Sensitivity of the Relative-Rate Test to Taxonomic Sampling

Marc Robinson,[1] *Manolo Gouy, Christian Gautier, and Dominique Mouchiroud*

UMR CNRS 5558—''Biométrie, Génétique et Biologie des Populations,'' Université Claude Bernard, Lyon, France

Relative-rate tests may be used to compare substitution rates between more than two sequences, which yields two main questions: What influence does the number of sequences have on relative-rate tests and what is the influence of the sampling strategy as characterized by the phylogenetic relationships between sequences? Using both simulations and analysis of real data from murids (APRT and LCAT nuclear genes), we show that comparing large numbers of species significantly improves the power of the test. This effect is stronger if species are more distantly related. On the other hand, it appears to be less rewarding to increase outgroup sampling than to use the single nearest outgroup sequence. Rates may be compared between paraphyletic ingroups and using paraphyletic outgroups, but unbalanced taxonomic sampling can bias the test. We present a simple phylogenetic weighting scheme which takes taxonomic sampling into account and significantly improves the relative-rate test in cases of unbalanced sampling. The answers are thus: (1) large taxonomic sampling of compared groups improves relative-rate tests, (2) sampling many outgroups does not bring significant improvement, (3) the only constraint on sampling strategy is that the outgroup be valid, and (4) results are more accurate when phylogenetic relationships between the investigated sequences are taken into account. Given current limitations of the maximum-likelihood and nonparametric approaches, the relative-rate test generalized to any number of species with phylogenetic weighting appears to be the most general test available to compare rates between lineages.

Introduction

Substitution rates between two species are routinely compared using relative-rate tests (Sarich and Wilson 1973; Wu and Li 1985). Since we have no direct knowledge of the ancestor sequence of the two modern sequences and little knowledge of their time of divergence, evolutionary rates cannot be compared directly. Instead, the basic idea of the relative-rate tests is to compare distances between each of the two species and a reference outgroup.

Several authors have proposed relative-rate tests using more than three sequences (Li and Bousquet 1992; Takezaki, Rzhetsky, and Nei 1995). It is thus possible to compare the substitution rates between two lineages, each consisting of many sequences, using another lineage of many sequences as an outgroup. But what are the advantages of these tests, compared with a simple relative-rate test on three sequences?

Clearly, to detect branches that depart from the clock in a phylogeny (Takezaki, Rzhetsky, and Nei 1995), all sequences used in the phylogeny must be included in the test. But the answer is less obvious concerning the classical relative-rate test: do rates differ between two lineages? Intuitively, it seems probable that taking into account more information should always improve results. But how is this influence characterized? Is the gain from a more complete taxonomic sampling significant? Does sampling of the ingroups and the outgroups play the same role? What bias can unbalanced sampling introduce?

Concerning the last question, in previously presented generalizations of the relative-rate test to the comparison of groups of sequences (Li and Bousquet 1992; Takezaki, Rzhetsky and Nei 1995), mean differences between groups of sequences are computed without taking into account phylogenetic relations between them. If the tree is unbalanced, this leads to overweighting highly represented sublineages. We propose a simple weighting scheme to take this into account. We also study possible cases of paraphyly.

We considered the two published approaches to the comparison of rates between two groups of more than one sequence per group, that of Li and Bousquet (1992), which is a straightforward generalization of the three-species test of Wu and Li (1985), and that of Takezaki, Rzhetsky, and Nei (1995). The latter test differs in using the method of Bulmer (1991) to estimate covariances without *a priori* knowledge of the phylogeny. The test of Li and Bousquet (1992) was extended to the use of several outgroup sequences, and the weighting scheme proposed here was implemented for both approaches.

Influence of sampling and importance of taxonomic weighting were investigated with simulation data and with two real data sets. Exons 2–6 of the LCAT gene have been sequenced for 20 rodents (Robinson et al. 1997), mostly murids, making it a good model for the influence of taxonomic sampling in a group represented in many studies by rat and mouse (see Li et al. 1996 for a review). In a recent study of APRT introns in closely related murids (Fieldhouse, Yazdani, and Golding 1997), three-species relative-rate tests gave contradictory results depending on the choice of outgroup among six species, thus providing a second example of the influence of taxonomic sampling on relative-rate tests.

Materials and Methods

Simulations

In each simulation, an ancestral sequence was randomly drawn, each of the four bases having a probabil-

[1] Present address: Department of Zoology, George S. Wise Faculty of Life Science, Tel Aviv University, Israel.

Key words: APRT, LCAT, molecular clock, murid, phylogeny, relative-rate test, sampling, substitution rate, topology, weighting.

