



# Parsimony overcomes statistical inconsistency with the addition of more data from the same gene

Kurt M. Pickett<sup>1,\*</sup>, Greg L. Tolman<sup>2</sup>, Ward C. Wheeler<sup>1</sup> and John W. Wenzel<sup>3</sup>

<sup>1</sup>*Division of Invertebrate Zoology, American Museum of Natural History, 79th Street at Central Park West, New York, NY 10024, USA;*

<sup>2</sup>*295 Bennett Avenue, 5E, New York, NY 10040, USA;* <sup>3</sup>*Museum of Biological Diversity, Department of Entomology,*

*The Ohio State University, 1315 Kinnear Road, Columbus, OH 43212, USA*

Accepted 17 August 2005

---

## Abstract

Many authors have demonstrated that the parsimony method of phylogenetic analysis can fail to estimate phylogeny accurately under certain conditions when data follow a model that stipulates homogeneity of the evolutionary process. These demonstrations further show that no matter how much data are added, parsimony will forever exhibit this statistical inconsistency *if the additional data have the same distributional properties as the original data*. This final component—that the additional data must follow the same distribution as the original data—is crucial to the demonstration. Recent simulations show, however, that if data evolve heterogeneously, parsimony can perform consistently. Here we show, using natural data, that parsimony can overcome inconsistency if new data from the same gene are added to an analysis already exhibiting a condition indistinguishable from inconsistency.

© The Willi Hennig Society 2005.

---

All methods of phylogenetic inference are statistically inconsistent when their underlying assumptions are violated. False groupings are particularly likely when two terminal branches are long relative to, in the four taxon case, the internal branch and the two other terminal branches. Under these conditions, phylogenetic methods will erroneously interpret convergent character states on the terminal branches as similar due to ancestry when these outnumber the true shared derived characters that actually unite the taxa (Felsenstein, 1978). This so-called *long-branch attraction* (LBA, originally *long edge attraction* (Hendy and Penny, 1989)) is a theoretical phenomenon with cogent mathematical underpinnings (Felsenstein, 1978). When data are simulated under conditions that assure they are independent and identically distributed (i.i.d.), parsimony can fall prey to LBA (Huelsenbeck and Hillis, 1993), whereas Maximum Likelihood (ML) methods can perform better when the model makes the i.i.d. assumption. Mathematical theory

also demonstrates, however, that ML models assuming homogeneity of the evolutionary process will suffer from inconsistency when the data under consideration violate, even weakly, the i.i.d. assumption (Chang, 1996). Recent work supports this theory, showing that when ML imposes the i.i.d. assumption on simulated data which were generated according to a heterogeneous process, ML can converge to the wrong answer with 100% certainty, but parsimony is not misled (Kolaczkowski and Thornton, 2004). Here, we do not address the accuracy or relevance of Kolaczkowski and Thornton's demonstration that ML performs poorly under these conditions; others have begun this (Spencer et al., 2005; Steel, 2005; Gaucher and Miyamoto, in press). Furthermore, it has long been conceded that if the assumed ML model is incorrect for the data, ML is not guaranteed to be consistent (Swofford et al., 1996). Here, we focus only on this later finding of Kolaczkowski and Thornton (2004)—that parsimony is not misled by data that do not satisfy the i.i.d. assumption.

Although the consistency of a phylogenetic estimator is not directly relevant given finite data sets (Siddall and Kluge, 1997; Sanderson and Kim, 2000), the claim that

---

\*Corresponding author.

E-mail address: kpickett@amnh.org

ML is consistent and parsimony is inconsistent (both of which are false statements when worded so generally) has been, by and large, the lone criterion offered for preferring ML over parsimony. The claim has begun to diminish with increasing evidence (see Goloboff, 2003) but persists nonetheless. We do not believe consistency to be an important criterion for choice among methods of analysis. However, if consistency is to be a criterion, the critical choice among methods hinges on the nature of the data under analysis. That is, it is trivial to say one method is consistent or not, without specifying conditions. Parsimony may perform at its poorest when data follow the i.i.d. assumption, as has usually been the case in simulation studies, but many have argued that natural data do not evolve homogeneously and violate the i.i.d. assumption (Farris, 1973; Siddall and Kluge, 1997; Tuffley and Steel, 1998; Siddall and Whiting, 1999; Steel et al., 2000; Kolaczkowski and Thornton, 2004), and many studies support these claims (Fitch and Markowitz, 1970; Fitch, 1976; Miyamoto and Fitch, 1995; Lockhart et al., 1998; Pollock et al., 1999; Philippe and Lopez, 2001; Lopez et al., 2002; Misof et al., 2002; Pupko and Galtier, 2002). Mathematical assertions and simulations that show parsimony performs poorly when analysing data generated under model homogeneity do not inform whether parsimony will perform well when data are generated under conditions of model heterogeneity—that is, when biological evolution has taken place.

