized Nucleotide Replacement Models

Table 2 shows the results of various LRTs involving the +gwF method for the mtDNA and HIV-1 data sets described above. Qualitatively equivalent results are obtained whether or not the models tested include the +Γ option (i.e., top/bottom half of table 2). For the mtDNA sequences, the +gwF models do not give a significant improvement over their corresponding +F versions. For the HIV-1 sequences, the K2P+gwF and K2P+gwF+Γ models do lead to significant improvements over the HKY and HKY+Γ substitution models, respectively. In both cases, however, the corresponding REV models are significantly better still.

Consequently, neither the mtDNA or HIV-1 data set gives an example of a case in which a +gwF model would be statistically the best description of the sequences' evolution. For this case to arise we need, for example, a data set where K2P+gwF+Γ is preferred to HKY+Γ and where REV+Γ is not significantly better than K2P+gwF+Γ.

Our use of $\tilde{\pi}_j = 1/4$ for all nucleotides j may be one of the reasons that the +gwF nucleotide substitution models do not perform well. Whereas for the amino acid models the effect of the database values $\tilde{\pi}_j$ is to incorporate information on an underlying distribution of amino acids common to all protein sequences, we have not yet included equivalent information into the nucleotide models. Perhaps in future the estimation of $\tilde{\pi}_j$ from large nucleotide databases, and indeed the (empirical) estimation of $\tilde{Q}$ to derive replacements for the commonly used (parametrically inspired) values of 1 or κ for s_{ij}, will lead to greater success. Because the +gwF models of nucleotide substitution are subsumed by the REV model, which it is computationally feasible to optimize for most data sets, it is not clear that the +gwF models will ever offer any important advantage.

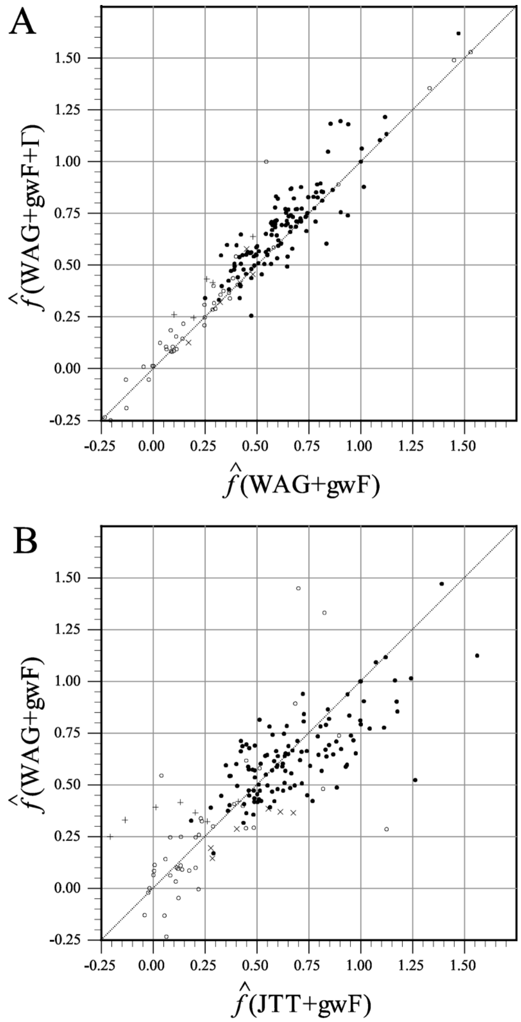

FIG. 3.—Correlation of estimates $\hat{f}$ under different +gwF models. *A,* comparison of $\hat{f}$ estimated under WAG+gwF and WAG+gwF+Γ models. Cases distinguished are: ○, neither model significantly better than its +F version; ×, WAG+gwF significantly better than WAG+F; +, WAG+gwF+Γ significantly better than WAG+F+Γ; ●, both models significantly better than their +F versions. *B,* comparison of $\hat{f}$ estimated under JTT+gwF and WAG+gwF models. Cases distinguished are: ○, neither model significantly better than its +F version; ×, JTT+gwF significantly better than JTT+F; +, WAG+gwF significantly better than WAG+F; ●, both models significantly better than their +F versions.