Address for correspondence and reprints: Marc Robinson, Department of Zoology, George S. Wise Faculty of Life Science, Tel Aviv University, Ramat-Aviv 69978, Israel. E-mail: marc@kimura.tau.ac.il.

Mol. Biol. Evol. 15(9):1091–1098. 1998

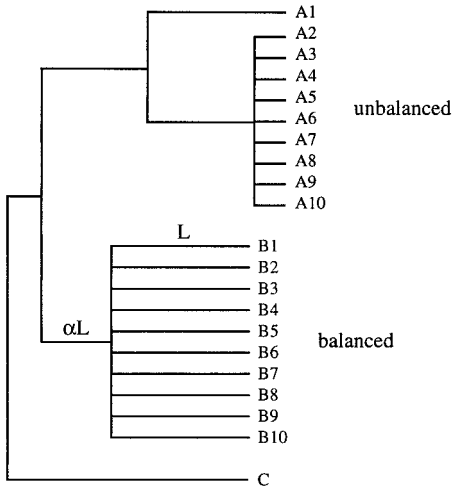

FIG. 1.—Example of a phylogeny used for simulations, in which one ingroup has unbalanced sampling and one has balanced sampling, while the outgroup is represented by only one sequence. α = ratio of the length of the internal branch (αL) to that of the external branches (L).

ity of occurence of 25%. Substitutions were then generated following the Jukes and Cantor (1969) one-parameter model. Parameters of the simulation were sequence length of 1,000 bp, mean distance to the root of 0.5 substitutions per site, a rate difference between the two compared lineages varying from 0 to 0.2 substitutions per site, and a ratio of lengths of the internal branch defining a group to length of the terminal branches varying from 0.1 to 10 (fig. 1). Moreover, strongly unbalanced topologies were used, either in an ingroup or in the outgroup, to study influence of topological weighting (fig. 1).

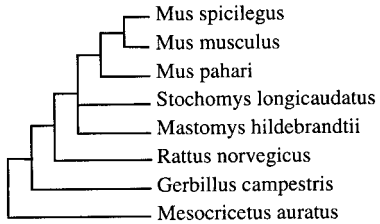

FIG. 3.—Phylogenetic relationships between eight murids for which APRT introns are known (modified from Fieldhouse, Yazdani and Golding 1997). Branch lengths are arbitrary. Accession numbers are X70112, U28961, U28720–U28723, M11310, L04970.

Murid Data Sets

The LCAT data set contains 205 aligned codons in 13 murids, 7 other rodents and 3 other eutherian mammals (fig. 2). Uncertainties about rodent and mammal phylogeny (Graur, Hide, and Li 1991; Luckett and Hartenberger 1993; Graur, Duret, and Gouy 1996) do not allow rooting. Murids are a clear monophyletic group, with the deepest branching separating two genera of subterranean murids (*Spalax* and *Rhizomys*) from the other murids (Robinson et al. 1997). Synonymous substitution rates were computed as in Li (1993) and Pamilo and Bianchi (1993).

The APRT data set contains four introns, of which three (I, II, and IV) are complete and form a data set of 351 aligned sites in the two ingroup species (*Mus musculus* and *Mus spicilegus*) and six outgroup species (fig. 3). Intron II is incomplete in one species (*Mastomys*

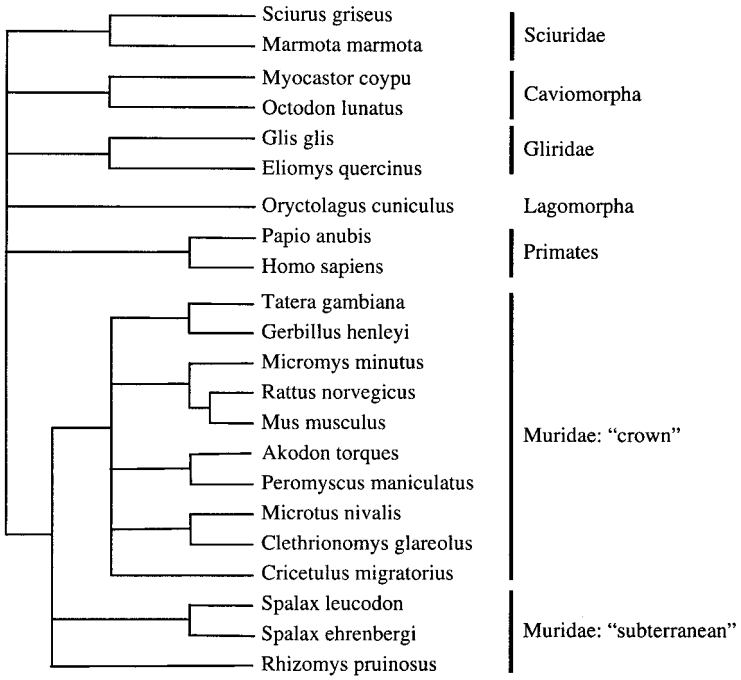

FIG. 2.—Current knowledge of phylogenetic relationships between 23 mammals, of which 13 are murids, for which LCAT coding sequences are known (modified from Robinson et al. 1997). Branch lengths are arbitrary. Accession numbers are X04981, L08633, J05154, D13668, X54096, U72293–U72326.

