



A review of long-branch attraction

Johannes Bergsten

Department of Ecology and Environmental Science, Umeå University, SE-90187 Umeå, Sweden

Accepted 14 February 2005

Abstract

The history of long-branch attraction, and in particular methods suggested to detect and avoid the artifact to date, is reviewed. Methods suggested to avoid LBA-artifacts include excluding long-branch taxa, excluding faster evolving third codon positions, using inference methods less sensitive to LBA such as likelihood, the Aguinaldo et al. approach, sampling more taxa to break up long branches and sampling more characters especially of another kind, and the pros and cons of these are discussed. Methods suggested to detect LBA are numerous and include methodological discordance, RASA, separate partition analyses, parametric simulation, random outgroup sequences, long-branch extraction, split decomposition and spectral analysis. Less than 10 years ago it was doubted if LBA occurred in real datasets. Today, examples are numerous in the literature and it is argued that the development of methods to deal with the problem is warranted. A 16 kbp dataset of placental mammals and a morphological and molecular combined dataset of gall wasps are used to illustrate the particularly common problem of LBA of problematic ingroup taxa to outgroups. The preferred methods of separate partition analysis, methodological discordance, and long branch extraction are used to demonstrate detection methods. It is argued that since outgroup taxa almost always represent long branches and are as such a hazard towards misplacing long branched ingroup taxa, phylogenetic analyses should always be run with and without the outgroups included. This will detect whether only the outgroup roots the ingroup or if it simultaneously alters the ingroup topology, in which case previous studies have shown that the latter is most often the worse. Apart from that LBA to outgroups is the major and most common problem; scanning the literature also detected the ill advised comfort of high support values from thousands of characters, but very few taxa, in the age of genomics. Taxon sampling is crucial for an accurate phylogenetic estimate and trust cannot be put on whole mitochondrial or chloroplast genome studies with only a few taxa, despite their high support values. The placental mammal example demonstrates that parsimony analysis will be prone to LBA by the attraction of the tenrec to the distant marsupial outgroups. In addition, the murid rodents, creating the classic “the guinea-pig is not a rodent” hypothesis in 1996, are also shown to be attracted to the outgroup by nuclear genes, although including the morphological evidence for rodents and Glires overcomes the artifact. The gall wasp example illustrates that Bayesian analyses with a partition-specific GTR + Γ + I model give a conflicting resolution of clades, with a posterior probability of 1.0 when comparing ingroup alone versus outgroup rooted topologies, and this is due to long-branch attraction to the outgroup.

© The Willi Hennig Society 2005.

Introduction to long-branch attraction

Until rather recently, long-branch attraction (LBA), the erroneous grouping of two or more long branches as sister groups due to methodological artifacts, was merely considered hypothetical, and it was doubted if

it affected real data. A few cases were suggested to be examples of LBA (Carmean and Crespi, 1995; Huelsenbeck, 1997), but these were criticized (Whiting, 1998; Siddall and Whiting, 1999) and even in a recent student textbook in systematics it was written that “There was, and still is, some question of whether ‘long-branch attraction’ actually occurs...” (Schuh, 2000, p. 140). The controversy, especially over the claim that Halteria (Diptera + Strepsiptera) is due to long-branch attraction was certainly appropriate since the claim was partly

Corresponding author.