Conclusions

We have shown that the new +gwF versions of two standard amino acid replacement models lead to a significantly improved fit to observed data in approximately 70% of 182 multiple sequence alignments analyzed. Evidently it can be useful to use both database-derived amino acid frequencies ($\tilde{\pi}_j$) and data-derived frequencies (π_j) in the more general manner indicated in equations (2)–(4). The assumptions of the fundamental nature of the exchangeabilities $\tilde{s}_{ij}$ and that rates of replacement be linearly proportional to the (observed) π_j are not necessarily correct.

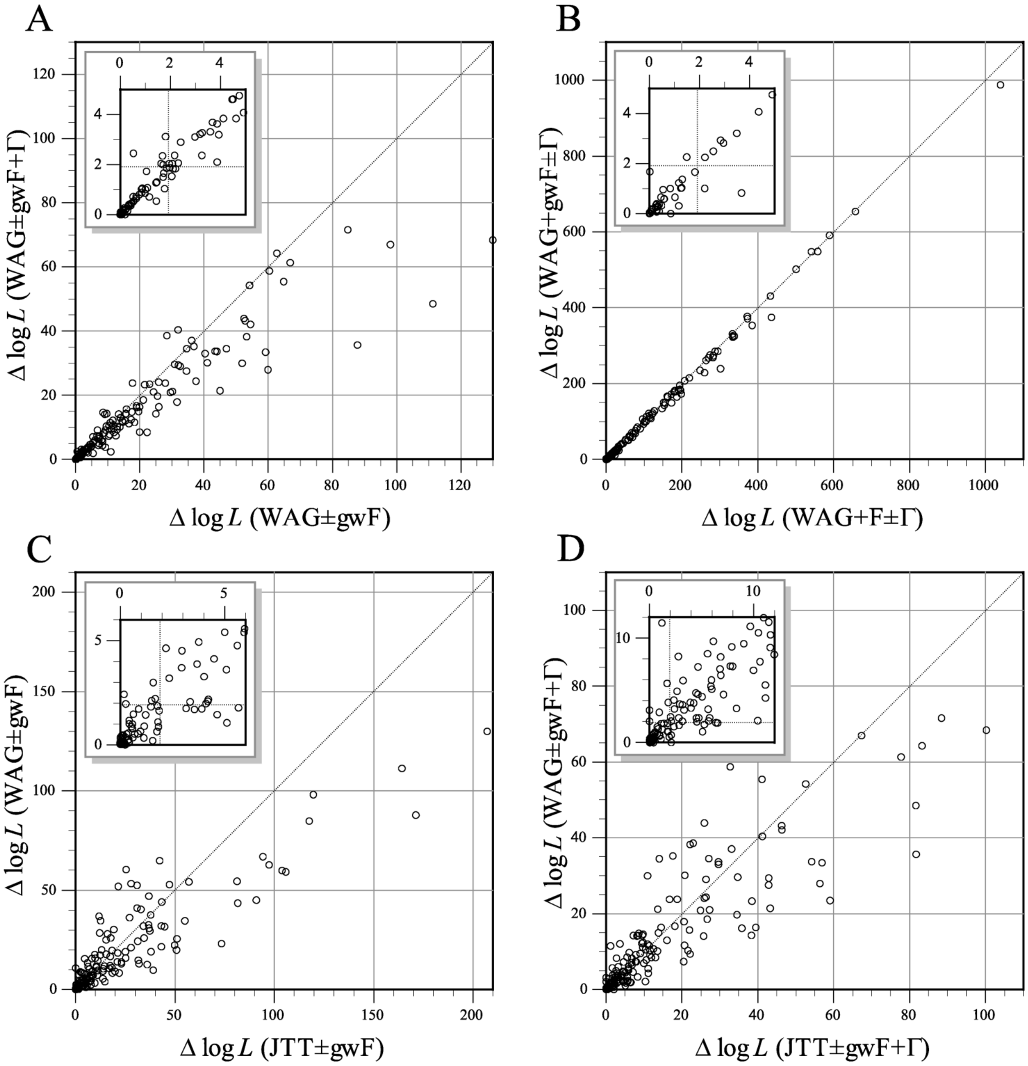

FIG. 4.—Effects of other model components on likelihood improvements afforded by +gwF and +Γ. Insets are expanded versions of the most densely populated areas of the plot. *A,* effect of +Γ on likelihood improvement given by +gwF with WAG base model. *B,* effect of +gwF on likelihood improvement given by +Γ with WAG base model. *C,* effect of choice of base model on likelihood improvement given by +gwF. *D,* effect of +Γ and choice of base model on likelihood improvement given by +gwF.

