J Mol Evol (2004) 58:661–672
DOI: 10.1007/s00239-004-2588-9

JOURNAL OF MOLECULAR EVOLUTION

Bayesian Estimation of Positively Selected Sites

John P. Huelsenbeck,[1] Kelly A. Dyer[2]

[1] Section of Ecology, Behavior and Evolution, Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093-0116, USA
[2] Department of Biology, University of Rochester, Rochester, NY 14627, USA

Received: 29 October 2002 / Accepted: 9 December 2003

Abstract. In protein-coding DNA sequences, historical patterns of selection can be inferred from amino acid substitution patterns. High relative rates of nonsynonymous to synonymous changes ($\omega = d_N/d_S$) are a clear indicator of positive, or directional, selection, and several recently developed methods attempt to distinguish these sites from those under neutral or purifying selection. One method uses an empirical Bayesian framework that accounts for varying selective pressures across sites while conditioning on the parameters of the model of DNA evolution and on the phylogenetic history. We describe a method that identifies sites under diversifying selection using a fully Bayesian framework. Similar to earlier work, the method presented here allows the rate of nonsynonymous to synonymous changes to vary among sites. The significant difference in using a fully Bayesian approach lies in our ability to account for uncertainty in parameters including the tree topology, branch lengths, and the codon model of DNA substitution. We demonstrate the utility of the fully Bayesian approach by applying our method to a data set of the vertebrate β-globin gene. Compared to a previous analysis of this data set, the hierarchical model found most of the same sites to be in the positive selection class, but with a few striking exceptions.

Correspondence to: John P. Huelsenbeck; *email:* johnh@biomail.ucsd.edu

Key words: Bayesian estimation — Codon model — Hierarchical Bayes — Markov chain Monte Carlo

Introduction

Selection at the molecular level can be detected by comparing the rates of nonsynonymous and synonymous substitutions across protein-coding DNA sequences. Equal rates of both types of substitutions signify neutral evolution, while an overabundance of synonymous substitutions, which result in few amino acid changes, indicates purifying selection. An excess of nonsynonymous substitutions is unequivocal evidence for positive selection; these events are especially important to identify because they record historical episodes of adaptive evolution. A few loci where positive selection has been identified in this manner are the human major histocompatibility complex (Hughes and Nei 1988), the HIV-1 envelope gene (Nielsen and Yang 1998), abalone sperm lysins (Lee et al. 1995), and primate stomach lysozymes (Messier and Stewart 1997).

Until recently, models identifying events of positive selection assumed equal selection pressures on all amino acids in a sequence. However, functional constraints can change across a gene, thus rendering this assumption biologically unreasonable. In fact, in almost all proteins where positive selection has been identified only a few sites were found responsible for the adaptive evolution (Hughes and Nei 1988; Yokoyama and Yokoyama 1996). Failure to ac-

commodate varied selection pressures among amino acid sites may result in overlooking or underestimating positive selection (Nielsen 1997).

Nielsen and Yang (1998) developed a codon-based model of substitution that incorporated varying selective intensities among amino acid sites. In their model, the nonsynonymous/synonymous rate ratio ($\omega = d_N/d_S$) for a site could fall into one of three categories: $\omega = 0$ for purifying selection (all substitutions are synonymous), $\omega = 1$ for neutral evolution (the synonymous and nonsynonymous rates are equal), and $\omega > 1$ for positive selection (nonsynonymous substitutions occur at a higher rate than synonymous substitutions). They also consider a model in which the nonsynonymous/synonymous rate ratio across sites is distributed between 0 and 1 using a truncated gamma distribution. This is the first model we are aware of that allows the nonsynonymous/synonymous rate ratio to be chosen from a distribution and thus to vary across sites. Yang et al. (2000) considered a large number of alternative models for allowing constraints to change across the sequence. Many of these models were quite complex and computationally intensive. However, the ability to identify positively selected sites was robust to choice of model.

The identification of sites under positive selection has relied on the empirical Bayes method (Nielsen and Yang 1998); the probability that each site is in the positively selected category is calculated conditional on the information at the tips of the tree. The conditioning is carried out using Bayes' rule and the free parameters of the model are fixed at reasonable values (the maximum likelihood value for each parameter). Here, the parameters of the model include a tree with branch lengths and parameters of the substitution process, such as the transition/transversion rate ratio, the nonsynonymous/synonymous rate ratio, and the frequencies of the 61 amino- acid-coding codons. The empirical Bayes method has a few disadvantages. Among them are the facts that it does not account for uncertainty in the model parameters and the prior is a function of the observations (Berger 1985).

In this paper, we use a fully Bayesian method rather than an empirical Bayes framework to identify positively selected sites. In a fully Bayesian analysis, the parameters of the model are integrated over a prior distribution, thereby accounting for uncertainty in the parameters. The high-dimensional summations and integrations required in a Bayesian analysis are impossible to perform analytically; we use Markov chain Monte Carlo (MCMC) to approximate posterior probabilities. Using a fully Bayesian approach we determine the probability that each amino acid site is under positive selection (i.e., that it falls into the positively selected class). We apply the method to aligned DNA sequences of the β-globin gene from vertebrates (Yang et al. 2000).

Materials and Methods

Data

We assume that aligned DNA sequences are available for homologous protein coding genes. The data are contained in the matrix $\mathbf{X} = \{x_{ij}\}$, where $i = 1, 2, \ldots, s$ and $j = 1, 2, \ldots, c$; s is the number of sequences and c is the number of codons in each sequence. The information at the jth site is contained in the vector $\mathbf{x}_j = (x_{1j}, x_{2j}, \ldots, x_{sj})'$. As an example, consider the following aligned DNA matrix of $s = 4$ sequences:

1 AAT CGA
2 AAT CGA
3 AAC CGC
4 AAC AGC

Here, $\mathbf{x}_1 = (AAT, AAT, AAC, AAC)'$ and $\mathbf{x}_2 = (CGA, CGA, CGC, AGC)'$. There are two sites in the sequence ($c = 2$); at the first site, there is at least a single synonymous change (all codons code for asparagine), whereas at the second site there has been at least one nonsynonymous and one synonymous change (CGA and CGC code for arginine and AGC codes for serine).

We examine the β-globin sequences from $s = 17$ vertebrate species. Each sequence is 432 nucleotide sites in length (or $c = 144$ codon sites in length). The data were originally collected by Yang et al. (2000) from the EMBL and GenBank databases. Yang et al. (2000) found evidence of positive selection at from 2 to 6% of the 144 sites, depending upon the selection model used. Specifically, they found sites 7, 42, 48, 50, 54, 67, 85, and 123 to be under diversifying selection.

Model

We assume that substitutions occur according to a time-homogeneous Markov process. The instantaneous rate matrix of this process, $\mathbf{Q}$, is given by

$$\mathbf{Q} = \{q_{ij}\} = \begin{cases} \omega\kappa\pi_j: & \text{nonsynonymous transition} \\ \omega\pi_j: & \text{nonsynonymous transversion} \\ \kappa\pi_j: & \text{synonymous transition} \\ \pi_j: & \text{synonymous transversion} \\ 0: & i \text{ and } j \text{ differ at more than one position} \end{cases}$$

where ω is the nonsynonymous/synonymous rate ratio, κ is the transition/transversion rate ratio, and π_j is the stationary frequency of codon j (Goldman and Yang 1994; Muse and Gaut 1994). This matrix specifies the rate of change from codon i to codon j. The stationary codon frequencies are constrained to sum to one and are contained in the vector π [$\pi = (\pi_{AAA}, \pi_{AAC}, \pi_{AAG}, \ldots, \pi_{UUU})$]. The diagonal of the rate matrix $\mathbf{Q}$ is specified such that the row sums equal zero. Moreover, we rescale the matrix such that $-\sum\pi_i q_{ii} = 1$; this means that the branch lengths are in terms of expected number of substitutions per codon site, v. The rate matrix is 61×61 because the three stop codons are excluded. Also, note that the rate matrix is time reversible, as $\pi_i q_{ij} = \pi_j q_{ji}$ for all i and j. Practically speaking, this means that the phylogenetic tree can be arbitrarily rooted without changing the likelihood. The transition probabilities are calculated as $\mathbf{P}(v) = \{p_{ij}(v, \omega, \kappa, \pi)\} = e^{\mathbf{Q}v}$. This is the same model proposed by Nielsen and Yang (1998) and is simply a more parameter-rich version of the original model proposed by Goldman and Yang (1994) and Muse and Gaut (1994).

