Comparison of Bayesian, maximum likelihood and parsimony methods for detecting positive selection

(Research Article)

Daniel Merl
Applied Mathematics & Statistics
University of California Santa Cruz

Ananías Escalante
Department of Biology
Emory University

Raquel Prado
Applied Mathematics & Statistics
University of California Santa Cruz

Corresponding Author:

Raquel Prado
Applied Mathematics and Statistics
SOE-2 University of California
1156 High St.
Santa Cruz, CA 95064, USA
E-mail:raquel@ams.ucsc.edu

Keywords: positive selection, parsimony, maximum likelihood, Bayesian estimation

Running head: Comparison of methods for detecting selection

Abbreviations: dN: rate of non-synonymous substitutions per non-synonymous site; dS: rate of synonymous substitutions per synonymous site; ω: non-synonymous to synonymous rates ratio (ω = dN/dS); dT: transitions rate; dR: transversions rate; PAML: Phylogenetic Analysis by Maximum Likelihood

Abstract

From a variety of vantage points, ranging from epidemiological to statistical, the problem of identifying the effects of natural selection at the molecular level is a fascinating one. Recent years have seen an explosion of model based methods for inferring such effects, with particular emphasis on detection of positive selection; some of the most popular of which are the maximum likelihood based method of Yang implemented in PAML, the parsimony based method of Suzuki and Gojorobi implemented in ADAPTSITE, and the hierarchical Bayesian method of Huelsenbeck and Ronquist implemented in MRBAYES. Although each of these three methodologies has appeared in the literature in the analyses of various sequence data, there have been no cross comparison studies of the performance of these methods when applied to the same data, in terms of the methods' abilities to predict amino acid sites influenced by positive selection. To this end, we employed the three methods to detect the presence of positively selected sites in the following sequence data, where each data set was chosen to represent a different level of phylogenetic uncertainty: a previously analyzed abalone sperm lysin alignment, three alignments of the Avian infectious bronchitis (AIB) virus S gene, and two alignments of the homologous S protein of the SARS coronavirus. The results shown here demonstrate important strengths and drawbacks of each method when dealing with data of different levels of phylogenetic uncertainty.

Introduction

Natural selection may be the cornerstone process of evolution, but there has been a prolific controversy over how it affects the observed diversity at the molecular level (Kimura, 1983; Gillespie, 1991). The neutral theory of evolution considers that natural selection operates mostly by eliminating deleterious mutations (negative selection), thereby regarding natural selection primarily as a constraint to the maximum allowable level of genetic polymorphism. In contrast to the neutral theory, so called positive selection is expected to maintain genetic polymorphism and accelerate divergence between homologous proteins.

The most widely used tests for detecting positive selection are based on the ratio of the rate of non-synonymous substitution (dN) to the rate of synonymous substitution (dS), herein referred to as ω (Hughes, 2000). Synonymous substitutions are implicitly neutral insofar as they do not change the amino acid composition of the protein, while non-synonymous substitutions may be subject to non-neutral (i.e. positive or negative) selective pressures (Kimura, 1983; Hughes and Nei, 1990). Thus, under the neutral theory, ω is expected to be less than or equal to one, where $\omega < 1$ implies negative selection and $\omega = 1$ is expected only in the case of selectively neutral polymorphisms. Tests for positive selection based on ω are considered conservative; indeed, for this metric to be of use, positive selection must be strong enough to allow for the accumulation of non-synonymous substitutions at a higher rate than synonymous substitutions (Kreitman and Akashi, 1995). The task of detecting positive selection is further complicated by the fact that in scenarios of no selection (as in the case of a pseudogene), or when synonymous substitutions are non-neutral (as in cases of extreme

codon bias), non-synonymous substitutions will accumulate faster than would otherwise be expected. Thus, important assumptions of these tests are that selection, negative or positive, operates only on non-synonymous substitutions in a functional gene, and synonymous changes are expected to be selectively neutral.

There are basically two major groups of statistical methods for testing departures from the null hypothesis of $\omega = 1$. Methods in the first group compute dN and dS by pairwise comparison in a set of aligned sequences of a polymorphic gene. Significance tests on the observed dN/dS ratio are performed using one of several available methods (see Nei and Kumar, 2000 for a review on these methods). Usually, regions within the gene identified a priori —such as functional domains or immunogenic regions— are compared in order to determine the impact of natural selection on different regions the protein (Hughes, 2000).

Methods in the second group estimate codon specific ω using phylogeny-based models. In other words, these models rely on the overall pattern of genetic divergence captured by the phylogeny in order to estimate ω for each codon. Conceptually, these methods focus on fixed differences rather than on genetic polymorphisms, although the basic models can also be used on pairwise comparisons (Escalante *et al.*, 2004; Tzeng *et al.*, 2004). The advantage of these models is that they account for differences in nucleotide composition and unequal codon usage (Goldman and Yang, 1994). They also allow the identification of specific codons, rather than regions, that are influenced by positive selection. A disadvantage is that these codon-based models are highly parameterized and consequently estimation procedures within this framework are often computationally demanding. In addition, it has been found that

some of these methods may become unreliable when short sequences are used (Tzeng *et al.*, 2004).

Of the dozens of studies recently published on the topic of identifying residues under positive selection, one of the most commonly used methods is the maximum likelihood based approach developed by Yang and collaborators (Goldman and Yang, 1994; Yang, 1997; Nielsen and Yang, 1998; Yang *et al.*, 2000; Anisimova *et al.*, 2002) and implemented in the software package PAML (Yang, 2003). PAML provides a powerful framework for investigating the presence of codon-level positive selection via stochastic models of sequence evolution. Codons are subdivided into categories based on their estimated rates of synonymous and non-synonymous substitutions (see Materials and Methods), where these rates are estimated via maximum likelihood (ML). PAML predicts the individual sites affected by positive selection (i.e., having $\omega > 1$) using an empirical Bayes approach. Additionally, PAML offers the possibility of formal comparison of nested evolutionary models using likelihood ratio tests (Nielsen and Yang, 1998; Anisimova *et al.*, 2002).

Popular alternative methods for detecting positive selection include the parsimony based methodology (Fitch *et al.*, 1997; Suzuki and Gojobori, 1999) implemented in ADAPTSITE (Suzuki *et al.*, 2001), and the hierarchical Bayesian approach of Huelsenbeck and Ronquist implemented in MRBAYES (Huelsenbeck and Ronquist, 2001; Huelsenbeck *et al.*, 2001). ADAPTSITE uses maximum parsimony methods to reconstruct ancestral sequences. The method then counts the changes along the phylogenetic tree at each site in order to identify those codons with an excess of non-synonymous substitutions. As in the case of the

ML methods, no phylogenetic uncertainty is considered in the estimation of the number of substitutions, and the reconstructed ancestral sequences are assumed error-free. In the fully Bayesian approach of Huelsenbeck and Ronquist, all the parameters of a stochastic model of sequence evolution similar to that of the ML method are estimated within a Bayesian framework. This allows for the consideration of uncertainty in the phylogeny, which is an important feature when dealing with lowly divergent sequence data for which the phylogenetic structure is poorly defined.