All the analyses performed indicate that inference under +gwF amino acid models is well-behaved, with no undue interactions between the new parameter f and other parameters and having stability with respect to other variations in models. This suggests that the +gwF models are reacting to evolutionary information in aligned sequence data sets that is not detected by existing models. Various analyses of CIs for $\hat{f}$ and likelihood scores show that f may be reliably estimated for a range of data sets representative of those used in typical phylogenetic studies.

In the expectation that improved evolutionary models may lead to improved understanding of the process of evolution, it seems justified to recommend that +gwF versions of the Dayhoff (Dayhoff, Schwartz, and Orcutt 1978), JTT (Jones, Taylor, and Thornton 1992) and WAG (Whelan and Goldman 2001) models be considered when analyzing amino acid sequences. The same recommendation can also be extended to replacement models derived specifically for mitochondrial (Adachi and Hasegawa 1996; Yang, Nielsen, and Hasegawa 1998) and chloroplast (Adachi et al. 2000) amino acid sequences. Given an inferred value of $\hat{f}$ for a particular data set we may now make an inference about the relative importance of selective or mutational forces acting on that protein, as described in the *Methods* section above.

We have not yet implemented the +gwF method within any codon models of sequence evolution (e.g., Goldman and Yang 1994; Muse and Gaut 1994; Nielsen and Yang 1998; Yang et al. 2000). These are effectively +F models, with instantaneous rates of codon substitutions typically being proportional to frequencies of replacing codons, estimated either (1) directly from the observed frequencies of each of the sense codons or (2) from the products of the observed frequencies of each nucleotide at each codon position. We also note that these models disallow instantaneous changes of more than one nucleotide within a codon, meaning that observed changes of two or more nucleotides within one

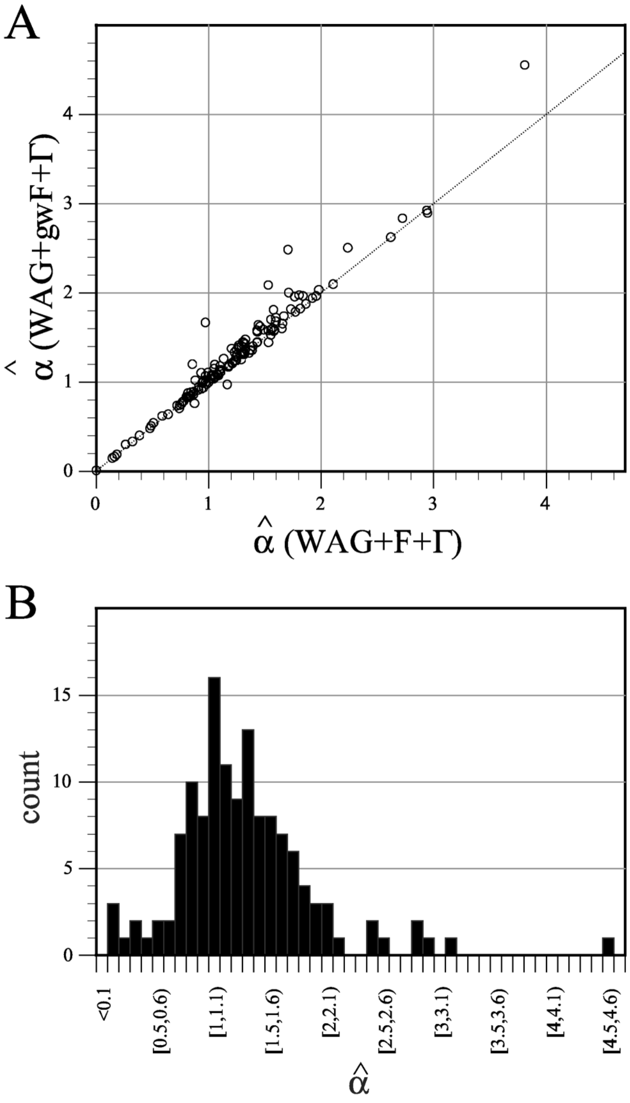

FIG. 5.—Estimates of the parameter α of the Γ distribution used to model rate heterogeneity. *A*, comparison of estimates under WAG+F+Γ and WAG+gwF+Γ models. *B*, distribution of estimates $\hat{\alpha}$ from 133 amino acid sequence alignments that indicate significant rate heterogeneity. See text for further details. The distribution has mean = 1.31, median = 1.25, SD = 0.61, interquartile range = 0.62.

