



OXFORD
BIOLOGY



the **TIMETREE** *of* **LIFE**

edited by **S. BLAIR HEDGES** *and* **SUDHIR KUMAR**
foreword by James D. Watson

Timetrees: beyond cladograms, phenograms, and phylograms

John C. Avise

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697, USA (javise@uci.edu)

Abstract

For several historical reasons discussed herein, until recently the absolute temporal dimension of many phylogenetic trees has been relatively ignored whereas the branching (cladistic) aspect typically has been the focus of most phylogeny-reconstruction efforts. This unfortunate neglect of “timetrees” is now being remedied, as this book will attest. Many scientific benefits can emerge from superimposing robust estimates of geological time on cladograms, including opportunities to: improve phylogenetic reconstructions of phenotypic evolution; illuminate causal geological or other events and processes in the history of life; and develop a universal time-standardized framework for biological classification that will facilitate studies in comparative evolution.

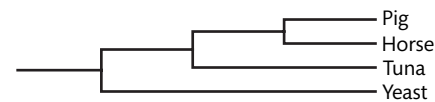
... the extent of variation of the primary structure [of proteins] ... may give rough approximations of the time elapsed since the lines of evolution leading to any two species diverged.

—Emanuel Margoliash (1963).

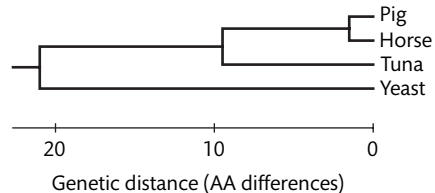
The notion that genetic differences between species tend to increase with time has been prevalent since the inception of the field of molecular evolution. Margoliash’s comment above was in reference to early empirical data for cytochrome *c* showing that horse and pig differ at only three amino acid sites, whereas horse and tuna display 19 amino acid substitutions and horse and yeast display at least 44 such changes. From such observations, Margoliash (1) concluded: “relatively closely related species show few [genetic] differences... phylogenetically distant species exhibit wider dissimilarities.” Zuckerkandl and Pauling (2) had noted similar kinds of outcomes in four members of the hemoglobin protein family in 1962, and in 1965 they coined the term “molecular clock” to encapsulate the notion that protein

sequences appear to diverge with some regularity across evolutionary time (3). During the ensuing four decades, the general time-dependent nature of molecular evolution (but not always any great precision for particular molecular clocks) gained voluminous support from studies on a wide variety of proteins and subsequently of DNA sequences (reviewed in 4).

Stepped cladogram



Distance phenogram



Evolutionary timetree

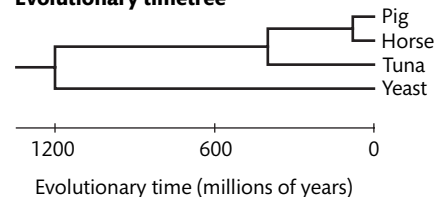


Fig. 1 Alternative phylogenetic depictions for pig, horse, tuna, and yeast based on the cytochrome *c* sequence data considered by Margoliash (1; see text). Top panel: stepped cladogram showing only the cladistic order of phylogenetic nodes (branch lengths have no meaning). Middle panel: distance phenogram showing branch lengths (amino acid changes) in addition to branching topology (thereby making this depiction a phylogram also; see text). Bottom panel: an evolutionary timetree showing estimates of nodal dates in addition to the branching topology.

The word phylogeny, from the Greek roots “phyl” meaning tribe and “geny” meaning origin, refers to the genealogical history of life. Given the early recognition of molecular clocks and the vast popularity of molecular systematics since the mid-1960s (5), it might seem that phylogeneticists would have been preoccupied with dating evolutionary trees, that is, in capitalizing upon the wealth of temporal (as well as cladistic information) seemingly inherent in the primary structures of proteins and nucleic acids. Ironically, however, phylogenetic chronograms (“timetrees,” for short) have been relatively neglected until recently in most molecular studies in favor of efforts instead to reconstruct merely the branching topologies (cladistic structures) of phylogenetic trees.