There has been, and continues to be, a great deal of controversy over the decision to use one method over the others. Although ML, parsimony, and Bayesian methodologies have each appeared in a number of recent analyses, no single study has been conducted to compare the effectiveness and appropriateness of these three methods when applied to common data. In fact, what evidence exists suggests very little consistency of results across methods. Suzuki and Nei (2004) consider that the ML method implemented in PAML may be misleading, since it may falsely identify positively selected sites where none should exist (Suzuki and Nei, 2004). Indeed, recent studies on the Sig1 protein of Thalassiosire weissflogii (Sorhannus, 2003) using both PAML and ADAPTSITE revealed that ML methods detect more residues under selection than the parsimony based methods, though whether or not these additional residues identified by PAML were in fact false positives is unknown. On the other hand, ADAPTSITE is often considered to have low statistical power to detect codons under positive selection (Wong *et al.*, 2004). In three separate analyses of the hemaglutinin gene of the influenza A virus, performed by the authors of each method, the methods again appeared to yield

disparate results, though no formal cross method comparison was conducted. (Suzuki and Gojobori, 1999; Yang *et al.*, 2000; Huelsenbeck *et al.*, 2001).

We formally investigate the bases of these discrepancies by applying each of these methods to the following sequence data, chosen to represent a range of levels of sequence divergence, and therefore a range of levels of phylogenetic uncertainty: a previously studied alignment of abalone sperm lysin, three new alignments of the Avian infectious bronchitis (AIB) virus S gene, and two new alignments of the homologous S protein in the SARS coronavirus. The lysin alignment was previously analyzed by Yang and collaborators (Yang *et al.*, 2000; Yang and Swanson, 2002), and provides a well understood instance of positive selection in action. In the AIB virus S gene, only negative and neutral selection is expected to operate. The SARS S gene represents an extreme case where very little is known about the dominant selective pressures, and the overall level of sequence divergence is very low.

Materials and Methods

Materials

Abalone sperm lysin. This alignment codes for a 122 residue region of the sperm lysin protein for 25 species of California abalone. Abalone reproduction involves species specific sperm-egg recognition in which the sperm lysin binds to and dissolves a complementary vitelline envelope (VE) surrounding the egg cell. This species specific interaction is considered targeted by positive selection of some 23 residues in the lysin protein, as it compensates for ongoing

genetic drift in the VE receptor (Galindo *et al.*, 2003). These data provide a well studied example of positive selection acting on individual amino acid sites, and are included in the PAML distribution. The sequences are sufficiently divergent, with a total tree length of 8.2 nucleotide substitutions per codon, which provides a good number of synonymous substitutions for comparison with the non-synonymous substitutions. Additionally, the crystalline structure of the molecule can be used to support or refute claims of positively selected amino acid sites (Yang *et al.*, 2000). Leading and trailing sites containing gaps were removed prior to analysis and so, residue numbers 1-122 correspond to residues 10-131 in Yang *et al.* (2000).

The spike (S) glycoprotein of the avian infectious bronchitis (AIB) virus. The AIB virus is a single-stranded RNA virus in the coronavirus family. Sequence data for the complete 1,148 residue coding region is only available for 11 isolates however, data for a 520 residue subregion of the gene exist for 38 isolates. Thus, we consider the following alignments: the complete 1,148 residues from 11 isolates, the 520 residue subregion from 38 isolates, and the 520 residue subregion from the 11 isolates for which the complete residue coding region is available. Gap regions were removed prior to analysis. Despite substantial polymorphism between sequences, there is much uncertainty regarding phylogeny which, in this case, only increases with additional sequences. In other words, the divergence among sequences does not carry enough phylogenetic signal in these sequences. A comparison between the sites under positive selection identified in the complete S gene alignments and those sites identified in the partial S gene alignments, for varying sample sizes, provides a way to investigate the effect that phylogenetic uncertainty may have in assessing positive selection.

"Severe Acute Respiratory Syndrome" coronavirus (SARS virus). In order to demonstrate the hazards of attempting to detect positive selection from data with low levels of divergence among sequences, we consider two alignments of the homologous spike protein from the SARS coronavirus. The first alignment consists of a 1,228 residue coding region from 37 isolates, while the second alignment contains only the 18 most divergent sequences from among the original 37. Once again, phylogenetic uncertainty is large, and increases with the number of additional (non-divergent) sequences.

Methods

The influence of positive natural selection on the previously described genes was determined in the following ways. Initial neighbor joining phylogenies for the AIB and SARS S gene alignments were created in PHYLIP by assuming a Kimura 2-parameter nucleotide substitution model. It is worth mentioning again that for the AIB and the SARS S gene alignments, there is a great deal of phylogenetic uncertainty and the neighbor joining method merely provides a simple and fast way to obtain initial phylogenetic estimates. For the lysin alignment, the phylogenetic structure is more resolved, and in this case, the phylogeny of Lee *et al.* (1995) (also provided in the PAML distribution), was used instead of a neighbor joining tree. These phylogenies were used by PAML and MRBAYES as initial values for the estimation algorithms and by ADAPTSITE during reconstruction of ancestral sequences. Following the recommendation of Suzuki and Nei (2004), an additional neighbor joining tree was created for each alignment, based on Suzuki's p-distance metric (proportion of different nucleotide

sites), for use in the ADAPTSITE analyses.

PAML. Using the `codeml` program from the PAML distribution (version 3.14), codon substitution models M0 (null), M1 (neutral), M2 (selection), M3 (discrete), M7 (β) and M8 ($\beta + \omega$) were fitted to each sequence alignment. Likelihood ratio tests (LRTs) between models M2 and M1 and between models M8 and M7, were used to test the null hypothesis that all selection acting on each gene is either purifying or neutral. The LRT between nested models is conducted by comparing twice the difference in log-likelihood values (i.e. $X = 2$[lnL(M2)-lnL(M1)] or $X = 2$[lnL(M8)-lnL(M7)]) against a χ^2_η (chi-squared) distribution, with degrees of freedom equal to the difference in the number of parameters between models ($\eta = 2$ for both of these pairs of models). If the p-value ($1 - Pr[X \leq t]$, with $X \sim \chi^2_\eta$) is less than a given threshold (e.g. 0.05, 0.01), we reject M1 or M7 in favor of M2 or M8. Rejection of models M1 or M7 in favor of models M2 or M8, is usually taken as indicative of the presence of positively selected sites, although more correctly it means that a proportion of sites are estimated to have evolved under $\omega > 1$. We conducted an additional LRT between models M1 and M3 (in this case we compared against a χ^2_η distribution with $\eta = 3$). Provided that at least one of M3's inferred ω values is greater than 1, this can also be considered a test for positive selection. This test has the advantage of not forcing the existence of a neutral selection category with $\omega = 1$. This additional degree of freedom can be important in the case of a gene for which some proportion of the sites are undergoing positive selection in an environment of varying degrees of negative selection.