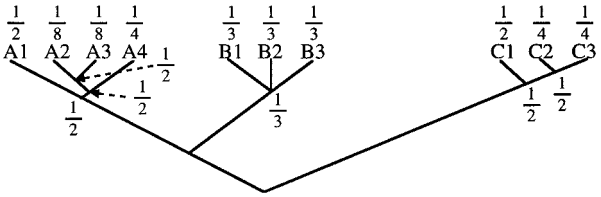

FIG. 4.—Example of topological weighting on a rooted tree when all groups under investigation (A, B, and C) are monophyletic to each other. Weights are indicated at each node and for each sequence.

hildebrandtii), which allows using a data set either of 486 sites in eight species, or of 897 sites in seven species.

For all relative-rate tests, we excluded all sites which contained a gap or an unresolved base in any of the sequences considered. Accession numbers to all sequences used in this study are available by anonymous ftp://biom3.univ-lyon1.fr/pub/datasets/MBE98/.

Resampling

To investigate the effect of species sampling, we computed the relative-rate test on all subsets of our data for each lineage. This was done by resampling all combinations of P sequences among N of a given lineage for all values of P between 1 and N and computing the relative-rate test with the other lineages complete. The influence of sequence relatedness was also investigated through the ratio of internal to external branches in simulations (fig. 1). For different values of this ratio inside a given group, the relative-rate test was computed for each sequence of the group. Thus, for each value of the branch length ratio, a distribution of N observed values (mean difference or standard deviation) was obtained and characterized by its standard deviation. The latter represents sensibility of the relative-rate test to sampling for a given branch length ratio.

All programs used in this work were written in ANSI C and are available by anonymous ftp://biom3.univ-lyon1.fr/pub/datasets/MBE98/.

Weighting Scheme

The basic idea of our weighting scheme is to take into account the topology of the phylogenetic tree between studied sequences (fig. 4). For this, we attribute a weight to each node of the tree, which is a function of the number of branches at this node: 1/(number of branches −1). This results in giving a weight of 0.5 to a dichotomous node and lower weights to polytomies. Each sequence being a terminal leaf of the tree, a ''topological weight'' p can be calculated as the product of the weights of the nodes separating a sequence from the root of its lineage (fig. 4). This results in three interesting properties: (1) for any monophyletic group of sequences (lineage), the sum of the topological weights is 1; (2) the same weight is given to sister lineages, whatever the number of sequences sampled for each of them; and (3) weights are stable to the addition of new sequences.

Resolution

In some methods of distance estimation, the number of useful sites for a given sequence differs depending on which other sequence it is compared to. This is the case with numbers of synonymous and nonsynonymous sites. To take this into account, we also weight by the number of sites l_{ij} compared between sequences i and j, as in Li and Bousquet (1992). The mean number of sites compared between two groups of sequences A and C, taking into account the phylogeny, is $\Sigma_{m=1}^{n_A} \Sigma_{n=1}^{n_C} p_m p_n 1_{mn}$. Thus, to each pair of sequences (A_i, C_j), we associate the final weight:

$$\pi_{ij} = p_i p_j \frac{l_{ij}}{\sum_{m=1}^{n_A} \sum_{n=1}^{n_C} p_m p_n l_{mn}}. \quad (1)$$

Since the sum of these weights for the $n_A n_C$ pairs is 1, the weighted mean distance between groups A and C is:

$$K_{AC} = \sum_{i=1}^{n_A} \sum_{j=1}^{n_C} \pi_{ij} K_{A_i C_j}. \quad (2)$$

The variance of this mean is the sum of the variances of $\pi_{ij} K_{A_i C_j}$ and of all the covariances between pairs of distances. As the variance of $\pi_{ij} K_{A_i C_j}$ is its covariance with itself, this simplifies to:

$$\mathrm{var}(K_{AC}) = \sum_{i=1}^{n_A} \sum_{j=1}^{n_C} \sum_{k=1}^{n_A} \sum_{r=1}^{n_C} \pi_{ij} \pi_{kr} \mathrm{cov}(K_{A_i C_j}, K_{A_k C_r}). \quad (3)$$

There are two ways of estimating the covariances between two distances. It is generally admitted that the distances along distinct branches of the phylogenetic tree are independent, that is, of covariance zero (Wu and Li 1985). With this assumption, we have:

$$\mathrm{cov}(K_{A_i C_j}, K_{A_k C_r}) = \mathrm{var}(K_{O_{ik} O_{jr}}), \quad (4)$$

in which O_{ik} is the last common ancestor of A_i and A_k, and O_{jr} is the last common ancestor of C_j and C_r. Covariances can also be estimated directly from sequence data (Bulmer 1991; Takezaki, Rzhetsky, and Nei 1995) using the number of shared substitutions.

