J Mol Evol (2000) 50:296–301
DOI: 10.1007/s002399910034

JOURNAL OF MOLECULAR EVOLUTION

The Power of Relative Rates Tests Depends on the Data

Lindell Bromham,[1–3] David Penny,[1] Andrew Rambaut,[3] Michael D. Hendy[4]

[1] Institute for Molecular BioSciences, Massey University, P.O. Box 11-222, Palmerston North, Aotearoa (New Zealand)
[2] Department of Zoology, University of Queensland, Brisbane 4072, Australia
[3] Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK
[4] Institute of Fundamental Sciences, Massey University, P.O. Box 11-222, Palmerston North, Aotearoa (New Zealand)

Received: 9 June 1999 / Accepted: 22 October 1999

Abstract. One of the most useful features of molecular phylogenetic analyses is the potential for estimating dates of divergence of evolutionary lineages from the DNA of extant species. But lineage-specific variation in rate of molecular evolution complicates molecular dating, because a calibration rate estimated from one lineage may not be an accurate representation of the rate in other lineages. Many molecular dating studies use a "clock test" to identify and exclude sequences that vary in rate between lineages. However, these clock tests should not be relied upon without a critical examination of their effectiveness at removing rate variable sequences from any given data set, particularly with regard to the sequence length and number of variable sites. As an illustration of this problem we present a power test of a frequently employed triplet relative rates test. We conclude that (1) relative rates tests are unlikely to detect moderate levels of lineage-specific rate variation (where one lineage has a rate of molecular evolution 1.5 to 4.0 times the other) for most commonly used sequences in molecular dating analyses, and (2) this lack of power is likely to result in substantial error in the estimation of dates of divergence. As an example, we show that the well-studied rate difference between murid rodents and great apes will not be detected for many of the sequences used to date the divergence between these two lineages and that this failure to detect rate variation is likely to result in consistent overestimation the date of the rodent–primate split.

Key words: Molecular clock — Substitution rate — Phylogeny — Molecular date estimates — Tajima test — Rate variation

Correspondence to: Lindell Bromham; *e-mail:* LBromham@zoology.uq.edu.au

Introduction

Many examples of lineage-specific rates of molecular evolution have been described in the literature, and evidence is growing that such rate variation is widespread (e.g., Bousquet et al. 1992; Mooers and Harvey 1995; Bromham et al. 1996). This presents a problem for the accurate reconstruction of molecular phylogenies, particularly where genetic divergence is used to estimate time since the separation of lineages. Some molecular phylogenetic studies attempt to avoid the problem of rate variation biasing date estimates by using a "clock test" to detect sequences that vary in substitution rate between lineages, which can then be removed from the analysis. The remaining nonrejected sequences are then treated as effectively rate-uniform to estimate dates of divergence by converting genetic distance to time using a calibration rate (e.g., Easteal et al. 1995; Hedges et al. 1996; Kumar and Hedges 1998). These clock tests are often based on the relative rates test (Sarich and Wilson 1973; Wu and Li 1985), which compares the distance from each member of a pair of taxa to an outgroup taxon. The reliability of triplet tests depends on the taxa and sequences chosen,

so they do not always provide a suitable filter for removing rate variable data from an analysis.

As an illustration of the way the reliability of clock tests depends on the data, we present a power test of a commonly used triplet clock test, the Tajima test. Although Tajima (1993) presented a power test for sequences of 1000 positions, and stated that the test may be conservative, this test is frequently applied to sequences with considerably fewer positions which are free to vary. It is therefore important to examine the effectiveness of this relative rates test for detecting sequences of fewer than 1000 sites free to vary, which differ in rate between lineages. We demonstrate that the Tajima test does not provide an adequate filter for removing rate-variable sequences from many molecular phylogenetic data sets. Other clock tests will be similarly limited in power. The distance-based relative rates test (Wu and Li 1985) has similar power to the Tajima test (Tajima 1993; see also Fig. 3). Likelihood ratio tests are also limited in power to detect rate variation for shorter sequences (Muse and Weir 1992; Tajima 1993; Rambaut and Bromham 1998).