Like Nielsen and Yang (1998), we assume that the nonsynonymous/synonymous rate ratio (ω) is a random variable. Specifically, we assume the "M3" model of Yang et al. (2000): With probability p_1, ω is equal to ω_1, with probability p_2 ω is equal to ω_2, and with probability $p_3 = 1 - p_1 - p_2$ ω takes the value ω_3.We con-

strain $\omega_1 < \omega_2 < \omega_3$. Categories with $\omega > 1$ model sites under positive selection.

We assume that the sequences are related by a bifurcating phylogenetic tree, τ. Trees are labeled 1, 2, ..., $B(s)$, where $B(s) = (2s-5)!/2^{s-3}(s-3)!$ is the number of possible unrooted trees for s species. The tips of this tree are labeled $n_1, n_2, \ldots, n_s$ and the interior nodes of the tree are labeled $n_{s+1}, \ldots, n_{2s-2}$ in a postorder traversal of the tree (*i.e.*, from the tips of the tree to the root). The tree is rooted at tip node n_s. The length of the ith branch is denoted v_i and is in terms of expected number of substitutions per codon. The branch lengths are contained in the vector $\mathbf{v} = (v_1, \ldots, v_{2s-2})$. The ancestor of node k on the tree is denoted $\sigma(k)$. The ancestor of node n_{2s-2} is $\sigma(2s-2) = s$ and the ancestor of node n_s is $\sigma(s) = \emptyset$. The probability of observing the data at the ith codon position ($\mathbf{x}_i$) is a sum over all possible codon assignments to the interior nodes of the phylogenetic tree. Let $\mathbf{y} = \{yk\}$ for $k = s+1, \ldots, 2s-1$ be a generic data vector for the interior nodes of the phylogenetic tree. The probability of observing the data at the ith codon site given the tree, substitution parameters, and the selection category ($K_i = 1, 2$ or 3) is then

$$f(\mathbf{x}_i|\tau, \mathbf{V}, \kappa, \pi, \omega K_i, K_i) = \sum_{\mathbf{y}} \pi_{x_s} \left(\prod_{k=1}^{s-1} p_{y_{\sigma(k)} x_k}(v_k, \omega_{K_i}, \kappa, \pi) \right) \left(\prod_{k=s+1}^{2s-2} p_{y_{\sigma(k)} y_k}(v_k, \omega_{K_i}, \kappa, \pi) \right)$$

Felsenstein (1981) described a pruning algorithm for efficiently performing the summation over ancestral assignments of codons.

Assuming independence of the substitutions among codons, the probability of observing the full sequence data set, $\mathbf{X}$, is

$$f(\mathbf{X}|\tau, \mathbf{v}, \kappa, \pi, \omega_1, \omega_2, \omega_3, p_1, p_2) = \prod_{i=1}^{c} \left(\sum_{K_i=1}^{3} f(\mathbf{x}_i|\tau, \mathbf{v}, \kappa, \pi, \omega_{K_i}, K_i) p_{K_i} \right)$$

Bayesian Analysis

In a Bayesian analysis, parameter estimates are based upon the posterior probability distribution of the parameter. For example, consider a statistical model involving two parameters, called θ_1 and θ_2 . The joint posterior probability of the parameters is $f(\theta_1, \theta_2|X)$, and is the joint probability of the parameters conditioned on the observations, X. If one were only interested in the first parameter, θ_1, then the standard approach is to base inferences on the marginal posterior probability distribution of the first parameter, $f(\theta_1|X)$. The marginal probability distribution of θ_1 is calculated by integrating over all possible values for θ_2:

$$f(\theta_1|X) = \frac{f(X|\theta_1) f(\theta_1)}{\int_{\theta_1} f(X|\theta_1) f(\theta_1) d\theta_1}$$

where

$$f(X|\theta_1) = \int_{\theta_2} f(X|\theta_1, \theta_2) f(\theta_2) d\theta_2$$

($f(X|\theta_1)$ is a marginal likelihood with respect to θ_2). Note that the likelihood, $f(X|\theta_1)$, is obtained by integrating over all possible values for θ_2, weighting each possible value by its prior probability density. In an empirical Bayes analysis, on the other hand, an estimate is substituted for θ_2, and the marginal distribution of θ_1 is calculated conditional on θ_2 taking some value:

$$f(\theta_1|X, \hat{\theta}_2) = \frac{f(X|\theta_1, \hat{\theta}_2) f(\theta_1, \hat{\theta}_2)}{\int_{\theta_1} f(X|\theta_1, \hat{\theta}_2) f(\theta_1, \hat{\theta}_2) d\theta_1}$$

The likelihood is calculated conditional on the parameter θ_2 taking a specific value (in this case, the maximum likelihood estimate of θ_2, $\hat{\theta}_2$).

In this study, we pursue a fully Bayesian approach and integrate over uncertainty in the model parameters, such as the tree topology, branch lengths, and substitution parameters. Specifically, all unrooted trees are considered *a priori* equally probable; branch lengths are assumed to be exponentially distributed with parameter 10, the category probabilities are drawn from a flat Dirichlet distribution, and the stationary codon frequencies are drawn from a flat Dirichlet distribution. The prior probability distribution on the transition/transversion rate ratio and on the nonsynonymous/synonymous rate ratios required more consideration. We took the approach advocated by Zwickl and Holder (2003), considering the prior on these ratio parameters to be the ratio of two random variables. Specifically, we assumed that the transition and transversion rates are random variables drawn from the same exponential distribution. The prior probability distribution on the transition transversion rate ratio [$f(\kappa)$], then, is the ratio of two exponentially distributed random variables:

$$f(\kappa) = \frac{1}{(1+\kappa)^2}$$

(the specific value for the parameter of the common exponential distribution for the transition and transversion rates does not matter). The prior probability distribution for the three nonsynonymous/synonymous rate ratios is calculated in a similar manner. Here, we have three nonsynonymous rates (d_{N1}, d_{N2}, d_{N3}) and one synonymous rate of substitution (d_S). The three nonsynonymous rates are all scaled to the synonymous rate, such that $\omega_1 = d_{N1}/d_S$, $\omega_2 = d_{N2}/d_S$, and $\omega_3 = d_{N3}/d_S$. Again, we assume that the four rates are all independent and ordered draws from the same exponential distribution. The joint prior probability distribution for the three ω's is then

$$f(\omega_1, \omega_2, \omega_3) = \frac{36}{(1+\omega_1+\omega_2+\omega_3)^4}$$

(There is an extra factor of 3! = 6 in the numerator because there are six different ways that one could observe ω's in the three infinitesimal intervals.)

The main criticism of Bayesian approaches concerns the specification of prior probability distributions on parameters of a model. We used the priors, discussed above, because they are either fiat over reasonable biological values for the parameters or only weakly informative. Some of the priors were weakly informative. For example, we use an exponential (10) prior distribution on branch lengths. We think that an exponential prior that places more weight on smaller values for the branch lengths is reasonable; the very fact that we have been able to align the sequences in the first place can be taken as evidence that the branch lengths are not so long that the sequences are saturated. Moreover, in practice the exponential prior on branch lengths can be easily overcome by the data if, in fact, a branch is long. Similarly, the priors on the transition/transversion rate ratio and the nonsynonymous/synonymous rate ratios were weakly informative. Importantly, for the ratio parameters a prior that is the ratio of two random variables does not bias the results toward large (or small) values for the parameters (Zwickl and Holder 2003). A uniform or flat prior was used for the other parameters of the model. For example, the codon frequencies and selection category frequencies both had flat Dirichlet priors. In both cases, there does not seem to be any reason to prefer one combination of frequencies over another, so we give all combinations of frequencies equal prior probability. Finally, we use a flat prior on phylogeny. Although it is true that a lot is known about the phylogeny of many groups before hand (from taxonomies, for example), this information is currently difficult to incorporate into a Bayesian framework. A flat prior on topology seems a safe approach to take until a scheme that allows different prior probabilities to be placed on trees is worked out; a flat prior does not bias the results toward any specific tree.