If distances are corrected for multiple hits, they are additive on the tree, and the least squares estimate of the length of the internal branch $O_{ik}O_{jr}$ is:

$$K_{O_{ik}O_{jr}} = \frac{K_{A_i C_j} + K_{A_i C_r} + K_{A_k C_j} + K_{A_k C_r} - 2K_{A_i A_k} - 2K_{C_j C_r}}{4}. \quad (5)$$

For most usual distance estimators between sequences, the analytical estimator of the variance can be deduced from the value of the distance estimate. Thus, $\mathrm{var}(K_{O_{ik}O_{jr}})$ can be deduced from $K_{O_{ik}O_{jr}}$, and $\mathrm{cov}(K_{A_i C_j}, K_{A_k C_r})$ can be obtained. This computation also needs the number of sites useful on the internal branch, estimated by:

$$l_{O_{ik}O_{jr}} = \frac{l_{A_i C_j} + l_{A_i C_r} + l_{A_k C_j} + l_{A_k C_r}}{4}. \quad (6)$$

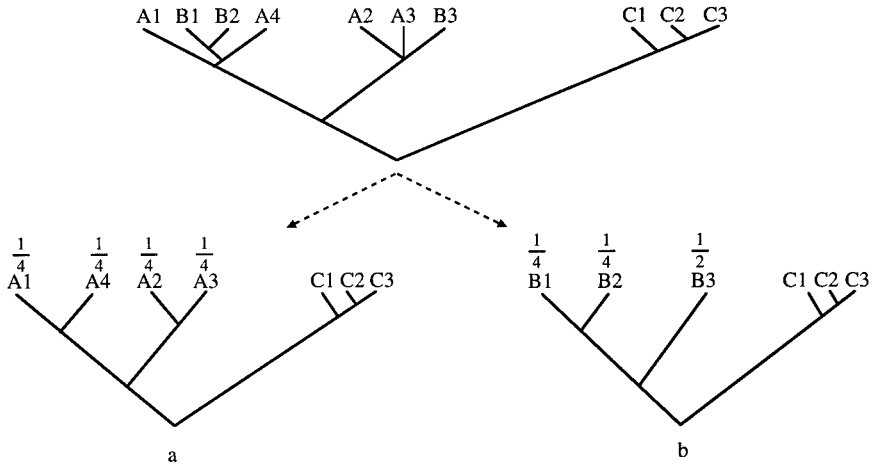

FIG. 5.—Rooted tree with ingroups A and B paraphyletic to each other. The two subtrees used to compute topological weights of sequences from groups A (tree a) and B (tree b) are shown.

In the relative-rate test, we are interested in the difference between K_{AC} and K_{BC}, and in its variance, which is:

$$\text{var}(K_{AC} - K_{BC}) = \text{var}(K_{AC}) + \text{var}(K_{BC}) - 2\,\text{cov}(K_{AC}, K_{BC}), \quad (7)$$

where

$$\text{cov}(K_{AC}, K_{BC}) = \sum_{i=1}^{n_A}\sum_{j=1}^{n_C}\sum_{k=1}^{n_B}\sum_{r=1}^{n_C} \pi_{ij}\pi_{kr}\text{cov}(K_{A_iC_j}, K_{B_kC_r}). \quad (8)$$

Covariance cov ($K_{A_iC_j}$, $K_{B_kC_r}$) is calculated in the same way as ($K_{A_iC_j}$, $K_{A_kC_r}$).

The test is done by computing the standardized difference:

$$\frac{|K_{AC} - K_{BC}|}{\sqrt{\text{var}(K_{AC} - K_{BC})}}.$$

The 5% level of this test value is 1.96.

Paraphyletic Groups of Sequences

Paraphyly of a group of sequences is here defined with regard to the other sequences included in the study. That is, a group is paraphyletic when the last common ancestor to all sequences of the group is also an ancestor to sequences which are not included in it, but are included in the study. Paraphyly of one or both of the ingroups (fig. 5) affects the topological weighting scheme only, since all calculations of means, variances and covariances are not dependent on monophyly of the groups considered. In this case, the weights for each group should be implemented by taking into account only the nodes which separate sequences from the considered group (fig. 5). Otherwise, computations are unchanged.