Assessing the power of a clock test for a given set of data is critical to avoid introducing significant bias into molecular date estimates. Contrary to common expectation, sequences that "pass" a relative rates test cannot always be considered sufficiently clock-like for the purposes of date estimation. Instead, the fewer variable sites that are considered, the greater the Type II error of the test (failure to spot a difference in rate), the larger the magnitude of rate variation that will remain undetected, and the greater the expected error in date estimates. If species-specific rate differences show a consistent pattern, then we expect the inclusion of rate variable sequences to introduce consistent bias to date estimates. We illustrate this point by considering one of the most frequently estimated molecular date divergences: the rodent/primate split. We show that relying on clock tests with insufficient power to reject sequences that evolve more rapidly in rodents than in primates can result in consistent overestimation of the date of the split between rodents and primates.

Analysis

The Tajima test is based on the expectation that under a uniform rate of substitution, the number of sites at which the amino acid or nucleotide state is shared by the outgroup and only one of the two ingroups should be equal for both ingroups (Tajima 1993). We use the Hadamard conjugation (Hendy and Penny 1993) to generate the probabilities for patterns of substitutions for a triplet of sequences where one member of the ingroup has a higher rate of substitution than the other. The Tajima test considers a triplet of sequences, so at any given sequence position (site) there are at most three states different present. The informative sites for a Tajima test are those in which the outgroup and one of the ingroups share a character state, but the other ingroup has a different state (Tajima 1993). So for a column of data with the outgroup state listed first and the two ingroups in the second and third positions, the informative patterns are 001 and 010 (and 101 and 110: the zeros and ones stand for any state and serve merely to show whether two states are the same or different). For example, ACA is an informative Tajima site (010) but AAA, ACC, and ACG are not. So whether the Tajima test is performed on 2-state data (for example, pyrimidine and purine nucleotides), 4-state data (A, C, T, G), or 20-state data (amino acids), the informative sites will be those that have only two different states. The test ignores the sites where three different states occur because such sites are not informative for testing for an excess of substitutions in one of the ingroups (Tajima 1993).

Because only sites with two states present are considered, and because the test treats each type of substitution equally, the Tajima test effectively works only on two-state patterns, however many character states are represented in the data. So we can use the two-state Hadamard conjugation to assess whether samples selected from these probabilities were sufficient to allow the Tajima test on Tajima's 1D, which is equivalent to the Jukes–Cantor model, to detect the rate variation. Unless the sequences are near saturation, the frequency of sites with three different states will be small, so the conclusions also apply to a four-state model. Our analysis is conservative because the Tajima test will have the highest power for two-state data: for three- and four-state data, more sites will have three character states, and thus be omitted from the test, effectively reducing the number of sites considered.

We selected sample sequence data from the expected distribution of sequence patterns analytically derived by Hadamard conjugation (Hendy and Penny 1993) from a defined tree with given edge length parameters. This distribution is described functionally below. The Hadamard conjugation allows us to predict the frequency of patterns in sequences that have been generated by a given tree. To generate a sequence of length L, we sample L sites from the distribution with replacement. This enables us to efficiently construct large numbers of samples without having to "grow" each sample stochastically.

Given a triplet, suppose that the taxa are numbered 1, 2, and 3, where 3 is the outgroup, with the root at time t before present (see Fig. 1), and the divergence between 1 and 2 at time ft before present ($0<f<1$). The parameters of our model are the expected numbers of substitutions (the edge lengths) of each edge of the generating tree. Suppose that there is a base substitution rate α on the lineages to taxa 1 and 3 and an acceleration (or deceleration) of the rate to $k\alpha$ on the lineages to taxon 2. Then the expected number of substitutions on the lineages to taxa 1 and 2 are $q_1 = ft\alpha$ and $q_2 = ftk\alpha$, respectively, while the expected number of substitutions between the base of the ingroup and taxon 3 is $q_3 = (2-f)t\alpha$.