E-mail address: johannes.bergsten@emg.umu.se

based on naïve (Carmean and Crespi, 1995) or at best inadequate evidence (Huelsenbeck, 1997) where critical morphological data were ignored, taxonomic sampling was extremely poor and the result actually hinged upon a single character in the alignment (Whiting, 1998; Siddall and Whiting, 1999; see also Whiting et al., 1997; Wheeler et al., 2001). Presently suggested examples of LBA are no longer merely a few; 112 hits were recovered in a search of the Web of Science database on “long-branch attraction”, many of these being case studies (43, according to Andersson and Swofford, 2004). This does not reflect the true number of LBA situations, since they are often disguised under expressions such as “not taking rate-heterogeneity into account”, “model misperfection” and the like. To name a few, organism groups and taxonomic levels in which LBA problems have been suggested range from species-level phylogenies of *Daphnia* (Omilian and Taylor, 2001), bees (Schwarz et al., 2004) and beetles (Bergsten and Miller, 2004) through genus and family level studies on teleost fishes (Tang et al., 1999; Clements et al., 2003), iguanid lizards (Wiens and Hollingsworth, 2000) orthopteroid insects (Flook and Rowell, 1997), ordinal-level phylogenies of insects (Carmean and Crespi, 1995; Huelsenbeck, 1997; Steel et al., 2000), mammals (Philippe, 1997; Sullivan and Swofford, 1997; Waddell et al., 2001; Lin et al., 2002a,b) and birds (Garcia-Moreno et al., 2003), class- and phylum level trees of metazoans (Aguinaldo et al., 1997; Kim et al., 1999), angiosperm plants (Soltis and Soltis, 2004), seed plants (Sanderson et al., 2000) and red algae (Moreira et al., 2000), up to the basal kingdom relationships of Eukaryota (Moreira et al., 1999; Archibald et al., 2002; Dacks et al., 2002; Inagaki et al., 2004), Bacteria (Bocchetta et al., 2000) and the very tree of life (Brinkmann and Philippe, 1999; Lopez et al., 1999; Philippe and Forterre, 1999; Gribaldo and Philippe, 2002). Likewise, a few examples will suffice to illustrate the wide range of data involved, from nuclear (CCTalpha: Archibald et al., 2002; EF-1 α : Moreira et al., 1999), mitochondrial (Cyt b: Kennedy et al., 1999; Wiens and Hollingsworth, 2000; CO1: Bergsten and Miller, 2004; ND4: Wiens and Hollingsworth, 2000) and chloroplast (*psaA*, *psbB*: Sanderson et al., 2000) protein coding genes, analyzed as nucleotides as well as transformed into amino acid data (Lin et al., 2002a,b; Nikaido et al., 2003; Goremykin et al., 2003; see Soltis and Soltis, 2004), nuclear (18S: Aguinaldo et al., 1997; Kim et al., 1999; 28S: Philippe and Germot, 2000; Omilian and Taylor, 2001) and mitochondrial (12S, 16S: Tang et al., 1999) ribosomal genes to complete mitochondrial (Philippe, 1997; Sullivan and Swofford, 1997; Lin and Penny, 2001) and chloroplast genomes (Soltis and Soltis, 2004), and even morphological characters (Wiens and Hollingsworth, 2000; Lockhart and Cameron, 2001), as well as word data in linguistic applications of cladistic methods (Rexova et al., 2003).

Finally, these suggested LBA results are not confined to any particular inference method of phylogeny, but include the use of parsimony (Aguinaldo et al., 1997; Kennedy et al., 1999; Tang et al., 1999), neighbor-joining (Moreira et al., 1999; Philippe and Forterre, 1999), likelihood (Brinkmann and Philippe, 1999; Sanderson et al., 2000; Omilian and Taylor, 2001), and Bayesian methods (Schwarz et al., 2004); the latter methods with varying complexities of substitution models, from JC69 to GTR + Γ + I.

Although admittedly, a few of these LBA examples might again be found to be suggested on weak grounds, the majority of cases have shown with varying tests that LBA is the most corroborated and least refuted hypothesis of the data. In addition, the empirical examples rest on a firm background knowledge from analytical and simulational work of the phenomenon. The claim that LBA has never been proven in real datasets should be of no interest to any falsificationist. Accordingly it is no longer justified, productive or even scientific to claim that LBA is purely hypothetical, although in a specific case it can naturally be justified to question it. Rather, systematists are better off being aware of and trying to deal with the problem. I agree with Siddall and Kluge (1997) that the truth is intractable and that consistency is of rather shallow interest since we are all dealing with finite datasets. However, finite datasets are what, in real examples, are suffering from LBA, and demonstrating that LBA is the least refuted hypothesis of a spurious outcome is not intractable with the combination of tests suggested to date. I therefore briefly outline the past history of LBA and then review and comment on suggested and previously used methods to detect and avoid LBA. By way of example, I re-analyze the hitherto largest nuclear dataset of placental mammals (16 kbp; Murphy et al., 2001b) as well as a combined analysis of gall wasps (Nylander et al., 2004) to demonstrate some of the reviewed methods for the detection and avoidance of LBA. Finally I draw some general conclusions from the review, and suggest what steps should be taken as standard practice in phylogenetic analyses.

