

Frequent False Detection of Positive Selection by the Likelihood Method with Branch-Site Models

Jianzhi Zhang

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor

Positive Darwinian selection promotes fixations of advantageous mutations during gene evolution and is probably responsible for most adaptations. Detecting positive selection at the DNA sequence level is of substantial interest because such information provides significant insights into possible functional alterations during gene evolution as well as important nucleotide substitutions involved in adaptation. Efficient detection of positive selection, however, has been difficult because selection often operates on only a few sites in a short period of evolutionary time. A likelihood-based method with branch-site models was recently introduced to overcome such difficulties. Here I examine the accuracy of the method using computer simulation. I find that the method detects positive selection in 20%–70% of cases when the DNA sequences are generated by computer simulation under no positive selection. Although the frequency of such false detection varies depending on, among other things, the tree topology, branch length, and selection scheme, the branch-site likelihood method generally gives misleading results. Thus, detection of positive selection by this method alone is unreliable. This unreliability may have resulted from its over-sensitivity to violations of assumptions made in the method, such as certain distributions of selective strength among sites and equal transition/transversion ratios for synonymous and nonsynonymous substitutions.

Introduction

Detection of positive Darwinian selection at the DNA sequence level is of substantial interest to evolutionary biologists, as such information helps to reveal the molecular basis of adaptation. Comparison of synonymous and nonsynonymous substitution rates has been widely used for this purpose, because synonymous substitutions are largely neutral but nonsynonymous substitutions may be subject to positive selection (Nei and Kumar 2000). A significantly higher number of nonsynonymous nucleotide substitutions per nonsynonymous site (d_N) than that of synonymous substitutions per synonymous site (d_S) is usually taken as strong evidence for the action of positive selection (Hughes and Nei 1988; Hughes 1999; Nei and Kumar 2000). However, a simple comparison of average d_N and d_S for entire gene sequences across a long evolutionary time is likely to have low power in detecting positive selection because positive selection often acts on only a small number of codons in a short period of evolutionary time. Several methods have been developed either to test selection on an individual branch of a phylogenetic tree for the entire gene sequences (Yu and Irwin 1996; Messier and Stewart 1997; Zhang, Kumar, and Nei 1997; Zhang, Rosenberg, and Nei 1998; Yang 1998) or to test it on individual codon sites over an entire phylogenetic tree (Nielsen and Yang 1998; Suzuki and Gojobori 1999; Yang et al. 2000). These methods are likely to be more powerful than the previous ones. Recently, Yang and Nielsen (2002) introduced a new method that tests positive selection on individual codons for specific branches. This method, referred to as the branch-site likelihood method, is believed to be even more powerful. Indeed, reports of positive selection have been made by numerous authors using this new method (Knipple et al. 2002; Yang et al. 2002; Bailly et al.

2003; Barkman 2003; Kinsella et al. 2003; Martinez-Castilla and Alvarez-Buylla 2003; Mathews, Burleigh, and Donoghue 2003; Rodriguez-Trelles, Tarrío, and Ayala 2003). Notably, in several of these cases, positive selection was detected on ancient branches, such as in the early stage of eukaryotic evolution before the divergence of plants and animals (Rodriguez-Trelles, Tarrío, and Ayala 2003) or during land plant evolution before the diversification of angiosperms (Mathews, Burleigh, and Donoghue 2003). Because synonymous substitutions that occurred in the distant past are difficult to infer because of saturation, these results suggest that either the new method is exceptionally powerful or it makes erroneous inferences. Here I use computer simulation to examine how reliable the method is. The results show that it detects positive selection in 20%–70% of cases even when positive selection is absent, calling for reevaluation of the past claims of positive selection that are primarily based on results from the branch-site likelihood method.

Methods

The Likelihood Method with the Branch-Site Models

The branch-site model of the likelihood method for detecting positive selection at individual sites along a specific lineage (Yang and Nielsen 2002) was developed on the basis of earlier likelihood methods (Nielsen and Yang 1998; Yang 1998; Yang et al. 2000). In the new method, a set of aligned coding sequences is analyzed based on a predetermined phylogenetic tree of the sequences. Tree branches are separated into two groups. Those branches tested for positive selection are referred to as “foreground” branches, whereas all other branches are referred to as “background” branches. This separation is designed to improve the detection of positive selection when it occurs only in the foreground branches. Two tests, A and B, are designed. In test A, the null model assumes that the evolutionary pattern (with regard to $\omega = d_N/d_S$) is the same for background and foreground branches and that there are two classes of sites with $\omega_0 = 0$ (class 0) and $\omega_1 = 1$ (class 1), with the fraction of each class being estimated

Key words: likelihood, computer simulation, positive selection, branch-site model, molecular evolution.

E-mail: jianzhi@umich.edu.