We are specifically interested in the probability that a given codon site is in the positive selection category. The proba-

bility that the ith codon site is in the positively selected class, $K_i = 3$, is

$$f(K_i = 3|\mathbf{X}) = \frac{f(\mathbf{X}|K_i = 3)p_3}{\sum_{j=1}^{3} f(\mathbf{X}|K_i = j)p_j}$$

and $f(\mathbf{X}|K_i = j)$ is obtained by summing over all possible trees and integrating over all possible combinations of branch lengths, transition/transversion rate biases, codon frequencies, and nonsynonymous/synonymous rate ratios for the purifying and positively selected classes. Hence, we place second-stage priors on parameters such as the tree topology, branch lengths, and substitution parameters. In contrast, in the empirical Bayes analysis of Nielsen and Yang (1998), the probability of a site being in the positive selection category is conditioned on the tree, branch lengths, and parameters of the substitution model taking specific values. Berger (1985) pointed out several advantages of a hierarchical approach, with perhaps the most important being that errors in the hyperparameters are included in the posterior distribution of interest automatically. (The hyperparameters are the parameters of the prior distribution.)

Markov Chain Monte Carlo

The summations and integrals required to calculate the posterior probability of a site being under diversifying selection are impossible to evaluate analytically. We approximate the posterior probability using Markov chain Monte Carlo (MCMC). Specifically, we use the Metropolis–Hastings algorithm (MH; Metropolis et al. 1953; Hastings 1970) to perform the numerical integration. Assume that the desired posterior probability distribution is $f(\theta|X) = f(X|\theta)\ f(\theta)/f(X)$. The probability of observing the data, $f(X)$, is an integral over all possible values for θ and is typically difficult or impossible to calculate analytically. The MH algorithm constructs a Markov chain with the state space being the parameter of interest (θ, in this case) and a stationary distribution which is the posterior probability distribution of the parameter. The samples drawn from this chain when at stationarity are valid (albeit dependent) draws from the posterior distribution of interest. In fact, the Markov chain law of large numbers states that the posterior probability distribution can be validly estimated from the long-run sample frequencies (Tierney 1994). The Markov chain is constructed as follows: (1) Initialize the chain with a value for the parameter θ. The initial value for θ might be drawn from the prior probability distribution for θ. The current state of the chain is denoted θ. (2) Propose a new value for the parameter, denoted θ'. The probability of proposing the new state given the current state is $f(\theta'|\theta)$. The probability of the reverse move, which is not actually made, is also calculated. (3) Calculate the acceptance probability, R, for the new state where

$$R = \min\left[1, \frac{f(\theta'|X)}{f(\theta|X)} \times \frac{f(\theta|\theta')}{f(\theta'|\theta)}\right]$$

$$= \min\left[1, \frac{f(X|\theta')f(\theta')/f(X)}{f(X|\theta)f(\theta)/f(X)} \times \frac{f(\theta|\theta')}{f(\theta'|\theta)}\right]$$

$$= \min\left[1, \frac{f(X|\theta')}{f(X|\theta)} \times \frac{f(\theta')}{f(\theta)} \times \frac{f(\theta|\theta')}{f(\theta'|\theta)}\right]$$

The acceptance probability is simply the product of three ratios that can be readily calculated: the likelihood ratio times the prior ratio times the proposal ratio (also called the "Hastings' factor"). Note that the factor that was difficult to calculate in the first place, $f(X)$, cancels. (4) Draw a uniformly distributed random variable on the interval [0,1]. If this number is less than R, then the proposed state is accepted and $\theta = \theta'$. If this random number is greater than R, then the proposed state is rejected, and the chain remains in its current state. (5) Return to step 2. This procedure is repeated a large number of times. The sequence of states visited forms a Markov chain. The fraction of the time the chain visits any particular value for the parameter is a valid approximation of the posterior probability distribution.

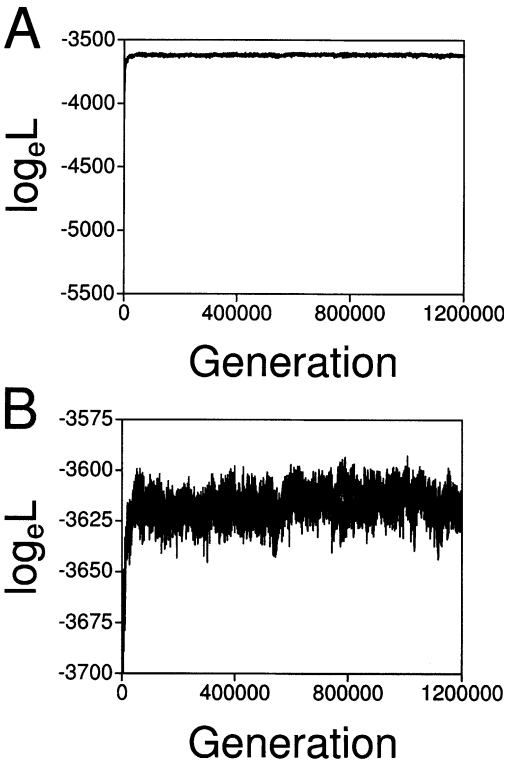

Fig. 1. Plots showing the log likelihood of the current state over the course of the MCMC analysis. The chain was started at a random tree and set of branch lengths; the initial state had a poor likelihood. As the chain is run, it quickly found parameters that best explain the data, and a plateau in the log likelihood is reached. **A** The log likelihoods for all states visited over the course of the analysis. **B** The log likelihoods for states visited when the chain was at apparent stationarity.

We construct a Markov chain that has as its state space phylogenetic trees (τ), branch lengths ($\mathbf{v}$), transition/transversion rate ratios (κ), codon frequencies (π), probability of a site being in the first, second, and third selection classes (respectively, p_1, p_2, and p_3), and ω values for the first, second, and third selection classes (ω_1, ω_2, and ω_2). The chain is constructed by randomly choosing a parameter to change (say the transition/transversion rate bias), proposing a new state for the parameter, and deciding whether the new state is accepted or rejected. The Appendix describes the proposal mechanisms used in this study.

Results

The program MrBayes v3.0 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003; http://morphbank.ebc.uu.se/mrbayes3/) implements the method described in this paper in a user-friendly program. The specific commands for setting up the model were 'lset nucmodel = codon nst = 2 omegavar = m3'. Two independent chains were run, each of which started from different randomly chosen trees. Each chain was run for 1.2 million update cycles, and