Long-branch attraction basics and history

My usage of the term LBA is equivalent to that of Sanderson et al. (2000, p. 782) “conditions under which bias in finite dataset[analyse]s and/or statistical inconsistency arise due to a combination of long and short branches” [in brackets my addition] and that by Andersson and Swofford (2004, p. 441) “any situation in which similarity due to convergent or parallel changes produces an artifactual phylogenetic grouping of taxa due to an inherent bias in the estimation procedure” in that they deal with finite datasets. Thus inconsistency, the property of a method to converge on the wrong

answer as infinite amount of data are gathered, is possible but not a prerequisite for LBA. This might have been somewhat confusing, because in the theoretical literature LBA has often been dealt with in association with methodological inconsistency. Bias means here that when analyzed by a method, mistaken inferences of relationship are not random, but certain incorrect topologies are preferred over others, and this is due to a factor not allowed for by the inference method. Although perhaps somewhat misleading, I include the equal branch length example of Kim (1996) in the concept of LBA, although the “combination of long and short branches” according to Sanderson et al.’s (2000, p. 782) definition here must be interpreted as “a combination of long and short paths”.

LBA is a phenomenon of molecular data in particular. An “A” or “ala” at a certain position inherited from a common ancestor in two lineages look identical to if the “A” or “ala” had been independently acquired. Since the number of different nucleotides is limited to four (and amino acids 20), the convergent indistinguishable evolution of characters is deemed to be common. Thus two long-enough non-sister branches, separated by a short enough internode, will by chance independently have acquired more identical bases, that will be judged as synapomorphies in a parsimony analysis, than the few number of inherited changes on the short internode grouping one of the long branches with its true relative (Felsenstein, 2003). The most parsimonious solution would, in such a case, be to erroneously group the two long branches as sister groups resulting in LBA. Morphological characters, although in theory not immune to LBA, should not be so commonly affected (Grant and Kluge, 2003). Firstly, because a much larger possible number of character states exist in morphology, as opposed to the limited four possible different states in DNA sequence data (Jenner, 2004). From this follows that convergent evolution will, to a higher degree, already be detected as not homologous at the character scoring state, and homoplasy thereby avoided to a larger degree than is possible with nucleotides. There also exists empirical evidence that morphological datasets in general experience less homoplasy than molecular datasets (Baker et al., 1998). Although not impossible, the cases thus far suggested of LBA from morphological data, by Lockhart and Cameron (2001) and Wiens and Hollingsworth (2000) are not convincing and this paper deals foremost with phylogenetic analysis using molecular data, if not otherwise stated.

Long-branch attraction was first demonstrated in theory by Felsenstein (1978), based on a four-taxon case, with unequal evolutionary rates and parsimony or maximum compatibility as optimality criterion (recently further explored by Schulmeister, 2004). Hendy and Penny (1989) expanded the theoretical conditions under which parsimony can become inconsistent because of

LBA to include cases with equal rates along all lineages. They looked at a five-taxon case and concluded that it is not necessarily unequal rates, but unequal branch-lengths that can cause LBA, and unequal branch-lengths can be caused by either unequal rates, or as a consequence of a non-symmetric topology. In phylogenetic analysis the latter can be real and due to differential speciation/extinction rates along lineages of the study group, or simply a consequence of incomplete taxon sampling. Finally, Kim (1996) showed that parsimony can be inconsistent, even if all branches are of the same length, although the tree needs to be unbalanced (asymmetrical).