Mol. Biol. Evol. 21(7):1332–1339. 2004

doi:10.1093/molbev/msh117

Advance Access publication March 19, 2004

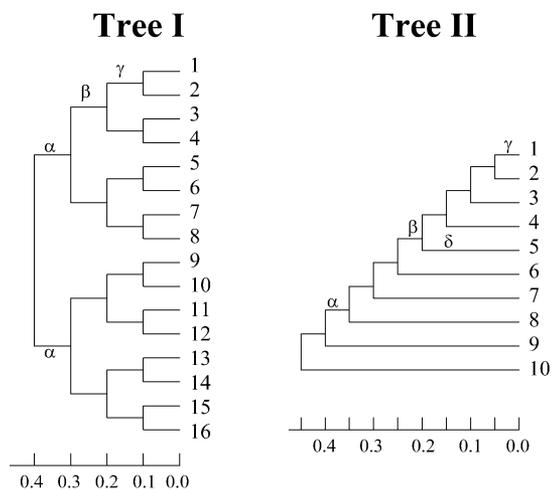


FIG. 1.—Model trees used for computer simulation. The trees are unrooted. Branch lengths are drawn in scale, in terms of the number of synonymous substitutions per site. Greek letters indicate individual foreground branches used in the simulation.

by likelihood. The alternative model allows an additional class of sites with ω_2 (class 2) for foreground branches, and the fraction of this class of sites is estimated by likelihood. Additionally, the alternative model assumes that the ratio of the number of class 0 sites to that of class 1 sites is the same for background and foreground branches. The likelihoods of the null and alternative models are then computed. Positive selection for foreground branches is inferred if the estimated $\omega_2 > 1$ and the alternative model is found to fit the data significantly better than the null model by a likelihood ratio test. Test B is the same as test A, except that now ω_0 and ω_1 are estimated from the data instead of being set a priori. In both tests A and B, when $\omega_2 > 1$, the posterior probability that a codon belongs to class 2 is computed for each codon. Those with high posterior probabilities (e.g., >95%) are regarded as being under positive selection with high statistical confidence.

Computer Simulation

Two unrooted model trees were used in the computer simulation conducted in this study (fig. 1). Both trees followed the molecular clock in terms of the number of synonymous substitutions per site. Model tree I had 16 taxa, with a symmetrical phylogenetic arrangement. The length of each branch was 0.1 synonymous substitutions per site, except for the deepest branch shown in the tree, which had a length of 0.2 synonymous substitutions per site. Model tree II had 10 taxa, arranged asymmetrically. The branch length for each interior branch was 0.05 synonymous substitutions per site. Simulation was conducted with coding sequences of 200 codons. Three evolutionary schemes with regard to ω values were used (table 1). 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

Table 1
The ω Values Used in Generating the DNA Sequences

Codon Sites	Selection Schemes		
	X	Y	Z
1–20	1.00	1.00	1.00
21–40	1.00	1.00	1.00
41–60	0.80	1.00	1.00
61–80	0.80	0.90	1.00
81–100	0.50	1.00	1.00
101–120	0.50	0.75	1.00
121–140	0.20	1.00	1.00
141–160	0.20	0.60	1.00
161–180	0.00	1.00	1.00
181–200	0.00	0.50	1.00

sites having $\omega = 1$. Note that there were no sites under positive selection in any of the three schemes. In each simulation, an ancestral DNA sequence of 200 codons with equal nucleotide frequencies was randomly generated for an interior node of a model tree. Random point mutations were then generated with the transition/transversion ratio of 2. All synonymous mutations were accepted, while a certain proportion of nonsynonymous mutations were accepted according to the evolutionary scheme used. Nonsense mutations were not allowed. Sequence evolution was thus simulated following a model tree. In the simulation, one branch was chosen as a foreground branch and all other branches were considered background branches. The evolutionary scheme X was always used for background branches, whereas either X, Y, or Z was used for foreground branches. For each condition examined, two hundred simulation replications were conducted. After the generation of the terminal sequences in the tree, the PAML software package (Yang 1997) was used to conduct the A and B tests, as described above, with the use of the known tree, alignment, and foreground branch. A likelihood ratio test was regarded as significant if the tail probability was lower than 5%.

Results

Tree I—Branch α

The first set of simulations was conducted using model tree I. The deepest branch in the tree, branch α , was chosen to be the foreground branch and all other branches were set as background branches. The evolutionary scheme X was used for both background and foreground branches (table 1). In other words, there is no change in ω values between branches. When test A was used, the correct null hypothesis of equal ω values between background and foreground branches was rejected in 45.5% of the 200 replications at the 5% significance level (table 2). At this level of significance, we should expect about 5% incorrect rejections (false positives) of the true hypothesis if the statistical test is unbiased. Apparently, test A was excessively liberal under this condition. Positive selection for the foreground branch was inferred for 1.5% of the replications (i.e., $\omega_2 > 1$ and the alternative model fits the data significantly better than the null model does). This error rate appears acceptable. When test B was

