J Mol Evol (1997) 44:112–119

JOURNAL OF MOLECULAR EVOLUTION

Estimating the Transition/Transversion Ratio from Independent Pairwise Comparisons with an Assumed Phylogeny

Andy Purvis,* Lindell Bromham

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, United Kingdom

Received: 22 August 1995 / Accepted: 26 July 1996

Abstract. A method is presented for estimating the transition/transversion ratio (TI/TV), based on phylogenetically independent comparisons. TI/TV is a parameter of some models used in phylogeny estimation intended to reflect the fact that nucleotide substitutions are not all equally likely. Previous attempts to estimate TI/TV have commonly faced three problems: (1) few taxa; (2) nonindependence among pairwise comparisons; and (3) multiple hits make the apparent TI/TV between two sequences decrease over time since their divergence, giving a misleading impression of relative substitution probabilities. We have made use of the time dependency, modeling how the observed TI/TV changes over time and extrapolating to estimate the ''instantaneous'' TI/TV—the relevant parameter for phylogenetic inference. To illustrate our method, TI/TV was estimated for two mammalian mitochondrial genes. For 26 pairs of cytochrome *b* sequences, the estimate of TI/TV was 5.5; 16 pairs of 12s rRNA yielded an estimate of 9.5. These estimates are higher than those given by the maximum likelihood method and than those obtained by averaging all possible pairwise comparisons (with or without a two-parameter correction for multiple substitutions). We discuss strengths, weaknesses, and further uses of our method.

Key words: Transition/transversion ratio — Independent comparisons — Phylogeny — Cytochrome b — 12s rRNA — Mammals

* *Present address:* Department of Biology, Imperial College, Silwood Park, Ascot, SL5 7PY, United Kingdom
Correspondence to: A. Purvis

Introduction

The use of sequence data to reconstruct phylogenies requires a model of sequence evolution (Felsenstein 1988) which, by necessity, simplifies complex biological processes. Models can most obviously be improved by increasing their complexity to more accurately reflect evolutionary processes. However, improvements can also be made in the way current models operate by determining the parameter values that maximize the fit between model and reality. Determining an appropriate value for the transition/transversion ratio (TI/TV) is an example of such an improvement.

Transitions (a change from one purine [A, G] to the other, or one pyrimidine [T, C] to the other) occur more frequently than transversions (purine to pyrimidine or vice versa), even though for any given nucleotide position there are twice as many possible transversions as transitions. The transition-to-transversion ratio (TI/TV) is an important aspect of models of sequence evolution because it expresses the relative probabilities of different types of nucleotide changes and thus is needed in order to correct measures of genetic distance, to weight character changes in parsimony, or for inclusion in a maximum likelihood model (see Wakeley 1996 for a recent review). Even though the choice of tree can be affected by the value of TI/TV that is used, the value is rarely specified in published molecular phylogenies, suggesting that the default value for the program may have been used (e.g., 2 in PHYLIP: Felsenstein 1993).

Empirical estimation of TI/TV is complicated by the way the observed TI/TV between two sequences depends upon divergence time: Transitions occur more frequently but with time are obscured by multiple hits until, at the

limit, no transition bias can be observed (Wakeley 1996). The value of TI/TV that is used should ideally reflect the evolutionary process—the relative frequencies with which transitions and transversions actually occur—rather than the pattern of differences that are seen between sequences. We are therefore aiming to estimate the ratio of transitions to transversions occurring at any instant in time. In order to do this, we make use of the way observed TI/TV changes over time. We use nonlinear regression of observed TI/TV for independent pairs of mitochondrial gene sequences against estimated divergence time to predict an instantaneous TI/TV of 5.5 for cytochrome *b* and 9.5 for 12s rRNA. By comparison, the means of all pairwise comparisons are 1.2 and 2.5, respectively. Correcting for multiple substitutions (Kimura two-parameter model) changes the estimates to 1.6 and 1.9, whereas the maximum likelihood estimates lie between 2 and 3 for both genes.

Methods

We used complete or nearly complete (at least 800 bp) mammalian cytochrome *b* and 12s rRNA sequences available from the GenBank database. These are the most widely sequenced genes for mammals. We located suitable cyt *b* sequences from 88 species and 12s sequences from 56. Our approach requires an independently derived estimate of the phylogeny of these species, including dates of splits. We compiled an estimate from the literature (Fig. 1). Where more than one estimate was available for the age of a node, we used the mean, as we had no objective means of judging which estimates were more accurate. As Table 1 shows, estimates often differed considerably. We were unable to date some of the nodes. Our phylogeny, especially the timescale, is doubtless incorrect in some details: We return to this point in the Discussion.

