Points of View

System. Biol. 48(2):396–399, 1999

Is the Rate of Molecular Evolution Inversely Related to Body Size?

JOSEPH B. SLOWINSKI[1] AND BRIAN S. ARBOGAST[2]

[1] *Department of Herpetology, California Academy of Sciences, Golden Gate Park, San Francisco, California 94118-4599, USA; E-mail: jslowins@cas.calacademy.org*

[2] *Department of Biology, Winston Hall, Wake Forest University, Winston-Salem, North Carolina 27109, USA*

The wealth of molecular data generated over the past two decades has led to the documentation of highly disparate rates of molecular evolution among different taxa. Among the more intriguing studies is that of Martin and Palumbi (1993), who presented data (their Table 2) derived from a variety of vertebrate taxa purporting to show that the rate of molecular evolution is inversely correlated with body size (and its correlates: metabolic rate, generation time, etc.). Their hypothesis has been cited as a possible explanation for variation in rate of molecular evolution among different taxa (e.g., Hafner et al., 1994; Rand, 1994; Stewart and Baker, 1994). We argue, however, that no such relationship is evidenced from Martin and Palumbi's data and that the problem lies in the fact that their distance-based estimates for the rate of molecular evolution were undercorrected for multiple substitutions.

In studies designed to estimate the rate of molecular evolution based on DNA sequence distance data, it is critical that the distances be adequately corrected for superimposed substitutions. Otherwise, estimates of the rate of molecular evolution will show a decline through time. In their study, Martin and Palumbi (1993) sampled a series of estimated rates (their Table 2) of mitochondrial DNA (mtDNA) evolution for various vertebrate taxa from the literature. The rates they sampled were based on restriction fragment length polymorphism (RFLP) distances corrected for multiple substitutions by use of equation 8 or 9 from Nei and Li (1979). These equations provide distances that correct for multiple substitutions under the Jukes–Cantor model (JC; Jukes and Cantor, 1969) of nucleotide substitution, which assumes that every type of substitution is equally likely. Brown et al. (1979), however, had shown that RFLP distances treated in this manner increase curvilinearly through time rather than linearly, demonstrating that these distances are undercorrected for superimposed substitutions.

From an examination of Martin and Palumbi's Table 2, it became clear to us that the large-bodied pairs of taxa they sampled tended to be associated with older divergence times (linear regression of body size on time: homeotherms, $P = 0.016$; poikilotherms, $P = 0.029$). This led us to suspect that the inverse relationship between body size and substitution rate observed by Martin and Palumbi might simply reflect the propensity for rates of evolution based on JC distances to decrease through time.

Brown et al. (1979) argued that the major reason for the time-dependence of JC-based rates of evolution is that the JC model ignores variation among nucleotide sites in substitution rate. Such rate variation is ubiquitous, and the failure to account for it will lead to underestimated evolutionary rates (Yang, 1996). To investigate Martin and Palumbi's data in this context requires a model that describes mathematically how JC-based rates will decline through time if

there is a quantifiable difference among nucleotide sites in the substitution rate. Under the JC model, the mean substitution rate per site per million years for two sequences $R_{JC} = 2\lambda$ is

$$R_{JC} = \frac{-\frac{3}{4}\ln\left(-\frac{4}{3}p_t + 1\right)}{t}, \quad (1)$$

where p_t is the probability that a site will vary between the two sequences after t units (millions of years) of time have elapsed since the common ancestral sequence. The JC model can be generalized to allow λ to vary among nucleotide sites. Yang (1993) found that the gamma distribution provides a reasonable description of rate variation among sites for observed molecular sequence data. Under the gamma-JC model, $p_t = 0.75 - 0.75[\alpha/(\alpha + 2.67\bar{\lambda}t)]^{\alpha}$ (Golding, 1983; Jin and Nei, 1990), where α describes the shape of the gamma distribution and $\bar{\lambda}$ is the average rate of substitution over all sites. Substituting this equation for p_t in Equation 1 leads to:

$$R_{JC} = \frac{-\frac{3}{4}\ln\left[\frac{\alpha}{(\alpha+\frac{8}{3}\bar{\lambda}t)^{\alpha}}\right]}{t}, \quad (2)$$

We performed nonlinear least-squares regressions of R_{JC} on t under Equation 2. Because the homeotherms (entries 1–9 in their Table 2) sampled by Martin and Palumbi have clearly experienced a higher rate of molecular evolution than the poikilotherms (entries 10–17), we analyzed the data for these two groups separately, as did Martin and Palumbi. Whenever a range of values was presented in Martin and Palumbi's Table 2, we used the midpoint. The nonlinear regressions provided separate mean estimates ($\bar{\alpha}$ and $\bar{\bar{\lambda}}$) for the α and $\bar{\lambda}$ parameters for homeotherms and poikilotherms. The estimated values for $\bar{\alpha}$ were then used to infer specific values of $\bar{\lambda}$ for each pair of taxa by using equation 2. These estimated specific $\bar{\lambda}$ values were then regressed against body size separately for homeotherms and poikilotherms to test Martin and Palumbi's hypothesis.