### Heterogeneity and homogeneity

When we speak of heterogeneity of the model process, we are referring to a general and relaxed type of rate heterogeneity in which sites are permitted to have variable rates through evolutionary time. Essentially, this type of heterogeneity permits the process of evolution to change. This general type of rate variation can be modeled in more or less restrictive ways. Some very general models (Farris, 1973; Tuffley and Steel, 1997) allow each character a potentially unique rate for every branch on the tree, not demanding any to be invariable. An earlier method, called the *covarion* model (Fitch and Markowitz, 1970) permits some sites to be variable and others to be invariable at a given time, but later, which sites are permitted to vary and which are prohibited can change. Both of these frameworks permit a general type of rate heterogeneity, which has also been termed *heterotachy* (Lopez et al., 2002), even though the covarion model allows a more assumption-laden flavor of heterotachy. Heterotachy-permissive models violate the “identically distributed” component of the i.i.d. assumption because they permit *within*-site rate heterogeneity (Steel et al., 1993; Lopez et al., 2002). Therefore models that permit heterotachy do not assume a stationary substitution process in which sites maintain the same

relative rates through time (but see Tuffley and Steel, 1998 for a covarion-style, i.i.d. conforming formulation, discussed below). A more assumption-laden restriction of this general form of heterogeneity corresponds to the familiar discrete approximation of the gamma distribution (Yang, 1994). In this formulation, although rates are permitted to vary at different sites across a tree, at a given place in the tree, the rates are the same or proportional (Chang, 1996; Gaucher et al., 2001); that is, a given site is not allowed to have a variable rate through time (this type of rate accommodation is sometimes called a “rates-across-sites”, or *RAS* model). Typical ML models—including all homogeneous Markov process models and the gamma approximation—assert this assumption of *homotachous* (Lopez et al., 2002) evolution. The assumption of homotachy is also true of the (usual) invariants model, and this model is even more restrictive than the gamma approximation.

Selection and functional constraints, among other phenomena, cause data to violate the i.i.d. assumption, and so it should be uncontroversial to state that many forms of natural data, especially protein-coding loci, evolve heterotachously. This was the basis upon which Farris (1973) asserted that ML models should permit model heterogeneity and not employ parametric, “typical” statistics. Despite this, Felsenstein (1978) generated a model framework that does precisely what Farris admonished as unrealistic (assumed homogeneity of the evolutionary process). For some time following this, parsimony has been criticized repeatedly for not being consistent under the conditions set forth by Felsenstein (1978), despite the general knowledge that natural data do not conform to those unrealistic assumptions. Since, LBA of parsimony under the unrealistic assumption of data homotachy has been used as a general argument against parsimony when applied to real data.

### Detecting LBA

Assertions of LBA are common, but the use of methods designed to detect LBA are rare. Suggestions have been offered (Carmean and Crespi, 1995; Huelsenbeck, 1997), but these rely either on the notion that parsimony is suspect of LBA when ML provides different results (Huelsenbeck, 1997), or that parsimony is suspect of LBA when resampling support is high (Carmean and Crespi, 1995). The former criterion was predicted and dismissed in the original demonstration that parsimony can be inconsistent (Felsenstein, 1978), despite the false assertion therein that ML “entirely avoid[s] the problem of statistical inconsistency” (p. 409); nevertheless, it has been the justification for the conclusion of LBA in by far the vast majority of papers claiming examples of LBA (*contra* Bergsten, 2005). The latter criterion is difficult to reconcile with

the notion of support itself (Siddall and Whiting, 1999). Given this and the recent evidence of the conditions leading to the inconsistency of ML (Chang, 1996; Swofford et al., 1996; Kolaczkowski and Thornton, 2004), showing that ML and parsimony differ in results or support cannot indicate which, if either, is inconsistent. See Siddall and Whiting (1999), Grant and Kluge (2003), and Bergsten (2005) for a comprehensive discussion of other proposed tests of LBA.

One method, however, permits the logical detection of LBA without reference to the relative merits of competing methods (Siddall and Whiting, 1999). Because taxa can attract each other only if they are simultaneously present in the analysis, putatively attracting taxa can be removed to see if they influence the placement of apparently attracted taxa (“long branch extraction” of Siddall and Whiting, 1999; henceforth, *extraction test*). The remaining taxa must change position to infer that they were in the wrong place due to LBA. This is a minimum criterion for LBA, but it *must be the case* if LBA is present.

Here we employ the extraction test to attempt to refute LBA in an empirical data set of social wasps. As with all tests, the extraction test cannot prove LBA, only refute it. Therefore, when the extraction test is performed, we will refer to the failure to reject (and support for) the null  $H_{LBA}$ , and its rejection. After extensive testing, we fail to refute LBA in the data set of consideration. In addition to these tests, we provide evidence relating to the prediction that analyses suffering from LBA will recover the phenomenon with increasing certainty as more data (from the same distribution) accumulate. There have been at least two interpretations of this prediction: (1) the bootstrap support for clades suspected of LBA increases (Carmean and Crespi, 1995); and (2) the frequency of clades suspected of LBA across replicates increases (Kolaczkowski and Thornton, 2004). We accommodate both interpretations here. Our approach, however, is unique in its reliance on empirical data alone. After presenting these results, we show that the simple addition of more data from the same gene removes the error.