codon have to be assumed to take place only via one or more intermediate codons. If these intermediates are rarely observed (in case (1) above), or consist of a combination of nucleotides rarely observed at those codon positions (in case (2) above), the observed changes are deemed virtually impossible by the model despite their perhaps common occurrence in a data set. This could lead to a poor fit of evolutionary model to observed data, misinterpretation of the evolutionary signal in a data set, and inaccurate phylogenetic inferences. These effects could be more pronounced in cases of unusual or extreme codon usage. In contrast, +gwF models could allow codon substitution rates to be inversely related to observed codon frequencies (cf. eq. 3 above with $f > 0$) and so could explain observed changes of two or three nucleotides within a codon by permitting the appearance and then rapid loss of the necessary but rarely observed intermediates. Future work will investigate the utility of codon models incorporating the +gwF method.

Table 2
Likelihood Ratio Tests of +gwF Models of DNA Substitution

		ΔLOG*L* FOR THE FOLLOWING DATA SET:	
BASE MODEL	COMPARED WITH	mtDNA	HIV-1
FEL (JC+F)	JC+gwF	0.799	0.513
HKY (K2P+F)	K2P+gwF	1.285	**3.592**
K2P+gwF	REV	—[a]	12.820
FEL+Γ (JC+F+Γ).	JC+gwF+Γ	0.002	0.880
HKY+Γ (K2P+F+Γ)	K2P+gwF+Γ	0.004	**1.997**
K2P+gwF+Γ	REV+Γ	—[a]	11.318

NOTE.—Values in bold indicates comparisons where the model incorporating +gwF is preferred to the competing model. All statistical tests compare Δlog*L* with a $\chi^2_1/2$ distribution except those involving the REV model, which use a $\chi^2_3/2$ distribution.

[a] Comparison not performed because the base model is rejected for this data set.

Because of the large scale of the analyses we have performed, only one topology could be considered for each amino acid sequence alignment in our database. Consequently, it has not been possible to study the effect the new models have on inferred topologies. The analysis we have reported of branch length estimates under different models does not lead us to expect systematic changes in branch length estimates or different inferred topologies. Until such indications are found, the new +gwF models seem to be contributing mostly to improved fit of models to data and thus to our understanding of the processes and patterns of amino acid sequence evolution.

Acknowledgments

The authors are supported by the Wellcome Trust. We thank two anonymous referees for comments that inspired substantial improvements to the article.

LITERATURE CITED

ADACHI, J., and M. HASEGAWA. 1996. Model of amino acid substitution in proteins encoded by mitochondrial DNA. J. Mol. Evol. **42**:459–468.

ADACHI, J., P. J. WADDELL, W. MARTIN, and M. HASEGAWA. 2000. Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast DNA. J. Mol. Evol. **50**:348–358.

ANISIMOVA, M., J. P. BIELAWSKI, and Z. YANG. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol. Biol. Evol. **18**:1585–1592.

BROWN, W. M., E. M. PRAGER, A. WANG, and A. C. WILSON. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18**:225–239.

CAO, Y., J. ADACHI, A. JANKE, S. PÄÄBO, and M. HASEGAWA. 1994. Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of a tree based on a single gene. J. Mol. Evol. **39**:519–527.

CUNNINGHAM, C. W., H. ZHU, and D. M. HILLIS. 1998. Best-fit maximum-likelihood models for phylogenetic inference: empirical tests with known phylogenies. Evolution **52**:978–987.

DAYHOFF, M. O., R. M. SCHWARTZ, and B. C. ORCUTT. 1978. A model of evolutionary change in proteins. Pp. 345–352 *in* M. O. DAYHOFF, ed. Atlas of protein sequence and structure, Vol. 5, Suppl. 3. National Biomedical Research Foundation, Washington, DC.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

GOLDMAN, N. 1993*a.* Simple diagnostic statistical tests of models for DNA substitution. J. Mol. Evol. **37**:650–661.

———. 1993*b*. Statistical tests of models of DNA substitution. J. Mol. Evol. **36**:182–198.

GOLDMAN, N., J. P. ANDERSON, and A. G. RODRIGO. 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. **49**:652–670.

GOLDMAN, N., J. L. THORNE, and D. T. JONES. 1998. Assessing the impact of secondary structure and solvent accessibility on protein evolution. Genetics **149**:445–458.

GOLDMAN, N., and S. WHELAN. 2000. Statistical tests of gamma-distributed rate heterogeneity in models of sequence evolution in phylogenetics. Mol. Biol. Evol. **17**:975–978.