It is important to note that any of these positive selection models can still be preferred

to its neutral nested counterpart, even when the inferred ω is only marginally greater than one (e.g. 1.01). In such a case sites associated with this positive selection category would more aptly be considered neutrally selected. This is not a failure of the LRT, but a problem with the definition of positive selection. The choice of when to regard a site category with $\omega > 1$ as truly indicative of positive selection, as opposed to neutral selection with stochastic effects, is poorly understood. We report sites associated with any $\omega > 1$ as positively selected, regardless of how close to 1 the value is, for purposes of comparison between methods. For the M2, M3, and M8 analyses, we report the number of sites with posterior probabilities greater than 0.95 and 0.99 of belonging to a positive selection category.

ADAPTSITE. ADAPTSITE (version 1.3) was used to identify positively selected sites via the maximum parsimony ancestral reconstruction method of Suzuki and Gojorobi (Suzuki and Gojobori, 1999; Suzuki *et al.*, 2001; Suzuki and Nei, 2004). For each alignment we conducted 16 experiments with ADAPTSITE, using as input all combinations of the following four phylogenies: the initially estimated neighbor joining phylogenies, the M3 and M8 maximum likelihood phylogenies, and the p-distance neighbor joining phylogeny; and Kimura 2-parameter substitution matrices using each of the following four transition to transversion ratios (κ): $\kappa = 1, 2, 4$ and κ^*, where κ^* is the MLE of κ obtained from PAML.

For each trial, ADAPTSITE constructs separate neutrality tests for each codon using the observed and expected numbers of synonymous and non-synonymous changes in the maximum parsimony reconstruction of ancestral codons. For each codon, the binomial probabilities of obtaining the observed numbers of synonymous and non-synonymous substitutions

assuming selective neutrality are computed by ADAPTSITE, and one-tailed and two-tailed p-values are reported. When one of these p-values for an individual codon is below a given threshold, and the substitution counts show a higher rate of non-synonymous substitutions than synonymous, positive selection is inferred. For the purpose of this discussion, we report sites to be under selection when the two-tailed p-values are below 0.05 and 0.01.

Following the methodology proposed by Wong *et al.* (2004) for comparing results between PAML and ADAPTSITE, for each alignment we also consider the modified Bonferroni procedure (Simes, 1986) for testing the selective neutrality of the gene as a whole. The modified Bonferroni test is conducted by comparing each rank-ordered p-value to a quantity equal to $(\text{rank} \times \delta)/n$, where n is the number of p-values and δ is the threshold used previously to identify positively selected sites. The null hypothesis that only neutral selection acts on the gene is rejected if any p-value is less than the corresponding specified quantity. If there exists such a p-value, we say the test passes, and the presence of positively or negatively selected sites is inferred. The modified Bonferroni test is therefore a test for variation in selective pressures acting on the gene.

MRBAYES. We searched for positively selected codons using the hierarchical Bayesian method developed by Huelsenbeck and Ronquist and implemented in the program MRBAYES (version 3.0B4). PAML codon substitution models M2 and M3 are implemented in MRBAYES, though in the comparison presented here we only report the results obtained with model M3. MRBAYES uses Markov chain Monte Carlo (MCMC) methods to generate draws from the posterior distribution of the parameters of model M3, incorporating

the user's choice of prior distributions. Specifically, the parameters of M3 requiring specification of prior distributions are the following: the transition to transversion rates ratio $\kappa = \mathrm{dT/dR}$; the codon frequencies $\pi_1, \ldots, \pi_{61}$; the three non-synonymous to synonymous rates ratios ω_1, ω_2 and ω_3, with their corresponding frequencies p_1, p_2 and p_3; the parameters associated with the tree topology that we denote by $\mathbf{T}$; and finally, a parameter related to the branch lengths of the tree denoted by B. The following prior distributions were considered: $\mathrm{dT}/(\mathrm{dT} + \mathrm{dR}) \sim \mathrm{Beta}(1, 1)$; $\pi_1, \ldots, \pi_{61} \sim \mathrm{Dirichlet}(1, \ldots, 1)$; $\mathrm{B} \sim \mathit{Exponential}(10)$; a uniform prior on all the bipartite trees of a given taxa was assumed on $\mathbf{T}$; and an exponential prior distribution, was imposed on $\mathrm{dN}_1, \mathrm{dN}_2, \mathrm{dN}_3$ and dS, where $\mathrm{dN}_1, \mathrm{dN}_2$ and dN_3 are the rates of non-synonymous substitutions for each or the three ω categories and dS the rate of synonymous substitutions, with the assumption that $\mathrm{dN}_1 < \mathrm{dN}_2 < \mathrm{dN}_3$ and with $\omega_j = \mathrm{dN}_j/\mathrm{dS}$ for $j = 1, 2, 3$. The rate parameter for the exponential distribution on $\mathrm{dN}_1, \mathrm{dN}_2, \mathrm{dN}_3$ and dS, is specified by MRBAYES and its value is not relevant as these rates are only used through their ratios (see MRBAYES help manual).

The MCMC algorithm is run for a sufficiently large number of iterations so that convergence of the Markov chain is achieved. As this is an extremely computationally intensive task, we limited the number of MCMC iterations to 100,000 for each alignment considered. As in all Bayesian analyses via MCMC, establishing MCMC convergence is of supreme importance (see Huelsenbeck *et al.*, 2002 for a discussion of this in the context of MRBAYES). In our analyses MCMC convergence was established informally, through visual inspection of the MCMC traces for certain model parameters and for functions of the parameters such

as the likelihood function. Convergence seemed to be achieved after the first 20,000-50,000 iterations, except where otherwise noted. The values sampled during MCMC iterations prior to convergence (burn-in values) were discarded. The values sampled after convergence are correlated draws from the joint posterior distribution of parameters. Therefore, we base our inferences on a sample of posterior draws constructed by subsampling the remaining post-convergence MCMC samples every 100th iteration in order to reduce correlation.

During each iteration of the MCMC algorithm, MRBAYES computes the probability that each codon belongs to each selection category, and for each codon, the sum of these probabilities for each selection category with $\omega > 1$ is reported. Thus in MRBAYES, there is a choice to be made as to the operational definition of a positively selected site. We use a definition based on the posterior mean of the probabilities of positive selection for each site, reporting a site as positively selected when this posterior mean probability is greater than 0.95 or 0.99.

