Evaluation of an Improved Branch-Site Likelihood Method for Detecting Positive Selection at the Molecular Level

Jianzhi Zhang,* Rasmus Nielsen,† and Ziheng Yang‡

*Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor; †Centre for Bioinformatics, University of Copenhagen, Denmark; ‡Department of Biology, University College London, London, United Kingdom

Detecting positive Darwinian selection at the DNA sequence level has been a subject of considerable interest. However, positive selection is difficult to detect because it often operates episodically on a few amino acid sites, and the signal may be masked by negative selection. Several methods have been developed to test positive selection that acts on given branches (branch methods) or on a subset of sites (site methods). Recently, Yang, Z., and R. Nielsen (2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol. **19**:908–917) developed likelihood ratio tests (LRTs) based on branch-site models to detect positive selection that affects a small number of sites along prespecified lineages. However, computer simulations suggested that the tests were sensitive to the model assumptions and were unable to distinguish between relaxation of selective constraint and positive selection (Zhang, J. 2004. Frequent false detection of positive selection by the likelihood method with branch-site models. *Mol. Biol. Evol.* **21**:1332–1339). Here, we describe a modified branch-site model and use it to construct two LRTs, called branch-site tests 1 and 2. We applied the new tests to reanalyze several real data sets and used computer simulation to examine the performance of the two tests by examining their false-positive rate, power, and robustness. We found that test 1 was unable to distinguish relaxed constraint from positive selection affecting the lineages of interest, while test 2 had acceptable false-positive rates and appeared robust against violations of model assumptions. As test 2 is a direct test of positive selection on the lineages of interest, it is referred to as the branch-site test of positive selection and is recommended for use in real data analysis. The test appeared conservative overall, but exhibited better power in detecting positive selection than the branch-based test. Bayes empirical Bayes identification of amino acid sites under positive selection along the foreground branches was found to be reliable, but lacked power.

Introduction

Positive Darwinian selection at the DNA sequence level is often tested by estimating the ratio (ω) of the rate of nonsynonymous (d_N) nucleotide substitutions to that of synonymous (d_S) substitutions between homologous protein-coding gene sequences (Nei and Kumar 2000). An ω value significantly higher than 1 is interpreted as evidence for positive selection, $\omega < 1$ suggests purifying selection (selective constraints), and $\omega = 1$ indicates neutral evolution. Detecting positive selection is generally difficult because positive selection often acts on a few sites and in a short period of evolutionary time, and the signal may be swamped by the ubiquitous negative selection. Recent years have seen the development of several methods for testing positive selection on an individual branch, or a set of branches, of a phylogenetic tree (i.e., branch methods) (Yu and Irwin 1996; Messier and Stewart 1997; Zhang, Kumar, and Nei 1997; Yang 1998; Zhang, Rosenberg, and Nei 1998) or on individual codon sites (i.e., site methods) (Nielsen and Yang 1998; Suzuki and Gojobori 1999; Yang et al. 2000).

More recently, Yang and Nielsen (2002) introduced a branch-site method for testing positive selection on individual codons along specific lineages. In this test, branches of the tree are divided *a priori* into foreground and background lineages, and a likelihood ratio test (LRT) is constructed by comparing a model that allows positive selection on the foreground lineages with a model that does not allow such positive selection. However, computer simulations showed that this test was sensitive to the underlying model, and, when the model assumptions were violated, could lead to frequent rejection of the null hypotheses of no selection, producing false positives (Zhang 2004). If some sites evolve under negative selection on the background lineages, but experience a relaxation of constraints on the foreground lineages, the test may be misled to incorrectly reject the null neutral model.

In this paper, we describe a simple modification to a branch-site model proposed by Yang and Nielsen (2002) and use it to construct two new LRTs, referred to as test 1 and test 2. We use computer simulation to evaluate the performance of the tests, examining the false-positive rate (type I error), the power, as well as the robustness. The results suggest that the branch-site test of positive selection has low false-positive rate and also more power than the branch-based test of Yang (1998). We also apply the new tests to the four data sets analyzed by Yang and Nielsen (2002). Our reanalysis led to the same conclusions as in the previous analysis.

Methods
LRTs and Identification of Positively Selected Sites

We refer to the reviews of Yang and Bielawski (2000) and Yang (2002) as well as the original papers for details of the likelihood method for detecting positive selection. The modified branch-site model, referred to as model A, is summarized in table 1. It is assumed that the branches on the phylogeny are divided *a priori* into foreground and background lineages. Only foreground lineages may have experienced positive selection. The model assumes four classes of sites. Site class 0 includes codons that are conserved throughout the tree, with $0 < \omega_0 < 1$ estimated. Site class 1 includes codons that are evolving neutrally throughout the tree with $\omega_1 = 1$. Site classes 2a and 2b include codons that

Table 1
Modified Branch-Site Model A

Site Class	Proportion	Background	Foreground
0	p_0	$0 < \omega_0 < 1$	$0 < \omega_0 < 1$
1	p_1	$\omega_1 = 1$	$\omega_1 = 1$
2a	$(1 - p_0 - p_1)\, p_0/(p_0 + p_1)$	$0 < \omega_0 < 1$	$\omega_2 > 1$
2b	$(1 - p_0 - p_1)\, p_1/(p_0 + p_1)$	$\omega_1 = 1$	$\omega_2 > 1$