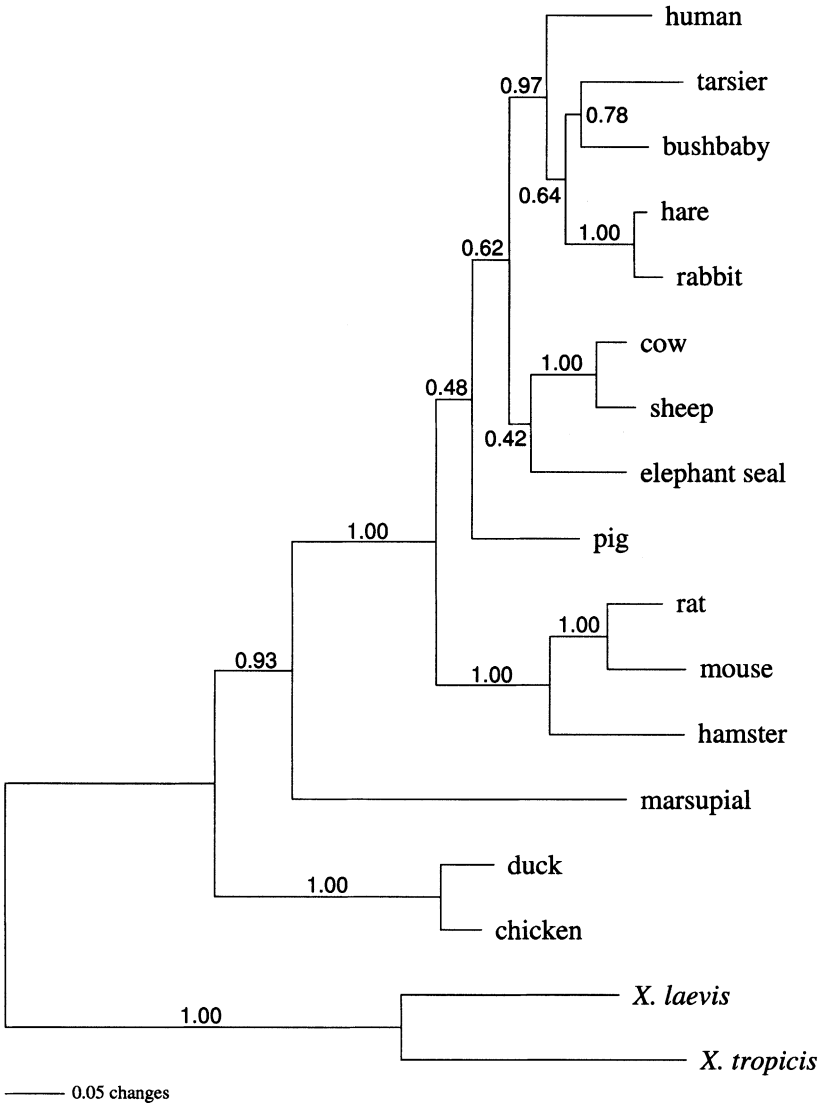

Fig. 2. A majority rule consensus tree summarizing uncertainty in the phylogeny of the $s = 17$ vertebrates. The majority rule consensus tree was based on the trees sampled after the burn-in period (the burn-in of the chain was all samples taken before cycle 200,000). The numbers at the interior nodes of the tree are interpreted as the posterior probability that the particular clade is correct.

the states of the chain were sampled every 100th cycle. Inferences were based on samples taken in the last one million cycles; samples taken during the first 200,000 cycles were discarded as the "burn-in" of the Markov chain. All of the results from this paper are based on the combined samples taken from the two chains. Figure 1 shows the log probability of observing the data for each sampled state through time for all four chains. Note that the chains started with combinations of parameter values that poorly explained the data but quickly improved in likelihood, eventually reaching a plateau. The plateau was reached relatively quickly. Also, note that there was no discernible difference in the log likelihoods for the two chains. Moreover, inspection of the posterior probability distributions of individual parameters based on each chain overlap. Together, these observations suggest that the MCMC algorithm is successfully sampling from the posterior probability distribution of the parameters.

There was a large degree of uncertainty in some of the parameters of the evolutionary model used in this study. For example, although there was strong support for some relationships (*e.g.*, that rat and mouse are each others closest relatives), the posterior probabilities of some of the other clades was quite low (Fig. 2). Similarly, there was uncertainty in the parameters of the substitution model (Fig. 3). Table 1 summarizes the mean and 95% credible interval for each of the parameters of the substitution model used in this analysis whereas Figure 3 shows the full marginal posterior probability density distributions for a number of the substitution parameters.

The probability that each site was under diversifying selection was calculated each time the chain was sampled. Figure 4A shows the average probability that each site was under positive selection. The mean and 95% credible interval of the probability for each site being under positive selection is also summarized in Table 2 (also see Fig. 6). The highest mean pos-

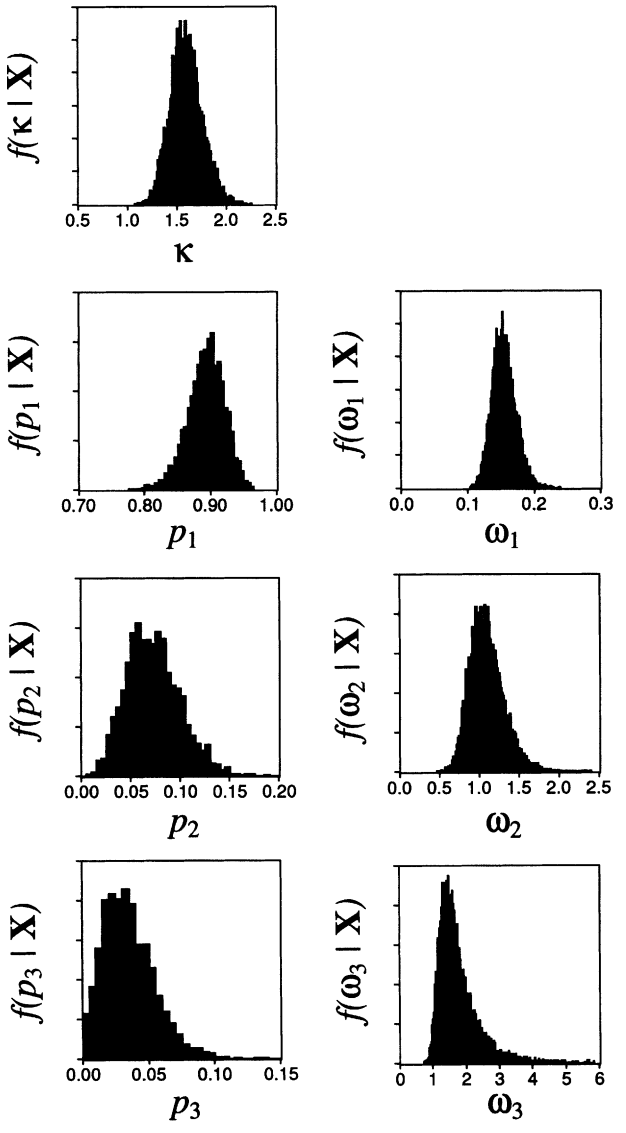

Fig. 3. The marginal posterior probability densities of several parameters of the substitution model. The parameters include p_1, p_2, and p_3 (the prior probabilities of being in the purifying, neutral, and positively selected categories, respectively), κ (the transition/transversion rate bias), and ω_1, ω_2, and ω_3 (the nonsynonymous/synonymous rate ratios for categories 1, 2, and 3, respectively).

terior probability was 0.934 for codon 67. Other sites that had among the highest posterior probabilities of being under diversifying selection were 67 (0.934), 50 (0.909), 48 (0.871), 7 (0.757), 74 (0.714), 85 (0.691), 123 (0.676), 110 (0.640), 54 (0.640), 11 (0.637), and 42 (0.632). Many of these are the same sites identified by Yang et al. (2000) as being under positive selection; however, the posterior probabilities for many of the sites in their study were higher.