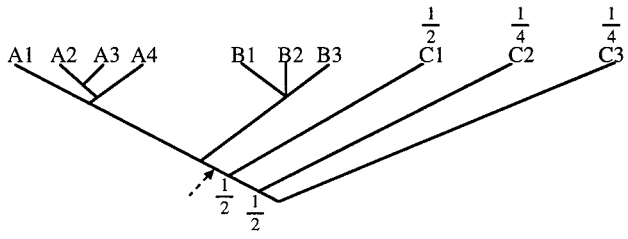

FIG. 6.—Rooted tree with outgroup C paraphyletic to the ingroups. The branch separating the ingroups from the outgroup is indicated by an arrow. Topological weights are indicated for sequences from the outgroup and for the relevant nodes.

Paraphyly of the outgroup (fig. 6) occurs when the root is on a branch separating different sequences of the outgroup. The relative-rate test described here does not depend on the rooting of the tree as far as computation is concerned, so computation occurs as if the root were on the branch separating the outgroup from the ingroups (indicated by an arrow on fig. 6). Interestingly, this amounts to giving lower weights to sequences which are farther from the ingroup.

Computation does not depend on the position of the root, but biological conclusions do. To compare rates, distances must be compared between groups of sequences with the same divergence time. This is the case both when the ingroups are paraphyletic compared to each other and when the outgroup is paraphyletic to the ingroups, since times of divergence from the two ingroups to the outgroup remain the same. The relative-rate test is no longer valid only if the ingroups as a whole are not monophyletic relative to the outgroup.

Results

Influence of Unbalanced Topology

The large inbalance imposed on some simulations (fig. 1) strongly misleads unweighted relative-rate tests, even when computation of covariances is independent of the phylogeny, as in the test of Takezaki, Rzhetsky, and Nei (1995). When the highly sampled subgroup (A2–A10) is slower than the single other sequence (A1) (fig. 1), and the ''true'' (simulated) difference between the fast group A and the slow group B is 0.05 substitutions per site, unweighted tests give a mean difference of 0.030 ($P > 0.05$). Weighting the test clearly improves results on these simulation data: the estimated difference

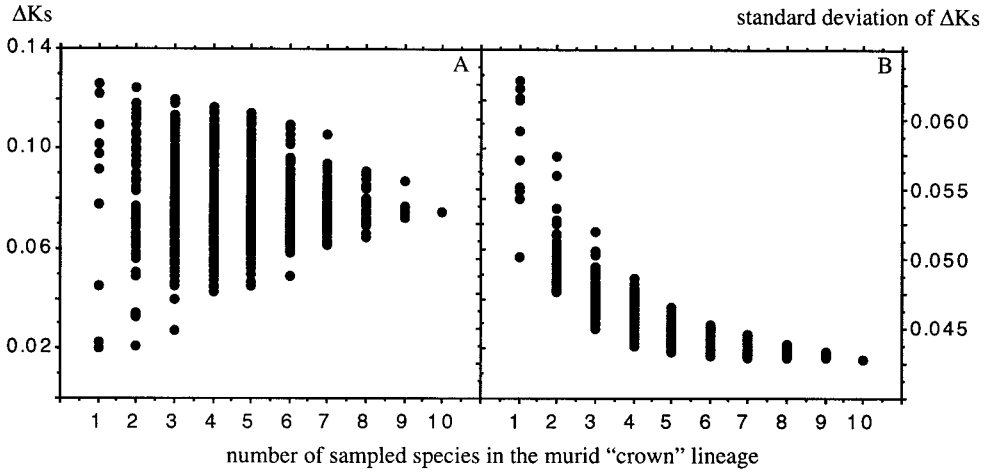

FIG. 7.—Influence of species sampling on relative-rate test results for LCAT coding sequences. For the ingroup of ''other murids'' (nonsubterranean), all combinations of *P* sequences among the *N* available have been made for all values of *P* between 1 and *N*. All available sequences are used for the other ingroup (''subterranean'' murids) and for the outgroup. A value of the difference in synonymous rates Δ*K*s (*A*) and a value of the standard deviation of this difference (*B*) are obtained for each sample of sequences.

becomes 0.048 ($P < 0.05$). Similar results are obtained when the outgroup is strongly unbalanced: the observed rate difference is 0.053 ($P < 0.05$) when weighting is used but 0.043 ($P > 0.05$) otherwise, again for a true value of 0.05.