The Hadamard conjugation gives the functional relationship between these q_i values and the probabilities of each of the possible patterns of states at any given site. These probabilities are easily described in terms of exponential functions E_1, E_2, and E_3 of the q_i values. Let s_{ABC} be the probability of obtaining the pattern of states ABC at a site; then we find

$$s_{XXX} = \frac{(1 + E_1 + E_2 + E_3)}{4} \quad \text{(the probability of either 000 or 111)}$$

$$s_{XXY} = \frac{(1 - E_1 - E_2 + E_3)}{4} \quad \text{(the probability of either 001 or 110)}$$

$$s_{XYX} = \frac{(1 + E_1 - E_2 - E_3)}{4} \quad \text{(the probability of either 010 or 101)}$$

$$s_{YXX} = \frac{(1 - E_1 + E_2 - E_3)}{4} \quad \text{(the probability of either 011 or 100)}$$

where

$$E_1 = \exp(-2(q_1 + q_3)) = \exp(-4t\alpha)$$
$$E_2 = \exp(-2(q_2 + q_3)) = \exp(-2(2 + f(k-1))t\alpha)$$
$$E_3 = \exp(-2(q_1 + q_2)) = \exp(-2(1 + k)ft\alpha)$$

To choose realistic sequence parameters, we focused on estimating the

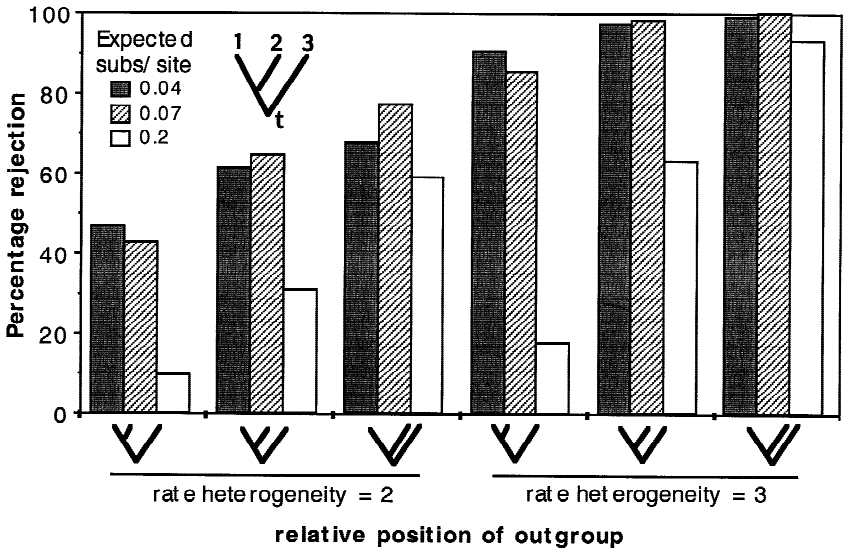

Fig. 1. Predicted percentage of cases for which the Tajima test will reject rate constancy when one member of the ingroup has a rate of substitution that is twofold and threefold that of the other ingroup taxon, for different substitution rates, and relative positions of ingroup to outgroup: the ingroup is one-third, one-half, and two-thirds the distance from the tip to the root of the triplet (sequence length = 700).