The timescale enables us to plot the observed TI/TV between two sequences against the time (*t*) since they diverged. The phylogeny, inasmuch as it is correct, lets us ensure that our observations of TI/TV are mutually independent (Felsenstein 1985). If pairs of species whose sequences are being compared are linked by lines drawn on the phylogeny, the pairs are independent if the lines neither meet nor cross (Felsenstein 1985:13; Burt 1989). There are very many ways of arranging the species into independent pairwise comparisons and no objective way to judge which is best (Burt 1989). We only made comparisons between species for which we had an estimate of time since divergence. Within this constraint, we chose pairs so as to maximize the number of independent comparisons. As well as the inherent advantages of large samples, this procedure maximizes the number of TI/TV values from recent divergences. Given that we are extrapolating from the observations to estimate TI/TV at $t = 0$, recent divergences are likely to yield the most pertinent information. Because of polytomies and undated nodes in our estimate of phylogeny, there was often a choice of species to compare: In such cases, we chose at random. We were left with 26 independent pairs of species for cyt *b* and 16 pairs for 12s.

Sequences were obtained from GenBank and aligned by eye (alignments available on request). Some regions of 12s showed too much variation or too many insertions and deletions for confident alignment, and so were excluded (around 10% of the data set). The observed TI/TV for each pair was computed by MEGA (Kumar et al. 1993); the values are in Table 1.

Figure 2A and B shows the plot of the proportion of differences between sequences that are transversions against *t* for cyt *b* and 12s, respectively. We fitted a negative exponential curve to these data in order to estimate the instantaneous transversion proportion (given by the y-intercept). The proportion of observed nucleotide changes that are transversions, *p*, rises from this value, *a*, to an asymptote (*s*). The equation therefore has the form:

$$p = a + (s - a)\,(1 - e^{-kt}) \tag{1}$$

where *k* is a constant indicating the speed with which the curve approaches the asymptote. The value of the asymptote (*s*) is the expected proportion of transversions in random sequences and varies with the base frequencies as follows:

$$s = \frac{(\mathrm{T} + \mathrm{C})(\mathrm{A} + \mathrm{G})}{\mathrm{A}^*\mathrm{G} + \mathrm{C}^*\mathrm{G} + (\mathrm{T} + \mathrm{C})(\mathrm{A} + \mathrm{G})} \tag{2}$$

For both cyt *b* and 12s the calculated value of *s* was 0.67.

The variance of *p* in equation (1) depends on its true value and on the total number of transitions and transversions according to:

$$V(p) = \frac{p(1 - p)}{n} \tag{3}$$

where n = total number of differences apparent between the two sequences. Note that the sequence length is not directly relevant, because we are interested in the proportion of changes that are transitions rather than exactly how many transitions there have been. As a further complication, the *p*'s in equation (3) are the ''true'' values, not subject to sampling error as observed *p*'s are. We used the SPSS statistics package (SPSS, release 4) to perform iterative-weighted least-squares regression. We used the observed *p*'s to calculate initial weights (weighting by the reciprocal of the square root of the variance given by equation 3). After the first regression, new weights were computed from the predicted *p*'s. The cycle of regression and reweighting continued until the parameter estimates converged.

To test whether the two genes had significantly different y-intercepts, we analyzed the two genes together. We introduced a dummy variable, *g*, into equation (1) to separate the genes and tested whether its coefficient, *c*, differed significantly from zero in:

$$p = a + c.g + (s - (a + c.g))\,(1 - e^{-kt}) \tag{4}$$

To place our results in context, we also estimated TI/TV by three other methods discussed in Wakeley's (1996) review. First, we found the average observed TI/TV across all possible pairwise comparisons for each gene. Second, we used MEGA's implementation of Kimura's (1980) two-parameter model to correct the observed TI/TV for multiple substitutions before averaging all possible pairwise comparisons. Finally, we analyzed the data set using FastDNAml (Olsen et al. 1994), varying TI/TV from 0.5 to 10 in increments of 0.5 to see which value gave the highest likelihood.

Results

The regressions give values for *a* of 0.154 for cyt *b* and 0.0949 for 12s (see Table 2 for confidence intervals and regression statistics). These values correspond to a TI/TV of 5.5 for cyt *b* and 9.5 for 12s. The different codon positions of cyt *b* all give similar estimates of *a*. Although the confidence intervals for 12s are very wide, the usual default value of 2 lies below the confidence interval for both genes. The confidence intervals all overlap, and the statistical test described above indicates that the

Table 1. Independent pairs of cytochrome *b* and 12s rRNA sequences[a]