A number of studies followed to test the predictions of earlier work and to compare different phylogeny reconstruction methods. In general, experimental laboratory-generated viral phylogenies were well recovered by most methods, but did not really include cases where LBA-artifacts might provide difficulties (Hillis et al., 1992, 1994a; Bull et al., 1993). Multiple simulation studies with finite datasets and the now classical four-taxon unrooted tree indeed showed that inference was difficult in what has been termed the “Felsenstein-zone”, where two long-branched non-sister taxa grouped together rather than with their true shorter branched sister (Huelsenbeck and Hillis, 1993; Hillis et al., 1994a,b; Huelsenbeck, 1995). Maximum likelihood was less sensitive to LBA than parsimony. However, the tree space modeled was always one in which the two long branches never occurred as sister taxa (Huelsenbeck, 1995), and a number of studies have shown that ML can also become inconsistent and suffer from LBA artifacts when the model assumptions are violated (Gaut and Lewis, 1995; Chang, 1996; Lockhart et al., 1996; Sullivan and Swofford, 1997; see also Farris, 1999). When Siddall (1998) instead modeled a tree space where the long branches were sister taxa, parsimony outperformed likelihood methods, which instead, according to Siddall suffered from “long-branch repulsion” (see also Pol and Siddall, 2001). It seemed thus as if there was no difference in the overall performance of recovering the true phylogeny between the two competing methods in the four-taxon case. However Swofford et al.’s (2001) simulation study traveled on a tree-axis between the Felsenstein-zone and the “long-branch repulsion” zone (instead of only moving within one of these as in the previous studies) and on this occasion, maximum likelihood again performed best overall (Swofford et al., 2001, their Fig. 7). Although Pol and Siddall’s (2001) simulation on a 10-taxon model tree revealed new differences in performance between methods dealing with long branches, earlier studies have by now taught the reader how dependent the outcome of such comparisons is on the choice of model tree and branch lengths (in this case long branches were either sister taxa or far apart in the tree, thus favoring parsimony). The

conclusion of likelihood having a problem in correctly placing a tree with a single long branch, while parsimony had no such problem (Pol and Siddall, 2001), was because the correct position was now on the second longest branch of the tree, and if this was changed, the superiority of the methods was reversed (Andersson and Swofford, 2004). Likewise, a recent simulation study by Kolaczkowski and Thornton (2004) seems to be another example of a well chosen simulation set up that favored parsimony in the comparison. In trying to mimic heterogeneous sequence evolution, Kolaczkowski and Thornton simulated two independent partitions on two different four-taxon trees where long and short branched terminals were reversed, and concatenated the two alignments into one. The partitions being equal in size and the reversed branch lengths being balanced, for parsimony, LBA artifacts equaled out by the two partitions, while for likelihood-based methods, the estimated branch lengths being an average, violates the true model in both partitions. Consequently, these types of simulation now seem to have been exhaustively investigated, with the simple conclusion that the larger bias of parsimony to group long-branched taxa together will yield positive results if they are sister groups in the model tree, while it will yield negative results if they are not.

All the above simulations were based on a single model tree, whereas Huelsenbeck and Lander (2003) approached the problem by asking how often do conditions (trees) arise under a simple model of cladogenesis (linear birth-death process), where parsimony will be inconsistent. Bifurcating trees of seven different total lengths with five to eight species were generated under the model and under three sampling schemes where these five to eight species represent all, 10% or 1% of the real number of species. After the probability of all possible character patterns had been calculated for each tree, assuming a Jukes–Cantor process of DNA substitution, the proportion of occasions in which parsimony was inconsistent could be estimated. Interesting and important trends can be read from the result; as expected, terminal branch lengths increased and so did the inconsistent estimate of the tree when the taxon sampling proportion decreased. With complete taxon sampling or with short total tree length, parsimony is rarely inconsistent. It is interesting to note that the inconsistent estimate increases with the number of species, which is the same conclusion reached by Kim (1996). The eight-taxon tree is about 10-fold more likely to be inconsistently estimated compared to the five-taxon tree. This means that the conditions under which the well studied four-taxon tree is inconsistently estimated are more severe than they need to be for larger trees. Under the worst conditions for the eight-taxon tree, with taxon sampling being 1% and a total tree length of one expected substitution per site from root to tip, parsimony was inconsistent 13% of the time. For calculative reasons

Huelsenbeck and Lander's (2003) study was limited to parsimony. I have summarized these studies only because they are relevant to how researchers have identified possible LBA artifacts in real cases, and whether the methods to detect LBA are justified or not. I do not intend to argue for or against parsimony or maximum likelihood in general, since that includes a range of other considerations apart from sensitivity to LBA: calculation efficiency, dealing with morphological data and scientific philosophical considerations being a few. However, the most important conclusions of the simulation studies with regard to LBA are:

1. Although in the beginning LBA was only discussed as a potential problem for parsimony, results from distance analysis as well as maximum likelihood can likewise be subject to LBA artifacts, in particular when the model assumptions are violated (Gaut and Lewis, 1995; Huelsenbeck, 1995; Chang, 1996; Lockhart et al., 1996), but also with the correct model but finite datasets (Yang, 1997). That is, no method is perfect under all conditions.