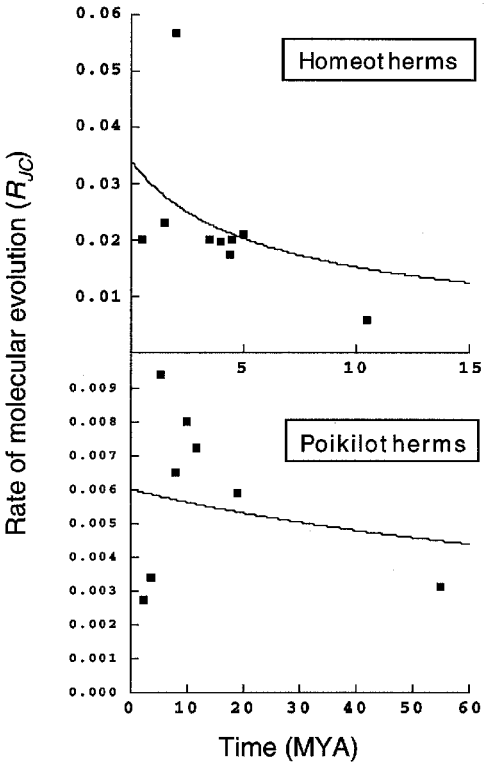

FIGURE 1. Nonlinear least-squares regression between R_{JC} and t of Martin and Palumbi's (1993:Table 2) data under the gamma-JC model, demonstrating the dependence of R_{JC} on t for this data (MYA = millions of years ago).

RESULTS

For the homeotherms, nonlinear regression (Fig. 1) estimated $\bar{\alpha}$ as 0.140 and $\bar{\bar{\lambda}}$ as 0.017 substitutions per site per million years. The low value for $\bar{\alpha}$ demonstrates strong rate variation among sites, but this is typical for mtDNA (Yang, 1996). Substituting the values for $\bar{\alpha}$ and $\bar{\bar{\lambda}}$ into Equation 2 results in $[R_{JC} \times t] = 1.47 + 0.105[\ln(0.140 + 0.045t)]$. To test the significance of the regression, we transformed the data for R_{JC} and t as indicated in the square brackets and performed a linear regression.

This analysis was found to be significant ($P = 0.003$). For the poikilotherms, nonlinear regression (Fig. 1) estimated $\bar{\alpha}$ as 0.583

and $\bar{\bar{\lambda}}$ as 0.003 substitutions per site per million years. A linear regression performed in the same manner as before was not significant ($P = 0.135$). However, visual inspection of this regression shows a regular, but curvilinear, relationship. The lack of significance is due to the curvilinearity, which indicates that even the gamma-JC model fails to adequately correct for multiple substitutions. Although more parameter-rich models would solve this problem, such models cannot be applied to RFLP data because the underlying sequences are not available.

Clearly, the R_{JC}'s sampled by Martin and Palumbi decline with time in a manner predicted by the gamma-JC model. Furthermore, linear regression of estimated specific $\bar{\lambda}$'s against body size was not significant for either homeotherms or poikilotherms (homeotherms, $P = 0.302$; poikilotherms, $P = 0.418$), demonstrating that Martin and Palumbi's data do not support their conclusion of an inverse relationship between body size and rate of molecular evolution.

Discussion

Our reanalysis of Martin and Palumbi's (1993) study underscores the importance of adequately correcting distance-based estimates of rates of molecular evolution for rate variation among nucleotide sites. Failure to do so leads to several problems (Yang, 1996), including misestimated times of divergence (Arbogast and Slowinski, 1998), misestimated substitution parameters (Wakeley, 1996), and spurious correlations between rate of molecular evolution and life-history parameters (this paper).

The obvious question that emerges from our note is: Why were the large-bodied taxa associated with older times of divergence? Two possibilities exist: Either this is a coincidence due to sampling error, or it reflects a real evolutionary pattern. If the latter is correct, possibly large-bodied taxa experience a lower rate of diversification, and hence times of divergence among pairs of large-bodied taxa are on average greater than for small-bodied taxa. In fact, some evidence indicates that small-bodied taxa do indeed have a higher rate of diversification than large-bodied taxa (Dial and Marzluff, 1988).

The point of our note has been to emphasize the importance of considering among-site variation in studies of molecular evolutionary rates, and to show how the failure to do so led Martin and Palumbi (1993) to erroneously infer an inverse relationship between body size and the rate of molecular evolution (within the context of their data). For this purpose, our analyses suffice. However, we point out that when investigating whether a relationship exists between the rate of molecular evolution and a life-history parameter for a set of taxa, investigators must take the phylogeny into account to avoid a possible spurious relationship attributable to phylogenetic effects (Felsenstein, 1985; Harvey and Pagel, 1991). It is not immediately clear to us how this could be done in the present context; nonetheless, the alternative of not taking the phylogeny into account is unacceptable.