## Data

Using standard PCR protocols, we amplified and sequenced 1261 bases from the Cytochrome Oxidase subunit I gene for the nine nominal species of the swarm-founding social wasp genus *Apoica* and eight exemplars of relatives. The sequence was obtained in two pieces: The first fragment obtained, herein referred to as the *small COI fragment*, extends from alignment position 704 to 1261; it was amplified using the primers *COI-P1 Klompen* (5'-TTGATTTTTGGTCAYCCWGAAGT-3'; positions 2172–2195 in the *Drosophila* mitochondrial

genome) and *COI-4 Klompen* (5'-CCWVYTARDCC-TARRAARTGTTG-3'; positions 2749–2763 in the *Drosophila* mitochondrial genome). The second fragment, herein referred to as the *large COI fragment*, extends from alignment position 1 to 703; it was amplified using the primers *COI-LCO* (5'-GGTCAACAAATCA-TAAAGATATTGG-3'; positions 1489–1514 in the *Drosophila* mitochondrial genome) and *COI-HCOoutout* (5'-GTAAATATATGRTGDGCTC-3'; positions 2330–2348 in the *Drosophila* mitochondrial genome). Sequencing was performed via the dideoxy termination method with dye-labeled terminators using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA polymerase and run on the ABI Prism 3700 DNA analyzer (Perkin-Elmer). Complementary strands were combined and edited with the computer program Sequencher 3.1 (Gene Codes Corporation). Obtained sequences were aligned using default parameters in the multiple alignment program ClustalX (Thompson et al., 1997); a static alignment was necessary for the resampling procedure outlined below. The data are free of indel events, and thus the Clustal alignment is perfectly concordant with the implied alignment (Wheeler, 2003) obtained via POY (Wheeler et al., 2002). All sequences are deposited in GenBank (Accession numbers: AY918908–AY918924).

## Phylogenetic analyses

In an initial empirical analysis of social wasps employing the small COI fragment<sup>1</sup>, we applied parsimony analysis via the following commands in TNT (Goloboff et al., 2003): “mult=tbr replic 50 hold 10”. These commands implement 50 random taxon additions by way of the Wagner tree-building algorithm (Kluge and Farris, 1969) and subjects the resulting trees to tree-bisection reconnection swapping (TBR), with 10 trees held during each round of swapping. In the results of this analysis, four putative long branches appeared to be involved in LBA (Fig. 1a). Of the 17 taxa analyzed, two members of the genus *Apoica* grouped near two morphologically, behaviorally, and geographically distinct tribes (Fig. 1a). *Apoica albimacula* and *Apoica arborea* have the longest branches of any *Apoica* sp. (two-tailed z-test:  $P < 0.001$ ). Ropalidiini and Mischocyttarini have the longest branches of the remaining taxa ( $P < 0.001$ ).

The paraphyly of *Apoica* is without precedent. The phylogeny of the social wasp tribe Epiponini

<sup>1</sup>We made no decision *a priori* to partition the COI data. The reason the data were analyzed separately was simply because we obtained the small COI fragment first, and analyzed it. The results prompted the acquisition of the large COI fragment. In other words, the “partitions” were simply the results of using conserved primers.

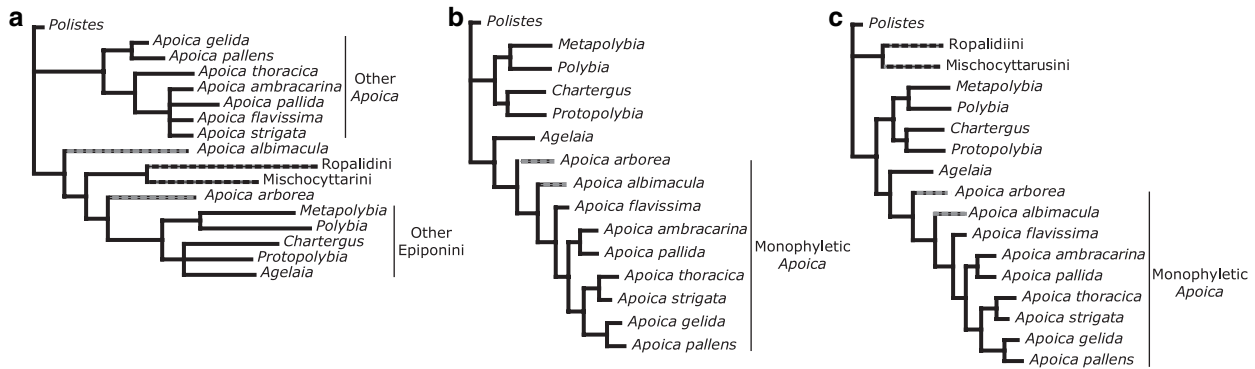


Fig. 1. Apparent Long Branch Attraction in social wasps. (a) Strict consensus of the 12 equally parsimonious solutions for the small COI fragment. *Apoica albimaculata* and *Apoica arborea* (gray with black stripes) group with Ropalidiini and Mischocyttarini (black with gray stripes), rendering *Apoica* paraphyletic. These four branches are unusually long ( $P < 0.001$ ; see text) and are suspected of LBA. (b) The resulting cladogram when Ropalidiini and Mischocyttarini are removed from the matrix and parsimony analysis conducted again. As is required by the predictions of  $H_{LBA}$ , *Apoica albimaculata* and *Apoica arborea* move to a position that recovers the monophyly of *Apoica*. (c) The single most parsimonious tree resulting from parsimony analysis of 1261 bases of the COI gene. *Apoica* is monophyletic.