Table 2
Numbers and Frequencies of Cases in Which the Likelihood Method A Indicates Different ω 's for Foreground (fg) and Background (bg) Branches and Positive Selection for Foreground Branches

Foreground branches	Simulation with bg = X, fg = X ^a		Simulation with bg = X, fg = Y ^a		Simulation with bg = X, fg = Z ^a	
	bg ≠ fg ^b	bg ≠ fg and $\omega_2 > 1^c$	bg ≠ fg ^b	bg ≠ fg and $\omega_2 > 1^c$	bg ≠ fg ^b	bg ≠ fg and $\omega_2 > 1^c$
Tree I- α	91 (0.455)	3 (0.015)	200 (1.000)	64 (0.320)	200 (1.000)	107 (0.535)
Tree I- β	56 (0.280)	6 (0.030)	199 (0.995)	74 (0.370)	200 (1.000)	91 (0.455)
Tree I- γ	58 (0.280)	5 (0.025)	196 (0.980)	73 (0.365)	200 (1.000)	93 (0.465)
Tree II- α	64 (0.032)	1 (0.005)	165 (0.825)	38 (0.190)	179 (0.895)	41 (0.205)
Tree II- β	46 (0.230)	6 (0.030)	180 (0.900)	61 (0.305)	192 (0.960)	49 (0.245)
Tree II- γ	33 (0.165)	7 (0.035)	177 (0.885)	96 (0.480)	188 (0.940)	107 (0.535)
Tree II- δ	72 (0.360)	3 (0.015)	199 (0.995)	77 (0.385)	200 (1.000)	106 (0.530)

^a See table 1 for the evolutionary schemes X, Y, and Z.

^b The column shows the number of replications (with frequency in parentheses) where the null hypothesis of identical ω 's between bg and fg is rejected at 5% significance level by the likelihood method.

^c The column shows the number of replications (with frequency in parentheses) where positive selection in fg is inferred at the 5% significance level by the likelihood method.

used, the null hypothesis of equal ω values was rejected in six of the 200 cases (table 3), indicating that test B was quite good under this condition. Among these six cases, two of them showed $\omega_2 > 1$, and both of them had $\omega_1 < 1$ from the null model. Thus, the two cases were inferred by test B to have positive selection specifically for the foreground but not background branches.

Note that although the null hypothesis of equal ω 's between foreground and background branches was correct in this set of simulations, the evolutionary scheme X was different from the substitution models assumed by the likelihood method. For instance, only two to three classes of sites (with respect to ω) were assumed by the likelihood method, while in fact five different classes existed in the simulated data (table 1). The failure of test A suggests that it is sensitive to violation of the assumption made in the likelihood ratio test (Zhang 1999). Figure 2 shows the distributions of log likelihood ratio $LR = 2(\ln L_1 - \ln L_0)$ for the 200 replications of tests A and B, where L_0 and L_1 are likelihoods of the null and alternative models, respectively. When the tests are unbiased, it is expected that LR follows the χ^2 distribution with 2 degrees of freedom. One can see that for test A, the empirical

distribution of LR is quite flat, resulting in more frequent occurrences of large LR values than expected. For test B, the empirical distribution is slightly left-shifted, indicating that it is somewhat more conservative than expected. The likelihood method also estimates the proportion of sites that have altered ω values between background and foreground branches and denotes it as $p_2 + p_3$. Here p_2 and p_3 are proportions of sites that belong to class 0 and 1 in background branches, respectively, but belong to class 3 in foreground branches. Because the background and foreground branches have exactly the same ω 's here, the true value of $p_2 + p_3$ is 0. However, the distribution of $p_2 + p_3$ estimated by the likelihood test A was bimodal, with peaks at 0 and 0.5 (fig. 3A), and the mean estimate of $p_2 + p_3$ was equal to 0.381 ± 0.015 . For test B, the distribution of estimated $p_2 + p_3$ was L-shaped, the mean being 0.125 ± 0.012 . Therefore, both tests were found to overestimate $p_2 + p_3$.

The second set of simulations was conducted using scheme X for the background branches but Y for the foreground branch. Note that Y represents a partial relaxation of purifying selection, but does not have any positive selection (table 1). Because ω values are different

Table 3
Numbers and Frequencies of Cases in Which the Likelihood Method B Indicates Different ω 's for Foreground (fg) and Background (bg) Branches and Positive Selection for Foreground Branches