Gene	Order	Taxon 1	Taxon 2	Div. time (MY)
cyt b	Marsupialia	*Didelphus virginiana*	*Monodelphus brevicaudata*	40
		Planigale gilesi	*Planigale ingrami*	2
		Sminthopsis murina	*Planigale maculata*	15
	Primates	*Pan paniscus*	*Pan troglodytes*	2.35
		Homo sapiens	*Gorilla gorilla*	8.29
	Primates/Chiroptera	*Pongo pygmaeus*	*Chiroderma salvini*	90
	Carnivora	*Panthera tigris*	*Panthera leo*	1.2, 1.9
	Pinnipedia/Carnivora	*Odobenus rosmarus*	*Thalarctos maritimus*	39
	Pinnipedia	*Leptonychotes weddelli*	*Phoca groenlandica*	15
	Cetacea	*Physeter catodon*	*Stenella longirostris*	30, 35
		Balaenoptera bonaerensis	*Balaena mysticetus*	20, 32
	Proboscidea/Sirenia	*Loxodonta africana*	*Dugong dugon*	70
	Perissodactyla	*Equus caballus*	*Equus grevyi*	3.8
	Perissodactyla/Artiodactyla	*Diceros bicornis*	*Tayassu tajacu*	85
	Artiodactyla	*Dama dama*	*Cervus nippon*	4, 4.5
		Odocoileus hemionus	*Antilocapra americana*	19.4, 28
		Bos jarvanicus	*Bos taurus*	0.7
		Nemorhaedus caudatus	*Ovis aries*	1.65, 4.5, 6.1
		Giraffa camelopardalis	*Tragulus javanicus*	38, 42
		Camelus dromedarius	*Lama guanicoe*	11, 17
	Rodentia	*Sciurus alberti*	*Sciurus niger*	3
		Cratogeomys tylorhinus	*Cratogeomys gymnurus*	0.7
		Cratogeomys merriami	*Cratogeomys castanops*	3
		Cavia porcellus	*Hystrix africaeaustralis*	46
		Rattus norvegicus	*Mus musculus*	21, 12
	Rodentia/Lagomorpha	*Geomys bursarius*	*Oryctolagus cuniculus*	68
12s	Insectivora	*Atelerix albiventris*	*Blarina brevicauda*	43
	Primates	*Pan paniscus*	*Pan troglodytes*	2.35
		Homo sapiens	*Gorilla gorilla*	8.29
	Primates/Pinnipedia	*Pongo pygmaeus*	*Phoca vitulina*	97
	Cetacea	*Stenella coeruleoalba*	*Balaenoptera musculus*	32, 42
	Artiodactyla	*Bos taurus*	*Tragelaphus imberbis*	7, 15, 19
		Damiliscus dorcas	*Aepyceros melampus*	6, 7.5
		Cephalophus maxwelli	*Kobus ellipsiprymnus*	13
		Gazella thomsoni	*Madoqua kirki*	9, 11.2
		Capra hircus	*Oryx gazella*	17, 23, 25
		Cervus unicolor	*Muntiacus reevesi*	9
		Odocoileus virginianus	*Hydropotes inermis*	20
		Antilocapra americana	*Tragulus napu*	38, 42
	Rodentia	*Cavia porcellus*	*Hystrix africaeaustalis*	46
		Mus musculus	*Rattus norvegicus*	16.5
		Proechimys longicaudatus	*Thryonomys swinderianus*	28

[a] Binomial names taken from the published sequences. Order-level taxonomy after Corbet and Hill (1991). Observed *p* is the proportion of all nucleotide changes that are transversions. The sources of divergence times were as follows: 1. Reig et al. 1987; 2. Baverstock et al. 1982; 3. Archer 1984; 4. Purvis 1995; 5. Novacek 1992; 6. Wayne et al. 1989; 7. Garland and Janis 1993; 8. Macdonald 1984; 9. Carrol 1989; 10. Forstèn 1992; 11. C.M. Janis, personal communication; 12. Hartl et al. 1988; 13. Hafner et al. 1989; 14. Randi et al. 1991; 15. Garland et al. 1993; 16. Stanley et al. 1994; 17. Hafner 1984; 18. Honeycutt and Williams 1982; 19. Sarich 1986; 20. Brownell 1983; 21. Catzeflis et al. 1992; 22. Kingdon 1982; 23. Georgiadis et al. 1990. The divergence time for the *Leptonychotes weddelli/Phoca groenlandica* comparison is an estimate of the upper bound

estimates of *a* for the two genes are not significantly different.

The results from our method are in marked contrast with those given by other approaches. The mean TI/TV of all possible pairwise comparisons is 1.18 ($n = 1327$, SE = 0.03) for cyt *b* and 2.49 ($n = 497$, SE = 0.08) for 12s. Correcting for multiple substitutions change these means to 1.57 (SE = 0.09) and 1.86 (SE = 0.05), respectively. Figure 3A and B shows how the log likelihood varies with TI/TV for cyt *b* and 12s. For cyt *b*, the peak is at around 2.5; for 12s, it is in the region of 2.