NOTE.—This branch-site model is the alternative hypothesis for two LRTs. The null model for test 1 removes site classes 2a and 2b and is the site model M1a, assuming two site classes with $0 < \omega_0 < 1$ and $\omega_0 = 1$ for both background and foreground branches. The null model for test 2 is the branch-site model A with $\omega_2 = 1$ fixed. Test 2 is also referred to as the branch-site test of positive selection.
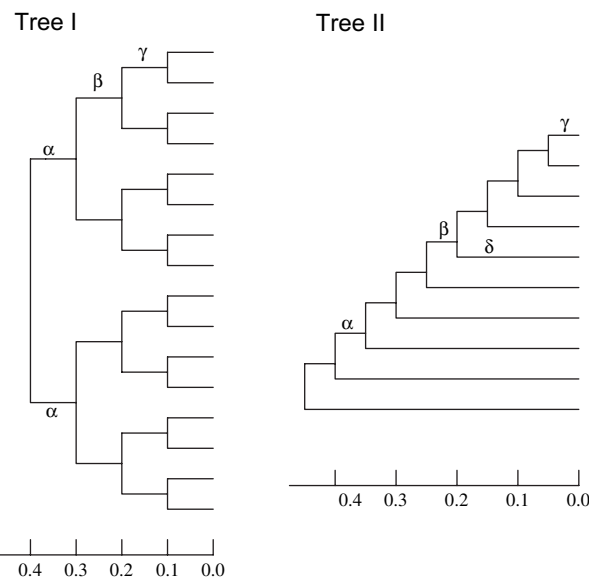

FIG. 1.—Two unrooted model trees used in computer simulation. Branch lengths are drawn in scale, in terms of the number of synonymous substitutions per synonymous site. Greek letters indicate individual foreground branches used in the simulation.

are conserved or neutral on the background branches, but become under positive selection on the foreground branches with $\omega_2 > 1$, estimated from the data. The model involves four parameters in the ω distribution: p_0, p_1, ω_0, and ω_2. This is a slight modification of the old branch-site model A of Yang and Nielsen (2002), which has $\omega_0 = 0$ fixed. The old model is very unrealistic as it does not allow conserved sites with $0 < \omega < 1$ and is thus replaced by the model of table 1. The old branch-site model B of Yang and Nielsen (2002) has both ω_0 and ω_1 estimated from the data. We found it advantageous to fix $\omega_1 = 1$ so that sites under weak constraint (with ω close to but less than 1) are lumped into this class of neutral sites rather than being falsely claimed to be under positive selection.

We use branch-site model A (table 1) to construct two LRTs. Model A is the alternative hypothesis in both tests. The null hypothesis for test 1 is the site model M1a (Yang et al. 2000; Yang, Wong, and Nielsen 2005), which assumes two site classes with $0 < \omega_0 < 1$ and $\omega_0 = 1$ for all branches. Significance of the test can be caused either by relaxed selective constraint on the foreground branch (e.g., presence of sites with $\omega < 1$ on the background branches and $\omega = 1$ on the foreground branches) or by positive selection along the foreground branch. Because the alternative hypothesis has two more parameters (p_2 and ω_2) than the null hypothesis, we use χ_2^2 to perform the test. However, we note that the regularity conditions for the χ^2 approximation are not satisfied, and the correct asymptotic distribution is unknown; that is, $p_2 = 0$ is at the boundary of the parameter space under the alternative hypothesis, and when $p_2 = 0$, ω_2 is not identifiable. In test 2, the null hypothesis is the branch-site model A (table 1) but with $\omega_2 = 1$ fixed. This null model allows sites evolving under negative selection on the background lineages to be released from constraint and to evolve neutrally on the foreground lineages. The test is thus a direct test for positive selection on the foreground lineages, and we refer to it also as the branch-site test of positive selection. As the alternative model A has $\omega_2 \geq 1$ constrained and the test is one-sided, the asymptotic null distribution is a 50:50 mixture of point mass 0 and χ_1^2, with the critical values to be 2.71 and 5.41, at the 5% and 1% levels, respectively (Chernoff 1954; Self and Liang 1987). However, to guide against small sample sizes and against violations of model assumptions, we use χ_1^2 to perform the test.

If the test of positive selection rejects the null hypothesis, the alternative model (model A of table 1) can be used to identify the sites under positive selection along the fore-ground lineages. The empirical Bayes approach can be used to calculate the posterior probabilities that each site belongs to the site class of positive selection on the foreground lineages. The procedure used by Yang and Nielsen (2002), known as naïve empirical Bayes (NEB), uses maximum likelihood estimates (MLEs) of parameters (such as proportions of site classes and the ω ratios) but ignores their sampling errors. A Bayes empirical Bayes (BEB) procedure was described recently by Yang, Wong, and Nielsen (2005), which accommodates uncertainties in the MLEs properly.

Computer Simulation

Computer simulations were conducted as described in Zhang (2004). The same two unrooted trees were used (fig. 1). In both trees, the molecular clock (rate constancy) holds for the synonymous substitution rate. Five evolutionary schemes with regard to ω values for sites were used (table 2). Scheme X represented normal gene evolution, with varying degrees of purifying selection at different sites. Scheme Y represented a partial relaxation of functional constraints, with some sites having higher ω values than those in scheme X. Scheme Z represented a complete relaxation of functional constraints, with all sites having $\omega = 1$. Schemes X, Y, and Z were used in the simulation of Zhang (2004) and assume no sites under positive selection. Two new schemes, U and V, contain some positively selected sites. Scheme U differed from X in that some negatively selected sites in X became positively selected. Scheme V differed from X in a more complex manner, with some sites having higher ω and some having lower ω.