GOLDMAN, N., and Z. YANG. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11**:725–736.

HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**:160–174.

HUELSENBECK, J. P., and B. RANNALA. 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. Science **276**:227–232.

JONES, D. T., W. R. TAYLOR, and J. M. THORNTON. 1992. The rapid generation of mutation data matrices from protein sequences. CABIOS **8**:275–282.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. MUNRO, ed. Mammalian protein metabolism, Vol. 3. Academic Press, New York.

KENDALL, M., and A. STUART. 1979. The advanced theory of statistics. 4th edition. Vol. 2. Charles Griffin, London.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

KUHNER, M. K., and J. FELSENSTEIN. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol. Biol. Evol. **11**:459–468 (see also: Erratum. Mol. Biol. Evol. **12**:525 [1995]).

LIÒ, P., and N. GOLDMAN. 1998. Models of molecular evolution and phylogeny. Genome Res. **8**:1233–1244.

MUSE, S. V., and B. S. GAUT. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol. Biol. Evol. **11**:715–724.

NIELSEN, R., and Z. YANG. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148**:929–936.

OTA, R., P. J. WADDELL, M. HASEGAWA, H. SHIMODAIRA, and H. KISHINO. 2000. Appropriate likelihood ratio tests and marginal distributions for evolutionary tree models with constraints on parameters. Mol. Biol. Evol. **17**:798–804.

PHILIPPE, H., and A. GERMOT. 2000. Phylogeny of eukaryotes based on ribosomal RNA: long-branch attraction and models of sequence evolution. Mol. Biol. Evol. **17**:830–834.

POSADA, D., and K. A. CRANDALL. 2001. Simple (wrong) models for complex trees: a case from the Retroviridae. Mol. Biol. Evol. **18**:271–275.

SOKAL, R. R., and F. J. ROHLF. 1994. Biometry. 3rd edition. W. H. Freeman and Co., New York.

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, and D. M. HILLIS. 1996. Phylogenetic inference. Pp. 407–514 *in* D. M. HILLIS, C. MORITZ, and B. K. MABLE, eds. Molecular systematics. Sinauer, Sunderland, Mass.

TAKEZAKI, N., and T. GOJOBORI. 1999. Correct and incorrect vertebrate phylogenies obtained by the entire mitocohndrial DNA sequences. Mol. Biol. Evol. **16**:590–601.

THORNE, J. L. 2000. Models of protein sequence evolution and their applications. Curr. Opin. Genet. Dev. **10**:602–605.

THORNE, J. L., N. GOLDMAN, and D. T. JONES. 1996. Combining protein evolution and secondary structure. Mol. Biol. Evol. **13**:666–673.

WHELAN, S., and N. GOLDMAN. 1999. Distributions of statistics used for the comparison of models of sequence evolution in phylogenetics. Mol. Biol. Evol. **16**:1292–1299.

———. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. **18**:691–699.

WHELAN, S., P. LIÒ, and N. GOLDMAN. 2001. Molecular phylogenetics: state of the art methods for looking into the past. Trends Genet. **17**:262–272.

YANG, Z. 1994*a.* Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. **39**:306–314.

———. 1994*b*. Estimating the pattern of nucleotide substitution. J. Mol. Evol. **39**:105–111.

———. 1996. Among-site rate variation and its impact on phylogenetic analysis. TREE **11**:367–372.

YANG, Z., and J. P. BIELAWSKI. 2000. Statistical methods for detecting molecular adaptation. TREE **15**:496–503.

YANG, Z., N. GOLDMAN, and A. FRIDAY. 1994. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. Mol. Biol. Evol. **11**:316–324.

———. 1995. Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. Syst. Biol. **44**: 384–399.

YANG, Z., R. NIELSEN, N. GOLDMAN, and A.-M. K. PEDERSEN. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155**:431–449.

YANG, Z., R. NIELSEN, and M. HASEGAWA. 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. Mol. Biol. Evol. **15**:1600–1611.

YODER, A. D., and Z. YANG. 2000. Estimation of primate speciation dates using local molecular clocks. Mol. Biol. Evol. **17**:1081–1090.

ZHANG, J. 1999. Performance of likelihood ratio tests of evolutionary hypotheses under inadequate substitution models. Mol. Biol. Evol. **16**:868–875.

MANOLO GOUY, reviewing editor

Accepted June 2, 2002