Results

Abalone sperm lysin alignment. A summary of the key results obtained from PAML and MRBAYES appears in Table 1. LRTs reject the neutral/negative selection models M1 and M7 in favor of positive selection models M2, M3 and M8 and so, the presence of positively selected sites is inferred. Note that for these data, the ML estimates of the parameters that are most important for detecting positively selected sites were essentially equal for models M2, M3 and M8, as were the maximum log-likelihood values, indicating that these three models

Model	lnL	$\hat{\kappa}$	$\hat{\omega}_{ps}$	$\hat{p}_{ps}$	number of positively selected sites at 0.95 (0.99) significance level
M2	-3932.87	1.635	3.249	0.243	19 (13)
M3	-3932.65	1.621	3.080	0.254	20 (14)
M8	-3933.16	1.605	2.981	0.255	19 (13)
MB	-4009.18	1.669	3.227	0.299	17 (16)

Table 1: PAML and MRBAYES (MB) results for abalone sperm lysin.

appear to fit the data equally well. M2 and M8 identified exactly the same 19 positively selected sites at the 0.95 level, and the same 13 sites at the 0.99 level. The one additional positively selected site identified by M3 at the 0.95 and 0.99 levels had high probability of belonging to the positive selection category in M2 and M8 (greater than 0.9) and so, this discrepancy is not cause for concern. These results are consistent with those of Yang *et al.* (2000).

Results from MRBAYES largely agree with those from PAML. For each parameter, we report the mean of the marginal posterior distribution, computed empirically using the posterior samples obtained via MCMC. The posterior mean estimates of κ, the ω for the positive selection category and its frequency are slightly larger than the corresponding ML estimates, but the ML values are well within the central regions of the posterior distributions of these parameters. The positively selected sites identified by MRBAYES agreed with those obtained in PAML. MRBAYES identified 17 sites at the 0.95 level, and 16 sites at the 0.99 level. All 17 sites were among those identified by M2, M3 and M8 in PAML.

For ADAPTSITE, the modified Bonferroni test failed for all trials at the 0.95 and 0.99 significance levels and so, based on these results no selection (positive or negative) can be

inferred by this method. In contrast to the results obtained from PAML and MRBAYES, this suggests that all sites are undergoing strictly neutral selection. Even ignoring the Bonferroni results, no sites met the criteria for positive selection based on the estimated rates of substitution and the 2-tailed p-values. Only 7 sites were identified as negatively selected at the 0.95 level, but again, in the light of the Bonferroni test, these results are not statistically significant. The discrepancies between these results and those obtained through PAML and MRBAYES emphasize the fact that ADAPTSITE is far more conservative than PAML and MRBAYES in detecting sites under positive selection.

Complete and Partial Avian S Gene Alignments. First we consider the alignment consisting of 11 complete S gene encodings. PAML and MRBAYES results are summarized in Table 2. LRTs appeared to yield conflicting information: The LRT between M2 and M1 fails to reject M1; however, the LRT between M3 and M1 does reject M1, and the LRT between M8 and M7 rejects M7. The apparent contradiction is that the first LRT (M2 vs M1) appears to support the hypothesis that only neutral or negative selection acts on the gene (which in fact is what biological insight suggests), while the second two LRTs (M3 vs M1 and M8 vs M7) appear to suggest the presence of positively selected sites. This confusion is due to the misinterpretation of the LRT. Given that M3 and M8 infer a category for which $\omega > 1$, they do identify a class of putative positively selected sites. However, the ω values related to such class, $\omega_{ps} = 1.012$ and $\omega_{ps} = 1.063$ for M3 and M8 respectively, are not strong evidence for positive selection. Rather, the superior fit provided by M3 and M8 is due to these models' greater flexibility in describing the degrees of negative selection acting on these data. Note

that for M2 the value of ω_{ps} was apparently unidentifiable, with ω_{ps} growing unreasonably large and p_{ps} going to 0. In model M3, 37 sites were associated with the positive selection category at the 0.95 level, while only 16 sites were associated with the positive selection category for M8. The 19 sites identified in M3 but not in M8 all had fairly high probabilities (> 0.8) of belonging to the positive selection category with $\omega_{ps} = 1.063$.

Model	lnL	$\hat{\kappa}$	$\hat{\omega}_{ps}$	$\hat{p}_{ps}$	number of positively selected sites at 0.95 (0.99) significance levels
M1	-14883.74	2.558	none	none	none
M2	-14883.74	2.558	23.825	0.000	0 (0)
M3	-14827.79	2.450	1.012	0.093	37 (15)
M7	-14840.70	2.403	none	none	none
M8	-14828.57	2.450	1.063	0.081	16 (4)
MB	-15431.57	2.128	1.730	0.091	45 (29)

Table 2: PAML and MRBAYES (MB) results for 11 sequence complete AIB S gene alignment.

The analysis with MRBAYES apparently revealed stronger evidence for positive selection than any of the PAML analyses, with a mean posterior value $\omega_{ps} = 1.730$. This led to 45 sites being identified as positively selected at the 0.95 level, and 29 at the 0.99 level. However, cross methods comparisons (see Table 4) show that as many as 31 of these sites belonged to the positive selection category in PAML with $\omega_{ps} = 1.012$.

Table 3 summarizes ADAPTSITE results. Here, although the modified Bonferroni tests failed to detect the presence of any non-neutrally selected sites, a number of individual sites were identified as positively selected, where that number ranged from 0 to 43 for different choices of phylogeny and κ. Assuming results obtained using the MLE $\kappa = 2.5$ will best represent the data, the number of positively selected sites varies between 29 sites under the

ML phylogeny, and 1 site under the p-distance tree. Table 4 shows that very few of these positively selected sites were identified with the positive selection category in either PAML or MRBAYES. In fact, 14 of the 25 sites identified using both the initial neighbor joining tree, as well as the ML tree, contained either no substitutions or only silent substitutions in the actual sequence alignment. The identification of sites as positively selected can only be explained by the presence of non-synonymous substitutions at these codons in ADAPTSITE's maximum parsimony reconstruction of ancestral sequences, as no information suggests that the influence of positive selection on these sites was present in the sequence alignment. This is an important instance in which adhering to the Bonferroni test results prevents the misidentification of a large number of codons.

phylogeny	κ	modified Bonferroni	number of positively selected sites at 0.95 (0.99) significance levels
NJ-g	1	fail (fail)	12 (0)
	2	fail (fail)	20 (0)
	2.5	fail (fail)	27 (0)
	4	fail (fail)	43 (2)
M3-ML	1	fail (fail)	14 (0)
	2	fail (fail)	24 (0)
	2.5	fail (fail)	29 (1)
	4	fail (fail)	43 (6)
M8-ML	1	fail (fail)	14 (0)
	2	fail (fail)	24 (0)
	2.5	fail (fail)	29 (1)
	4	fail (fail)	43 (6)
NJ-p	1	fail (fail)	0 (0)
	2	fail (fail)	1 (0)
	2.5	fail (fail)	1 (0)
	4	fail (fail)	1 (0)

Table 3: ADAPTSITE results for 11 sequence complete AIB S gene alignment.

	M3	M8	Adpt (NJ-g,$\hat{\kappa}$)	Adpt (M8-ML,$\hat{\kappa}$)	Adpt (NJ-p,$\hat{\kappa}$)
M8	16 (4)				
Adpt (NJ-g,$\hat{\kappa}$)	2 (0)	0 (0)			
Adpt (M8-ML,$\hat{\kappa}$)	3 (0)	1 (0)	25 (0)		
Adpt (NJ-p,$\hat{\kappa}$)	1 (0)	1 (0)	0 (0)	1 (0)	
MB	31 (10)	12 (3)	2 (0)	3 (0)	1 (0)

Table 4: Numbers of positively selected sites commonly identified in the 11 sequence complete AIB S gene alignment.