Discussion

The method we present has two major advantages: The comparisons it uses are independent, and it controls for the dependence of TI/TV on divergence time. It provides

Table 1. Continued

Avg. div. time (MY)	Observed TS:TV	Observed *P*	Nucleotides compared	Div. time source
40	1.086	0.479	1,140	1
2	2.316	0.302	926	2, 3
15	1.707	0.369	926	2, 3
2.35	18.667	0.051	1,140	4
8.29	7.000	0.125	1,140	4
90	1.057	0.486	1,140	5
1.6	13.375	0.070	1,140	6, 7
39	1.812	0.356	1,140	6
15	4.917	0.169	1,140	6
33	2.356	0.298	1,140	8, 9
26	7.538	0.117	1,140	8, 9
70	1.351	0.425	1,137	5
3.8	12.571	0.074	1,140	10
85	1.040	0.490	1,140	5
4.3	3.028	0.248	1,140	7, 11
23.7	1.818	0.355	1,140	7, 8
0.7	8.636	0.104	1,140	12
4.1	7.647	0.116	1,140	12, 7, 14
40	1.526	0.396	1,140	15, 8
14	3.789	0.209	1,140	7, 16
3	4.382	0.186	1,134	17
0.7	3.350	0.230	1,140	18
3	2.892	0.257	1,140	18
46	1.066	0.484	1,140	19
16.5	1.179	0.459	1,116	20, 21
68	1.013	0.497	1,140	5
43	1.09	0.479	864	9
2.35	12.00	0.077	870	4
8.29	10.67	0.086	867	4
97	1.26	0.443	857	5
37	4.80	0.172	872	8, 9
13.67	5.40	0.156	869	7, 15, 22
6.75	5.27	0.186	872	7, 22
13	8.20	0.248	879	22
10.1	6.67	0.109	872	7, 23
21.67	5.89	0.145	879	7, 22
9	10.33	0.088	880	11
20	4.33	0.188	878	11
40	2.15	0.317	873	15, 8
46	1.39	0.418	717	19
16.5	1.88	0.348	802	20, 21
28	0.89	0.470	692	8

an empirical estimate of TI/TV which may be contrasted to the estimates derived by other commonly used methods. In addition, the parameters of the regression equation can provide information on the relative rates of sequence evolution, which may vary both among genes and among taxa.

Nonindependence of comparisons causes the confidence intervals associated with parameter estimates to be too narrow, with an associated increase in the type I error rate when testing hypotheses (Felsenstein 1985). Additionally, when all pairwise comparisons among sequences are averaged to estimate TI/TV, branches near the base of the phylogeny are given too much weight because they are included in disproportionately many pairwise comparisons (Felenstein 1992). The dependence of TI/TV on divergence time will thwart any attempt to estimate TI/TV that does not explictly model the effect of time since divergence. It is therefore unsurprising that the mean observed TI/TV of all possible comparisons is an underestimate of the true TI/TV. More surprisingly, the two-parameter model fails to correct adequately for multiple substitutions and maximum likelihood also underestimates TI/TV, findings that merit further investigation. If only independent comparisons are used, the two-parameter model produces an estimate of TI/TV for cyt *b* closer to that yielded by our method (mean = 4.99, SE = 0.96), but the 12s estimate was again very low (mean = 2.41, SE = 0.31).

The reliance on an independently derived phylogeny and data estimates is an obvious limitation of our approach. First, this mode of analysis can be applied only when a plausible phylogeny of at least some of the taxa