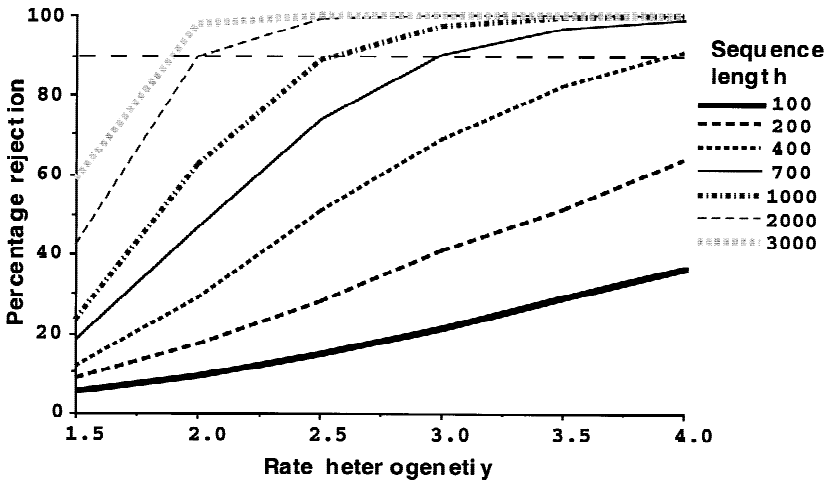

Fig. 2. Expected percentage of cases for which the Tajima test will reject rate constancy when one member of the ingroup has a rate of substitution that is 1.5 to 4.0 times faster than the other ingroup taxon, for a range of sequence lengths (substitution rate = 0.04 substitution/site; ratio of ingroup (rate-constant) to outgroup = 1/6).

date of the primate–rodent divergence using nuclear genes (e.g., Easteal et al. 1995; Hedges et al. 1996; Kumar and Hedges 1998). Not only is this one of the most frequently estimated dates of divergence (see Bromham et al. 1999), but also it is one of the most intensely studied examples of lineage-specific rates of molecular evolution (e.g., Gu and Li 1992; Li et al. 1996; Yang and Nielsen 1998). Simulations were conducted using six levels of rate variation between members of the ingroup (k = 1.5, 2, 2.5, 3, 3.5, 4) and three different numbers of substitutions [$2\alpha t$ = 0.04, 0.07, and 0.2 substitution per site, values chosen to give a broad range of rates, using Yang and Nielsen's (1998) study of substitution rates in primates and rodent as a guide].

Results

The results of these tests indicate that the Tajima test is unlikely to detect moderate levels of rate variation for sequences with 400 or fewer sites free to vary (Fig. 2). Sequences of at least 1000 positions that are equally free to vary are needed to reliably detect threefold rate differences and above, and much longer sequences (>2000 positions free to vary) are needed to detect twofold rate differences. A conservative estimate of the ratio of rodent to primate rates is 1.5 (Wu and Li 1985; Gu and Li 1992; Easteal et al. 1995; Li et al. 1996; Yang and Nielsen 1998). None of the sequence lengths tested were sufficient to allow reliable detection of a rodent rate 1.5 times the primate rate. Studies that collate the maximum number of available sequences will tend to be biased toward shorter sequences, frequently fewer than 500 variable positions (e.g., Hedges et al. 1996; Wang et al. 1998), and in this situation the Tajima test cannot reliably be used to reject sequences with relatively large rate differences.

The effectiveness of the Tajima test is also reduced when applied to saturated data, so both the substitution rate and the relative distance from the ingroup to the outgroup are important (Fig. 1). Because the branch leading to the outgroup is likely to be most affected by saturation, the decrease in power caused by a high number of substitutions per site can be partially ameliorated by selecting a less distant outgroup. Selecting a very distant

outgroup [e.g., birds as outgroup to rodent/primate (Hedges et al. 1996; Kumar and Hedges 1998)] reduces the chance of detecting rate differences (Tajima 1993; Robinson et al. 1998). Because the Tajima test considers only sites which differ in one of the ingroup branches, it may also be ineffective at very low rates of substitution where few changes in the ingroup have occurred.

Example: A Tale of Mice and Men

We illustrate the importance of power testing using a recent example. Kumar and Hedges (1998) present molecular estimates of a range of mammalian taxa, supporting the divergence of many orders in the Cretaceous, long before the dinosaurs' extinction. As with their earlier study (Hedges et al. 1996), they screened a large number of genes for rate constancy using the Tajima test. After rejecting 22% of the genes as rate variable, the remaining genes were analyzed on the assumption that they were rate constant.