Reducing the length of the sequences to a 520 residue region yielded similar results for these 11 sequences. PAML and MRBAYES results appear in Table 5. LRTs for these analysis failed to reject M1 for M2, but did rejected M1 in favor of M3, and M7 in favor of M8. The ω_{ps} values for the positive selection category in M3 and M8 increased compared to the previous analyses, as did the p_{ps} values. This may indicate a stronger signal for positive selection in this subregion, despite M2's estimate of $\omega_{ps} = 1.000$, or it may be another instance of the superior fit of M3 and M8 arising through greater flexibility in describing the varying degrees of negative selection. Here, M3 identified 25 positively selected sites at the 0.95 level, while M8 detected only 3. These numbers fall to 10 and 1, respectively, at the 0.99 level.

Model	lnL	$\hat{\kappa}$	$\hat{\omega}_{ps}$	$\hat{p}_{ps}$	number of positively selected sites at 0.95 (0.99) significance levels
M1	-7243.50	2.328	none	none	none
M2	-7243.50	2.328	1.000	0.085	0 (0)
M3	-7203.00	2.218	1.114	0.130	25 (10)
M7	-7210.67	2.165	none	none	none
M8	-7205.08	2.226	1.299	0.087	3 (1)
MB	-7510.94	1.963	1.745	0.149	31 (10)

Table 5: PAML and MRBAYES results for 11 sequence partial AIB S gene alignment.

MRBAYES produced an estimate for ω_{ps} almost identical to the previous MRBAYES estimate for the complete S gene, and a slightly larger estimate of p_{ps}, leading to the identification of 31 positively selected sites at the 0.95 level. This number falls to 10 at the 0.99 level, 8 of which are also identified as positively selected by PAML's M3. Thus at the 0.99 level, the MRBAYES results seem fairly consistent with those obtained by PAML. All ADAPTSITE trials failed the Bonferroni test, and no sites met the criteria for positive selection.

Increasing the number of partial S gene sequences to 38 apparently strengthened the positive selection signal yet again, as the PAML and MRBAYES results in Table 6 demonstrate. The outcome of the LRTs was the same as in each of the previous AIB S gene alignments, but the ω_{ps} value in M3 increased to 1.261, with $p_{ps} = 0.091$ and for M8 these values increased to $\omega_{ps} = 1.600$ and $p_{ps} = 0.060$. These values lead to 37 positively selected sites under M3 and 18 sites under M8, at the 0.95 level.

Model	lnL	$\hat{\kappa}$	$\hat{\omega}_{ps}$	$\hat{p}_{ps}$	number of positively selected sites at 0.95 (0.99) significance levels
M1	-15159.20	2.455	none	none	none
M2	-15159.20	2.455	1.000	0.000	0 (0)
M3	-15018.35	2.398	1.261	0.091	37 (28)
M7	-15034.90	2.327	none	none	none
M8	-14992.44	2.417	1.600	0.060	18 (10)
MB	-15460.45	1.970	1.725	0.183	35 (21)

Table 6: Summary of PAML and MRBAYES results for 38 sequence partial AIB S gene alignment.

MRBAYES inferred the presence of a category of positively selected sites, with posterior mean values $\omega_{ps} = 1.725$ and $p_{ps} = 0.183$. MRBAYES identified 35 positively selected sites at the 0.95 probability level, 29 of which were also identified as positively selected by

PAML's M3 (Table 7). Of the 21 sites identified by MRBAYES at the 0.99 level, 18 were also identified by PAML's M3 at the 0.99 level. Thus, once again, there is a satisfactory degree of consistency between PAML and MRBAYES.

	M3	M8	Adpt (NJ-g,$\hat{\kappa}$)	Adpt (M8-ML,$\hat{\kappa}$)	Adpt (NJ-p,$\hat{\kappa}$)
M8	18 (10)				
Adpt (NJ-g,$\hat{\kappa}$)	2 (0)	1 (0)			
Adpt (M8-ML,$\hat{\kappa}$)	2 (0)	1 (0)	3 (0)		
Adpt (NJ-p,$\hat{\kappa}$)	0 (0)	0 (0)	0 (0)	0 (0)	
MB	29 (18)	16 (9)	3 (0)	3 (0)	0 (0)

Table 7: Numbers of positively selected sites commonly identified in the 38 sequence partial S gene alignment.

ADAPTSITE results for the 38 partial S gene alignment appear in Table 8. All the Bonferroni tests passed for this alignment, indicating the presence of non-neutrally selected sites. Using the MLE for κ, $\hat{\kappa} = 2.4$, the same 3 positively selected sites were identified at the 0.95 level when the initial neighbor joining tree or the maximum likelihood tree were used. Using the p-distance tree, no sites were identified as positively selected unless κ was increased to 4. Of these 3 positively selected sites, 2 were identified by M3, 1 by M8, and all 3 were identified by MRBAYES (see Table 7).

Table 9 compares results across all three AIB S gene alignments by showing the numbers of simultaneously identified sites for each pair of selected analyses. For this purpose we chose the following three analyses from among all those conducted for each alignment: M3 in PAML, MRBAYES, and ADAPTSITE using the initial neighbor joining tree and the

phylogeny	κ	modified Bonferroni	number of positively selected sites at 0.95 (0.99) significance levels
NJ-g	1	pass (pass)	0 (0)
	2	pass (pass)	3 (0)
	2.4	pass (pass)	3 (0)
	4	pass (pass)	8 (0)
M3-ML	1	pass (pass)	1 (0)
	2	pass (pass)	3 (0)
	2.4	pass (pass)	3 (0)
	4	pass (pass)	9 (0)
M8-ML	1	pass (pass)	1 (0)
	2	pass (pass)	3 (0)
	2.4	pass (pass)	3 (0)
	4	pass (pass)	9 (0)
NJ-p	1	pass (pass)	0 (0)
	2	pass (pass)	0 (0)
	2.4	pass (pass)	0 (0)
	4	pass (pass)	1 (0)

Table 8: ADAPTSITE results for 38 sequence partial AIB S gene alignment.

MLE of κ. Between different alignments, the most consistency was observed between like-methods: for example, of the 25 sites identified by M3 in the 11 partial sequences, 24 were previously identified by M3 in the 11 complete sequences, and of the 31 sites identified by MRBAYES in the 11 partial sequences, 30 were previously identified in the 11 complete sequences. After reducing the length of the 11 original sequences to the 520-residue region, the number of positively selected sites decreased. After increasing the number of partial sequences from 11 to 38, the number of positively selected sites increased, although this increase did not recover the sites lost after the reduction in sequence length. Since ω_{ps} and p_{ps} values were not drastically different for the various alignments, the discrepancies between PAML and MRBAYES between alignments are presumably due primarily to different estimates of

branch lengths for the phylogenies. The differences in the number of identified sites may seem disturbing, but in general the sites identified by PAML and not by MRBAYES (and vice versa) had high probabilities (> 0.8 in most cases) of positive selection in the other method, though not at the 0.95 or 0.99 level. The differences in numbers do emphasize the sensitivity of results to the definition of positive selection. The differences between these methods and ADAPTSITE would be more disturbing if the Bonferroni tests had indicated that the ADAPTSITE results for individual sites were statistically significant, however, note that none of the 3 sites identified by ADAPTSITE in the 38 partial sequences were among the 22 identified in the corresponding region of the 11 partial sequences.