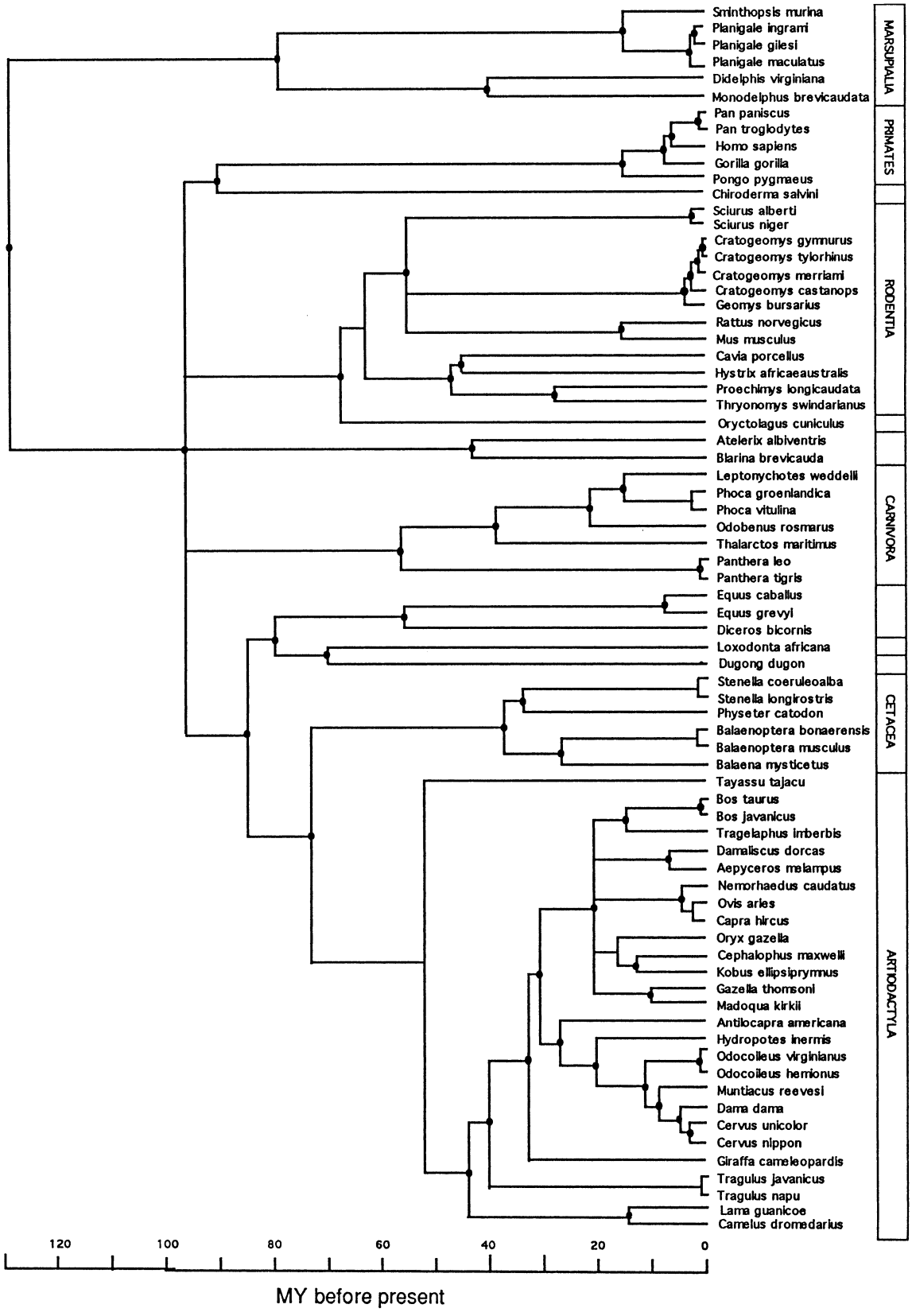

Fig. 1. Assumed phylogeny of the species in this study. Binomial names taken from the published sequences. The *bar down the right-hand side* is divided into orders; not all orders are shown. Order-level taxonomy after Corbet and Hill (1991). Divergence time (million years before present) indicated by *scale bar*. Nodes for which dates were available are marked with a *solid circle* (●). Topology and divergence time estimates came from the following sources: Allard et al. 1992; Archer 1984; Baverstock et al. 1982; Brownell 1983; Butler 1988; Carroll 1989; Catzeflis et al. 1989, 1992; Eisenberg 1981; Forstén 1992; Garland et al. 1993; Garland and Janis 1993; Georgiadis et al. 1990; Geraads 1992; Hafner 1984; Hartl et al. 1988, 1990; Honeycutt and Williams 1982; Janis 1988; Janis personal communication; Kingdon 1982; Macdonald 1984; Novacek 1992; Purvis 1995; Randi et al. 1991; Reig et al. 1987; Sarich 1986; She et al. 1990; Stanley et al. 1994; Wayne et al. 1989.

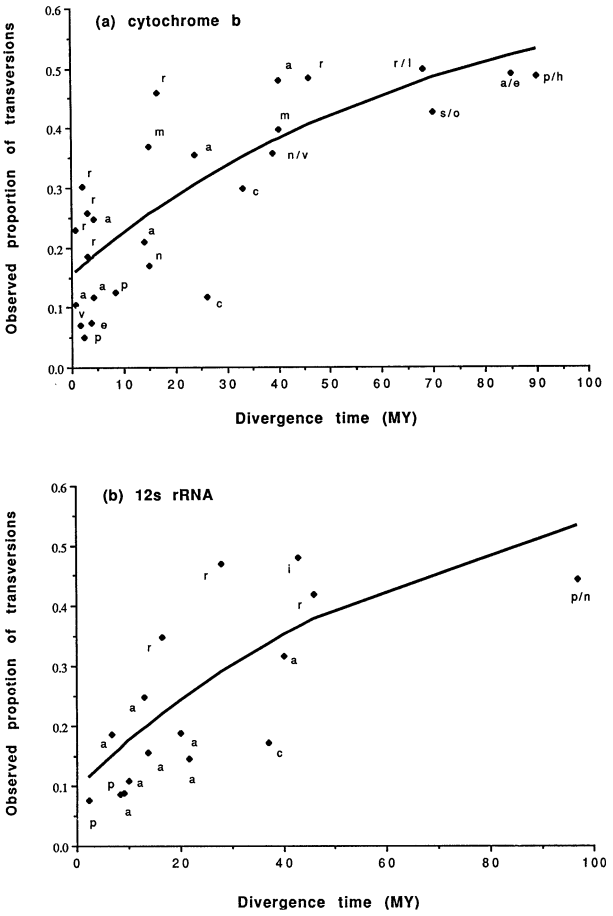

Fig. 2. Observed proportion of transversions (TV/TV + TS) against divergence time for pairs of sequences for **(A)** cytochrome *b* and **(B)** 12s rRNA. *Solid line* represents weighted nonlinear regression equation fitted to the data. Points are identified by order: *a* = Artiodactyla, *c* = Cetacea, *e* = Perissodactyla, *h* = Chiroptera, *i* = Insectivora, *l* = Lagomorpha, *m* = Marsupialia, *n* = Pinnipedia, *o* = Proboscidea, *p* = Primates, *r* = Rodentia, *s* = Sirenia, *v* = Carnivora.