In support of this contention, a recent compendium on the “tree of life” (6) included nearly 200 phylogenetic branching diagrams, fewer than 10 of which (ca. 5%) provided explicit indications of the absolute dates of evolutionary nodes. If tallies likewise were to be conducted on the phylogenetic representations published during the last four decades in leading journals such as *Systematic Biology*, *Molecular Phylogenetics and Evolution*, or *Evolution*, a similar neglect of evolutionary dates in phylogenetic trees would undoubtedly be evidenced. This state of affairs is highly ironic, because timetrees can, in principle, encapsulate far more information about evolutionary history than do cladograms or phenograms (Fig. 1).

Here, I will speculate on why timetrees have been relatively neglected in biological systematics (at least until very recently), and, more importantly, why this situation can and should be remedied.

Past neglect of timetrees

Historical reasons for the relative disinterest in timetrees (compared to the enthusiasm for cladograms) probably entail both scientific and sociological factors. Here I conjecture on why timetree reconstructions have not always been pursued with vigor.

Cladistic–phenetic distractions

The rise of molecular biology that gave birth to molecular phylogenetics in the early 1960s happened to coincide with two seminal developments in traditional (i.e., nonmolecular) systematics: the publication in 1963 of *Principles of Numerical Taxonomy* by Sokal and Sneath (7), and the 1966 translation into English of Hennig’s

Phylogenetic Systematics (8). These books gave rise, respectively, to the polarized schools of phenetics and cladistics that were to dominate philosophical discourse in systematics for more than two decades (see chapter 4 in 9). Pheneticists argued that organisms should be grouped and classified according to their overall similarity as measured by defined rules using as many quantifiable traits as possible. Results were summarized as phenograms depicting phenetic similarities (not necessarily phylogenetic relationships) among taxa. Cladists countered that organisms should be grouped and classified according explicitly to their evolutionary relationships as evidenced by (even a few) shared-derived traits, that is, synapomorphies. Results were summarized as cladograms depicting cladistic topologies in phylogenetic trees.

Especially in the 1960s through the 1980s, the young field of molecular evolution found itself distracted by (and sometimes immersed in) the cladistic–phenetic wars. On one battlefront, some molecular systematists were forced to defend their approaches against hard-core cladists who automatically discredited any “phenetic” method (e.g., DNA–DNA hybridization) that merely yielded genetic similarity or distance estimates between taxa. Many cladists also impugned any statistical clustering algorithms for phylogenetic inference (e.g., UPGMA or neighbor-joining, 10) that employed composite genetic distance estimates. On entirely another battlefront, molecular systematists were sometimes forced to counter the notion by a few hard-core pheneticists that phylogenetic reconstruction itself was not one of the primary achievable goals of systematics.

Interestingly, neither the pheneticists nor the cladists devoted much attention to how absolute time of evolutionary separation between taxa might be extracted from empirical information (molecular or otherwise). Thus, the relatively few timetrees estimated from molecular or other data were not readily accepted into either the traditional phenetic or cladistic camps. This is ironic for the following reasons.

In one important sense, various molecular approaches approximate a cladistic ideal, because the data are genetic and the volume of molecular information can be so vast as to provide a strong collective signal regarding branching topologies (as well as branch lengths) in phylogenetic trees. In another sense, however, various molecular approaches also closely approximate a phenetic ideal, because the number of assayed traits can be huge (up to millions of nucleotide positions in some current DNA sequence comparisons) and the data are inherently

quantifiable in terms of overall similarities or distances among taxa. Furthermore, given that evolutionary convergences or reversals at individual nucleotide positions can be rather common (in part because only four interconvertible character states exist per site), any cladistic analysis that focuses unduly on any few presumptive shared-derived characters in molecular data could be inappropriate.

Thus, molecular phylogenetic approaches are neither cladistic nor phenetic exclusively, but rather they can encompass some of the best of both worlds. Irrespective of the method of analysis (e.g., via distance-based, parsimony, maximum likelihood, or Bayesian methods), large molecular data sets can yield phylograms that portray both the cladogenetic and anagenetic components of phylogeny. However, the main thesis of this chapter is that phylogeneticists should now strive for even more information by estimating absolute dates for evolutionary nodes. To the extent that this task is successfully accomplished, the resulting timetrees will convey much more information than traditional cladograms, phenograms, or phylograms.