		11 partial sequences			38 partial sequences		
		M3	Adpt (NJ-g,$\hat{\kappa}$)	MB	M3	Adpt (NJ-g,$\hat{\kappa}$)	MB
11 complete sequences	M3	24 (10)	0 (0)	23 (9)	22 (7)	2 (0)	20 (5)
	Adpt (NJ-g,$\hat{\kappa}$)	2 (0)	0 (0)	2 (0)	3 (0)	0 (0)	7 (1)
	MB	19 (8)	0 (0)	30 (18)	19 (12)	2 (0)	18 (5)

		38 partial sequences		
		M3	Adpt (NJ-g,$\hat{\kappa}$)	MB
11 partial sequences	M3	19 (4)	1 (0)	16 (2)
	Adpt (NJ-g,$\hat{\kappa}$)	0 (0)	0 (0)	0 (0)
	MB	16 (9)	2 (0)	15 (5)

Table 9: Numbers of positively selected sites identified in common for each AIB S gene alignment.

SARS Spike Protein Alignments. Table 10 summarizes PAML and MRBAYES results for the 38 sequences SARS S gene alignment. LRTs favored models M2, M3 and M8 over models M1

and M7. Estimates of model parameters κ, ω_{ps} and p_{ps} were remarkably consistent between M2, M3 and M8, with each model identifying the same 9 strongly positively selected sites at the 0.95 level. MRBAYES, however, had great difficulty with this alignment. Initializing the MCMC with the same neighbor joining phylogeny provided to PAML and ADAPTSITE caused MRBAYES to crash and so, we were forced to start the analysis with a randomly chosen phylogeny. After 100,000 iterations of sampling, the MCMC did not appear to have reached convergence, and so the chain was run for an additional 100,000 iterations. Even after these additional iterations, the chain still did not show convergence. The problem was related to the parameters ω_3 and p_3 (which are usually related to the positive selection category), with the sampled values of ω_3 growing to values between 38 and 78 and the values of p_3 decreasing to 0. We believe this to be due to an identifiability problem, similar to that encountered with PAML's M2 in the analysis of the complete AIB S gene sequences. Consequently, MRBAYES did not identify any positively selected sites.

Model	lnL	$\hat{\kappa}$	$\hat{\omega}_{ps}$	$\hat{p}_{ps}$	number of positively selected sites at 0.95 (0.99) significance levels
M2	-5617.66	4.811	12.493	0.052	9 (2)
M3	-5617.66	4.811	12.501	0.052	9 (2)
M8	-5618.69	4.819	13.120	0.048	9 (2)
MB	-5881.29	2.846	60.195	0.040	0 (0)

Table 10: PAML and MRBAYES (MB) results for 38 sequence SARS S gene alignment.

ADAPTSITE results for this alignment appear in Table 11. Here not only did every Bonferroni test indicate the presence of non-neutrally selected sites, but also a huge number of sites met the criteria for positive selection. Of the 153 sites identified as positively selected

for every choice of phylogeny (see Table 12), 142 did not contain any non-synonymous substitutions in the sequence alignment. Of the remaining 11 sites, only 2 were among those identified by M2, M3 and M8.

phylogeny	κ	modified Bonferroni	number of positively selected sites at 0.95 (0.99) significance levels
NJ-g	1	pass (pass)	113 (82)
	2	pass (pass)	133 (109)
	4	pass (pass)	153 (125)
	4.8	pass (pass)	163 (127)
M3-ML	1	pass (pass)	113 (83)
	2	pass (pass)	132 (112)
	4	pass (pass)	153 (125)
	4.8	pass (pass)	163 (127)
M8-ML	1	pass (pass)	113 (82)
	2	pass (pass)	132 (109)
	4	pass (pass)	153 (125)
	4.8	pass (pass)	163 (127)
NJ-p	1	pass (pass)	112 (85)
	2	pass (pass)	130 (109)
	4	pass (pass)	163 (127)
	4.8	pass (pass)	164 (132)

Table 11: ADAPTSITE results for 38 sequence SARS S gene alignment.

Removal of the 20 most similar sequences from this alignment improved the behavior of MRBAYES and caused increased numbers of positively selected sites in PAML (see Table 13). Again LRTs favored models M2, M3 and M8 over M1 and M7, and all three of these preferred models yielded virtually equivalent inferences. The number of sites identified by the three PAML models rose to 51 at the 0.95 and 0.99 level. For these data MRBAYES appeared to achieve convergence, although the parameter values as well as the number of positively selected sites were both much lower than in PAML. The 51 sites identified by PAML

	M3	M8	Adpt (NJ-g,$\hat{\kappa}$)	Adpt (M8-ML,$\hat{\kappa}$)	Adpt (NJ-p,$\hat{\kappa}$)
M8	9 (2)				
Adpt (NJ-g,$\hat{\kappa}$)	2 (2)	2 (2)			
Adpt (M8-ML,$\hat{\kappa}$)	2 (2)	2 (2)	163 (127)		
Adpt (NJ-p,$\hat{\kappa}$)	2 (2)	2 (2)	153 (121)	153 (121)	
MB	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

Table 12: Numbers of positively selected sites commonly identified in the 38 sequence SARS S gene alignment.

were the only 51 sites displaying any non-synonymous substitutions in this alignment. Thus, every site displaying a single non-synonymous substitution in this alignment was identified as positively selected, even when the overall incidence of non-synonymous substitutions was less than 10%. Incidentally, this was not the case in the 38 sequence alignment; in that alignment there were non-positively selected sites with higher incidence of non-synonymous substitutions than those identified as positively selected. This highlights the influence of phylogeny on these inferences.

Model	lnL	$\hat{\kappa}$	$\hat{\omega}_{ps}$	$\hat{p}_{ps}$	number of positively selected sites at 0.95 (0.99) significance levels
M2	-5482.75	4.287	7.833	0.140	51 (51)
M3	-5482.75	4.287	7.830	0.140	51 (51)
M8	-5482.74	4.287	7.830	0.140	51 (51)
MB	-5605.62	4.582	5.709	0.002	1 (1)

Table 13: PAML and MRBAYES results for 18 sequence SARS S gene alignment.