2. Nevertheless, parsimony undoubtedly has a stronger bias towards grouping long branches together (rightfully as well as wrongfully), than methods trying to account for unequal rates or branch lengths and then correcting for unobserved changes (Pol and Siddall, 2001; Swofford et al., 2001).

3. We should be worried about LBA in real datasets, especially when taxon sampling is poor.

4. The outcome of comparative simulation studies are highly dependent on the subjective choice of model tree and branch length space under investigation.

Methods to avoid LBA artifacts

A number of methods have been suggested and applied in order to diminish the risk of having an outcome affected by LBA. To use a method less sensitive to LBA, like various modifications of parsimony (Lake, 1987; Willson, 1999) or, more commonly, maximum likelihood (ML) is one proposed (Swofford et al., 1996; Huelsenbeck, 1997) and widely used approach (Huelsenbeck, 1998; Kennedy et al., 1999; Kim et al., 1999). However, since ML is more resistant to LBA but not immune, it can never serve as a valid approach by itself. The plethora of different models available is problematic and using the likelihood ratio test (Posada and Crandall, 1998) is in no way a guarantee against LBA. In addition, the problem of choosing accelerates as multiple genes are simultaneously analyzed; any model for one gene could be combined with any other model applied to the other gene, etc. (Ronquist and Huelsenbeck, 2003). Nevertheless, model improvements such as taking rate-heterogeneity across sites into account with a discrete gamma-distribution parameter

have proved very useful in overcoming LBA (Sullivan and Swofford, 1997; Nikaido et al., 2003).

Exclusions

It is common that third codon positions are excluded based on the notion they are too fast evolving, saturated or randomized (Swofford et al., 1996; Sullivan and Swofford, 1997), and this may be an adequate way of reducing LBA artifacts (Lyons-Weiler and Hoelzer, 1997). Indeed such an approach may reduce LBA artifacts, since greater evolutionary rates cause parsimony at least to increase the probability of inferring the wrong tree (Huelsenbeck and Lander, 2003). However, this method may simultaneously pay a very high cost of reduction in resolution (Källersjö et al., 1999) and can as a consequence not be recommended in general. Neither is deleting data a very valiant scientific endeavor in general. To simply exclude long branched taxa, identified on various grounds, to avoid “confounding effects” is practiced by some (Hanelt et al., 1996; Lyons-Weiler and Hoelzer, 1997; Farias et al., 2001; Dacks et al., 2002) and although this approach might be successful in reducing LBA artifacts, it is certainly of no help if the relationship of those same LB taxa within the study group is of interest.

A related approach, as suggested by Aguinaldo et al. (1997), and followed by Kim et al. (1999) and Brinkmann and Philippe (1999) is to select taxon representatives of larger clades based on their evolutionary rate, and only to include in the analysis low-rate representative taxa. Indeed this approach had a drastic effect in the study by Aguinaldo et al. (1997), where the nematode clade represented by both slow and fast evolving genera grouped basally in the metazoan tree (attracted to the cnidarian outgroup), while when represented only by a slowly evolving genus, the nematode found a place next to arthropods, in agreement with an Ecdysozoa clade. Taxa are thus similarly excluded in this approach, but the slight difference being that in the “simply exclude long branches” approach, a test identifies which branches are long and excludes them, while Aguinaldo et al. (1997) first identified which major clades are of interest and then kept the single representative of each major clade with the slowest rate. However, representing higher taxa with single representatives runs counter to advice from both theoretical work (Yeates, 1995) and simulation studies (Wiens, 1998) based on exemplar methods. In addition, excluding taxa runs counter to a large number of studies which have concluded that accuracy generally increases with taxonomic sampling, and not the opposite (Hillis, 1996, 1998; Graybeal, 1998; Poe, 1998; Rannala et al., 1998; Pollock et al., 2002; Zwickl and Hillis, 2002; Huelsenbeck and Lander, 2003; but see also Kim,

1996 and Poe and Swofford, 1999). As a consequence, the exclusion of taxa cannot be advised as a general remedy for LBA.