Reservations about molecular clocks

A second set of historical obstacles to greater enthusiasm for timetrees involved some misunderstandings about molecular clocks. On the conceptual side, the notion of molecular clocks meshed well with an emerging neutrality theory predicting that molecular evolutionary rates were driven by (indeed were equitable with) rates of neutral mutation (11). Accordingly, praise or condemnation of molecular clocks often hinged on a researcher's philosophical stance in the broader neutralist–selectionist debate. But this situation was inappropriate because molecular clocks are also compatible with most selectionist scenarios. For example, if large numbers of genes are acted upon by multifarious selection pressures over long periods of time, then any short-term fluctuations in selection intensity might average out such that the overall genetic distances among taxa correlate well with elapsed times since common ancestry.

On the empirical side, any hopes for a universal molecular timepiece of great precision were dashed as empirical data accumulated showing significant variation in molecular rates at several levels: among nucleotide positions within codons, among nonhomologous genes within a lineage, among various classes of DNA such as coding and noncoding, and between full genomes such as nuclear and mitochondrial (review in 12). Seemingly

most damning to universal clock arguments were rate differences also reported at homologous loci among different taxonomic lineages (e.g., 13–16). Some systematists interpreted such findings to doom all efforts to estimate timetrees from molecular data, but this reaction was overly negative because, in principle and often in practice, rate variation (as well as complicating phenomena such as saturation effects in the nucleotide substitution process) can be recognized and accommodated in timetree reconstructions (e.g., 17–19). Thus, contrary to a widespread sentiment, attempts to date phylogenetic nodes do not necessarily hinge critically on the precise ticking of molecular clocks. Furthermore, many interesting questions in evolutionary biology can be answered (at least to a first approximation) with even ballpark temporal estimates.

In addition, pronounced rate variation is likely to be greater on local than on global genomic scales. A useful analogy might be to the gas laws in physics. Much as the individual molecules within a gas have idiosyncratic and unpredictable movements in submicroscopic space, individual nucleotide sites and loci within a genome can display idiosyncratic and unpredictable changes over evolutionary time. But large collections of gas molecules also have consistent composite properties (summarized by the gas laws of physics), and likewise large collections of DNA nucleotides may have fairly consistent conglomerate behaviors such as mean rates of sequence evolution. In other words, with regard to temporal signal, DNA sequences are detail-noisy yet composite-rich.

A final and oft-overlooked point is that estimating timetrees is not an endeavor exclusively for the field of molecular evolution. To the contrary, nodes in phylogenetic trees are most securely dated when multiple lines of evidence (from paleontology, historical geography, comparative biology, etc.) are also thoroughly integrated into the analysis. Indeed, every initial calibration of a molecular clock requires at least one absolute temporal landmark (e.g., from fossil or biogeographic evidence) independent of the molecular data. Thus, if objections are to be raised against the promotion of timetrees, they should not stem from a (misplaced) sentiment that traditional nonmolecular approaches to systematics are thereby somehow being excluded.

Rationales and prospects for timetrees

Having speculated on why time–time approaches have been relatively neglected, I want to suggest why this situation can and should change. First, molecular data

are now being gathered and phylogenetically analyzed for thousands of species, and major structural features plus many finer details of the Tree of Life are quickly emerging. Second, because these newest phylogenetic appraisals are typically based on unprecedented volumes of sequence information (sometimes of entire genomes), the resulting phylogenetic estimates should be almost as secure as might ever become possible. Third, cladogenetic topologies and anagenetic branch lengths (alone or together) paint incomplete phylogenetic pictures unless secure estimates of nodal dates are included in the representations as well. Thus, the time is right for biologists to begin exploring much more fully the absolute temporal dimensions of the phylogenetic trees they reconstruct. Secure timetrees will then offer many potential added benefits to biology, including the following.

Character mappings will be enriched

All branches of comparative biology could profit from better knowledge about the evolutionary histories of morphological, behavioral, physiological, biochemical, or other phenotypic traits. In recent years, phylogenetic character mapping (PCM) has become a wildly popular exercise wherein scientists use molecular trees as historical backdrops for deciphering the evolutionary pathways traversed by all sorts of phenotypic traits in a wide variety of plant, animal, and microbial taxa (review in 20). Well-supported phylograms are especially useful in PCM reconstructions via maximum likelihood analyses, often yielding insights about ancestral character states that would not be apparent from maximum parsimony reconstructions based on cladogram structures alone (e.g., 21). Furthermore, time is the common denominator in all rate estimates, so assessments of both absolute and relative evolutionary rates in various phenotypic (and molecular) traits are obviously facilitated when well-dated nodes in the relevant timetrees are available.