In the simulation, a DNA sequence of 200 codons was randomly generated for an interior node of a model tree, with equal frequencies of the 4 nucleotides. Random point mutations were then generated assuming a transition/transversion rate ratio of $\kappa = 4$. The expected number of

Table 2
The ω Values Used in Generating the DNA Sequences in Computer Simulation

Codon Sites	Selection Schemes				
	X	Y	Z	U	V
1–20 (domain 1)	1.00	1.00	1.00	1.00	1.00
21–40 (domain 2)	1.00	1.00	1.00	1.00	0.70
41–60 (domain 3)	0.80	1.00	1.00	4.00	4.00
61–80 (domain 4)	0.80	0.90	1.00	0.80	0.80
81–100 (domain 5)	0.50	1.00	1.00	2.00	2.00
101–120 (domain 6)	0.50	0.75	1.00	0.50	0.50
121–140 (domain 7)	0.20	1.00	1.00	0.20	0.30
141–160 (domain 8)	0.20	0.60	1.00	0.20	0.20
161–180 (domain 9)	0.00	1.00	1.00	0.00	0.10
181–200 (domain 10)	0.00	0.50	1.00	0.00	0.00
Average ω	0.50	0.88	1.00	0.97	0.96

accepted nonsynonymous substitutions and the expected number of accepted synonymous substitutions followed the evolutionary scheme used. Nonsense mutations were not allowed. One branch in the model tree was chosen as the foreground branch, and all other branches were background branches. The evolutionary scheme X was always used for the background branches, whereas X, Y, Z, U, or V was used for the foreground branches. After the generation of the extant sequences, the codeml program in the PAML package (Yang 1997) was used to perform the likelihood analysis and conduct the two tests, as described above. The correct tree topology and the correct identification of the foreground branches were assumed, although the branch lengths were estimated by maximum likelihood without assuming the clock. We use the 5% significance level for the LRTs.

In about 2%–3% of the simulated replicates, the alternative hypothesis had lower log likelihood than the null hypothesis, apparently because the program failed to find the global maximum likelihood. Such cases were discarded, and new data sets were simulated until 200 replicates with useful results were generated for each condition examined. For a few simulation schemes, we also used another strategy, conducting the same analysis twice by using different initial values for numerical optimization. The results were virtually identical between the two strategies.

Simulation Results
False-Positive Rate When the Null Model is Correct

We first conducted simulations under the null model to examine the reliability of the χ^2 approximation to the null distribution. Tree I of figure 1 is used to generate data, but the unit of evolution is changed so that 0.1 synonymous substitutions per synonymous site in the tree becomes 0.3 nucleotide substitutions per codon.

We first evaluate test 1. We simulated 1,000 data sets under site model M1a, with two site classes in proportions $p_0 = 0.7$ and $p_1 = 0.3$ and ω ratios $\omega_0 = 0.2$ and $\omega_1 = 1$. Other simulation conditions are as previously described. There are 200 codons in the sequence. The data are analyzed under the null model M1a (neutral) and under the alternative branch-site model A, with branch α used as the foreground branch. In 512 of the 1,000 replicate data sets, the test statistic was equal to 0. The estimated critical values

at the 10%, 5%, and 1% levels from the simulated empirical distribution are 2.33, 3.37, and 6.10, respectively. The critical values from the χ_2^2 distribution are 4.60, 6.63, and 9.21. Use of χ_2^2 is thus conservative, as expected.

Next, we examine test 2, the branch-site test of positive selection. We generate 1,000 data sets under the null model of the test. The 200 sites are drawn at random from four site classes in proportions $p_0 = 0.6$, $p_1 = 0.2$, $p_{2a} = 0.15$, and $p_{2b} = 0.05$, with ω ratios specified according to table 1 with $\omega_0 = 0.2$, $\omega_1 = 1$, and $\omega_2 = 1$. Branch α is the foreground branch. Each data set is analyzed under branch-site model A, either with $\omega_2 = 1$ fixed or with $\omega_2 \geq 1$ estimated. In 521 of the 1,000 data sets, the test statistic was equal to 0. Asymptotically, $2\Delta\ell$ should follow a 50:50 mixture of a point mass at 0 and the χ_1^2 distribution (Self and Liang 1987), so that the proportion of zeros should be about 50%. The estimated critical values at the 10%, 5%, and 1% levels from the empirical distribution are 1.75, 2.95, and 5.98, respectively. The critical values from the χ_1^2 distribution are 2.70, 3.84, and 5.99, while the critical values based on the 50:50 mixture of 0 and the χ_1^2 distribution are 1.64, 2.71, and 5.41. Thus, use of the χ_1^2 distribution makes the test too conservative, while use of the mixture makes the test slightly too liberal.

We suspect that the slight discrepancy between the simulated distribution and the mixture distribution is due to the small sample size because there are only 200 codons in the sequence. To confirm this, we conducted another set of simulations, increasing the sequence length to 3,000 codons. In 515 of the 1,000 data sets, the test statistic was equal to 0. Again, this is close to the expected proportion of 50%. The estimated critical values at the 10%, 5%, and 1% levels from the empirical distribution are 1.37, 2.57, and 4.62, respectively, which are close to but smaller than 1.64, 2.71, and 5.41, the expected values from the mixture distribution. Thus, for this sequence length, use of both the mixture distribution and the χ_1^2 distribution is conservative.

The simulations under the null models of the tests confirm that the asymptotic distribution for test 2 (the branch-site test of positive selection) is the 50:50 mixture of a point mass at 0 and the χ_1^2 distribution, and use of the χ_1^2 distribution makes the test conservative. Similarly, use of the χ_2^2 distribution makes test 1 conservative. In the rest of the paper, we use the χ_2^2 and χ_1^2 distributions for the two tests to guide against violations of model assumptions and small sample sizes.