(Hymenoptera: Vespidae; Polistinae), of which the genus *Apoica* is a member, is well supported by numerous sources of data (Carpenter, 1991; Wenzel, 1993; Wenzel and Carpenter, 1994; Noll et al., 2004). *Apoica* is unique among paper wasps both morphologically and behaviorally, and many aspects of their natural history, including their nocturnality, strongly indicate the unity of the genus. Accordingly, the monophyly of *Apoica* has never been doubted, and no data (save these) suggest that these highly eusocial, swarm founding wasps are close relatives of the primitively eusocial, independent founding Ropalidiini or Mischocyttarini.

To investigate LBA in this situation, we applied the extraction test to the two putatively attracting branches leading to Ropalidiini and Mischocyttarini. If LBA is absent, the two long-branched *Apoica* must stay in their same position relative to the remaining taxa when Ropalidiini and Mischocyttarini are extracted. If LBA is present, *Apoica albimaculata* and *Apoica arborea* are expected to move relative to the remaining taxa—a solution that was obtained (Fig. 1b).

However, when more COI data are added, matters change. When the large COI fragment was added to the small COI fragment and parsimony analysis conducted again, the cladogram does not appear to suffer from LBA (Fig. 1c). *Apoica* is monophyletic, and the two tribes Ropalidiini and Mischocyttarini are in a place concordant with results from morphology and behavior (Carpenter, 1991; Wenzel, 1993; Wenzel and Carpenter, 1994), and the simultaneous analysis of morphology, behavior and molecules (Pickett and Wenzel, in press). Minus Ropalidiini and Mischocyttarini, this tree (Fig. 1c) is topologically identical to the tree resulting from the removal of the attracting taxa in the analysis of the small COI fragment (Fig. 1b).

### Testing for LBA

The results from the extraction test above are, alone, strongly concordant with the predictions of  $H_{LBA}$ . But this apparent inconsistent behavior may merely be an example of character sampling error, and not actual inconsistency. Therefore, we employed a more rigorous test for LBA.

Data matrices were generated via randomly resampling, with replacement, the small COI fragment using Merrin (a PHP application developed by G.L. Tolman, available at <http://www.socialwasps.com/>). Thus we generated new data that were drawn randomly from precisely the same distribution as the original dataset, with all of its natural complexity. Each of the resampled data matrices was composed of either 2000, 4000, 6000, 8000, 10 000 or 100 000 nucleotide base characters. The characters from each preceding matrix were present at the beginning of the following, simulating the accumulation of data. Each of these resampled matrices were merged with the original small COI fragment matrix, creating six matrices of 2558, 4558, 6558, 8558, 10 558 and 100 558 characters. This procedure was implemented 100 times, such that 100 trials of six data matrices each were constructed. Each of the resulting 600 data matrices was subjected to parsimony analysis in TNT as before. Parsimony bootstrap proportions were calculated from 1000 pseudoreplicates for each of the 600 data matrices using the command “resample boot replications=1000” in TNT (see Appendix 1 for an example of the TNT scripts employed).

The extraction test was applied to matrices that resulted in trees containing apparently attracting pairs. Only two worst-case alternate topologies, involving *both* members of putatively attracted pairs, were considered suspect of  $H_{LBA}$  (Fig. 2). Putatively attracting long-branch taxa were extracted from the matrices using the “taxcode” command in TNT, and the parsimony search procedure

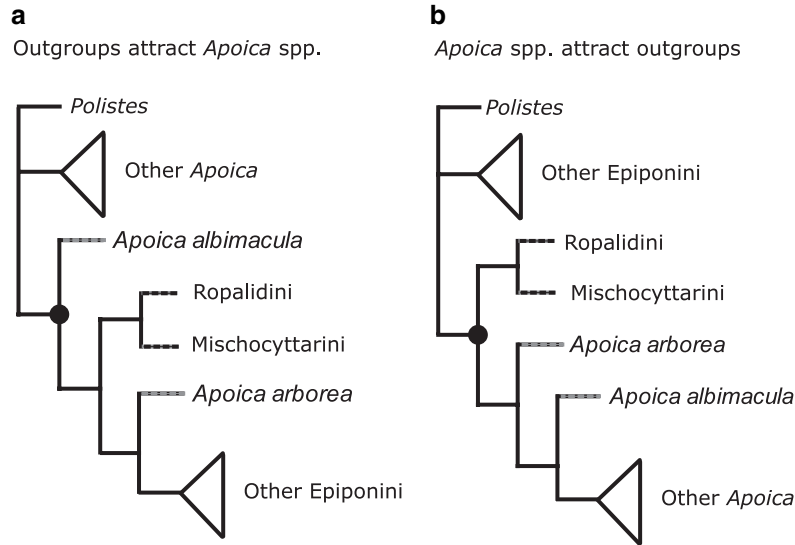


Fig. 2. Topologies considered suspect of LBA, resulting from analyses of resampled data. (a) *Apoica albimaculata* and *Apoica arborea* group with Ropalidiini and Mischocyttarini, rendering *Apoica* paraphyletic. (b) Ropalidiini and Mischocyttarini are sister to *Apoica*, apparently by *A. albimaculata* and *A. arborea* (the two basal-most *Apoica*), rendering Epiponini paraphyletic.