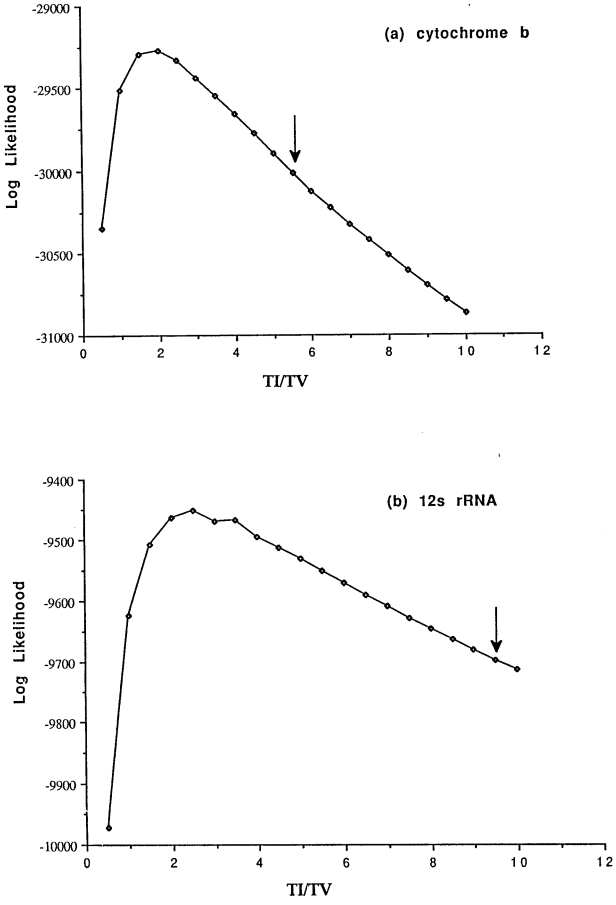

Fig. 3. Log likelihood for a range of values of the proportion of transversions for **(A)** cytochrome *b* and **(B)** 12s rRNA. *Arrows* indicate the estimates of instantaneous proportion of transversions obtained by our method.

Table 2. Regression statistics: estimates of instantaneous proportion of transversions (*a*), decay coefficient (*k*), 95% confidence intervals (CIs), and R^2

Sequence	Parameter	Estimate	95% C.I.s	R^2
cyt b (whole)	*a*	0.1540	0.0941–0.2140	0.652
	k	0.0146	0.0090–0.0203	
cyt b (1st pos.)	*a*	0.1142	0.0310–0.1781	0.584
	k	0.0114	0.0066–0.0163	
cyt b (2nd pos.)	*a*	0.1167	0.0528–0.1806	0.325
	k	0.0052	0.0018–0.0086	
cyt b (3rd pos.)	*a*	0.1536	0.0876–0.2197	0.676
	k	0.0185	0.0111–0.0259	
12s	*a*	0.0949	−0.0213–0.2110	0.557
	k	0.0149	0.0058–0.0240	

is available from other lines of evidence. Second, if the true phylogeny differs markedly from the estimated phylogeny used in the analysis, the results could be meaningless. (Note that this argument cannot be used to justify nonphylogenetic approaches: Pagel and Harvey 1992.) Examination of the residuals from the regressions provides some indication of the worth of our assumed phylogeny. If the phylogeny were wildly inaccurate, we should expect either a few marked outliers or a generally very poor fit to the model. The absence of obvious outliers in Fig. 2 (no standardized residuals greatly exceeded 2) and the reasonable R^2 values suggest that the phylogeny is adequate for our purposes.

An alternative approach to ours would be to compare only very recently diverged sequences: The estimate of instantaneous TI/TV might be either the highest of all the observed values (e.g., Hafner et al. 1994) or an average for a number of taxa (e.g., Reeder 1995). However, such closely related sequences will not have accumulated many differences, so the sampling error associated with TI/TV will be high (equation 3; see also Wakeley 1996). The highest observed TI/TV in our cyt *b* data set was over 18, but we are fairly confident that the true value is lower than this.

It is clear from Fig. 2 and Table 2 that the confidence intervals associated with the 12s data set are wider than those for cyt *b*. The difference is too great to be ascribed to effects of sequence length or number of taxa. One

reason may be that the 12s data set has fewer points between 0 MY and 5 MY: Estimating the instantaneous TI/TV therefore involves more of an extrapolation beyond the data than for cyt *b*. Also, inspection of the residuals suggests that the model in equation (1) fits 12s evolution less well than it fits cyt *b*, because the weighting procedure has not removed the heterogeneity of variance. One assumption of our model that is likely to be more nearly met by cyt *b* than by 12s is that substitutions are independent. Concerted evolution of paired sites (Wheeler and Honeycutt 1988) violates this assumption and may affect the weights used in the regression. Sequence evolution is also likely to differ between these two genes because cyt *b* is a protein-coding gene whereas 12s codes for an rRNA molecule.