Foreground Branches	Simulation with bg = X, fg = X ^a			Simulation with bg = X, fg = Y ^a			Simulation with bg = X, fg = Z ^a		
	bg ≠ fg ^b	bg ≠ fg, $\omega_2 > 1^c$	bg ≠ fg, $\omega_2 > 1, \omega_1 < 1^d$	bg ≠ fg ^b	bg ≠ fg, $\omega_2 > 1^c$	bg ≠ fg, $\omega_2 > 1, \omega_1 < 1^d$	bg ≠ fg ^b	bg ≠ fg, $\omega_2 > 1^c$	bg ≠ fg, $\omega_2 > 1, \omega_1 < 1^d$
Tree I- α	6 (0.030)	2 (0.010)	2 (0.010)	200 (1.000)	100 (0.500)	97 (0.485)	200 (1.000)	137 (0.685)	132 (0.660)
Tree I- β	12 (0.060)	6 (0.030)	6 (0.030)	183 (0.915)	105 (0.525)	103 (0.515)	199 (0.995)	129 (0.645)	128 (0.640)
Tree I- γ	6 (0.030)	2 (0.010)	2 (0.010)	160 (0.800)	111 (0.555)	107 (0.535)	198 (0.990)	125 (0.625)	124 (0.620)
Tree II- α	6 (0.030)	1 (0.005)	1 (0.005)	113 (0.562)	66 (0.330)	66 (0.330)	148 (0.740)	104 (0.520)	103 (0.515)
Tree II- β	7 (0.035)	4 (0.020)	4 (0.020)	142 (0.710)	90 (0.450)	89 (0.445)	174 (0.870)	122 (0.610)	117 (0.585)
Tree II- γ	7 (0.035)	4 (0.020)	4 (0.020)	160 (0.800)	81 (0.405)	77 (0.385)	178 (0.890)	103 (0.515)	102 (0.510)
Tree II- δ	6 (0.030)	2 (0.010)	2 (0.010)	199 (0.995)	114 (0.570)	114 (0.570)	200 (1.000)	110 (0.550)	110 (0.550)

^a See table 1 for the evolutionary schemes X, Y, and Z.

^b The column shows the number of replications (with frequency in parentheses) where the null hypothesis of identical ω 's between bg and fg is rejected at the 5% significance level.

^c The column shows the number of replications (with frequency in parentheses) where positive selection in fg is inferred at the 5% significance level.

^d The column shows the number of replications (with frequency in parentheses) where positive selection in fg, but not bg, is inferred at the 5% significance level.

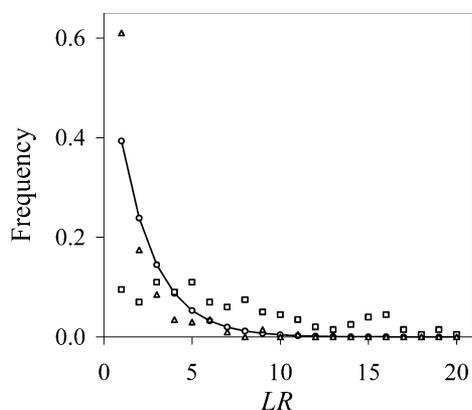


FIG. 2.—Distribution of $LR = 2(\ln L_1 - \ln L_0)$ from 200 replications of simulations. In the simulations, branch α of tree I is assigned as the foreground branch, and scheme X is used for both background and foreground branches. The circles and the line show the expected χ^2 distribution with the number of degrees of freedom as 2. Squares and triangles show empirical distributions derived from simulations of tests A and B, respectively.

between background and foreground branches in this set of simulations, it is not unexpected that both A and B tests rejected the null hypothesis of no difference in ω values in each of the 200 replications. However, tests A and B erroneously detected positive selection in 32% and 50% of the cases, respectively, for the foreground branch (tables 2 and 3). Interestingly, for all the cases in which test A detected positive selection, test B also detected positive selection. For test B, 48.5% of all cases were inferred to have positive selection specifically in the foreground ($\omega_2 > 1$) but not background branches ($\omega_1 < 1$ in the null model). These results show that both A and B tests frequently detect positive selection when there is some degree of reduction in purifying selection. In this set of simulations, the true value of $p_2 + p_3$ was 0.8. The distribution of estimated $p_2 + p_3$ was J-shaped for both tests A and B (fig. 3B), and the mean estimate of $p_2 + p_3$ was 0.820 ± 0.013 and 0.755 ± 0.015 for the two tests, respectively.

The number of codons (N_{PS}) that are inferred to be under positive selection with $>95\%$ posterior probabilities in the foreground branch was also examined. In both tests, N_{PS} was positively correlated with LR (correlation coefficient $r = 0.19$, $P = 0.005$ in test A; $r = 0.24$, $P = 0.0004$ in test B) and ω_2 ($r = 0.20$, $P = 0.004$ in test A; $r = 0.26$, $P = 0.0002$ in test B). N_{PS} was always 0 for cases where positive selection was not inferred. In contrast, it was sometimes 0 and sometimes higher than 0 when positive selection was inferred. The mean N_{PS} was 34.2 and 10.7 in tests A and B, respectively, for cases where positive selection was inferred. Interestingly, test A detected four cases in which all 200 codons were inferred to be under positive selection with $>95\%$ posterior probabilities, yet test B found only 0 to 21 positively selected codons in these same cases. Test B did not detect any cases in which all 200 codons were under positive selection. Because the evolutionary scheme Y included 40% of sites that are under purifying selection, the above results showed that