Acknowledgments

We thank R. Browne, J. Demastes, S. Edwards, M. Hafner, K. Kron, P. Lewis, D. Reed, T. Spradling, and M. White for their comments.

References

Arbogast, B. S., and J. B. Slowinski. 1998. Pleistocene speciation and the mitochondrial DNA clock. Science 282:1955a.

Brown, W. M., M. George, Jr., and A. C. Wilson. 1979. Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA 76:1967–1971.

Dial, K. P., and J. M. Marzluff. 1988. Are the smallest organisms the most diverse? Ecology 69:1620–1624.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791.

Golding, G. B. 1983. Estimates of DNA and protein sequence divergence: An examination of some assumptions. Mol. Biol. Evol. 1:125–142.

Hafner, M., P. D. Sudman, F. X. Villablanca, T. A. Spradling, J. W. Demastes, and S. A. Nadler. 1994. Disparate rates of molecular evolution in cospeciating hosts and parasites. Science 265:1087–1090.

Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford Univ. Press, Oxford, England.

Jin, L., and M. Nei. 1990. Limitations of the evolutionary parsimony method of phylogenetic analysis. Mol. Biol. Evol. 7:82–102.

Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pages 21–132 *in* Mammalian pro-

tein metabolism (H. N. Munro, ed.). Academic Press, New York.

MARTIN, A. P., AND S. R. PALUMBI. 1993. Body size, metabolic rate, generation time, and the molecular clock. Proc. Natl. Acad. Sci. USA 90:4087–4091.

NEI, M., AND W.-H. LI. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA 76:5269–5273.

RAND, D. M. 1994. Thermal habit, metabolic rate and the evolution of mitochondrial DNA. Trends Ecol. Evol. 9:125–131.

STEWART, D. T., AND A. J. BAKER. 1994. Patterns of sequence variation in the mitochondrial D-loop region of shrews. Mol. Biol. Evol. 11:9–21.

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, AND D. M. HILLIS. 1996. Phylogenetic inference. Pages 407–514 *in* Molecular systematics, 2nd edition (D. M. Hillis, C. Moritz, and B. K. Mable, eds.). Sinauer, Sunderland, Massachusetts.

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.

YANG, Z. 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates vary over sites. Mol. Biol. Evol. 10:1396–1401.

YANG, Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. 11:367–372.

Received 23 September 1997; accepted 15 October 1998
Associate Editor: D. Cannatella

Syst. Biol. 48(2):399-406, 1999

Apomorphy Distribution Is an Important Aspect of Cladogram Symmetry

PAUL N. PEARSON

Department of Earth Sciences, University of Bristol, Queens Road, Bristol BS8 1RJ, United Kingdom; E-mail: paul.pearson@bristol.ac.uk

In recent years there has been a great deal of interest in the balance of cladograms, which has generally come to be accepted as meaning the extent to which the internal nodes subtend clades of equal size (Fig. 1). In particular, several studies have suggested that a higher proportion of real cladograms culled from the literature are unbalanced (comb-shaped or pectinate) than would be expected if they were produced by a random Markovian branching process of speciation (Colless, 1982; Guyer and Slowinski, 1991, 1993; Heard, 1992; Mooers et al., 1995). One reason for the interest is methodological: If cladograms are statistically unbalanced, they will contain a higher proportion of long branch lengths than would otherwise be expected, which has implications for the accuracy with which they reflect true phylogeny (Rohlf et al., 1990). A more fundamental reason for the interest, however, is that the findings may provide important information about patterns and processes of evolution (Heard, 1992, 1996; Kirkpatrick and Slatkin, 1993; Mooers and Heard, 1997; Bond and Opell, 1998).

A variety of potential methodological artifacts might produce a tendency for imbalance in cladograms, even if no such pattern exists in the underlying phylogeny (Colless, 1982, 1995; Guyer and Slowinski, 1991; Mooers et al., 1995; Huelsenbeck and Kirkpatrick, 1996). Nevertheless, if the methodological problems can be resolved (e.g., Farris and Källersjö, 1998), or at least if their effects can be adequately quantified, it should be possible to study patterns of evolution through using cladogram shape. This field is in its infancy (Mooers and Heard, 1997). Perhaps the most fundamental question that cladograms may help resolve is whether evolution is largely stochastic and nonprogressive (e.g. Gould, 1988), or if phylogenies contain ingrained asymmetry, implying nonrandom differences in the evolutionary success of species.

An important aspect of cladogram symmetry has been overlooked in previous work on the subject, namely, the distribution of apomorphies. As discussed by Mindell et al.