LCAT synonymous rates were compared among rodent lineages under the assumption of rodent monophyly. For instance, comparing caviomorphs + murids to sciurids + glirids yields a nonsignificant ($P > 0.05$) 0.085 synonymous substitutions/site difference if topological weighting is used, but a very significant ($P < 0.01$) 0.131 synonymous substitutions/site difference if it is not. This is due to the fact that among these four rodent lineages, murids have the highest distance to the outgroup and also are the most sampled group; unweighted tests are then heavily influenced by this important sampling of one (fast) group compared with the others.

APRT introns are a good example of an unbalanced tree as far as outgroup species sampling is concerned (fig. 3), where the highest weight can be given to the nearest species; *Mus pahari* alone gets a weight of 0.5. Using all outgroup sequences together to compare complete introns, the rate difference (0.019) is significant with the weighted test *a la* Li and Bousquet (1992) (standardized difference = 2.16), but not if the test is unweighted (observed difference = 0.015; standardized difference = 1.66). It should be noted that despite the absence of topological weighting the test of Takezaki, Rzhetsky, and Nei (1995) detects the difference. While this test was only slightly more powerful than the approach of Li and Bousquet (1992) on simulated data, the difference appears to be important for certain real data sets.

With both real and simulated data and for both outgroup and ingroups, unbalanced sampling can bias results of the relative-rate test. The topological weighting we propose corrects for this bias. Although the main source of error is estimation of the rate difference, estimation of covariances through the method of Bulmer (1991), as in Takezaki, Rzhetsky, and Nei (1995), yields a more powerful test.

Influence of Ingroup Sampling

Sampling only one LCAT sequence to represent the 10 available species of the ''crown'' of murids results in wide variability of the observed rate difference with ''subterranean'' murids, ranging over an almost 10-to-1 range (fig. 7*A*). When more sequences are taken into account, this variability reduces, converging to a stable value when all sequences are used. When more sequences are used, there is also a strong decrease in the standard deviation associated with the test (fig. 7*B*). Thus, using more sequences yields a more powerful test.

The same patterns are obtained with simulated data (not shown), although with lower variability (twofold range of rate differences), which probably stems from the simplicity of the simulations as compared with real evolution. Simulated data make two important points: (1) The value toward which the observed difference converges is the real rate difference, whether this difference is very high or zero. Thus, using more sequences allows a more accurate test. (2) Sampling influence is stronger for more distantly related sequences (fig. 8). Indeed, the variability of parameter estimates due to sequence sampling decreases when sequences become more closely related.

Influence of Outgroup Sampling

Similar results can be observed sampling the outgroup sequences both for simulated and LCAT sequences (not shown), but whatever the number of outgroup sequences used, the observed difference is nearer to the real value and the standard deviation is smaller when the mean distance between outgroup and ingroup sequences is smaller. Simulation data also indicate that using the nearest outgroup sequence does not always yield the most significant result, but the nearest to the real difference, even if it is zero. Therefore, the test is both more accurate and more powerful when only the

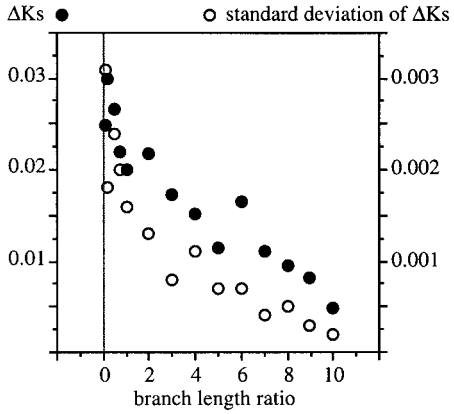

FIG. 8.—Influence of the relatedness of sequences on sampling effect. Abscissa: ratio of the length of the internal branch of a polytomy to the length of the external branches (α of fig. 1). Ordinates: variance among sampled sequences of estimated values of Δ*K* (closed circles) and of estimated values of standard deviation of Δ*K* (open circles).

nearest outgroup sequence is used than when all available outgroup sequences are used.

On APRT introns, Fieldhouse, Yazdani and Golding (1997) had already noted that the only outgroup yielding a significant difference alone was the nearest. And indeed, not only is the standardized difference higher with *M. pahari* alone (2.84) than with any other species (0.34 to 1.86), but it is higher than that with all species (2.16). In light of the simulation results, this strongly supports Fieldhouse, Yazdani, and Golding's conclusion that a rate difference indeed exists between these two species.