There was a large degree of variability in the posterior probabilities that sites were under positive selection. Figure 5 shows frequency histograms of the posterior probabilities that sites are under positive selection for the states visited by the Markov chains. Note that even for sites for which there is strong

Table 1. Parameter estimates for the vertebrate β-globin gene

Parameter	Mean (95% CI)
TL	2.8053 (2.4320, 3.2800)
κ	1.5896 (1.2971, 1.9208)
ω_1	0.1542 (0.1223, 0.1914)
ω_2	1.0898 (0.7319, 1.5779)
ω_3	1.7194 (1.0147, 3.2997)
π_1	0.8928 (0.8330, 0.9430)
π_2	0.0725 (0.0265, 0.1297)
π_3	0.0347 (0.0047, 0.0804)
π_{AAA}	0.0158 (0.0099, 0.0220)
π_{AAC}	0.0265 (0.0174, 0.0362)
π_{AAG}	0.0447 (0.0311, 0.0591)
π_{AAT}	0.0332 (0.0246, 0.0447)
π_{ACA}	0.0039 (0.0006, 0.0104)
π_{ACC}	0.0159 (0.0081, 0.0254)
π_{ACG}	0.0036 (0.0005, 0.0076)
π_{ACT}	0.0262 (0.0157, 0.0383)
π_{AGA}	0.0032 (0.0007, 0.0083)
π_{AGC}	0.0221 (0.0140, 0.0321)
π_{AGG}	0.0069 (0.0021, 0.0136)
π_{AGT}	0.0223 (0.0127, 0.0347)
π_{ATA}	0.0057 (0.0011, 0.0126)
π_{ATC}	0.0292 (0.0184, 0.0411)
π_{ATG}	0.0103 (0.0045, 0.0178)
π_{ATT}	0.0194 (0.0107, 0.0312)
π_{CAA}	0.0088 (0.0034, 0.0171)
π_{CAC}	0.0233 (0.0120, 0.0356)
π_{CAG}	0.0216 (0.0128, 0.0338)
π_{CAT}	0.0284 (0.0190, 0.0392)
π_{CCA}	0.0066 (0.0016, 0.0165)
π_{CCC}	0.0150 (0.0075, 0.0239)
π_{CCG}	0.0044 (0.0009, 0.0104)
π_{CCT}	0.0200 (0.0107, 0.0320)
π_{CGA}	0.0040 (0.0001, 0.0141)
π_{CGC}	0.0124 (0.0057, 0.0214)
π_{CGG}	0.0023 (0.0002, 0.0064)
π_{CGT}	0.0057 (0.0011, 0.0134)
π_{CTA}	0.0034 (0.0012, 0.0067)
π_{CTC}	0.0160 (0.0103, 0.0217)
π_{CTG}	0.0450 (0.0307, 0.0626)
π_{CTT}	0.0081 (0.0046, 0.0127)
π_{GAA}	0.0176 (0.0108, 0.0254)
π_{GAC}	0.0206 (0.0128, 0.0314)
π_{GAG}	0.0183 (0.0106, 0.0289)
π_{GAT}	0.0285 (0.0186, 0.0409)
π_{GCA}	0.0074 (0.0034, 0.0123)
π_{GCC}	0.0413 (0.0321, 0.0528)
π_{GCG}	0.0056 (0.0023, 0.0098)
π_{GCT}	0.0401 (0.0297, 0.0523)
π_{GGA}	0.0108 (0.0051, 0.0177)
π_{GGC}	0.0237 (0.0137, 0.0344)
π_{GGG}	0.0057 (0.0023, 0.0104)
π_{GGT}	0.0142 (0.0087, 0.0206)
π_{GTA}	0.0027 (0.0009, 0.0060)
π_{GTC}	0.0153 (0.0097, 0.0216)
π_{GTG}	0.0214 (0.0137, 0.0317)
π_{GTT}	0.0188 (0.0127, 0.0249)
π_{TAC}	0.0112 (0.0049, 0.0203)
π_{TAT}	0.0106 (0.0037, 0.0187)
π_{TCA}	0.0085 (0.0024, 0.0169)
π_{TCC}	0.0241 (0.0134, 0.0364)
π_{TCG}	0.0013 (0.0001, 0.0048)
π_{TCT}	0.0316 (0.0208, 0.0445)
π_{TGC}	0.0098 (0.0040, 0.0200)
π_{TGG}	0.0192 (0.0066, 0.0355)

(Continued)

Table 1. Continued

Parameter	Mean (95% CI)
π_{TGT}	0.0199 (0.0087, 0.0327)
π_{TTA}	0.0040 (0.0006, 0.0099)
π_{TTC}	0.0229 (0.0147, 0.0338)
π_{TTG}	0.0092 (0.0053, 0.0145)
π_{TTT}	0.0221 (0.0141, 0.0346)

Note. TL, the sum of the branch lengths on the tree (the tree length).

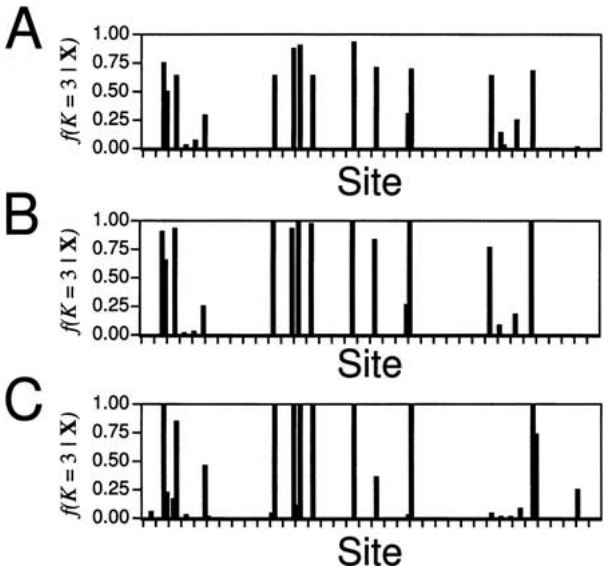

Fig. 4. **A** The posterior probability of each site being in the positively selected class. **B** The median of the posterior probabilities of each site being in the positively selected class. **C** The posterior probability of being in the positively selected class when maximum likelihood is used to estimate parameters.

evidence of diversifying selection, such as sites 7, 11, and 48, the posterior probabilities of being under positive selection varied substantially. In fact, for some sites the posterior probability of being under positive selection varied over the entire range of the parameter during the course of the chain. For example, for site 74 the average posterior probability was 0.714, which suggests that the site may have been under positive selection. However, the chain considered reasonable combinations of parameters that made the posterior probability for this site range from 0.00034 to 0.99969, with the chain spending 95% of the time in the interval ranging from 0.006 to 0.993 (the 95% credible interval).

Discussion

Earlier studies suggest that the ability to identify positively selected sites may be relatively insensitive to the choice of phylogeny. For example, Yang et al. (2000) found that inferences of positive selection were fairly robust to which of six trees were used for the

Table 2. Posterior probabilities of positive selection

Site	Mean (95% CI), median
1	0.000 (0.000, 0.000) 0.000
2	0.000 (0.000, 0.000) 0.000
3	0.001 (0.000, 0.010) 0.000
4	0.000 (0.000, 0.000) 0.000
5	0.000 (0.000, 0.000) 0.000
6	0.000 (0.000, 0.000) 0.000
7	0.757 (0.018, 0.997) 0.897
8	0.493 (0.000, 0.964) 0.656
9	0.000 (0.000, 0.000) 0.000
10	0.002 (0.000, 0.008) 0.001
11	0.637 (0.001, 0.996) 0.936
12	0.000 (0.000, 0.000) 0.000
13	0.000 (0.000, 0.000) 0.000
14	0.024 (0.000, 0.125) 0.010
15	0.000 (0.000, 0.000) 0.000
16	0.000 (0.000, 0.000) 0.000
17	0.067 (0.000, 0.365) 0.022
18	0.000 (0.000, 0.000) 0.000
19	0.000 (0.000, 0.000) 0.000
20	0.295 (0.000, 0.843) 0.245
21	0.000 (0.000, 0.000) 0.000
22	0.000 (0.000, 0.000) 0.000
23	0.000 (0.000, 0.000) 0.000
24	0.000 (0.000, 0.000) 0.000
25	0.000 (0.000, 0.000) 0.000
26	0.000 (0.000, 0.000) 0.000
27	0.000 (0.000, 0.001) 0.000
28	0.000 (0.000, 0.000) 0.000
29	0.000 (0.000, 0.000) 0.000
30	0.000 (0.000, 0.000) 0.000
31	0.000 (0.000, 0.000) 0.000
32	0.000 (0.000, 0.000) 0.000
33	0.000 (0.000, 0.000) 0.000
34	0.000 (0.000, 0.000) 0.000
35	0.000 (0.000, 0.000) 0.000
36	0.000 (0.000, 0.000) 0.000
37	0.000 (0.000, 0.000) 0.000
38	0.000 (0.000, 0.000) 0.000
39	0.000 (0.000, 0.000) 0.000
40	0.000 (0.000, 0.000) 0.000
41	0.002 (0.000, 0.012) 0.001
42	0.632 (0.000, 1.000) 0.997
43	0.000 (0.000, 0.000) 0.000
44	0.000 (0.000, 0.000) 0.000
45	0.000 (0.000, 0.000) 0.000
46	0.000 (0.000, 0.000) 0.000
47	0.000 (0.000, 0.000) 0.000
48	0.871 (0.275, 0.993) 0.929
49	0.002 (0.000, 0.010) 0.001
50	0.909 (0.189, 1.000) 1.000
51	0.000 (0.000, 0.000) 0.000
52	0.000 (0.000, 0.000) 0.000
53	0.000 (0.000, 0.000) 0.000
54	0.640 (0.001, 0.999) 0.965
55	0.000 (0.000, 0.000) 0.000
56	0.000 (0.000, 0.000) 0.000
57	0.000 (0.000, 0.000) 0.000
58	0.000 (0.000, 0.000) 0.000
59	0.000 (0.000, 0.000) 0.000
60	0.000 (0.000, 0.000) 0.000
61	0.000 (0.000, 0.000) 0.000
62	0.000 (0.000, 0.000) 0.000
63	0.000 (0.000, 0.000) 0.000
64	0.000 (0.000, 0.001) 0.000