repeated. If both of the putatively attracted taxa moved after removal of their apparent attracters, the trial was counted as support for  $H_{LBA}$ . If both did not,  $H_{LBA}$  was rejected. If parsimony analyses resulted in multiple equally parsimonious trees, for the trial to count as support for  $H_{LBA}$ , every tree had to show movement of the suspect taxa.

The results of this test further conform to the predictions of  $H_{LBA}$ . As sample size grew,  $H_{LBA}$  was increasingly supported, ultimately in 100% of trials (Fig. 3a). As the resampled matrices increased, so increased the bootstrap support for the clades supporting  $H_{LBA}$  (Fig. 3b). This result is expected from theory, which suggests that the inconsistent estimator will recover the incorrect result with increasing certainty as more data from the same distribution are added to the analysis (Felsenstein, 1978). As data increase, however,

bootstrap values might increase in general. Bootstrap values for clades supporting  $H_{LBA}$  were therefore divided by the average bootstrap support for the entire cladogram. These normalized values indicate that increasing bootstrap support for clades supporting  $H_{LBA}$  was not merely an artifact of generally increasing support (Fig. 3b). This is expected in an analysis that is being positively misled (*sensu* Felsenstein) by statistical inconsistency. All the evidence, taken together, supports the hypothesis that parsimony never overcame inconsistency, but was further confounded by it.

**Overcoming LBA**

Having failed to refute LBA in the analysis of the small CO1 fragment despite extensive testing, we investigated if

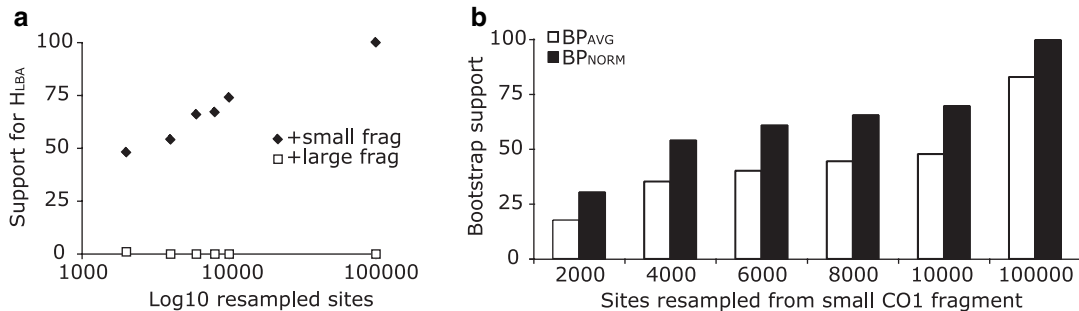


Fig. 3. Parsimony overcomes LBA. (a) Data resampled from the small CO1 fragment and added to the small CO1 fragment (+small frag) increasingly favor  $H_{LBA}$ , ultimately in 100/100 trials. Data resampled from the large CO1 fragment and added to the small CO1 fragment (+large frag) support  $H_{LBA}$  in 1/100 trials with 2558 bases, but in 0 trials with more data. (b) Average ( $BP_{AVG}$ ) and normalized ( $BP_{NORM}$ ) bootstrap proportions for clades supporting  $H_{LBA}$  (dots in Fig. 2) increase as data resampled from the small CO1 fragment are added to the small CO1 fragment (“+small frag” trials in a).

Table 1

*P*-values resulting from the implementation of the heterogeneity test discussed in Ané *et al* (2005). Significant values are in bold. The columns “1, 2”, “3” and “All” refer to base positions. See text for discussion

| Excluded clades               | Small CO1 frag |       |              | Large CO1 frag |              |              | All CO1 data |       |              |
|-------------------------------|----------------|-------|--------------|----------------|--------------|--------------|--------------|-------|--------------|
|                               | 1, 2           | 3     | All          | 1, 2           | 3            | All          | 1, 2         | 3     | All          |
| Ropalidiini + Mischocyttarini | 0.077          | 0.077 | 0.224        | <b>0.031</b>   | 0.522        | <b>0.029</b> | <b>0.050</b> | 0.705 | <b>0.036</b> |
| <i>Apoica</i>                 | <b>0.021</b>   | 0.114 | <b>0.009</b> | <b>0.026</b>   | <b>0.032</b> | 0.072        | <b>0.003</b> | 0.377 | <b>0.002</b> |

the addition of the large CO1 fragment, which need not exhibit the same distributional properties as the original data, had indeed salvaged the analysis. Although LBA was apparently absent when both small and large CO1 fragments were analyzed simultaneously (Fig. 1c), we were again concerned that this result may simply have been due to sampling error, and the addition of more data of the latter type might eventually also support  $H_{LBA}$ .