The power of the Tajima test is limited by sequence length and substitution rate. Using the sequence parameters supplied in an earlier study (Hedges et al. 1996), we simulated the evolution of amino acid sequences for a rodent, a primate, and a bird, giving the rodent lineage a substitution rate 1.5 times that of the primates [a conservative estimate of the expected difference in rate of sequence evolution (Wu and Li 1985; Gu and Li 1992; Easteal et al., Li et al. 1996; Yang and Nielsen 1998)]. With an average length of 400 amino acids (Hedges et al. 1996), we would expect the test to reject rate constancy for only 16% of genes evolving under this degree of rate variation (Fig. 3a). A similar result would be expected for a distance-based relative rates test (Fig. 3b). Note that if not all sites in the sequence are equally free to vary, then the power will be even lower. It therefore seems probable that not all rate-variable genes were excluded from the analysis.

What are the consequences of failing to reject sequences that show a moderate level of rate variation? It should not be assumed that any rate variation not detected by a clock test is unimportant for the purposes of molecular phylogenetic date estimation. If a relative rates test were used to select "clock-like" sequences to date the rodent–primate split, we would expect it to fail to reject most of the sequences of moderate length (700 or fewer variable sites) that vary in rate between rodents and primates by a factor of 1.5 to 3.0. What effect would the inclusion of these sequences in the analysis have on the accuracy of molecular date estimates of the primate–rodent split?

To test the effect of including sequences with realistic levels of rate variation that would remain undetected by the Tajima test, we simulated the evolution of DNA sequences (length = 700) along a triplet representing rodent and primate lineages and an outgroup (Fig. 4),

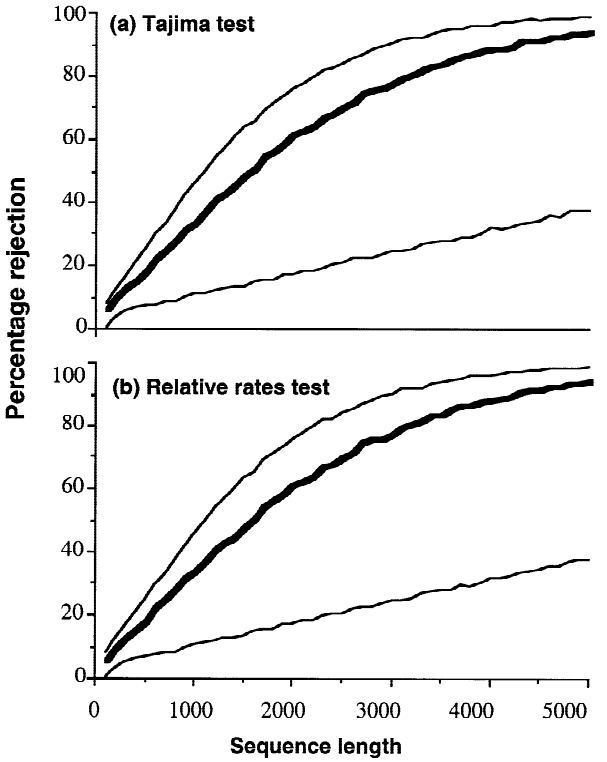

Fig. 3. Percentage of cases in which **(a)** the Tajima test and **(b)** the distance-based relative rates test reject rate constancy at different lengths of amino acid sequence. Simulations are based on data from Hedges et al. (1996), conducted using SeqGen (Rambaut and Grassly 1997), which draws the number of substitutions for each branch from a Poisson distribution (mean defined by the expected number of substitutions along each branch) and applies these substitutions to random sites. We set the rate of evolution of the rodent lineage to 1.5 times that of the primate lineage. Each point on the plot is the mean of 10,000 replicate simulations. For each simulation, the number of substitutions occurring on each lineage was Poisson distributed (an ideal molecular clock process), with a mean calculated from the substitution rates and divergence times reported by Hedges et al. (1996). For the *thick central line*, the primate and rodent had a mean number of substitutions per site of 0.044 (95 Myr, at a rate of 4.6×10^{-4} substitutions per site per Myr), but with 1.5 times more substitutions on the rodent lineage than on the primate lineage. The lineage leading from the divergence of the mammals to the bird had a mean of 0.24 substitution per site (525 Myr at a rate of 4.6×10^{-4}. Results for the maximum and minimum values of divergence for the rodent/primate comparisons (when the rate is 8.5×10^{-4} and 0.8×10^{-4}, respectively) are also shown (*thin lines*).