(Continued)

Table 2. Continued

Site	Mean (95% CI), median
65	0.000 (0.000, 0.000) 0.000
66	0.000 (0.000, 0.002) 0.000
67	0.934 (0.346, 1.000) 0.999
68	0.000 (0.000, 0.002) 0.000
69	0.000 (0.000, 0.000) 0.000
70	0.000 (0.000, 0.000) 0.000
71	0.001 (0.000, 0.007) 0.000
72	0.000 (0.000, 0.000) 0.000
73	0.000 (0.000, 0.000) 0.000
74	0.714 (0.006, 0.993) 0.830
75	0.000 (0.000, 0.002) 0.000
76	0.000 (0.000, 0.000) 0.000
77	0.000 (0.000, 0.000) 0.000
78	0.000 (0.000, 0.000) 0.000
79	0.000 (0.000, 0.000) 0.000
80	0.000 (0.000, 0.000) 0.000
81	0.002 (0.000, 0.009) 0.000
82	0.000 (0.000, 0.000) 0.000
83	0.000 (0.000, 0.000) 0.000
84	0.311 (0.000, 0.866) 0.270
85	0.691 (0.003, 1.000) 1.000
86	0.000 (0.000, 0.000) 0.000
87	0.000 (0.000, 0.000) 0.000
88	0.000 (0.000, 0.000) 0.000
89	0.000 (0.000, 0.000) 0.000
90	0.000 (0.000, 0.000) 0.000
91	0.000 (0.000, 0.000) 0.000
92	0.000 (0.000, 0.000) 0.000
93	0.000 (0.000, 0.000) 0.000
94	0.000 (0.000, 0.000) 0.000
95	0.000 (0.000, 0.000) 0.000
96	0.000 (0.000, 0.000) 0.000
97	0.000 (0.000, 0.000) 0.000
98	0.000 (0.000, 0.000) 0.000
99	0.000 (0.000, 0.000) 0.000
100	0.000 (0.000, 0.000) 0.000
101	0.000 (0.000, 0.000) 0.000
102	0.000 (0.000, 0.000) 0.000
103	0.000 (0.000, 0.000) 0.000
104	0.000 (0.000, 0.000) 0.000
105	0.000 (0.000, 0.000) 0.000
106	0.000 (0.000, 0.000) 0.000
107	0.000 (0.000, 0.000) 0.000
108	0.000 (0.000, 0.000) 0.000
109	0.001 (0.000, 0.008) 0.000
110	0.640 (0.002, 0.995) 0.768
111	0.000 (0.000, 0.000) 0.000
112	0.000 (0.000, 0.000) 0.000
113	0.136 (0.000, 0.526) 0.084
114	0.024 (0.000, 0.157) 0.004
115	0.004 (0.000, 0.025) 0.001
116	0.000 (0.000, 0.000) 0.000
117	0.000 (0.000, 0.000) 0.000
118	0.248 (0.000, 0.802) 0.182
119	0.006 (0.000, 0.050) 0.001
120	0.000 (0.000, 0.000) 0.000
121	0.000 (0.000, 0.000) 0.000
122	0.000 (0.000, 0.000) 0.000
123	0.676 (0.000, 1.000) 1.000
124	0.000 (0.000, 0.001) 0.000
125	0.000 (0.000, 0.000) 0.000
126	0.000 (0.000, 0.000) 0.000
127	0.000 (0.000, 0.000) 0.000
128	0.000 (0.000, 0.000) 0.000
129	0.000 (0.000, 0.000) 0.000
130	0.000 (0.000, 0.001) 0.000
131	0.000 (0.000, 0.000) 0.000
132	0.000 (0.000, 0.000) 0.000
133	0.000 (0.000, 0.000) 0.000
134	0.000 (0.000, 0.000) 0.000
135	0.000 (0.000, 0.000) 0.000
136	0.000 (0.000, 0.000) 0.000
137	0.013 (0.000, 0.106) 0.002
138	0.000 (0.000, 0.000) 0.000
139	0.000 (0.000, 0.000) 0.000
140	0.000 (0.000, 0.001) 0.000
141	0.000 (0.000, 0.000) 0.000
142	0.000 (0.000, 0.000) 0.000
143	0.000 (0.000, 0.000) 0.000
144	0.000 (0.000, 0.001) 0.000

(Continued)

Table 2. Continued

Site	Mean (95% CI), median

β-globin gene data. This is an encouraging result because it suggests that estimates of positive selection may be relatively insensitive to a major source of uncertainty in phylogenetic studies. However, the codon models typically used to detect diversifying selection are parameter rich, containing many free parameters besides the topology and branch lengths of the phylogeny; the model we used, for example, contained a total of 66 free parameters, excluding the tree and branch length parameters (κ, ω_1, ω_2, ω_3, p_1, p_2, and π). Uncertainty in any of these parameters can contribute to uncertainty in inferences of diversifying selection. Some of the parameters are expected to play an especially important role; the site category probabilities (here, p_1, p_2, and p_3) should strongly affect the posterior probability of being in the positively selected class because Bayes's rule directly uses these quantities. The estimates of these parameters had some degree of associated uncertainty; the credible intervals for p_1, p_2, and p_3 were (0.733, 0.943), (0.026, 0.129), and (0.005, 0.080), respectively. The greatest source of uncertainty in this particular study, however, can be attributed to the dN/dS parameters for each category. In particular, sometimes the Markov chain had all three parameters less than 1 (ω_1, ω_2, $\omega_3 < 1$), which means that the probability of any site being in positive selection for those samples was zero. Figure 5 shows this phenomenon; note that there is a spike of probability density on zero for the nine sites depicted in the figure. The mean posterior probability of finding a site in the positive selection category is decreased when all three selection classes have $\omega < 1$. One possible solution is to summarize the results of the MCMC analysis using the median posterior probabilities, as shown in Fig. 4B. Generally speaking, taking the median of the posterior probabilities of finding sites in the positive selection class makes the results more in line with the proba-