PCM exercises have illustrated two broader ironies about traditional systematics. First, for many taxa, extensive discussions and debates have often centered on fine details of alternative branching orders within particular cladograms, while the no-less-important temporal contexts of the phylogenies were often virtually neglected. Second, whereas phylogenetic treatments have traditionally focused on particular taxa one at a time, much of interest can emerge from comparative assessments (for which molecular data are uniquely well suited) across even disparate taxonomic groups. In short, the regular

presentation of explicit timetrees will open many conceptual worlds for fruitful discussions in comparative biology. Consider, for example, how the identical-stepped cladograms in Fig. 2 take on new meaning and raise novel questions when properly stretched to reflect their relative temporal dimensions.

Causal historical events will be illuminated

Another benefit of timetrees is that they will inevitably help to focus scientific attention on causal processes underlying phylogenetic diversity. Several examples are evident in this book, but to introduce the broader argument I will briefly mention two additional cases here. Interest has long centered on whether ancient vicariant events (related to plate tectonic movements) or subsequent over-water dispersals (e.g., by rafting) account for the presence of various terrestrial vertebrates in the West Indies. Geologic evidence indicates that the Greater Antilles formed in close proximity to North and South America during the mid-Cretaceous and that these islands began drifting from their continental partners at least 80 million years ago (Ma). The vicariance scenario thus predicts that sister clades on the islands vs. mainlands separated more than 80 Ma, whereas dispersal

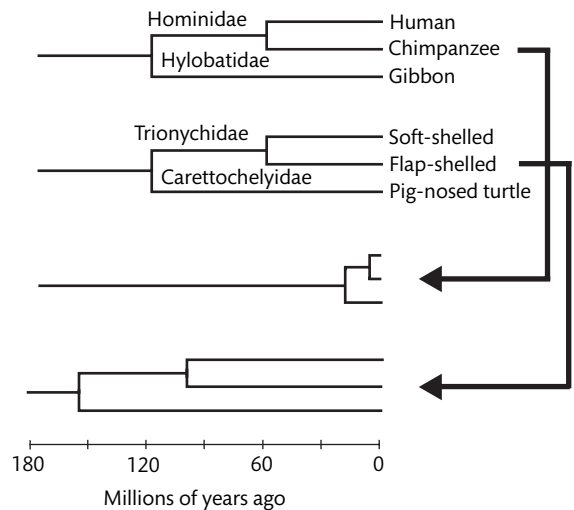


Fig. 2 Example of widely differing temporal frameworks for cladograms with identical branching sequence. Above: stepped cladograms for four species representing two taxonomic families of primates and for four species representing two families of marine turtles. Below: timetrees for these respective assemblages [estimates of nodal evolutionary dates are provisional].

scenarios predict that the separations were more recent and probably variable in time. From molecular phylogenetic appraisals of nodal dates for more than 35 relevant pairs of vertebrate taxa, Hedges (22) effectively falsified the ancient vicariance hypothesis for these faunas.

Another illustrative example of an informative time-tree is reproduced in Fig. 3. By comparing extensive sequence data from several nuclear and mitochondrial genes in species representing 139 genera and nearly all higher taxa of extant ants, and by incorporating dates of relevant fossils into the phylogenetic analyses, Moreau *et al.* (23) concluded that an early evolutionary radiation of ant lineages coincided with the great proliferation of angiosperm plants (as well as coleopteran and hemipteran insects) during the Late Cretaceous (100–66 Ma). These findings, which indicate that ants diversified much earlier than previously supposed, are generating exciting