where the rodent lineage has a substitution rate 1.0–3.0 times faster than the primate and outgroup rate, and the branch from the ingroup to the outgroup is three times the length of the primate branch. Simulations were performed with the SeqGen program (Rambaut and Grassly 1997) using a Kimura two-parameter model. Each simulated data set was subjected to the Tajima test, and those data sets that passed the Tajima test were analyzed as if rate constant. The K2P distance between the rodent and the primate lineages was estimated for each "rate-constant" simulated data set and halved to get the depth of the rodent–primate node. The depth of the primate–rodent split is overestimated by increasing amounts as

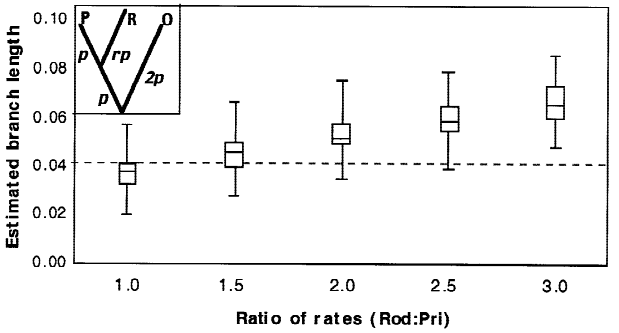

Fig. 4. The effect of variation in the rate of DNA sequence evolution on the accuracy of date estimates when a molecular clock is assumed is demonstrated by simulating data [SeqGen, Kimura two-parameter model (Rambaut and Grassly 1997)] along a triplet representing rodent and primate lineages and an outgroup (see **inset**). The length of the primate lineage is *p* (here, $p = 0.04$ substitution/site. This corresponds to the average *dN* from Yang and Nielsen (1998) and the average substitution rate from Hedges et al. (1996). The rodent lineage is longer than the primate lineage (i.e., has a faster rate of evolution) by a factor of *k*, where *k* is between 1.0 and 3.0. The length of the branch from the ingroup to the outgroup is 3*p*. One thousand sets of simulated sequence data were generated for each level of rate heterogeneity. Each simulated data set was analyzed using the Tajima test. Any data sets that were not rejected by the Tajima test as containing significant rate variation were analyzed as if they were rate constant. For each simulated data set that passed the Tajima test, the depth of the primate–rodent divergence was estimated by halving the K2P distance between R and P. In this plot, the length of the primate branch length estimated in this way is compared to the true length of the primate branch ($p = 0.04$; *dashed line*). Box plots represent mean, maximum, minimum, and quartiles of the estimates given by all datasets that passed the Tajima test at each level of rate heterogeneity. The true length of the primate branch *(p)* is overestimated by increasing amounts as lineage-specific differences in substitution increases.

lineage-specific differences in substitution rate increases. So if the nonrejected sequences were considered sufficiently clock-like, so that the distance between them was halved to give the depth of the divergence, then the age of the rodent–primate split would be consistently overestimated (Fig. 4).

Conclusions

The power of relative rates tests is clearly dependent on (a) the number of variable sites considered, (b) the rate of substitution, (c) the distance from the ingroup to the outgroup, and (d) the degree of variation in molecular evolution rate between members of the ingroup taxa. While we use the Tajima test as a convenient illustration, other clock tests (such as the distance-based relative rates test and the likelihood ratio test) are similarly limited in power, dependent on the number of variable sites. For many sequences used in molecular phylogenetic analyses, such tests will be of insufficient power to be a reliable means of detecting lineage-specific rate variation. In addition, relative rates tests can assess only "parallel rate equality" and are not able to detect concerted acceleration or deceleration in rate across multiple lineages (Gingerich 1986). Because of the high Type II error, relative rates tests can be used to demonstrate significant rate variation (e.g., Bromham et al. 1996; Bousquet et al. 1996) but can rarely be relied upon to prove a "molecular clock." These concerns have been raised before (e.g., Scherer 1989; Avise 1994; Robinson et al. 1998) but we are prompted to draw attention to the limited power of triplet tests because of their problematic application to provide evidence of rate constancy in molecular dating analyses.