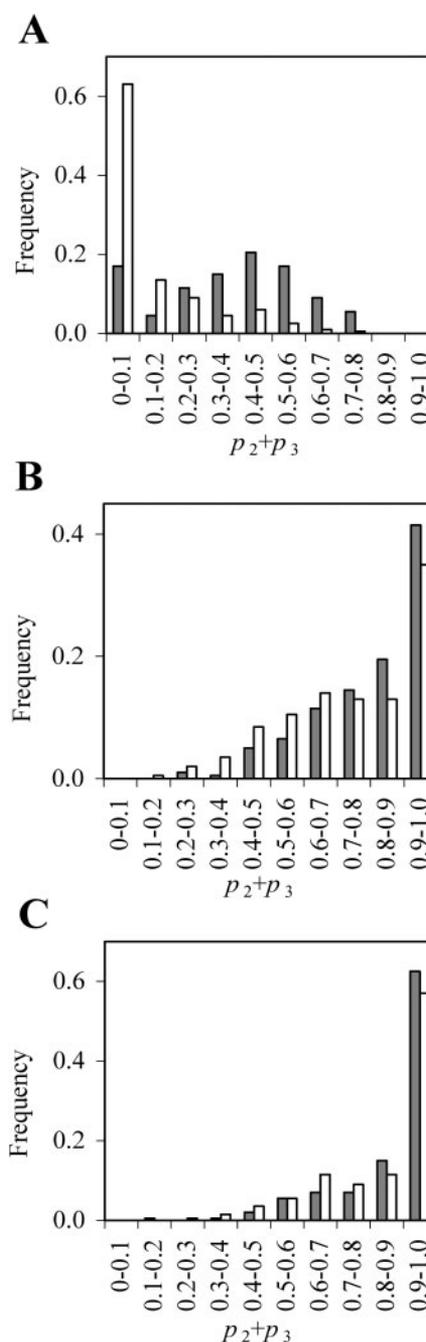


FIG. 3.—Distributions of the estimated proportion ($p_2 + p_3$) of codons that have altered ω in foreground branches. Evolutionary scheme X is used for background branches and X, Y, and Z are used, respectively, for foreground branches in panels A, B, and C. Filled and open bars show distributions from tests A and B, respectively. In the simulations, branch α of tree I is assigned as the foreground branch.

test A erroneously inferred positive selection even at sites that are under purifying selection.

Sequence evolution was also simulated using scheme X for the background branches and Z for the foreground branch. Z represents a complete relaxation of purifying selection but, again, involves no positive selection (table 1). It was found that both tests A and B correctly rejected the null hypothesis of no difference in ω values between the

background and foreground branches in all replications (tables 2 and 3). However, tests A and B erroneously detected positive selection in 53.5% and 68.5% of the 200 replications, respectively, for the foreground branch (tables 2 and 3). Again, for all cases in which test A detected positive selection, test B also detected it. For test B, 66% of the cases were inferred to experience positive selection specifically in the foreground but not background branches. The results of both tests A and B were worse when evolutionary scheme Z was used instead of Y for the foreground branch. Apparently, the method has difficulties distinguishing between relaxation of purifying selection and positive selection; this difficulty becomes more serious when the relaxation becomes more extensive. In the simulation, the true value of $p_2 + p_3$ was 0.8. The distribution of the estimated $p_2 + p_3$ was J-shaped, with the majority of estimates at 1 (fig. 3C). The mean of the distribution was 0.889 ± 0.011 and 0.856 ± 0.013 for tests A and B, respectively. Therefore, $p_2 + p_3$ was overestimated by the branch-site likelihood method. It was found that for this set of simulations N_{PS} had no significant correlation with either LR or ω_2 ($P > 0.05$). Again, N_{PS} was always 0 for cases in which positive selection was not inferred, whereas it was equal to or greater than 0 for cases in which positive selection was inferred. The mean N_{PS} was 15.7 and 18.6 in tests A and B, respectively, for cases in which positive selection was inferred. Test A detected two cases in which all 200 codons were inferred to be under positive selection with $>95\%$ posterior probabilities, but test B found only 0 and 20 positively selected codons, respectively, in these two cases. Test B detected four cases with all 200 codons under positive selection, but test A found only 0 to 16 positively selected codons in these cases.

In summary, the results from the above three sets of simulations show that while tests A and B are powerful in detecting a difference in ω values between branches, they are also extremely liberal and make frequent false detections of positive selection.

Tree I—Branches β and γ

The above results were for simulations conducted using branch α of tree I as the foreground branch. Simulations were also conducted when two other branches of tree I (β or γ ; fig. 1) were respectively chosen as the foreground branch. The three foreground branches examined (α , β and γ) differ in length and/or location in the tree. The simulations showed that results for branches β and γ are similar to those for branch α qualitatively, though they differ quantitatively (tables 2 and 3). The location and length of the foreground branch did not affect the performance of the likelihood method very much, and both tests A and B were overly liberal in detecting positive selection. However, test B was even more liberal than test A, as test B made more erroneous detections of positive selection than did A under all the conditions examined.