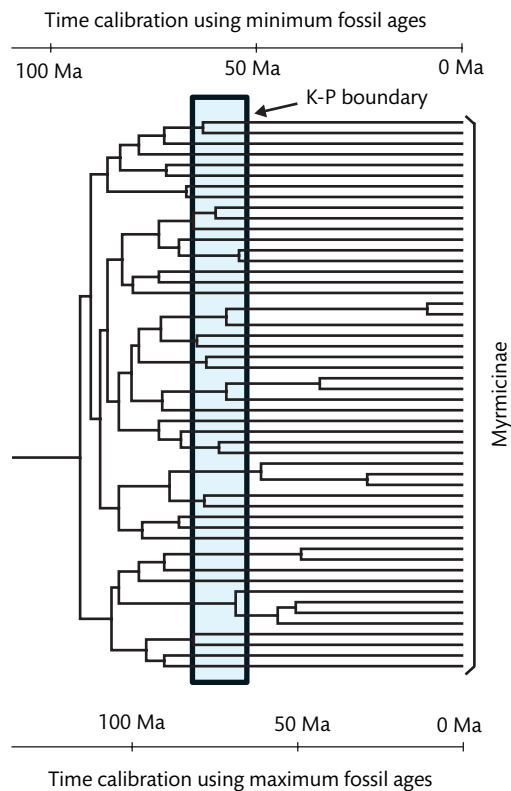


Fig. 3 Timetree for 52 genera (each right-terminal node is a different genus) in the ant subfamily Myrmicinae after Moreau *et al.* (23). This diagram includes only a representative subset of the total of 139 ant genera (in 19 subfamilies) included in the broader phylogenetic study by Moreau and colleagues. Note the evident proliferation of lineages before the Cretaceous–Paleogene (K-P) boundary (see text).

new hypotheses about what biological factors might have been responsible. For example, ground litter favorable for ants is highly diverse in angiosperm forests and may have provided many novel habitat opportunities; and/or, the expansion of herbivorous insects at about that same time may have offered diverse food resources ripe for exploitation by the newly diversifying ants.

Biological classifications can be universally standardized

If and when secure timetrees become widely available, opportunities will also arise to develop the first ever universally standardized scheme of biological classification. Unfortunately, current classifications provide no assurance that one taxonomic genus or family of mammals, for example, is comparable to another, much less to a genus or family of fishes, insects, or beetles. Indeed, no standards have been adopted by which such assurances might even be attempted. Another aspect of inconsistency in current classifications is that whereas many taxa are valid clades, others are polyphyletic or paraphyletic grades (see, e.g., the top of Fig. 1), and the nomenclature gives no indication which is which. Hennig (8) bemoaned these states of affairs when he wrote, “If systematics is to be a science it must bow to the self-evident requirement that objects to which the same label is given must be comparable in some way.” Others have echoed similar thoughts: “No scientific enterprise, least of all one that considers the promotion of nomenclatural universality as one of its primary objectives, can accept the inconsistencies and ambiguities current in biological taxonomy” (24). Although systematists readily admit that the biological classifications in use today are wildly nonuniform across disparate taxonomic groups, little has been accomplished to rectify this huge flaw.

This unfortunate situation stems in part from the lack of biological universality in the morphological, physiological, or other phenotypes that systematists traditionally use to classify organisms. For example, traits conventionally employed in fish systematics (number of lateral-line scales, fins placements, etc.) often have no useful counterparts in mammals or insects, one net result being that few systematists seem to have given much thought to how a universally standardized taxonomy might be erected or used. By contrast, many DNA and protein molecules are more or less universal and could in principle serve as common yardsticks for biological classifications. However, different lineages show rather variable evolutionary rates and patterns (see earlier), so

molecular sequences per se still do not provide an ideal standard for cross-taxon comparisons.

Instead, evolutionary time itself is the ultimate gold standard for developing any universally consistent framework for biological classification. Avise and Johns (25) detail the rationale for this sentiment and also suggest a logical means for implementing it (at least in principle). Their approach, following Hennig (8), involves the notion of “temporal banding.” For any securely dated timetree, temporal windows or bands of defined but arbitrary width (ideally sanctioned by a consensus of systematists) would be superimposed. By definition, each window would correspond to a particular rank in the conventional Linnaean hierarchy (such as genus, family, or order), and extant members of each clade would automatically be assigned a taxonomic rank empirically defined by the evolutionary window that contains the stem node.

Under this temporal-banding proposal, all named taxa would be clades, and all clades of a given taxonomic rank would have originated within the same window of evolutionary time. Such clades could be of mammals, fishes, insects, corals, or any other kinds of creature for which timetrees become securely estimated. Assuming that timetrees for many or most forms of life will eventually emerge from systematists’ collective efforts, an implementation of the temporal-banding protocol could yield biology’s first universally standardized classification. However, a major difficulty with this proposal (in its original formulation) concerns the large number of rank shifts and nomenclatural changes that likely would be entailed. Taxonomic stability is also very important in systematics (26), so any wholesale transformation of classification in the name of global standardization could be counterproductive if it complicated more so than facilitated the communication and retrieval of biological information.