Although our aim has been to estimate TI/TV, the regressions also provide information on rates of evolution. The decay coefficient, *k*, describes the rate at which the proportion of transversions drops from its initial value to the asymptote: It is a measure of how rapidly sequences saturate, and so can be used to compare evolutionary rates of, for instance, different codon positions or different lineages. The similarity between the *k* estimates for cyt *b* and 12s may indicate a similar average rate of evolution. The variation in *k* among codon positions of cyt *b* follows the expected pattern (Table 2); it is highest for third bases, where only a third of the possible base changes cause amino acid replacements. We would expect this effect to be even stronger in a comparison between nondegenerate, twofold degenerate, and fourfold degenerate sites.

Our approach also facilitates testing whether parameters vary among lineages. Inspection of Fig. 2 shows that, for instance, all the rodent comparisons lie above the line and all primate comparisons below. This could be because TI/TV varies among lineages or because the rate of approach to the asymptote varies among lineages, or both. An obvious possible correlate of the variation is body size (or some factor, like generation time, that covaries strongly with body size: Bromham et al. 1996). In a series of tests analogous to the test for homogeneity of *a* among genes, we split the comparisons into two groups: small-bodied (average < 3 kg) and large-bodied. When *a* was allowed to differ between the two groups (*k* held constant), the difference was significant for both genes ($P < 0.001$). Similarly, *k* differed significantly between groups ($P < 0.005$ for each gene) when *k* but not *a* was allowed to differ. When both *a* and *k* were allowed to vary between groups, *a* differed between groups only for cyt *b* ($P < 0.001$), and *k* showed no significant differences. The interpretation of these results is that at least one model parameter varies among lineages for each gene. For cyt *b*, the evidence suggests that it is *a* that varies, but our data do not allow us to tell whether the variation for 12s is in *a* or in *k*. Such heterogeneity in model parameters of course suggests that a model using a single value may not adequately reflect reality. Additional parameters could be introduced into the model to capture further complexities: As always, when using models, there is a tradeoff between generality and realism.

Phylogenetically independent comparisons are a valuable tool both for assessing the suitability of models and for estimating parameters in the model of choice. The method of independent pairwise comparisons holds promise for other similar problems in molecular evolutionary biology, such as estimating the relative frequencies of different nucleotide substitutions and determining the fraction of invariant sites in a gene.

Acknowledgments. We thank Eddie Holmes, Alan Grafen, Paul Harvey, Mark Pagel, Robin McCleery, and Andrew Rambaut for discussion help, and advice, and thank two anonymous referees for their very helpful comments on an earlier draft. This work was supported by the Natural Environment Research Council, U.K. (GR3/8515), the Rhodes Trust, and TechRentals.

References

Allard MW, Miyamoto MM, Jarecki L, Kraus F, Tennant MR (1992) DNA systematics and evolution of the artiodactyl family Bovidae. Proc Natl Acad Sci USA 89:3972–3976

Archer M (1984) The Australian marsupial radiation. In: Archer M, Clayton G (eds) Vertebrate zoogeography and evolution in Australia. Hesperian Press, Perth, pp 633–809

Baverstock PR, Archer M, Adams M, Richardson BJ (1982) Genetic relationships among 32 species of Australian dasyurid marsupials. In: Arch M (ed) Carnivorous marsupials. Royal Zoological Society of New South Wales, Sydney, pp 641–650

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

Brownell E (1983) DNA/DNA hybridization studies of muroid rodents: symmetry and rates of molecular evolution. Evolution 37:1034–1051

Burt A (1989) Comparative methods using phylogenetically independent contrasts. Oxf Surv Evol Biol 6:33–53

Butler PM (1988) Phylogeny of the insectivores. In: Benton MJ (ed) The phylogeny and classification of tetrapods, volume 2: mammals. Clarendon Press, Oxford, pp 117–142

Carroll RC (1989) Vertebrate paleontology and evolution. WH Freeman, New York

Catzeflis FM, Nevo E, Ahlquist JE, Sibley CG (1989) Relationships of the chromosomal species in the Eurasian mole rates of the *Spalax ehrenbergi* group as determined by DNA-DNA hybridization, and an estimate of the spalacid-murid divergence time. J Mol Evol 29:223–232

Catzeflis FM, Aguilar J-P, Jaeger J-J (1992) Muroid rodents: phylogeny and evolution. Trends Ecol Evol 7:122–126

Corbet GB, Hill JE (1991) A world list of mammalian species. Oxford University Press, Oxford

Eisenberg JF (1981) The mammalian radiations. The Athlone Press, London

Felsenstein J (1985) Phylogenies and the comparative method. Am Nat 125:1–15

Felsenstein J (1988) Phylogenies from molecular data: inference and reliability. Annu Rev Genet 22:521–565

Felsenstein J (1992) Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. Genet Res Camb 59:139–147

Felsenstein J (1993) PHYLIP. Version 3.5c. University of Washington, Seattle

Forstén A (1992) Mitochondrial-DNA time-table and the evolution of *Equus:* comparison of molecular and paleontological evidence. Ann Zool Fennici 28:301–309