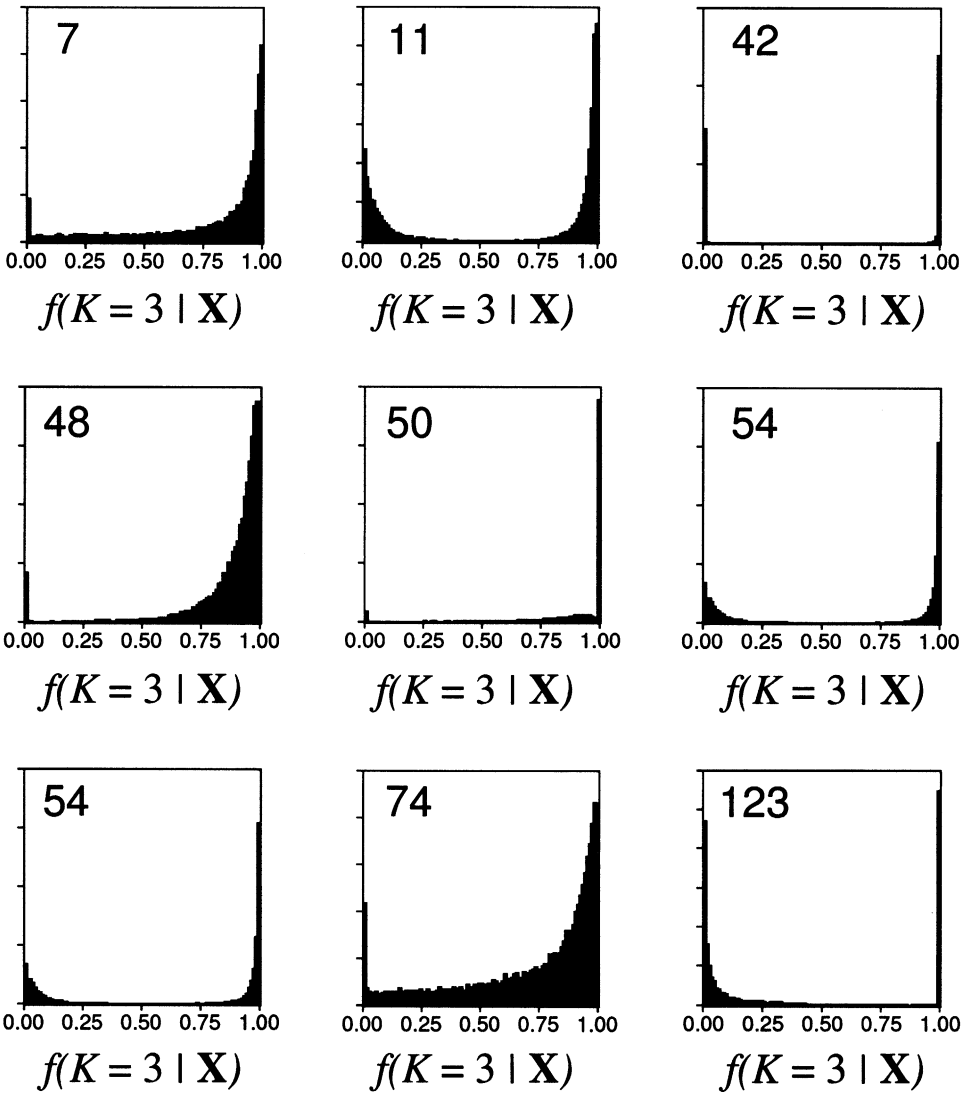

Fig. 5. The variation in the probability that several different sites were in the positively selected class.

bilities obtained using the empirical Bayes approach (Fig. 4C).

We found a number of sites with posterior probabilities greater than 0.70 of being under positive selection. Specifically, sites 7, 48, 50, 67, and 74 had high mean posterior probabilities. If the median posterior probability sampled by the MCMC algorithm is used instead, we identify four sites with posterior probabilities greater than 0.99 (42, 67, 85, and 123) and an additional three sites with posterior probabilities greater than 0.90 (11, 48, 54). We identified many of the same sites as being under diversifying selection as Yang et al. (2000). This is not unexpected, as the modeling assumptions used in both studies were nearly the same, with the main difference being the treatment of the nuisance parameters of the model.

The posterior probability of being in the positive selection category was typically lower for sites that were identified as being under diversifying selection in this study than analysis of the same gene suggested in an empirical Bayes analysis (using a program written by JPH that implements the M3 model and estimates parameters using maximum likelihood). Figure 8 shows the relationship between the posterior probabilities of being in the positive selection class for all 144 amino acid positions. A number of the sites had posterior probabilities near one using the empirical Bayes method; these same sites had a lower posterior probability of being under positive selection using the fully Bayesian method. The likely explanation concerns the use of averaging probabilities over MCMC samples versus taking the maximum likelihood estimate. In this study, we averaged the posterior probability of a site being under positive selection over all of the reasonable parameter values sampled by the Markov chain. Figure 5 shows the frequency histogram of the posterior probabilities for nine sites. For sites that are likely under diversifying selection, such as site 50, the distribution is strongly left-skewed;

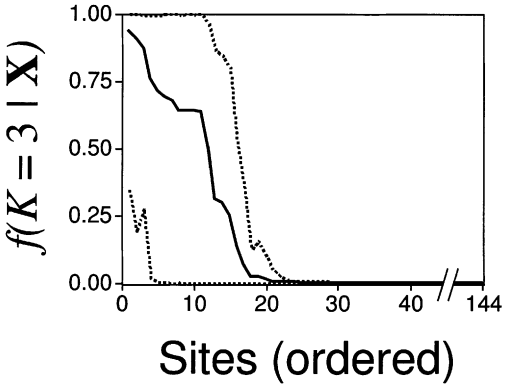

Fig. 6. The posterior probabilities of sites being in the positively selected class when the sites are ordered from highest to lowest posterior probability. The thick solid line indicates the average posterior probability of each site being under positive selection. The dashed lines show the 95% credible interval for each site being in the positive selection category.

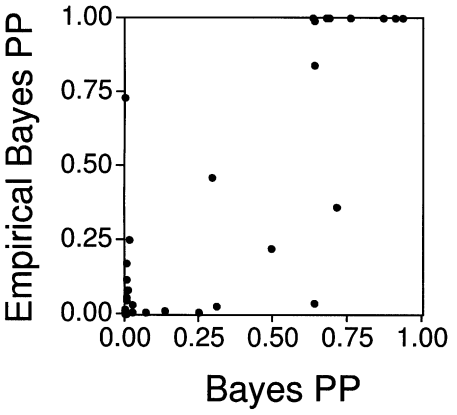

Fig. 8. The relationship between the posterior probability of a site being in the positive selection class as determined using the fully Bayesian method (Bayes PP) and the empirical Bayes method (Empirical Bayes PP).

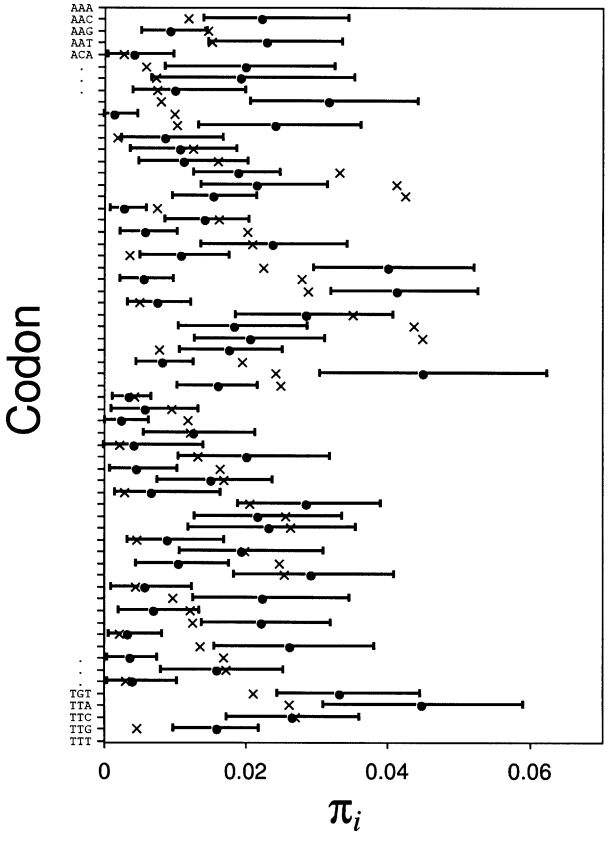

Fig. 7. The mean and 95% credible interval of the frequency for each of the 61 codons. The mean is denoted by the circle and the interval by the horizontal bar. The codon frequencies predicted by the frequencies of the nucleotides at the three codon positions are shown as the X. Note that in about half of the cases, the predicted frequencies are well outside of the 95% credible interval.

most of the states sampled by the Markov chain resulted in a high posterior probability of being in the positively selected class. The results for site 50 are similar for both the empirical and the fully Bayes approaches. However, the same is not true for site 48, where there is more uncertainty about the assignment of the site to the positive selection class. The Markov chain considered states that made the posterior probability of being under positive selection vary from zero to near one. In this case, taking the mean of the posterior probabilities over MCMC samples causes the probability of being in the positive selection class to be substantially lower under the fully Bayesian method as opposed to the empirical Bayes method (0.871 vs. 0.998).