To overcome this problem, we recently introduced an addendum to the original temporal-banding proposal (20). Rather than transmogrify existing classifications to conform to temporal windows, the temporal window for each clade could simply be indicated, in shorthand, by an appropriate time-clip attached to the existing taxon name. Each time-clip would refer to the specific temporal window (such as a geological period or epoch) within which the clade is thought to have arisen. Thus, the time-clip proposal would provide a practical way to retain the familiar Linnaean system and traditional taxon names yet simultaneously incorporate important temporal information into existing classifications.

Conclusion: Time is of the essence

For all of these reasons, the time seems right to focus much greater attention on the temporal component (in addition to cladistic component) of phylogenetic trees. The routine estimation and utilization of timetrees could add exciting new dimensions to biology, including enhanced opportunities to integrate large molecular data sets with fossil and biogeographic evidence (and thereby foster greater communication between molecular and traditional systematists); estimate not only ancestral character states but also evolutionary rates in numerous categories of organismal phenotype; help establish more reliable associations between causal historical processes and biological outcomes; develop a universally standardized scheme for biological classifications; and, in general, promote novel avenues of thought in many arenas of comparative evolutionary biology.

References

1. E. Margoliash, *Proc. Natl. Acad. Sci. U.S.A.* **50**, 672 (1963).
2. E. Zuckerkandl, L. Pauling, in *Horizons in Biochemistry*, M. Kasha, B. Pullman, Eds. (Academic Press, New York, 1962), pp. 189–225.
3. E. Zuckerkandl, L. Pauling, in *Evolving Genes and Proteins*, V. Bryson, H. J. Vogel, Eds. (Academic Press, New York, 1965), pp. 97–166.
4. S. Kumar, *Nat. Rev. Genet.* **6**, 654 (August 2005).
5. D. M. Hillis, C. Moritz, B. K. Mable, *Molecular Systematics* (Sinauer Associates, Sunderland, Mass., 1996).
6. J. Cracraft, M. J. Donoghue, *Assembling the Tree of Life* (Oxford University Press, New York, 2004).
7. R. R. Sokal, P. H. A. Sneath, *Principles of Numerical Taxonomy* (W. H. Freeman, San Francisco, 1963).
8. W. Hennig, *Phylogenetic Systematics* (University of Illinois Press, Urbana, 1966).
9. J. C. Avise, *Molecular Markers, Natural History, and Evolution* (Sinauer Associates, Sunderland, Mass., 2004).
10. M. Nei, S. Kumar, *Molecular Evolution and Phylogenetics* (Oxford University Press, New York, 2000).
11. M. Kimura, *Nature* **217**, 624 (1968).
12. W.-H. Li, *Molecular Evolution* (Sinauer Associates, Sunderland, Mass., 1997).
13. F. J. Ayala, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7776 (1997).
14. X. Gu, W.-H. Li, *Mol. Phylo. Evol.* **1**, 211 (1992).
15. S. Kumar, S. B. Hedges, *Nature* **392**, 917 (1998).
16. D. M. Rand, *Trends Ecol. Evol.* **9**, 125 (1994).
17. M. Hasegawa, J. L. Thorne, H. Kishino, *Genes Genet. Syst.* **78**, 267 (2003).
18. M. S. Springer, W. J. Murphy, E. Eizirik, S. J. O’Brien, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1056 (2003).

19. J. L. Thorne, H. Kishino, I. S. Painter, *Mol. Biol. Evol.* **15**, 1647 (1998).
20. J. C. Avise, *Evolutionary Pathways in Nature: A Phylogenetic Approach* (Cambridge University Press, Cambridge, 2006).
21. D. Schluter, T. Price, A. O. Mooers, D. Ludwig, *Evolution* **51**, 1699 (1997).
22. S. B. Hedges, *Ann. Rev. Ecol. Syst.* **27**, 163 (1996).
23. C. S. Moreau, C. D. Bell, R. Vila, S. B. Archibald, N. E. Pierce, *Science* **312**, 101 (2006).
24. K. deQueiroz, J. Gauthier, *Ann. Rev. Ecol. Syst.* **23**, 449 (1992).
25. J. C. Avise, G. C. Johns, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7358 (1999).
26. E. Mayr, *The Growth of Biological Thought: Diversity, Evolution, and Inheritance* (Belknap Press, Cambridge, Mass., 1982).