False-Positive Rate of the LRTs Under Models of Relaxed Constraints

We examined the false-positive rate of the two LRTs when the data were simulated under models of relaxed selective constraint using the schemes of table 2. The null hypotheses of the tests were thus violated even though the data were simulated without positive selection. Under such conditions, the old branch-site tests of Yang and Nielsen (2002) were found to have high false positives (Zhang 2004). The results for the two new tests are shown in table 3. The first set of simulations was conducted using model tree I (fig. 1). The deepest branch in the tree, α, was the foreground branch, and all other branches were set as background

Table 3
Numbers and Frequencies of Cases in Which Positive Selection for Foreground Branches Is Erroneously Identified by the Branch-Site Likelihood Method (Type I Error)

Foreground Branches	Simulation with bg = X[a], fg = X		Simulation with bg = X, fg = Y		Simulation with bg = X, fg = Z	
	Test 1	Test 2	Test 1	Test 2	Test 1	Test 2
Tree I-α	5 (0.025)	3 (0.015)	198 (0.990)	6 (0.030)	200 (1.000)	5 (0.025)
Tree I-β	3 (0.015)	4 (0.020)	179 (0.895)	2 (0.010)	194 (0.970)	9 (0.045)
Tree I-γ	2 (0.010)	0 (0.000)	183 (0.915)	5 (0.025)	198 (0.990)	13 (0.065)
Tree II-α	3 (0.015)	2 (0.010)	81 (0.405)	4 (0.020)	134 (0.670)	2 (0.010)
Tree II-β	5 (0.025)	5 (0.025)	122 (0.610)	8 (0.040)	168 (0.840)	8 (0.040)
Tree II-γ	4 (0.020)	4 (0.020)	138 (0.690)	11 (0.055)	173 (0.865)	15 (0.075)
Tree II-δ	6 (0.030)	4 (0.020)	199 (0.995)	7 (0.035)	200 (1.000)	10 (0.050)

NOTE.—The number of simulated replicates is 200. See legend to table 1 for descriptions of branch-site test 1 and test 2; bg, background; fg, foreground.

[a] See Table 1 for the evolutionary schemes X, Y, and Z.

branches. Evolutionary scheme X was used for both background and foreground branches; that is, there was no positive selection and no change in ω among branches (table 2). When test 1 was used with the null distribution χ_2^2, in 5 of the 200 replicates, the null hypothesis of no positive selection was rejected at the 5% level (table 3). At this level of significance, we should expect about 5% incorrect rejections (false positives) of the null hypothesis if the test is unbiased. Thus, test 1 is conservative. Test 2 is also conservative, as the null hypothesis was rejected in 3 of the 200 replicates. Similarly, when branches β or γ of tree I were designated the foreground branch or when tree II (fig. 1) was used with branches α, β, γ, or δ used as the foreground branch, both tests are conservative, with false-positive rates below the nominal 5% (table 3).

When schemes X and Y were used for the background and foreground branches, respectively, test 1 rejected the null hypothesis in between ~40% and 100% of the replicates. If we consider the test as a test of positive selection, these will be false-positive rates, which are unacceptably high. Alternatively, if we consider the test as a test of relaxed constraint on the foreground lineage, the positives are true positives, and the test had good power. The false-positive rate of test 2 (the branch-site test of positive selection) is at or below the nominal 5%. The results were very similar when scheme Z was applied to the foreground branch (table 3). Thus, this set of simulations showed that the type I error rate of test 2 is at or below the nominal level. However, test 1 should not be considered a reliable test of positive selection.

The above simulations examined the robustness of the LRTs to more complex variations of the ω ratios among sites than assumed in the models. We also conducted two simulations to examine the robustness of the LRTs to more complex variations of the ω ratio across lineages on the phylogeny. In particular, we examined whether the tests will be misled to produce high false positives when the foreground branch, which is not under positive selection, is surrounded on the tree by background branches that are under positive selection. We use branch β on tree II as the foreground branch (with length 0.05). All other branches are background branches, evolving under scheme X, except the ancestral branch and the interior descendant branch of branch β, which follow scheme V, with sites un-

der positive selection (see fig. 2). Branch β also follows scheme X. In other words, there are no sites under positive selection along branch β, but there are some on two background branches. At the 5% significance level, positive selection was falsely detected for branch β in 1, 5, and 6 cases by the branch test, branch-site test 1, and branch-site test 2, at the proportions 0.5%, 2.5%, and 3%, respectively. The second simulation was conducted in the same way except that all three background branches surrounding branch β (one branch ancestral and two branches descendent to branch β) evolve under scheme V, while branch β again follows scheme X. The false-positive rate at the 5% significance level was 0%, 2.5%, and 1.5% for the branch test, branch-site test 1, and branch-site test 2, respectively. These results suggest that all three tests were robust and not misled by positive selection on branches close to the foreground branch.

Power of LRTs in Detecting Positive Selection

We examined how often the two LRTs detected positive selection if it indeed occurred on the foreground branches (table 4). We used scheme X for background branches and used either U or V for foreground branches, which include positively selected sites. For comparison, we also used the branch test of Yang (1998), which ignores variation in ω among sites and detects positive selection only if the ω ratio for the whole sequence is significantly greater than 1 on the foreground lineages. In tree I, the two branch-site tests were significant in 7.5%–17% of the cases (table 4). Although these rates of detection are low, we note that the branch test detected positive selection in only 0%–2% of the cases, presumably because the average ω ratio across all sites is <1 for the foreground branch under both schemes U and V (table 2). Thus, the new branch-site tests are considerably more powerful than the branch test. It is also interesting to note that test 2 (the branch-site test of positive selection) had about the same power as test 1. Similar results were obtained in data simulated using tree II (table 4).