To address this concern, we randomly resampled characters with replacement from the large CO1 fragment. For this analysis, data matrices were generated and analyzed under the same resampling procedure used for the small CO1 fragment above. In this case, however, data added to the small CO1 fragment were resampled with replacement from the large CO1 fragment. In this way, additional data with the same properties as the large CO1 fragment could be added to the small CO1 fragment. If the apparent escape from LBA afforded by the simultaneous analysis of small and large CO1 fragments was an artifact of a small sample, as more data resampled from the large CO1 fragment are added, the analysis should increasingly favor  $H_{LBA}$ . The results, however, contradict this.  $H_{LBA}$  was supported in only 1 of 100 trials when 2558 bases were analyzed; the inclusion of additional data eliminated support for  $H_{LBA}$  entirely (Fig. 2b).

### Testing for heterotachy

The results above show that the two CO1 fragments exhibit different properties. To investigate some of the details of these differences, we employed a recently implemented test for data heterogeneity (in Ané *et al.*, 2005). In summary, this test examines if: (1) the data are better suited by ( $H_0$ ) a completely homogeneous model versus ( $H_A$ ) a homotachous RAS model (such as gamma), and then (2) if the data are better served by ( $H_0$ ) the homotachous RAS model or ( $H_A$ ) the fusion of a homotachous RAS accommodation plus the covarion-style model of Tuffley and Steel (1998). We say *covarion-style* because this is the language of Tuffley and Steel (1998), owing to the fact that—in choosing rates, i.i.d., from a distribution—their model falls short of the spirit of the covarion framework, which is to accommodate data that violate the i.i.d. assumption. As such, the covarion-style model may not be covarion-like

when applied to finite data. Nevertheless, one can ask if a RAS + covarion-style model is better than RAS alone.

The test of Ané *et al.* (2005) operates via a modification of the statistics set forth in Tuffley and Steel (1998) and Lockhart *et al.* (1998) (see details therein). The test investigates rate heterogeneity on branches subtending clades identified by the investigator. For this test, the topology deriving from all the data was used as the reference topology. We would ideally identify clades that include only the long-branch taxa. That can be accomplished for Ropalidiini and Mischocyttarini. However, the two long-branch *Apoica* do not form a clade, and so cannot be removed to the exclusion of other taxa. Instead, we present tests for the branch leading to the clade (Ropalidiini + Mischocyttarini) and the clade of all *Apoica*, the two basal taxa of which are the putative long branch taxa. The tests were performed for all 1st and 2nd codon position nucleotides (“1, 2” of Table 1), all third position nucleotides (“3” of Table 1), and all positions (“All”) for the small fragment alone, the large fragment alone, and all the data combined.

The first heterogeneity test (homogeneous versus RAS) showed significant heterogeneity for all branches and data partitions investigated (results not shown). The results from the second heterogeneity test (RAS versus RAS + covarion-style) appear in Table 1 (significant results in bold). With respect to the branches and nucleotide positions investigated, the small CO1 fragment exhibits less covarion-style heterotachy than the large CO1 fragment (2/6 tests versus 4/6 tests). In addition, considering all nucleotide positions per partition in the small CO1 fragment, the test favored heterotachy for the branch leading to *Apoica*, but not (Ropalidiini + Mischocyttarini); the opposite was true for the large CO1 fragment. Heterotachy for both branches was found in the first and second positions of the large CO1 fragment and when all positions from both partitions were analyzed simultaneously. These results therefore support the deduction that the differential behavior of the partitions implies differences in heterotachy, and that the heterotachy of the gene is maximized when all data are analyzed simultaneously.

### Conclusions

Our findings show that if a phylogenetic analysis using parsimony falls prey to a condition concordant with the

predictions of LBA, the addition of data that conform to the distributional properties of the original data are unable to salvage the analysis, as predicted by theory (Felsenstein, 1978). The addition of a relatively small amount of data that does not conform to the distributional properties of the original data, however, can rescue the analysis from LBA. This finding agrees with recent simulation (Kolaczowski and Thornton, 2004), which shows that the presence of heterotachy can permit the consistency of parsimony under branch-length conditions thought permissive of LBA. Furthermore, this necessary model heterogeneity exists within a single gene, one of the most commonly used in phylogenetic analyses.

Interestingly, the 558-base small CO1 fragment, though exhibiting some heterotachy, is sufficiently homogeneous to cause parsimony to suffer from a condition indistinguishable from LBA. This is only logical, given that smaller gene fragments will always be less heterogeneous than larger data samples. When both the small and large CO1 fragments of the gene are analyzed simultaneously, however, the data exhibit greater heterotachy than either fragment analyzed alone (see Table 1), and parsimony is not confounded. This is not surprising, as protein-coding loci are well known for their evolutionary heterogeneity (Fitch, 1976; Philippe and Lopez, 2001; Lopez et al., 2002; Pupko and Galtier, 2002). When multiple loci and morphology are combined in increasingly popular simultaneous analyses (Kluge, 1989; Nixon and Carpenter, 1996), there may be little reason to suspect data homogeneity. Given this, parsimony need not be generally overlooked for fear of statistical inconsistency.