While the Tajima test, or similar clock tests, may be used to reject a molecular clock, its use in selecting clock-like sequences for molecular dating analyses is inappropriate without first assessing that it has the desired level of power for a given data set. When such a test is applied to data for which its power to reject rate variable sequences is low, there is a significant risk of generating inaccurate date estimates. Ideally, long sequences (several kilobases) should be used in studies that aim to use clock tests to select sequences that approximate rate constancy. Where this is not possible, combining many short sequences or developing a "whole tree" approach that incorporates information for many lineages may improve the accuracy of the tests. An alternative to hunting for rate-constant sequences is to develop molecular dating methods that allow for variation in substitution rate between lineages (e.g., Sandersen 1997; Rambaut and Bromham 1998).

References

Avise JC (1994) Molecular markers, natural history and evolution. Chapman & Hall, New York

Bosquet J, Strauss SH, Doeksen AH, Price RA (1992) Extensive variation in evolutionary rate of rbcL gene sequences among seed plants. Proc Natl Acad Sci USA 89:7844–7848

Bromham LD, Rambaut AE, Harvey PH (1996) Determinants of rate variation in mammalian DNA sequence evolution. J Mol Evol 43: 610–621

Bromham LD, Phillips MJ and Penny D (1999) Growing up with dinosaurs: Molecular dates and the mammalian radiation. Trends Ecol Evol 14:113–118

Easteal S, Collett C, Betty D (1995) The mammalian molecular clock. RG Landes, Austin, TX

Gingerich PD (1986) Temporal scaling of molecular evolution in primates and other mammals. Mol Biol Evol 3:205–221

Gu X, Li W-H (1992) Higher rates of amino acid substitution in rodents than in man. Mol Phylogenet Evol 1:211–214

Hedges SB, Parker PH, Sibley CG, Kumar S (1996) Continental breakup and the ordinal diversification of birds and mammals. Nature 381:226–229

Hendy MD, Penny D (1993) Spectral analysis of phylogenetic data. J Class 10:5–24

Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. Nature 392:917–920

Li W-H, Ellesworth DL, Krushkal J, Chang BH-J, Hewett-Emmett D

(1996) Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. Mol Phylogenet Evol 5:182–187

Mooers AO, Harvey PH (1994) Metabolic rate, generation time and the rate of molecular evolution in birds. Mol Phylogenet Evol 3:344–350

Muse SV, Weir BS (1992) Testing for equality of evolutionary rates. Genetics 132:269–276

Rambaut AE, Bromham LD (1998) Estimating divergence dates from molecular sequences. Mol Biol Evol 15:442–448

Rambaut A, Grassly NC (1997) Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput Applic Biosci 13:235–238

Robinson M, Gouy M, Gautier C, Mouchirod D (1998) Sensitivity of relative rates tests to taxonomic sampling. Mol Biol Evol 15:1091–1098

Sanderson MJ (1997) A nonparamteric approach to estimating divergence times in the absence of rate constancy. J Mol Evol 14:1218–1231

Sarich VM, Wilson AC (1973) Generation time and genomic evolution in primates. Science 179:1144–1147

Scherer S (1989) The relative-rate test of the molecular clock hypothesis: a note of caution. Mol Biol Evol 6:436–441

Tajima F (1993) Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135:599–607

Wang DY-C, Kumar S, Hedges SB (1999) Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. Proc Roy Soc Lond B 266:163–171

Wu C-I, Li W-H (1985) Evidence for higher rates of nucleotide substitutions in rodents than in man. Proc Natl Acad Sci USA 82: 1741–1745

Yang Z, Nielsen R (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J Mol Evol 46:409–418