One may expect it to be easier to detect positive selection that occurred in recent lineages than in ancient lineages because the method can infer (probabilistically) recent substitutions more reliably than ancient substitutions. This
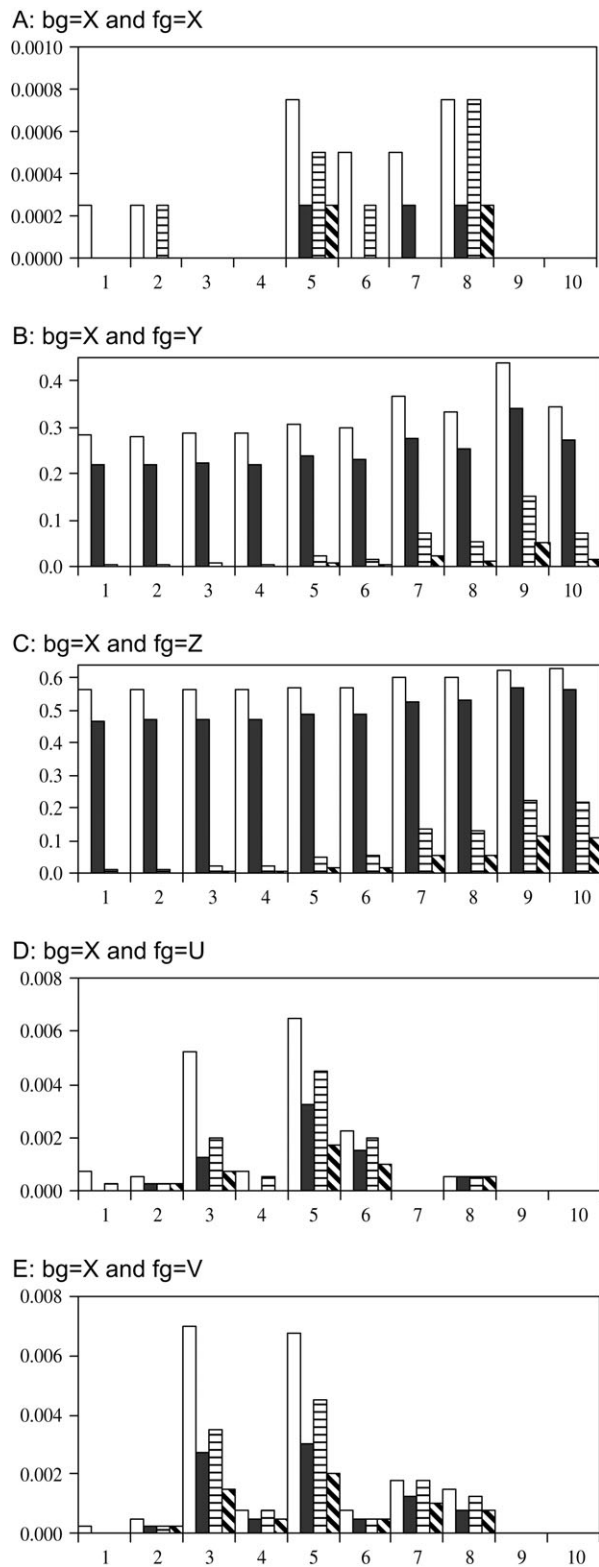
A: bg=X and fg=X

B: bg=X and fg=Y

C: bg=X and fg=Z

D: bg=X and fg=U

E: bg=X and fg=V

FIG. 2.—Frequency (*f*) of sites in different domains being identified as positively selected in the foreground branch by the NEB and BEB methods. Data were simulated using tree I of figure 1, with branch α being the foreground branch. Scheme X was used in background branches, whereas schemes (*A*) X, (*B*) Y, (*C*) Z, (*D*) U, and (*E*) V were used in the foreground branches, respectively. Note the difference in the scale of the *f* axis in different panels. The open, solid, horizontally hatched, and diagonally hatched bars represent *f* values calculated under the cutoffs of 90% NEB, 95% NEB, 90% BEB, and 95% BEB posterior probabilities,

expectation was supported overall although the difference was small and sometimes in the wrong direction. For example, in tree II, the power for both tests was higher for the recent branch γ than for the deep branches α or β, but in tree I, the power was very similar for branches β and γ (table 4). The length of the foreground branch had greater influence on the power. For example, the power was higher for the long branch α than for the short branches β or γ in tree I, while in tree II, the power was much higher for the long branch δ than for short branches α, β, or γ. To examine the effect of the foreground branch length, we increased the length of branch β in tree II from 0.05 to 0.15 or 0.45. The power for both tests increased by threefold to sevenfold. We expect that the power will eventually decrease when the foreground branch becomes too long. Similarly, when several branches are designated foreground branches and experience positive selection on the same sites, the power of the tests is expected to improve. This was indeed the case: when we applied the same selection scheme to both branches β and γ in tree II, the power was much higher than if either β or γ alone was the foreground branch (table 4). Nevertheless, we stress that this analysis assumes that all foreground branches are under the same selection scheme, with the same sites under positive selection, an assumption that may not be realistic for many data sets.