Our results support the theoretical concerns regarding the inconsistency of phylogenetic estimation using parsimony when datasets are small. Our results also demonstrate that when additional natural data are analyzed, heterotachy is enhanced, and the data no longer follow the strict requirements upon which those theoretical concerns are contingent. In short, a parsimony analysis can escape the Felsenstein Zone if only the investigator adds more data, which is in line with a quite general view of materialist investigation.

### Acknowledgments

This work was supported by The Ohio State University (J.W.W.), an OSU Presidential Fellowship (K.M.P.), the Theodore Roosevelt Fellowship from the American Museum of Natural History (K.M.P.), and NASA (NAG5-13028 to W.C.W.). We thank Julian Faivovich, James S. Farris, John V. Freudenstein, Mark Siddall, Mark P. Simmons, and Mike Steel for discussion and recommendations over the last four years. Pablo Goloboff kindly offered assistance with TNT macro language, and Cécile Ané offered generous

support with her heterogeneity program. Michael Engel and Diego Pol provided much needed encouragement. Mike Whiting and anonymous reviewers provided helpful comments on the manuscript.

### References

- Ané, C., Burleigh, J.G., McMahon, M.M., Sanderson, M.J., 2005. Covarion structure in plastid genome evolution: a new statistical test. *Mol. Biol. Evol.* 22, 914–924.
- Bergsten, J., 2005. A review of long-branch attraction. *Cladistics*, 21, 163–193.
- Carmean, D., Crespi, B.J., 1995. Do long branches attract flies? *Nature*, 373, 666.
- Carpenter, J.M., 1991. Phylogenetic relationships and the origin of social behavior in the Vespidae. In: Ross, K.G., Matthews, R.W. (Eds.), *Social Biology of Wasps*. Cornell University Press, Ithaca, pp. 7–32.
- Chang, J.T., 1996. Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. *Math. Biosci.* 134, 189–215.
- Farris, J.S., 1973. A probability model for inferring evolutionary trees. *Syst. Zool.* 22, 250–256.
- Farris, J.S., 1999. Likelihood and inconsistency. *Cladistics*, 15, 199–204.
- Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27, 401–410.
- Fitch, W.M., 1976. The molecular evolution of cytochrome c in eukaryotes. *J. Mol. Evol.* 8, 13–40.
- Fitch, W.M., Markowitz, E., 1970. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. *Biochem. Genet.* 4, 579–593.
- Gaucher, E.A., Miyamoto, M.M., in press. A call for likelihood phylogenetics even when the process of sequence evolution is heterogeneous. *Mol. Phyl. Evol.* 00, 000–000, in press. doi:10.1016/j.ympev.2005.03.027.
- Gaucher, E.A., Miyamoto, M.M., Benner, S.A., 2001. Function–structure analysis of proteins using covarion-based evolutionary approaches: Elongation factors. *Proc. Nat. Acad. Sci. USA*, 98, 548–552.
- Goloboff, P., 2003. Parsimony, likelihood, and simplicity. *Cladistics*, 19, 91–103.
- Goloboff, P., Farris, J.S., Nixon, K., 2003. TNT (Tree analysis using New Technology), Version 1.0. Published by the authors, Tucumán, Argentina.
- Grant, T., Kluge, A.G., 2003. Data exploration in phylogenetic inference: scientific, heuristic, or neither. *Cladistics*, 19, 379–418.
- Hendy, M.D., Penny, D., 1989. A framework for the quantitative study of evolutionary trees. *Syst. Zool.* 38, 297–309.
- Huelsenbeck, J.P., 1997. Is the Felsenstein zone a fly trap? *Syst. Biol.* 46, 69–74.
- Huelsenbeck, J.P., Hillis, D.M., 1993. Success of phylogenetic methods in the 4-taxon case. *Syst. Biol.* 42, 247–264.
- Kluge, A.G., 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* 38, 7–25.
- Kluge, A.G., Farris, J.S., 1969. Quantitative phyletics and the evolution of anurans. *Syst. Zool.* 18, 1–32.
- Kolaczowski, B., Thornton, J.W., 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature*, 431, 980–984.
- Lockhart, P.J., Steel, M.A., Barbrook, A.C., Huson, D.H., Charleston, M.A., Howe, C.J., 1998. A covarion model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. *Mol. Biol. Evol.* 15, 1183–1188.
- Lopez, P., Casane, D., Philippe, H., 2002. Heterotachy, an important process of protein evolution. *Mol. Biol. Evol.* 19, 1–7.