Tree II—Branches α , β , γ , and δ

Simulations were also conducted when four different branches of tree II were individually chosen as the

foreground branch (fig. 1). The four branches are located at different positions in the tree, and the length of branch δ is 4 times the length of α , β , or γ . Similar to the case of tree I, both A and B tests were found to be liberal in detecting positive selection in tree II. In addition, test A was found to make fewer mistakes than test B, particularly when the foreground branch was not an exterior branch (tables 2 and 3). However, when the foreground branch was a short exterior branch, test A performed worse than test B. In general, both tests performed worse when evolutionary scheme Z was used for foreground branches than when Y was used, although this was not a strict rule. The overall error rate of tests A and B in tree II was comparable to that in tree I (tables 2 and 3).

Discussion

The branch-site likelihood method for detecting positive selection uses the likelihood-ratio test to compare the null hypothesis of no positive selection with the alternative hypothesis of positive selection. According to theory, for the likelihood ratio test to be unbiased, the assumptions made in the likelihood computation should be correct (Zhang 1999). In reality, however, some assumptions are probably violated because many parameters of gene sequence evolution are difficult to know. Under such circumstances, it is important to understand what factors might affect the performance of a particular likelihood-ratio test and whether the test is robust against violations of certain assumptions (Zhang 1999). In this work, computer simulation was used to investigate the performance of test A and test B of the branch-site likelihood method. The two tests differ only in the ω values assigned to sites. The simulations were conducted with the absence of positive selection for the purpose of estimating the rate of false positives of the two tests. Two trees, seven branches, and three different evolutionary schemes were used in the simulation. Although the number of conditions examined was somewhat limited, due largely to the computational intensity of the simulations, this study revealed serious and consistent defects of the branch-site likelihood method that are worth attention. First, test A tends to reject the correct null hypothesis more often than expected when there is no difference in ω values between background and foreground branches. In addition, test A detects positive selection in 19%–54% of the cases (depending on the conditions) when there is a relaxation of purify selection but no positive selection in foreground branches. Second, although test B performs reasonably well when there is no difference in ω values between background and foreground branches, it performs worse than test A when there is a relaxation of purifying selection but no positive selection in foreground branches, as it erroneously detects positive selection 33%–66% of the time. Consequently, the excessively high error rate of the branch-site likelihood method calls for reevaluation of past claims of positive selection based on results generated from its use. Distinguishing positive selection from relaxed purifying selection is important, because rapid sequence evolution with an elevated ω can be caused by either of the two phenomena. However, the biological implications of

these two types of changes in natural selection are drastically different. The action of positive selection implies that gene function is altered in such a way that fitness is enhanced. In contrast, a relaxation of purifying selection implies that the importance of a particular gene (or gene region) has been reduced, without any fitness consequences. It has been debated, for instance, whether positive selection and/or relaxed purifying selection is responsible for frequent observations of elevated rates of nonsynonymous substitution after gene duplication (Kimura 1983; Li and Gojobori 1983; Zhang, Rosenberg, and Nei 1998; Hughes 1999; Van de Peer et al. 2001; Zhang, Zhang, and Rosenberg 2002; Zhang 2003), and it was suggested that the branch-site likelihood method might settle the controversy (Bielawski and Yang 2003). Unfortunately, the present study shows that the method is unable to distinguish positive selection and relaxed purifying selection.

The branch-site method was also used in a recent study to look for genes that have been under positive selection in human evolution after the separation of humans from chimpanzees (Clark et al. 2003). Among 7,645 genes tested, 667 (8.7%) were found to be under positive selection at 5% significance level by a new test resembling the hybrid of tests A and B. In this hybrid test, ω_0 is no longer fixed at 0 but is constrained to be between 0 and 1, whereas ω_1 is fixed at 1; also, the relative branch lengths of the tree were fixed in the likelihood computation. Although the performance of this hybrid test has not been examined by computer simulation, the fact that it is a hybrid of A and B suggests the possibility that it is likewise liberal in detecting positive selection and that some cases of relaxation of purifying selection may have been mistaken as positive selection, as the authors suspected (Clark et al. 2003). For example, the Clark et al. (2003) study detected positive selection in over 50% of human olfactory receptor (OR) genes (Clark et al. 2003). Although positive selection in some human OR genes is possible (Gilad et al. 2003a), massive pseudogenization of ORs in the human lineage has been well documented (Gilad et al. 2003b; Menashe et al. 2003). So, it is highly suspicious that over 50% of OR genes would actually be under positive selection. Given the high chance of false detection of positive selection by the branch-site method and the multiple comparisons involved (7,645 genes), it would have been better to use a more stringent P value than the 5% value used by the authors.