Obviously, the power of the tests should improve when more sites are under positive selection on the foreground branches or when the positive selection is stronger, with higher ω ratios. We conducted two additional simulations to demonstrate this effect. In the first set of simulations, we used branch β in tree II as the foreground branch (with length 0.15), on which evolution follows scheme V, while all other branches are background branches, evolving under scheme X. However, we simulated 600 codons in the sequence, with 60 (instead of 20) codons in each domain of table 2. At the 5% significance level, positive selection was detected for branch β in 0, 71, and 39 cases (with proportions 0, 0.355, and 0.195, respectively) for the branch test, branch-site test 1, and branch-site test 2, respectively. The corresponding proportions when there were 200 codons in the sequence are much lower, at 0.005, 0.135, and 0.110 (table 4). In the second simulation, we again used branch β in tree II as the foreground branch (with length 0.15 and scheme V), while the background branches evolve under scheme X. There are 200 codons in the sequence. However, we increased the ω ratios for domains 3 and 5 under scheme V (table 2) from 4 and 2 to 8 and 4, to model stronger positive selection on those domains. At the 5% significance level, positive selection was detected for branch β in 4, 98, and 100 cases (with proportions 0.02, 0.490, and 0.500, respectively) for the branch test, branch-site test 1, and branch-site test 2, respectively. These proportions are much

←

respectively. Note that only domains 3 and 5 in (*D*) and (*E*) are under positive selection, so detected sites in these domains are true positives. Other domains in (*D*) and (*E*) and all domains in (*A*), (*B*), and (*C*) are under neutral evolution or purifying selection, so detected sites in these domains are false positives.

Table 4
Numbers and Frequencies of Cases in Which Positive Selection for Foreground Branches Is Correctly Inferred by the Branch Method or Tests 1 and 2 of the Branch-Site Method

Foreground Branch[a]	Simulation with bg = X[b], fg = U			Simulation with bg = X, fg = V		
	Branch Method	Test 1	Test 2	Branch Method	Test 1	Test 2
Tree I-α (0.2)	2 (0.010)	27 (0.135)	29 (0.145)	0 (0.000)	34 (0.170)	23 (0.115)
Tree I-β (0.1)	0 (0.000)	23 (0.115)	22 (0.110)	0 (0.000)	24 (0.120)	15 (0.075)
Tree I-γ (0.1)	4 (0.020)	17 (0.085)	19 (0.095)	1 (0.005)	23 (0.115)	15 (0.075)
Tree II-α (0.05)	0 (0.000)	7 (0.035)	5 (0.025)	1 (0.005)	13 (0.065)	11 (0.055)
Tree II-β (0.05)	2 (0.010)	11 (0.055)	8 (0.040)	0 (0.000)	8 (0.040)	7 (0.035)
Tree II-γ (0.05)	4 (0.020)	21 (0.105)	19 (0.095)	1 (0.005)	32 (0.160)	24 (0.120)
Tree II-δ (0.2)	0 (0.000)	40 (0.200)	44 (0.220)	0 (0.000)	62 (0.310)	60 (0.300)
Tree II-β (0.15)[c]	1 (0.005)	24 (0.120)	26 (0.130)	1 (0.005)	27 (0.135)	22 (0.110)
Tree II-β (0.45)[c]	0 (0.000)	29 (0.145)	36 (0.180)	0 (0.000)	54 (0.270)	38 (0.190)
Tree II-δ and γ[d]	0 (0.000)	57 (0.285)	66 (0.330)	0 (0.000)	80 (0.400)	87 (0.435)

NOTE.—See legend to table 1 for descriptions of branch-site test 1 and test 2.
[a] The length of the foreground branch is shown in parentheses (see fig. 1).
[b] See Table 1 for the evolutionary schemes X, U, and V.
[c] Branch length of β is increased from 0.05 to 0.15 or 0.45.
[d] Both δ and γ are under positive selection and are specified as foreground branches.

higher than 0.005, 0.135, and 0.110 (table 4) under the weaker positive selection of scheme V in table 2.

Performance of NEB and BEB in Identifying Positively Selected Sites

When the LRT of positive selection on the foreground branches is significant, the NEB (Yang and Nielsen 2002) and BEB (Yang, Wong, and Nielsen 2005) procedures can be used to compute the posterior probability that a codon belongs to the class of positive selection. A codon with high posterior probability is likely to be under positive selection on the foreground lineages. Here, we examine the performance of both the NEB and the BEB approaches. In our simulations, the entire sequence was divided into 10 domains, each with 20 codons. Sites within a domain have the same ω. We tabulated the average frequency (f) at which a codon in a given domain is identified to be under positive selection on the foreground lineages by either NEB or BEB at either the 90% or 95% cutoffs. We refer to such frequencies as f_{N90}, f_{N95}, f_{B90}, and f_{B95}, corresponding to the method and cutoff. For domains that do not have any sites under positive selection, these frequencies will be the false-positive rates. Note that here we are evaluating the frequentist properties of the NEB and BEB methods: if the methods are good, they should satisfy $f_{N90}, f_{B90} \leq 0.1$ and $f_{N95}, f_{B95} \leq 0.05$ when there is no positive selection. However, we point out that NEB and BEB compute Bayesian posterior probabilities and do not explicitly control the type I error (or false positive) rate. For a discussion of the Bayesian and frequentist measures of performance in detecting positively selected sites, see Yang, Wong, and Nielsen (2005), who also demonstrated that the BEB procedure had excellent frequentist properties in site-based analysis.