Discussion

Influence of Species Sampling

Both accuracy (difference between observed and real rate differences) and power (standard deviation) of relative-rate tests increase with the number of sequences used in the analysis. These effects are stronger when sequences are more distantly related (fig. 8), which suggests that taxonomic sampling strategies should not focus on the number of species alone, but also on their distribution; to characterize ''murid'' rates, more of an improvement will be made by adding *Spalax* to a sample of *Rattus*, *Mus*, and *Cricetus* than by adding yet another ''crown'' murid (see fig. 2).

While the test is improved by increasing the number of ingroup or outgroup sequences, these two types of data do not play the same role. Increasing the number of ingroup sequences not only improves the test from a statistical viewpoint, it also ensures that the results have wider biological value. The outgroup sequences, on the other hand, have no relevance to the biological conclusion. They are used because we do not have access to the ancestor sequence O of compared (groups of) sequences A and B: AC-BC is an approximation of AO-BO. The best choice of outgroup sequences C is thus those which give the best approximation of AO-BO. As may be expected, and as results confirm, this is the one C sequence for which OC is the lowest, thus generating the least variance possible.

It should be noted that the number of substitutions separating the ingroups and the outgroup may be smaller for a slowly evolving outgroup than for a phylogenetically near outgroup, in which case the first should be preferred. It should also be noted that the most accurate estimation is obtained with the nearest outgroup sequence, but other sequences may result in apparently ''more significant'' values. This does not mean that the test is more powerful with these sequences. Indeed, simulations in which the real difference in substitution rates between the two ingroups is zero show that the more distant the outgroup sequence is, the more it tends to evidence false differences.

Methods for Rate Comparison on Phylogenies

Several tests have been proposed to compare rates between homologous sequences. They are based on three main statistical approaches and may ask two related biological questions. The first question, which we have investigated here, concerns the comparison of two given lineages. Relative-rate tests seek to answer this question. It is legitimate when the two lineages have a clear *a priori* definition, such as comparing rodents and primates or parasitic and nonparasitic organisms. In other cases, no such division in two groups is justified, and it is best to ask another question: are there, overall, significant rate differences in the considered phylogeny? Rate heterogeneity tests, also called tests of the molecular clock, aim to answer this question. These two types of tests should not be confused, since overall rejection of a regular molecular clock may not be due to differences between lineages of particular interest, while similar mean rates in two lineages do not garantee a molecular clock inside each lineage, and even less for the outgroup.

Least-squares (Felsenstein 1984, 1988) can be used both in molecular clock tests (Takezaki, Rzhetsky, and Nei 1995; Uyenoyama 1995) and in relative-rate tests (Wu and Li 1985; Li and Bousquet 1992; Takezaki, Rzhetsky, and Nei 1995). It is the approach used here and has several strong advantages: it allows easy implementation of a large class of models, it is computationally fast, and it can be easily generalized to large groups of sequences, even paraphyletic groups and even if the phylogeny is not totally resolved. We have seen that estimating covariances by the method of Bulmer (1991) leads to a more powerful test than does computing all covariances on the tree as in Li and Bousquet (1992).

A χ^2 test on likelihoods with and without a molecular clock (Felsenstein 1988) appears to be valid only if the observed sample size is sufficiently important compared with the number of possible states, the number of which increases very fast with the number of sequences (Goldman 1993). Goldman (1993) suggested the use of simulations to overcome this difficulty, but this quickly becomes computationally very expensive. The χ^2 test can be applied safely to three sequences, which allows a relative-rate test if one of the sequences is constrained to being an outgroup to the other two in a rooted phylogeny (Muse and Weir 1992). This test is more powerful than the classical least-squares approach

(Muse and Weir 1992; Muse and Gaut 1994) and allows using a more general model, notably for coding sequences (Muse and Gaut 1994), but it remains computationally expensive and limited to small data sets. We believe that further work should overcome the current problems of the likelihood ratio test.

The last approach is to implement a nonparametric relative-rate test (Gu and Li 1992; Tajima 1993). This approach does not take into account an explicit model, and thus ignores multiple events which may saturate or bias observed changes. It does not allow a simple generalization to more than three sequences. Its advantage is that it can be used for comparing rates for which no models are available, such as insertions and deletions (Gu and Li 1992).

Conclusion

Generalized relative-rate tests allow the minimization of sampling bias due to the choice of sequences. This is all the more important because this choice is often constrained by sample availability, experimental constraints, or availability of sequences in databases. Two aspects of sampling are taken into account. (1) By using all sequences together, one avoids choosing only one sequence to represent a group. This is illustrated by the important variability observed depending on sequence choice (fig. 7). (2) By weighting by tree topology, one avoids giving more weight to overrepresented lineages in the sample. As soon as homologous sequences are compared, they are phylogenetically structured. It is clearly important to always take into account phylogenetic relationships as soon as means are to be computed or compared between homologous structures (Harvey and Puvis 1991), and this also applies to sequences.