It is important to understand why the branch-site method does not perform well and how its poor performance might be corrected. The main problem appears to be due to the fact that certain key assumptions in the likelihood model are problematic and unrealistic. For instance, the assumption of only two classes (with respect to ω) of codons for background branches and three classes for foreground branches is overly simplistic, as real gene sequences contain sites with many different ω values. The evolutionary scheme used for the background branches contained five classes of sites and those for the foreground branches contained either five or one class of sites in the simulations conducted in this study. This seemingly small deviation from the assumption made by

the likelihood method causes both tests A and B to fail, suggesting that the tests are sensitive to the violation of this type of assumption. Another assumption made in the likelihood method is that $p_0/p_1 = p_2/p_3$, where p_0 and p_1 are proportions of sites belonging to classes 0 and 1, respectively, in both background and foreground branches, and p_2 and p_3 are the proportions of sites that belong to classes 0 and 1 in background branches, respectively, but belong to class 3 in foreground branches. This assumption does not have any biological foundation. It is violated in the simulations conducted here and is probably always violated for real genes. Another key assumption in the branch-site method that is nearly always violated in real data sets is an equal transition/transversion ratio (κ) for synonymous and nonsynonymous substitutions. This is unrealistic, because transitions are more likely to be synonymous, whereas transversions are more likely to be nonsynonymous (Nei and Kumar 2000). Therefore, κ should be different for synonymous and nonsynonymous substitutions. In this study, although only one κ was used in generating mutations, the κ for synonymous substitutions and the κ for nonsynonymous substitutions were different because of the action of natural selection. It is unknown how sensitive the likelihood method is to the violation of this assumption, as a likelihood model separating two κ 's is yet to be developed. It will be interesting to see whether the development of more realistic models with regard to the number of classes of sites, proportion of each class of sites, and κ will correct the problems of the branch-site likelihood method.

In addition to these problems, the likelihood method does not specify *a priori* sites that are potentially under positive selection. This is in contrast to the original method of comparing d_S and d_N (Hughes and Nei 1988), in which such sites are specified *a priori* using independent biological information such as protein structure and function. In the likelihood method, the average ω value is computed for each class of sites, which are grouped by the ω values. This circularity may be another reason why the method is excessively liberal, as chance variations in ω among sites may be erroneously amplified by grouping sites according to the ω (Suzuki and Nei 2004).

Interest in studying positive selection in gene evolution has risen over the past few years, due in large part to the development of statistical methods that identify selection at individual sites and/or branches (Yang 1998; Zhang, Rosenberg, and Nei 1998; Suzuki and Gojobori 1999; Yang et al. 2000; Yang and Nielsen 2002). The performance of these methods needs to be examined by both empirical data analysis and computer simulation, as has already been done for some of them (Zhang, Kumar, and Nei 1997; Anisimova, Bielawski, and Yang 2001, 2002; Suzuki and Nei 2001, 2002). Blind use of statistical methods, without understanding their properties or biological meanings of statistical results, is dangerous. Probably because adaptation is of interest to many biologists, statistical results indicating positive selection appear to be favored by some investigators. This results in preferential use of methods that liberally detect positive selection, rather than methods that are based on solid biological and/or statistical grounds but are more conser-

vative. For example, Sorhannus (2003) analyzed the DNA sequences of the sexual reproduction gene *Sig1* from the diatom *Thalassiosira weissflogii* using likelihood- and parsimony-based methods. While the former method indicates action of positive selection, the latter did not. Although there is no independent evidence for positive selection in *Sig1*, the author believed that *Sig1* should be under positive selection and thus concluded that the likelihood method is better than the parsimony method. In fact, detailed analyses of the data by Suzuki and Nei (2004) suggested that the likelihood result was probably a false positive. Likelihood methods are often perceived to be superior to other methods because they are statistically sound when the assumptions made in the methods are satisfied. However, the superiority of likelihood is not necessarily guaranteed because many of these assumptions on gene sequence evolution are either difficult to justify or simply wrong. Furthermore, the present and previous studies have shown that likelihood methods are sensitive to violations of certain assumptions (Zhang 1999; Suzuki and Nei 2001, 2002, 2004). In such cases, more conservative statistical methods are preferred to liberal ones.

Acknowledgments

I thank Masatoshi Nei, Alex Rooney, David Webb, the Associate Editor, and two anonymous reviewers for valuable comments. This work was supported by a startup fund from the University of Michigan and an NIH grant (GM67030) to J.Z.