Figure 2 shows the results obtained for tree I, with α being the foreground branch. Results obtained using other foreground branches or using tree II are similar and thus not presented. The NEB and BEB approaches were applied without first conducting the LRTs. In all cases, the posterior probability was estimated under branch-site model A (table 1). Note that only domains 3 and 5 in figure 2D and E are

under positive selection, while all other domains contain no sites under positive selection. When scheme X was used for both background and foreground branches, the false-positive rate f for both NEB and BEB was extremely low (<0.001) for all domains and for both cutoffs (fig. 2A). When we changed the evolutionary scheme to Y for the foreground branch, f_{N90} and f_{N95} are higher than 0.2 for all domains (fig. 2B), and the false-positive rate of NEB was excessively high. In contrast, for the BEB method we observed $f_{B90} < 0.1$ and $f_{B95} < 0.05$ in most domains (fig. 2B). The averages over the whole sequence were $f_{B90} = 0.04$ and $f_{B95} < 0.01$, suggesting that BEB had the false-positive rate below the nominal 10% or 5% levels. When scheme Z was used for the foreground branch, NEB again had very high false-positive rates (fig. 2C). BEB performed much better, even though the false-positive rate for domains 9 and 10 was $f_{B90} \approx 0.2$ and $f_{B95} \approx 0.1$, respectively. Averaged over the whole sequence, the false-positive rates were $f_{B90} = 0.09$ and $f_{B95} = 0.04$, again below the nominal 10% or 5%. We next examined f when scheme U or V was used for the foreground branch. Under these schemes, positive selection occurs in domains 3 and 5 (table 2). We note that the power of detecting positively selected codons by either NEB or BEB is very low, with $f < 1\%$. The false-positive rates for both NEB and BEB were also very low for domains other than 3 and 5.

Real Data Analysis

We apply the two branch-site tests to the four data sets analyzed by Yang and Nielsen (2002), to examine whether the new tests lead to different conclusions. The results are summarized in table 5. We provide a brief summary of the data sets here. For more details and for the phylogenetic trees used, see Yang and Nielsen (2002) and the original papers.

The first data set includes 19 primate lysozyme c gene sequences, originally analyzed by Messier and Stewart (1997). Positive selection is suspected along the branch ancestral to the clade of colobine monkeys, which is the foreground branch in our tests. Those leaf-eating primates

Table 5
LRT Statistics ($2\Delta\ell$) and *P* Values for Different Branch-Site Tests of Positive Selection

Data Set	Number of Sequences	Number of Codons	Tests of this Paper		Old Tests of Yang and Nielsen (2002)	
			Test 1	Test 2	M1–MA	M3–MB
Primate lysozyme *c* gene	19	130	3.36 ($P = 0.19$)	1.48 ($P = 0.22$)	3.36 ($P = 0.19$)	1.92 ($P = 0.38$)
Primate cancer gene BRCA1	8	1,160	8.10 ($P = 0.017$)	3.64 ($P = 0.056$)	8.60 ($P = 0.014$)	5.54 ($P < 0.063$)
Angiosperm phytochrome gene *phyA-C/F*	15	1,105	84.88 ($P < 0.001$)	19.88 ($P < 0.001$)	352.8 ($P < 0.001$)	27.72 ($P < 0.001$)
Angiosperm phytochrome gene *phyB-D/E*	11	1,105	57.90 ($P < 0.001$)	13.29 ($P < 0.001$)	325.1 ($P < 0.001$)	109.8 ($P < 0.001$)

NOTE.—Branch-site tests 1 and 2 are described in the legend to table 1. The two old tests are described in the text and in Yang and Nielsen (2002).

have foreguts where lysozyme is found and may have acquired a new digestive function (Stewart, Schilling, and Wilson 1987). The second data set is from Huttley et al. (2000), who analyzed sequences from exon 11 of the tumor-suppression gene BRCA1. The human and chimpanzee lineages are suspected to be under positive selection and are the foreground branches in our tests. The third and fourth data sets consist of sequences from the angiosperm phytochrome gene *phy* (Alba et al. 2000). Phytochromes are important proteins that regulate many events in plant development in response to light. The 15 sequences from the A and C/F subfamilies and the 11 sequences from the B/D and E subfamilies are analyzed as separate data sets. We test whether positive selection has played a role in driving functional divergence after gene duplication in this gene family. The foreground branch in our analysis is the branch separating the A and the C/F subfamilies in data set 3 and the branch separating the B/D and E subfamilies in data set 4.

Table 5 shows the results of the two LRTs discussed in this paper, in comparison with results obtained by Yang and Nielsen (2002) for the two old branch-site tests. Both old tests are similar to branch-site test 1. The M1-MA test compares the site model M1 (neutral) and the branch-site model A, with both models fixing $\omega_0 = 0$ for the conserved sites. The M3-MB test compares the site model M3 (discrete, with $K = 2$ classes) and the branch-site model B, with both models estimating ω_0 and ω_1 as free parameters. The old and new tests reached the same conclusions in these four data sets. There is no statistically significant support for positive selection in the primate lysozyme data set, marginally significant support in the primate BRCA1 data set, and overwhelming evidence for positive selection in the two angiosperm phytochrome data sets.

We also applied the BEB procedure to calculate the posterior probabilities that each site is under positive selection in the four data sets. In the primate lysozyme and primate BRCA1 data sets, no site had a posterior probability higher than 95%. In the phytochrome *phy* A-C/F data set, 27 sites were identified to be potentially under positive selection along the foreground lineage at the 95% cutoff. They are 55R, 102T, 105S, 117P, 130T, 147S, 171T, 216E, 227F, 252I, 304I, 305D, 321L, 440L, 517A, 552T, 560Y, 650T, 655S, 700A, 736K, 787N, 802V, 940T, 986M, 988Q, and 1087D (the amino acids refer to the *Zea mays phyA* sequence). In the phytochrome *phy* B/D-E data set, 19 sites were identified to be potentially under positive selection along the foreground branch at the 95% level. They are 2S, 216K, 287A, 400H, 446Y, 501Y, 535Q, 640E, 782I, 810N, 851A, 858A, 859S,

860A, 865A, 902S, 945S, 1032S, and 1046Y (the amino acids refer to the *Sorghum bicolour phyB* sequence).