- Misof, B., Anderson, C.L., Buckley, T.R., Erpenbeck, D., Rickert, A., Misof, K., 2002. An empirical analysis of mt 16S rRNA covarion-like evolution in insects: Site-specific rate variation is clustered and frequently detected. *J. Mol. Evol.* 55, 460–469.
- Miyamoto, M.M., Fitch, W.M., 1995. Testing the covarion hypothesis of molecular evolution. *Mol. Biol. Evol.* 12, 503–513.
- Nixon, K.C., Carpenter, J.M., 1996. On simultaneous analysis. *Cladistics*, 12, 221–241.
- Noll, F.B., Wenzel, J., Zucchi, R., 2004. Evolution of caste in neotropical swarm-founding wasps (Hymenoptera: Vespidae: Epiponini). *Am. Mus. Nov.* 3467, 1–24.
- Philippe, H., Lopez, P., 2001. On the conservation of protein sequences in evolution. *Trends Biochem. Sci.* 26, 414–416.
- Pickett, K.M., Wenzel, J.W., in press. Revision and cladistic analysis of *Apoica* Lepeletier (Hymenoptera: Vespidae: Polistinae, Epiponini). *Am. Mus. Nov.* 00, 000–000.
- Pollock, D.D., Taylor, W.R., Goldman, N., 1999. Coevolving protein residues: Maximum likelihood identification and relationship to structure. *J. Mol. Evol.* 287, 187–198.
- Pupko, T., Galtier, N., 2002. A covarion-based method for detecting molecular adaptation; applicaiton to the evolution of primate mitochondrial genomes. *Proc. R. Soc. Lond. B*, 269, 1313–1316.
- Sanderson, M.J., Kim, J., 2000. Parametric phylogenetics? *Syst. Biol.* 49, 817–889.
- Siddall, M.E., Kluge, A.G., 1997. Probabilism and phylogenetic inference. *Cladistics*, 13, 313–336.
- Siddall, M.E., Whiting, M.F., 1999. Long-branch abstractions. *Cladistics*, 15, 9–24.
- Spencer, M., Susko, E., Roger, A.J., 2005. Likelihood, parsimony, and heterogeneous evolution. *Mol. Biol. Evol.* 22, 1161–1164.
- Steel, M., 2005. Should phylogenetic models be trying to ‘fit an elephant’? *Trends Genet.* 21, 307–309.
- Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), *Molecular Systematics*, 2nd edn. Sinauer Associates, Sunderland, MA, pp. 407–514.
- Spencer, M., Susko, E., Roger, A.J., 2005. Likelihood, parsimony, and heterogeneous evolution. *Mol. Biol. Evol.* 22, 1161–1164.
- Steel, M., Huson, D., Lockhart, P., 2000. Invariable sites models and their use in phylogeny reconstruction. *Syst. Biol.* 49, 225–232.
- Steel, M., Szekely, L., Erdos, P.L., Waddell, P., 1993. A complete family of phylogenetic invariants for any number of taxa under Kimura’s 3ST model. *NZ J. Bot.* 31, 289–296.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acid. Res.* 25, 4876–4882.
- Tuffley, C., Steel, M., 1997. Links between maximum likelihood and maximum parsimony under a simple model of site substitution. *Bull. Math. Biol.* 59, 581–607.
- Tuffley, C., Steel, M., 1998. Modeling the covarion hypothesis of nucleotide substitution. *Math. Biosci.* 147, 63–91.
- Wenzel, J.W., 1993. Application of the biogenetic law to behavioral ontogeny – a test using nest architecture in paper wasps. *J. Evol. Biol.* 6, 229–247.
- Wenzel, J.W., Carpenter, J.M., 1994. Comparing methods: Adaptive traits and tests of adaptation. In: Eggleton, P., Vane-Wright, R.I., (Eds.), *Phylogenetics and Ecology*. Academic Press, London, pp. 79–101.
- Wheeler, W.C., 2003. Implied alignment: a synapomorphy-based multiple-sequence alignment method and its use in cladogram search. *Cladistics*, 19, 261–268.
- Wheeler, W.C., De Laet, J., Gladstein, D.S., 2002. POY: The optimization of alignment characters, Version 3.0.4. Program and Documentation. Available at <ftp://ftp.amnh.org/pub/molecular>. Documentation by D. Janies and W.C. Wheeler, New York, NY.
- Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J. Mol. Evol.* 39, 306–314.

## Appendix 1

```

loop 1 100
log/
log %1.homo_#1_2000.txt;
mxram 1000;
proc %1.homo_#1_2000.ss
hold 100000;
mult=tbr replic 50 hold 10;
nel;
export* %1.homo_#1_2000.nex;
resample boot replications 1000 savetrees;
taxcode -Apoica_albimacula -Apoica_arborea;
mult=tbr replic 50 hold 10;
nel;
xread-;
export* %1.homo_#1_2000_minus_aa.nex;
taxcode + Apoica_albimacula + Apoica_arborea
-Ropalidia_sp. -Mischoctytarus_sp.;
mult=tbr replic 50 hold 10;
nel;
xread-;
export* %1.homo_#1_2000_minus_rm.nex;
proc/
stop
//starts 100 loops of the following//
//close previous log file//
//outputs log files for each loop//
//sets RAM buffer, in MB//
//reads numbered data file//
//sets tree buffer//
//runs analysis//
//calculates strict consensus//
//saves matrix, trees to nexus//
//conducts bootstrap analysis//
//removes two long-branch taxa//
//re-runs analysis//
//calculates strict consensus//
//prints new matrix to screen//
//saves new matrix. trees to nexus//
//replaces two taxa, removes two others//
//re-runs analysis//
//calculates strict consensus//
//prints new matrix to screen//
//saves new matrix, trees to nexus//

```