ADAPTSITE results, shown in Table 14, were similar to those from the analysis of the 38 sequences. All Bonferroni tests indicated the presence of non-neutrally selected sites, and

phylogeny	κ	modified Bonferroni	number of positively selected sites at 0.95 (0.99) significance level
NJ-g	1	pass (pass)	46 (24)
	2	pass (pass)	73 (32)
	4	pass (pass)	100 (42)
	4.2	pass (pass)	101 (42)
M3-ML	1	pass (pass)	46 (24)
	2	pass (pass)	73 (32)
	4	pass (pass)	100 (42)
	4.2	pass (pass)	101 (42)
M8-ML	1	pass (pass)	46 (24)
	2	pass (pass)	73 (32)
	4	pass (pass)	100 (42)
	4.2	pass (pass)	101 (42)
NJ-p	1	pass (pass)	86 (41)
	2	pass (pass)	105 (54)
	4	pass (pass)	133 (63)
	4.2	pass (pass)	137 (64)

Table 14: ADAPTSITE results for 18 sequence SARS S gene alignment.

there were 93 sites identified as positively selected for all phylogenies (Table 15). However, given that the 51 sites identified by PAML were the only sites containing non-synonymous substitutions, and the overlap between PAML results and ADAPTSITE results is at most 8 sites (depending on phylogeny), this means that ADAPTSITE consistently identified at least 85 monomorphic sites as positively selected.

Table 16 compares results between SARS S gene alignments. In this case, reduction of the number of sequences considered reduced phylogenetic uncertainty and increased the number of positively selected sites identified. Of the 9 positively selected sites by PAML in the 38 sequence alignment, 8 of them were also among those 51 identified by PAML in the 18 sequence alignment. The one site identified by PAML in the 38 sequence alignment not

	M3	M8	Adpt (NJ-g,$\hat{\kappa}$)	Adpt (M8-ML,$\hat{\kappa}$)	Adpt (NJ-p,$\hat{\kappa}$)
M8	51 (51)				
Adpt (NJ-g,$\hat{\kappa}$)	7 (2)	7 (2)			
Adpt (M8-ML,$\hat{\kappa}$)	7 (2)	7 (2)	101 (42)		
Adpt (NJ-p,$\hat{\kappa}$)	8 (4)	8 (4)	93 (16)	93 (16)	
MB	1 (1)	1 (1)	0 (0)	0 (0)	0 (0)

Table 15: Numbers of positively selected sites commonly identified in the 18 sequence SARS S gene alignment.

		18 sequence alignment		
		M3	Adpt (NJ-g,$\hat{\kappa}$)	MB
38 sequence alignment	M3	8 (2)	2 (1)	1 (0)
	Adpt (NJ-g,$\hat{\kappa}$)	11 (7)	101 (42)	0 (0)
	MB	0 (0)	0 (0)	0 (0)

Table 16: Numbers of positively selected sites identified in common in the SARS S gene alignments.

identified in the 18 sequence alignment was one for which all non-synonymous substitutions present in the 38 sequence alignment were contained in the 20 sequences removed to create the 18 sequence alignment. Consistency between ADAPTSITE and PAML for the different alignments was similar to the consistency between those methods within the analysis of each alignment. ADAPTSITE's inferences were consistent between alignments, although once again we stress that many of the sites identified by ADAPTSITE are almost surely false positives.

Discussion and conclusions

Our experiences with PAML, ADAPTSITE, and MRBAYES made evident several problems with using these methods for detecting positive selection in divergent DNA sequences when little attention is paid to the assumptions made by the model regarding the data. The sequence data considered in these analyses were chosen to represent a spectrum of divergence levels. The lysin alignment is representative of the type of data for which these methods were expressly designed, having a strongly supported phylogenetic structure and a small number of well separated dominant selective pressures. Although it can be said that none of these methods were intended to be used with data of the type represented by the AIB S gene and SARS alignments, the question "does positive selection act on these proteins?" is an interesting and relevant question.

Our results highlight certain problems described in previous studies in which ADAPTSITE and PAML were the only methods being compared. In these studies ADAPTSITE tends to identify a much smaller number of codons (sometimes none) under selection than PAML. This has been taken as evidence that ADAPTSITE is more conservative and less prone to produce false positives than PAML (Suzuki and Nei, 2002; Suzuki and Nei, 2004). However, recent studies suggest that this is a consequence of lack of power in ADAPTSITE, rather than a good feature of the methodology (Wong *et al.*, 2004). Indeed, in our analysis of the lysin alignment, not a single codon was identified as positively selected by ADAPTSITE, while both PAML and MRBAYES identified a similar number of codons under positive selection.

Major problems were encountered in the analyses of data with low overall genetic diversity and, consequently, low phylogenetic signal.

The most obvious concern has to do with ADAPTSITE identifying sites as positively selected even in cases where not a single non-synonymous mutation was observed in the data. This phenomenon occurred in the analyses of both the AIB S gene alignment and the SARS S gene alignment. In such cases the maximum parsimony ancestral reconstruction process simply overestimates the number of non-synonymous changes, resulting in the false identification of positively selected sites. In such circumstances, ADAPTSITE does not behave as a conservative method. This highlights, in our opinion, the limitations of the parsimony based methods when used to count mutations in sequences with very low levels of divergence. The use of the modified Bonferroni test to assess the overall significance of the ADAPTSITE results can be useful in preventing the misidentification of positively selected sites, as was the case for the AIB S gene sequences, but even statistically significant results can contain false positives, as in the case of the SARS S gene.

Another major conclusion is that the number of codons identified as codons under selection strongly depends on the data in terms of the number and length of sequences included. This appears to be particularly critical in the data sets containing several lowly divergent sequences. In the case of the SARS S gene, the exclusion of the most lowly divergent sequences increased dramatically the number of codons identified as positively selected by PAML and ADAPTSITE. The number of sites identified by MRBAYES did not increase, however, this result should not be regarded as conclusive, given that problems with MCMC convergence

were evident. A conservative approach derived from these analyses would be to use only complete sequences, whenever possible, and carefully explain the sampling when extensive data sets are analyzed, specially in cases where the phylogenetic signal is not strong among the sequences.

We believe PAML to have two weaknesses. The first is the inability to report any estimation errors, which is critical for cases in which the likelihood surface is very flat. The second weakness is far more subtle. In order to make the process of fitting the evolutionary models to real data computationally tractable, PAML's evolutionary models involve a number of simplifying assumptions, the most problematic of which is that of fixing the number of selection categories. The problem arises in choosing the number of such categories. Throughout these analyses, we have assumed 3 selection categories for M3, in part for compatibility with MRBAYES, and in part because in published analyses using PAML, this is the number of categories often chosen. When this assumption does not adequately describe the degree of variation of the selective pressures in the data, interpreting the results becomes difficult. This is because each selection category, represented by a single ω value, in fact represents a range of ω values. A site then becomes associated with a particular selection category of the available ones, when the site is more likely to have evolved under the ω and p values representing that category than those of any other category. This is what makes the identification of weak positive selection (in contrast to neutral selection) so difficult. For instance, suppose that in a given data set just a few sites are undergoing positive selection. In the PAML analysis of those data, these sites would probably receive high posterior probabilities of belonging

to the w_{ps} category, where $w_{ps} = 1.01$. We would be unlikely to interpret that category of sites as positively selected and so, those sites would be missed. Conversely, suppose that the majority of the sites are undergoing negative selection. Specifically, assume that 50% of the sites are undergoing strong negative selection with $w = 0.1$, and another 30% with $w = 0.5$. Suppose then that another 10% of sites are undergoing neutral evolution with $\omega = 1$, and the remaining 10% are undergoing positive selection with $\omega = 2$. A likely ML scenario would be that the negative selection categories are inferred, but that the neutral and positive selection categories are lumped together with $w_3 = 1.5$. To avoid such problems, the only current solution seems to be attempting to arrive at the optimal number of site categories, through repeated analyses in PAML with increasing numbers of categories. LRTs can be conducted between models to select the optimal number of categories.