Although the question of phylogenetic inference is clearly distinct from that of substitution rate inference, our results have some relevance to the first field. Indeed, many molecular phylogenies are obtained using distance methods, which rely on the same information as relative-rate tests. Several studies have indicated that the number of species used is an important parameter in molecular phylogeny reconstruction (Lecointre et al. 1993; Cummings, Otto, and Wakeley 1995). Our results indicate that such studies should go beyond the simple assertion that ''more sequences is better'', and investigate the differential influence of sampling ingroups and outgroups, of sampling distantly or closely related species, and probably also of sampling far or near nodes of high interest.

The most important factor in the relative-rate test seems to be a good estimation of the rate difference. This is obtained by using as many distantly related ingroup sequences as possible with topological weighting and by using the nearest outgroup sequence alone.

Acknowledgments

We thank Nicolas Galtier for providing C code routines; Brian Golding for providing an alignment of APRT sequences; and Dan Graur, Marcy K. Uyenoyama, and an anonymous referee for useful comments.

LITERATURE CITED

BULMER, M. 1991. Use of the method of generalized least squares in reconstructing phylogenies from sequence data. Mol. Biol. Evol. **8**:868–883.

CUMMINGS, M. P., S. P. OTTO, and J. WAKELEY. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. Mol. Biol. Evol. **12**:814–822.

FELSENSTEIN, J. 1984. Distance methods for inferring phylogenies: a justification. Evolution **38**:16–24.

———. 1988. Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. **22**:521–565.

FIELDHOUSE, D., F. YAZDANI, and G. B. GOLDING. 1997. Substitution rate variation in closely related rodent species. Heredity **78**:21–31.

GOLDMAN, N. 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. **36**:182–198.

GRAUR, D., L. DURET, and M. GOUY. 1996. Phylogenetic position of the ordre Lagomorpha (rabbits, hares and allies). Nature **379**:333–335.

GRAUR, D., A. H. HIDE, and W. H. LI. 1991. Is the guinea-pig a rodent? Nature **351**:649–652.

GU, X., and W. H. LI. 1992. Higher rates of amino acid substitution in rodents than in humans. Mol. Phylogenet. Evol. **1**:211–214.

HARVEY, P. H., and A. PUVIS. 1991. Comparative methods for explaining adaptations. Nature **351**:619–624.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. MUNRO, ed. Mammalian protein metabolism. Academic Press, New York.

LECOINTRE, G., H. PHILIPPE, H. L. VAN LE, and H. LE GUYADER. 1993. Species sampling has a major impact on phylogenetic inference. Mol. Phylogenet. Evol. **2**:205–224.

LI, P., and J. BOUSQUET. 1992. Relative-rate test for nucleotide substitutions between two lineages. Mol. Biol. Evol. **9**: 1185–1189.

LI, W. H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36**: 96–99.

LI, W. H., D. L. ELLSWORTH, J. KRUSHKAL, B. H. J. CHANG, and D. HEWETT-EMETT. 1996. Rates of nucleotide substitution in Primates and Rodents and the generation-time hypothesis. Mol. Phylogenet. Evol. **5**:182–187.

LUCKETT, W. P., and J. L. HARTENBERGER. 1993. Monophyly or polyphyly of the order Rodentia: possible conflict between morphological and molecular interpretations. J. Mamm. Evol. **1**:127–147.

MUSE, S. V., and B. S. WEIR. 1992. Testing for equality of evolutionary rates. Genetics **132**:269–276.

MUSE, S. V., and B. S. GAUT. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol. Biol. Evol. **11**:715–724.

PAMILO, P., and N. O. BIANCHI. 1993. Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. Mol. Biol. Evol. **10**:271–281.

ROBINSON, M., F. M. CATZEFLIS, J. BRIOLAY, and D. MOUCHIROUD. 1997. Molecular phylogeny of rodents, with special emphasis on murids: evidence from nuclear gene LCAT. Mol. Phylogenet. Evol. **8**:423–434.

SARICH, V. M., and A. C. WILSON. 1973. Generation time and genomic evolution in primates. Science **179**:1144–1147.

TAJIMA, F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics **135**:599–607.

TAKEZAKI, N., A. RZHETSKY, and M. NEI. 1995. Phylogenetic test of the molecular clock and linearized trees. Mol. Biol. Evol. **12**:823–833.

UYENOYAMA, M. K. 1995. A generalized least-squares estimate for the origin of sporophytic self-incompatibility. Genetics **139**:975–992.

WU, C. I., and W. H. LI. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA **82**:1741–1745.

MARCY K. UYENOYAMA, reviewing editor

Accepted May 11, 1998