Garland TJ, Janis CM (1993) Does metatarsal/femur ratio predict maximal running speed in cursorial mammals? J Zool Lond 229:133–151

Garland TJ, Dickerman AW, Janis CW, Jones JA (1993) Phylogenetic analysis of covariance by computer simulation. Syst Biol 42:265–292

Georgiadis NJ, Kat PW, Oketch H, Patton J (1990) Allozyme divergence within the Bovidae. Evolution 44:2135–2149

Geraads D (1992) Phylogenetic analysis of the tribe Bovini (Mammalia: Artiodactyla). Zool J Linn Soc 104:193–207

Hafner DJ (1984) Evolutionary relationships of the Nearctic Sciuridae. In: Murie JO, Michener GR (eds) The biology of ground-dwelling squirrels: annual cycles, behavioural ecology, and sociality. pp 3–23

Hafner MS, Sudman PD, Villablanca FX, Spradling TA, Demastes JW, Nadler SA (1994) Disparate rates of molecular evolution in cospeciating hosts and parasites. Science 265:1087–1090

Hartl GB, Göltenboth R, Grillitsch M, Willing R (1988) On the biochemical systematics of the Bovini. Biochem Syst Ecol 16:575–579

Hartl GB, Burger H, Willing R, Suchentrunk F (1990) On the biochemical systematics of the Caprini and Rupicaprini. Biochem Syst Ecol 18:175–182

Honeycutt RL, Williams SL (1982) Genic differentiation in pocket gophers of the genus *Pappogeomys* with comments on intergeneric relationships in the subfamily Geomyinae. J Mamm 63:208–217

Janis CM (1988) New ideas in ungulate phylogeny and evolution. Trends Ecol Evol 3:291–297

Kimura M (1980) A simple method for estimating evolutionary rates of base substitution through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kingdon J (1982) East African mammals: an atlas of evolution. Academic Press, London

Kumar S, Tamura K, Nei M (1993) MEGA: molecular evolutionary genetics analysis. The Pennsylvania State University, University Park

Li WH, Gouy M, Sharp PM, O'hUigin C, Yang YY (1990) Molecular phylogeny of Rodentia, Lagomorpha, Primates, Artiodactyla and Carnivora and molecular clocks. PNAS 87:6703–6707

Macdonald DW (ed) (1984) The encyclopaedia of mammals. Unwin Hyman, London

Novacek MJ (1992) Mammalian phylogeny: shaking the tree. Nature 356:121–125

Olsen GJ, Matsuda H, Hagstrom R, Overbeek R (1994) Fast DNAml—a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput Appl Biosci 10:41–48

Pagel MD, Harvey PH (1992) On solving the correct problem: wishing does not make it so. J Theor Biol 156:425–430

Purvis A (1995) A composite estimate of primate phylogeny. Philos Trans R Soc Lond Biol 348:405–421

Randi E, Fusco G, Lorenzini R, Tosco S, Tosi G (1991) Allozyme divergence and phylogenetic relationships among *Capra, Ovis* and *Rupicapra* (Artiodactyla, Bovidae). Heredity 67:281–286

Reeder TW (1995) Phylogenetic relationships among phrynosomatid lizards as inferred from mitochrondrial ribosomal DNA sequences: substitutional bias and information content of transitions relative to transversions. Mol Phylogenet Evol 4:203–222

Reig OA, Kirsch JAW, Marshall LG (1987) Systematic relationships of the living and Neocenozoic American ''opossum-like'' marsupials (suborder Didelphimorphia), with comments on the classification of these and of the Cretaceous and Paleogene New World and European metatherians. In: Archer M (ed) Possums and opossums: studies in evolution. Surrey Beatty and the Royal Zoological Society of New South Wales, Sydney, pp 1–89

Sarich VM (1986) Rodent macromolecular systematics. In: Luckett WP, Hartenberger J-L (eds) Evolutionary relationships among rodents; a multidisciplinary analysis. pp 423–452

She JX, Bonhomme F, Boursot P, Thaler L, Catzeflis F (1990) Molecular phylogenies in the genus *Mus:* comparative analysis of electrophoretic, scnDNA hybridization, and mtDNA RFLP data. Biol J Linn Soc 41:83–103

Stanley HF, Kadwell M, Wheeler JC (1994) Molecular evolution of the family Camelidae: a mitochondrial DNA study. Proc R Soc Lond [Biol] 256:1–6

Wakeley J (1996) The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. Trends Ecol Evol 11:158–163

Wayne RK, Benveniste RE, Janczewski DN, O'Brien SJ (1989) Molecular and biochemical evolution of the Carnivora. In: Gittleman JL (ed) Carnivore behavior, ecology, and evolution. Chapman and Hall, London, pp 465–494

Wheeler WC, Honeycutt RL (1988) Paired sequence difference in ribosomal RNAs—evolutionary and phylogenetic implications. Mol Biol Evol 5:90–96