In our analyses, the MRBAYES results were affected by the same properties of the data that affected the results of PAML. This is to be expected, given that both methods are based on the same model of sequence evolution. MRBAYES has the capacity to produce more robust results when faced with phylogenetic and model parameter uncertainty, provided appropriate prior distributions are used, convergence is achieved, and that the three categories of selective pressure can adequately describe the data. The limitations of MRBAYES are that the software provides limited choice in prior distributions, convergence is difficult to assess (especially in terms of the phylogeny) and there is no straightforward mechanism implemented in the software for comparing negative/neutral selection models with positive selection models in a manner analogous to the LRT in PAML. We believe that the Bayesian

framework has the potential to yield tremendous insight into the positive selection problem, by allowing a more flexible definition of positive selection based on posterior distributions of the relevant parameters. Through the effective use of prior distributions, a Bayesian methodology could be made to yield useful results, with respect to the prior beliefs, even when divergence is low and only a small number of substitutions is observed. Although MRBAYES does not provide for model testing, we emphasize that it is possible to conduct model comparisons within the Bayesian framework.

Any apparent failure of the methods considered should not be seen as a failure of the methods in terms of doing what they are designed mathematically to do, which is to fit a specified model, but rather a failure for that model's definition of positive selection to coincide with our biological understanding of that process. Clearly, there are situations in which these definitions coincide. The situations in which they do not seem to arise as a result of certain simplifying assumptions made by the models, which are violated in different ways by different data. As there is usually no way to foresee such violations, the only available alternative is to regard the output of these methods with a great deal of scrutiny, carefully investigating any discrepancies between substitution counts and inferences, and the ability of the model to effectively capture the apparent diversity in selective pressures.

Acknowledgments

We would like to thank Ziheng Yang and John Huelsenbeck for their many helpful answers to our many questions. We also thank Omar E. Cornejo for his assistance. This investigation was supported by the NIH/NIGMS grant 1R01GM720030-01.

Literature Cited

Anisimova, M., Bielawski, J. and Yang, Z. (2002) Accuracy and power of Bayes prediction of amino acid sites under positive selection. Molecular Biology and Evolution, **19(6)**:950–958.

Escalante, A. A., Cornejo, O. E., Rojas, A., Udhayakumar, V. and Lal, A. A. (2004) Assessing the effect of natural selection in malaria parasites. Trends Parasitol., **20(8)**:388–395.

Fitch, W.M., Bush, R.M., Bender, C.A. and Cox, N.J. (1997) Long term trends in the evolution of H(3) HA1 human influenza type A. Proc Natl Acad Sci USA, **94(15)**:7712–7718.

Galindo, B., Vacquier, V. and Swanson, W. (2003) Positive selection in the egg receptor for abalone sperm lysin. PNAS, **100(8)**:4639–4643.

Gillespie, J. H. (1991) *The Causes of Molecular Evolution*. New York: Oxford University Press.

Goldman, N. and Yang, Z. (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol, **11(5)**:725–36.

Huelsenbeck, J. P., Ronquist, F., Nielsen, R. and Bollback, J.P. (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. Science, **294**:2310–2314.

Huelsenbeck, J.P., Larget, B., Miller, R.E. and Ronquist, F. (2002) Potential applications and pitfalls of Bayesian inference of phylogeny. Syst Biol, **51(5)**:673–688.

Huelsenbeck, J.P. and Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics, **17(8)**:754–755.

Hughes, A. L. (2000) *Adaptive Evolution of Genes and Genomes*. New York: Oxford University Press.

Hughes, A. L. and Nei, M. (1990) Evolutionary relationships of class ii major histocompatibility complex genes in mammals. Mol. Biol. Evol., **7**:491–514.

Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.

Kreitman, M. and Akashi, H. (1995) Molecular evidence for natural selection. Annu. Rev. Ecol. Syst., **26**:403–422.

Lee, Y. H., Ota, T. and Vacquier, V. D. (1995) Positive selection is a general phenomenon in the evolution of abalone sperm lysin. Mol. Biol. Evol., **12**:231–238.

Nei, M. and Kumar, S. (2000) *Molecular Evolution and Phylogenetics*. New York: Oxford University Press.

Nielsen, R. and Yang, Z. (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics, **148(3)**:929–936.

Simes, R. J. (1986) An improved Bonferroni procedure for multiple tests of significance. Biometrika, **73**:751–754.

Sorhannus, U. (2003) The effect of positive selection on a sexual reproduction gene in Thalassiosira weissflogii (Bacillariophyta): Results obtained from maximum-likelihood and parsimony-based methods. Mol. Biol. Evol., **20(8)**:1326–1328.

Suzuki, Y. and Gojobori, T. (1999) A method for detecting positive selection at single amino acid sites. Mol. Biol. Evol, **16(10)**:1315–1328.

Suzuki, Y., Gojobori, T. and Nei, M. (2001) ADAPTSITE: detecting natural selection at single amino acid sites. Bioinformatics, **17(7)**:660–661.

Suzuki, Y. and Nei, M. (2002) Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites. Mol. Biol. Evol., **19(11)**:1865–1869.

Suzuki, Y. and Nei, M. (2004) False-positive selection identified by ML-based methods: examples from the *sig1* gene of the diatom *thalassiosira weissfloggi* and the *tax* gene of a human T-cell lymphotropic virus. Mol. Biol. Evol., **21**:914–921.

Tzeng, Y. H., Pan, R. and Li, W. H. (2004) Comparison of three methods for estimating rates of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol., **21(12)**:2290–2298.

Wong, W. S. W., Yang, Z., Goldman, N. and Nielsen, R. (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences for identifying positively selected sites. Genetics, **168**:1041–1051.

Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci, **13(5)**:555–556.

Yang, Z. (2003) Phylogenetic analysis by maximum likelihood. Version 3.13. University College, London.

Yang, Z., Nielsen, R., Goldman, N. and Pedersen, A.M. (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics, **155(1)**:431–449.

Yang, Z. and Swanson, W. J. (2002) Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Molecular Biology and Evolution, **19(1)**:49–57.

Yang, Z., Swanson, W.J. and Vacquier, V.D. (2000) Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. Mol Biol Evol, **17(10)**:1446–55.