Given that the old tests had high false positives while the new test did not under some simulation conditions, we expect that the different tests may produce different results in some data sets. It may be worthwhile to use branch-site test 2 to confirm the results if positive selection is detected using the old tests, but the sequences are short and the positive selection is weak (with small estimates of p_2 and ω_2).

Discussion

In this paper, we used the modified branch-site model A (table 1) to construct two LRTs to detect positive selection affecting a small subset of sites along prespecified branches. Our computer simulation suggests that test 2 can accurately distinguish between relaxed constraint and positive selection and often has higher power in detecting positive selection than test 1. We thus recommend the use of test 2 in real data analysis. We note that this test was rather successful in our simulations and was in particular able to avoid high false positives in the presence of relaxed selective constraint, in contrast to the previous branch-site tests evaluated by Zhang (2004). Test 1, however, is not a reliable test of positive selection as it fails to distinguish between positive selection and relaxed constraint on the foreground branches.

For identifying codons under positive selection in the foreground branches, our simulation suggests that NEB can have high false-positive rates and is unusable. NEB was found to be unreliable in small data sets in site-based analysis as well (Yang, Wong, and Nielsen 2005). In contrast, BEB appeared to be reliable, with the false-positive rate below 10% or 5% when sites were identified at the 90% or 95% cutoff. However, the power of the analysis can be very low. In our simulations, it is common for the branch-site test of positive selection to provide significant support for the presence of positively selected sites on the foreground lineages, whereas no sites show a posterior probability higher than 95% according to BEB. We note that identifying sites under positive selection is intrinsically more difficult than testing whether such sites exist. When multiple sites are under positive selection with $\omega > 1$, the LRT combining evidence from all such sites may be able to infer their presence, but the information at any single site may not be strong enough for the BEB probability to reach high levels (especially as positive selection has affected only one lineage or a very few lineages on the tree). We suggest that it is still useful to be able to detect positive selection acting on the sequence even if the affected sites cannot be reliably

inferred. It was also noted that a site could be identified by BEB as being under positive selection with high posterior probability, but the branch-site test of positive selection applied to the same data failed to provide significant support for positive selection. We suggest that in such a case, the results should not be interpreted as providing evidence for positive selection.

The branch-site model (table 1) makes a number of assumptions about the evolutionary process. It is not entirely clear which of the assumptions are the most worrisome and have the greatest impact on the reliability of the tests. We considered the restrictive assumptions about the selective pressures to be important and thus designed the simulation experiment to evaluate the robustness of the analysis to violations of such assumptions. For example, the null and alternative models used in the tests assume no more than three site classes, while many more site classes were used in the simulation. Branch-site model A assumes one site class for positive selection, while schemes U and V have two classes of positively selected sites, with different ωs. Similarly, the tests allow only two kinds of branches on the tree, the foreground branches that are likely to be under positive selection and the background branches that are under constraint. We thus evaluated a few cases where some background branches are under positive selection. The simulation results suggest that the tests are robust to these violations of assumptions. Nevertheless, we caution that our simulation is limited in scope and it remains possible that the tests may fail under conditions not explored here. Identification of such conditions will be most useful in helping us improve the test.

Acknowledgments

Literature Cited

Alba, R., P. M. Kelmenson, M.-M. Cordonnier-Pratt, and L. H. Pratt. 2000. The phytochrome gene family in tomato and the rapid differential evolution of this family in angiosperms. Mol. Biol. Evol. 17:362–373.

Chernoff, H. 1954. On the distribution of the likelihood ratio. Ann. Math. Stat. 25:573–578.

Huttley, G. A., S. Easteal, M. C. Southey, A. Tesoriero, G. G. Giles, M. R. E. McCredie, J. L. Hopper, and D. J. Venter. 2000. Adaptive evolution of the tumour suppressor BRCA1 in humans and chimpanzees. Nat. Genet. 25: 410–413.

Messier, W. and C.-B. Stewart. 1997. Episodic adaptive evolution of primate lysozymes. Nature 385:151–154.

Nei, M. and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford University Press, Oxford.

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

Self, S. G. and K.-Y. Liang. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J. Am. Stat. Assoc. 82:605–610.

Stewart, C.-B., J. W. Schilling, and A. C. Wilson. 1987. Adaptive evolution in the stomach lysozymes of foregut fermenters. Nature 330:401–404.

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

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

Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. 15:568–573.

Yang, Z. 2002. Inference of selection from multiple species alignments. Curr. Opin. Genet. Dev. 12:688–694.

Yang, Z. and J. P. Bielawski. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. 15:496–503.

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

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

Yang, Z., W. S. W. Wong, and R. Nielsen. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22:1107–1118.

Yu, M. and D. M. Irwin. 1996. Evolution of stomach lysozyme: the pig lysozyme gene. Mol. Phylogenet. Evol. 5: 298–308.

Zhang, J. 2004. Frequent false detection of positive selection by the likelihood method with branch-site models. Mol. Biol. Evol. 21:1332–1339.

Zhang, J., S. Kumar, and M. Nei. 1997. Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. Mol. Biol. Evol. 14:1335–1338.

Zhang, J., H. F. Rosenberg, and M. Nei. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA 95:3708–3713.