It is difficult to assess the computational merits of the approach taken here. The analyses described in this paper took about a day on a fast desktop computer, whereas the same analysis using the empirical Bayes approach would take about an hour. Although the analysis took longer, we did obtain considerable information on the parameters (specifically, the joint and marginal posterior probability densities of the parameters) and were able to explicitly incorporate uncertainty in the parameter estimates when searching for positively selected sites. Moreover, the substitution model used in this analysis was more parameter rich than most codon models used to identify positive selection in that the codon frequencies were all allowed to vary freely (they were assigned a flat Dirichlet prior). Typically, in maximum likelihood approaches to the problem, the codon frequencies are calculated as a function of the frequencies of the four nucleotides A, C, G, and T at the first, second, and third codon positions ($f_i^{(k)}$ is the frequency of nucleotide i at codon position k). Hence, the frequency of codon ACT is calculated as $(f_A^{(1)} \times f_C^{(2)} \times f_T^{(3)})/x$ (x is a constant that takes into account the fact that only 61 of the 64 codons are used). This means that in the implementation of the codon model described here, there was more freedom for the model to explain the data by modifying substitution rates using the codon frequencies. Figure 7 illustrates how well a model that estimates codon frequencies using four nucleotide parameters explains the codon fre-

quencies visited in the course of the MCMC analysis. Specifically, Fig. 7 shows the mean and 95% credible interval for each of the 61 codon frequencies. The codon frequencies predicted by the product of the nucleotide frequencies at the three codon positions are shown as the X. In 31 of the 61 cases, the codon frequencies predicted using the nucleotide frequencies are outside of the 95% credible interval.

Many questions in evolutionary biology require the effective and efficient identification of the signature of natural selection at the molecular level. As computational limits become less cumbersome, biologically realistic models that attempt to do this are becoming more feasible. The codon models of Goldman and Yang (1994) and Muse and Gaut (1994) represent a significant advance in the field, as they allowed the exact partitioning of substitutions into those that change and those that do not change the protein sequence. The work of Nielsen and Yang (1998) also represents a significant advance because it allowed constraints to change across the sequence as well as the identification of sites under diversifying selection. A major advantage of the fully Bayesian approach described here is that it takes advantage of the earlier advances in codon models and accommodates uncertainty in parameters that in an empirical Bayes analysis must be hard-wired point estimates. This benefit may outweigh the associated computational costs.

Acknowledgments. We thank J. Bollback for comments on an earlier version of this paper. This work was supported by NSF Grants DEB-0075406 and MCB-0075404 awarded to J.P.H.

Appendix

Changing the Tree Topology and Branch Lengths

With probability ψ_1 a move is attempted that simultaneously changes the tree topology, τ, and branch lengths, $\mathbf{v}$. We used a proposal mechanism first described by Larget and Simon (1999). One of the $s-3$ internal branches of the unrooted tree was chosen at random. This branch will be the middle branch of a set of three contiguous branches that form the "backbone" of the move and is denoted e. There are two branches that are incident to either end of this branch. The branches incident to one end of the branch are randomly labeled a and b, whereas the branches incident to the other end are randomly labeled c and d. The three branches that make up the backbone of the move are branches a, e, and c. The sum of the tree branch lengths, or the path length, is $m = v_a + v_e + v_c$ A new path length is chosen as $m' = m \times e^{\lambda(U-1/2)}$, where U is a uniform (0,1) random number and λ is a tuning parameter that is fixed at the beginning of the analysis. All of the branch lengths in the back bone are modified by multiplying them by m'/m. Finally, either branch b or branch d is chosen at random and detached from the backbone. This branch is reattached to the backbone uniformly over the new length of the backbone. The acceptance probability for this move is

$$R = \min\left[1, (\text{LikelihoodRatio}) \times (\text{PriorRatio}) \times \left(\frac{m'}{m}\right)^2\right]$$

(Larget and Simon 1999).

Changing the κ, ω_1, ω_2, and ω_3

With probability ψ_2, ψ_3, ψ_4, and ψ_5 a move is attempted that changes κ, ω_1, ω_2, or ω_3, respectively. The current value of the parameter (generically, θ_i) was increased or decreased by adding a uniformly distributed random variable on the interval $[\theta_i - w_i, \theta_i + w_i]$, where $i = 2$, 3, or 4. The window is centered on the current value of the parameter. When a new value is proposed that is outside of the valid parameter space, then the excess is reflected back. The acceptance probability for this proposal mechanism is

$$R = \min[1, (\text{Likelihood Ratio}) \times (\text{Prior Ratio})]$$

Changing the Codon Frequencies and Category Probabilities

With probability ψ_6 and ψ_7 a move is attempted that changes the equilibrium codon frequencies, π, or category probabilities, p_1, p_2, p_3. New codon frequences are proposed from a Dirichlet distribution with expected values at the current values. The Dirichlet distribution has probability density

$$f(\pi|\alpha) = \frac{\Gamma(\alpha_0)}{\prod_{i\in S}\Gamma(\alpha_i)} \prod_{i\in S} \pi_i^{(\alpha_i - 1)}$$

where S is the state space (AAA, AAC, AAG, AAT, ACA, ACC, ACG, ..., TTT), α_i is the Dirichlet parameter for the ith codon, $\alpha_0 = \sum_{i\in S}\alpha_i$, and π_i is the frequency of the ith codon. New codon frequencies, π', are drawn from the Dirichlet distribution with $\alpha_i = \pi_i \alpha_0$. The acceptance probability for a move that changed codon frequencies is

$$R = \min\left[1, (\text{Likelihood Ratio}) \times (\text{Prior Ratio}) \times \prod_{i\in S} \frac{\Gamma(\pi_i\alpha_0)\pi_i^{(\pi'_i\alpha_0 - 1)}}{\Gamma(\pi'_i\alpha_0)\pi_i'^{(\pi_i\alpha_0 - 1)}}\right]$$

The same mechanism was used to propose new values for the category probabilities p_1, p_2, and p_3.

References

Berger JO (1985) Statistical decision theory and Bayesian analysis. Springer-Verlag, New York

Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. J Mol Evol 17:368–376

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

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109

Huelsenbeck JH, Ronquist F (2001) MrBayes: Bayesian inference of phylogeny. Bioinformatics 17:754–755

Hughes A, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–170

Larget B, Simon D (1999) Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol Biol Evol 16:750–759

Lee YH, Ota T, Vaquier V (1995) Positive selection is a general phenomenon in the evolution of abalone sperm lysin. Mol Biol Evol 12:231–238

Messier W, Stewart CB (1997) Episodic adaptive evolution of primate lysozymes. Nature 385:151–154

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. J Chem Phys 21:1087–1091

Muse SV, Gaut BS (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates with application to the chloroplast genome. Mol Biol Evol 11:715–724

Nielsen R (1997) The ratio of replacement to silent divergence and tests of neutrality. J Evol Biol 10:217–231

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

Ronquist F, Huelsenbeck JP (2003) MrBayes version 3.0: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Tierney L (1994) Markov chains for exploring posterior distributions. Ann Stat 22:1701–1762

Yang Z, Nielsen R, Goldman N, Pedersen A (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Yokoyama S, Yokoyama R (1996) Evolution of photoreceptor cells and visual pigments. Annu Rev Ecol Syst 27:543–567