Literature Cited

- Anisimova, M., J. P. Bielawski, and Z. Yang. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* **18**:1585–1592.
- . 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Mol. Biol. Evol.* **19**:950–958.
- Bailly, X., R. Leroy, S. Carney, O. Collin, F. Zal, A. Toulmond, and D. Jollivet. 2003. The loss of the hemoglobin H2S-binding function in annelids from sulfide-free habitats reveals molecular adaptation driven by Darwinian positive selection. *Proc. Natl. Acad. Sci. USA* **100**:5885–5890.
- Barkman, T. J. 2003. Evidence for positive selection on the floral scent gene isoeugenol-O-methyltransferase. *Mol. Biol. Evol.* **20**:168–172.
- Bielawski, J. P., and Z. Yang. 2003. Maximum likelihood methods for detecting adaptive evolution after gene duplication. *J. Struct. Funct. Genomics* **3**:201–212.
- Clark, A. G., S. Glanowski, R. Nielsen et al. (17 co-authors). 2003. Inferring nonneutral evolution from human-chimpanzee orthologous gene trios. *Science* **302**:1960–1963.
- Gilad, Y., C. D. Bustamante, D. Lancet, and S. Paabo. 2003a. Natural selection on the olfactory receptor gene family in humans and chimpanzees. *Am. J. Hum. Genet.* **73**:489–501.
- Gilad, Y., O. Man, S. Paabo, and D. Lancet. 2003b. Human specific loss of olfactory receptor genes. *Proc. Natl. Acad. Sci. USA* **100**:3324–3327.
- Hughes, A. L. 1999. *Adaptive evolution of genes and genomes*. Oxford University Press, New York.
- Hughes, A. L., and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**:167–170.
- Kimura, M. 1983. *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.
- Kinsella, R. J., D. A. Fitzpatrick, C. J. Creevey, and J. O. McInerney. 2003. Fatty acid biosynthesis in *Mycobacterium tuberculosis*: lateral gene transfer, adaptive evolution, and gene duplication. *Proc. Natl. Acad. Sci. USA* **100**:10320–10325.
- Knipple, D. C., C. L. Rosenfield, R. Nielsen, K. M. You, and S. E. Jeong. 2002. Evolution of the integral membrane desaturase gene family in moths and flies. *Genetics* **162**:1737–1752.
- Li, W. H., and T. Gojobori. 1983. Rapid evolution of goat and sheep globin genes following gene duplication. *Mol. Biol. Evol.* **1**:94–108.
- Mathews, S., J. G. Burleigh, and M. J. Donoghue. 2003. Adaptive evolution in the photosensory domain of phytochrome A in early angiosperms. *Mol. Biol. Evol.* **20**:1087–1097.
- Martinez-Castilla, L. P., and E. R. Alvarez-Buylla. 2003. Adaptive evolution in the Arabidopsis MADS-box gene family inferred from its complete resolved phylogeny. *Proc. Natl. Acad. Sci. USA* **100**:13407–13412.
- Menashe, I., O. Man, D. Lancet, and Y. Gilad. 2003. Different noses for different people. *Nat. Genet.* **34**:143–144.
- 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, New York.
- 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.
- Rodriguez-Trelles, F., R. Tarrío, and F. J. Ayala. 2003. Convergent neofunctionalization by positive Darwinian selection after ancient recurrent duplications of the xanthine dehydrogenase gene. *Proc. Natl. Acad. Sci. USA* **100**:13413–13417.
- 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**:1326–1328.
- Suzuki, Y., and T. Gojobori. 1999. A method for detecting positive selection at single amino acid sites. *Mol. Biol. Evol.* **16**:1315–1328.
- Suzuki, Y., and M. Nei. 2001. Reliabilities of parsimony-based and likelihood-based methods for detecting positive selection at single amino acid sites. *Mol. Biol. Evol.* **18**:2179–2185.
- . 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**:1865–1869.
- . 2004. False positive selection identified by ML-based methods: examples from the *Sig1* gene of the diatom *Thalassiosira weissflogii* and the *tax* gene of a human T-cell lymphotropic virus. *Mol. Biol. Evol.* **21**:914–921.
- Van de Peer, Y., J. S. Taylor, I. Braasch, and A. Meyer. 2001. The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes. *J. Mol. Evol.* **53**:436–46.
- Yang, J., J. Huang, H. Gu, Y. Zhong, and Z. Yang. 2002. Duplication and adaptive evolution of the chalcone synthase genes of *Dendranthema* (Asteraceae). *Mol. Biol. Evol.* **19**:1752–1759.
- Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**:555–556.

- . 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**:568–573.
- 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. Pedersen. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**:431–449.
- Yu, M., and D. M. Irwin. 1996. Evolution of stomach lysozyme: the pig lysozyme gene. *Mol. Phylogenet. Evol.* **5**:298–308.
- Zhang, J. 1999. Performance of likelihood ratio tests of evolutionary hypotheses under inadequate substitution models. *Mol. Biol. Evol.* **16**:868–875.
- . 2003. Evolution by gene duplication—an update. *Trends Ecol. Evol.* **18**:292–298.
- 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.
- Zhang, J., Y. P. Zhang, and H. F. Rosenberg. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat. Genet.* **30**:411–415.

Robin Bush, Associate Editor

Accepted February 27, 2004