Adding taxa to break up long branches

This leads to one of the most widely suggested remedies for LBA artifacts, i.e., to add more taxa to break up long branches (Hendy and Penny, 1989; Hillis, 1996, 1998; Swofford et al., 1996; Graybeal, 1998; Page and Holmes, 1998). This has been applied repeatedly, often with the effect that earlier results have been overthrown by the sampling of more taxa (Philippe, 1997; Halanych, 1998; Moreira et al., 1999; Chen et al., 2001; Jenkins and Fuerst, 2001; Soltis and Soltis, 2004). There are some limitations to this otherwise promising approach, and that is its inapplicability if LBA occurs even when all described living species in the study group have already been sampled (Bergsten and Miller, 2004). Extinctions may naturally make the sampling of critical taxa impossible. A second notion raised is that additional taxon sampling may also create new problems, or be of no benefit, depending on which taxa are sampled—more precisely their branch length and where on the tree they fit (Kim, 1996; Poe and Swofford, 1999). But with support from theoretical work and, as will be outlined in the concluding discussion, with support from several case studies, this is a highly recommended approach.

Adding data

A final approach suggests that LBA artifacts can be overcome by adding or combining more data (e.g., Xiang et al., 2002), preferably unlinked genes (Rokas et al., 2003) or morphological data to a molecular dataset (Bergsten and Miller, 2004). Initially, this might sound counter-intuitive since, if inconsistent, theory predicts that such a method will only reinforce the wrong tree with the addition of more data (Sullivan and Swofford, 1997). However, these predictions come from simulation studies and theories that always assume that adding more data means adding more data *of the same kind*, which is rarely the case. In real cases adding more data normally means adding another gene, perhaps from another genome, a morphological or behavioral dataset, which will have *different properties*. Theoretical predictions are scarce here, but arguments such as “adding data only reinforces the wrong tree” are certain to be a gross simplification when applied to real situations. The only simulation study that has touched upon the subject would be the recent work by Kolaczkowski and Thornton (2004), where two simulated partitions with differing properties were combined and showed the advantage for parsimony relative to likelihood-based methods, but the opposite when

partitions were analyzed alone. In addition, all methods, even if consistent, can infer the wrong tree with limited numbers of characters, in which case adding more characters can converge on the correct topology. The outcome from adding more data is unpredictable however, and in particular taxa with an elevated mutation rate for one gene might be likely to do the same for another gene. The greater disparity of data, the better presumably, but more theoretical work along the lines of that of Kolaczowski and Thornton (2004) is warranted. Since morphological data, as already stated, are less prone to LBA artifacts compared to molecular data (Grant and Kluge, 2003; Jenner, 2004; see also Baker et al., 1998), the combination should, from the view of the molecular partition, only be beneficial. Jenner (2004) cited a large number of papers where morphology has contributed positively in combined analyses to clade support values, “hidden clade support” or complementary resolution. “Hidden clade support”, the discovery of a secondary signal in a dataset that emerges only when combined with another dataset, I suspect, can often involve LBA artifacts (see examples below).

Methods of detecting LBA artifacts

Simply noting long branches (Bocchetta et al., 2000), possibly grouped together with high bootstrap support (Carmean and Crespi, 1995) is hardly appropriate, since it denies the possibility of close relationships between long branched taxa and the illogical consequence that the higher support the lower the trust in a group. There are good reasons to believe that a higher evolutionary rate in many phylogenetic studies is inherited, in fact a relevant potential synapomorphy, and that long branches do often belong together (Whiting et al., 1997; Siddall, 1998).

Separate partition analyses

Comparing molecular results with evidence from morphology is also a very widely used approach, either with a scored morphological matrix or simply in a discussion of previous morphology-based classifications. Where long branches are grouped together by molecular data but lack support from, and conflict with, morphological data, this has been taken at least in part as evidence of LBA, usually together with other evidence (Clements et al., 2003; Wiens and Hollingsworth, 2000; Bergsten and Miller, 2004). Although based on the logic of lesser sensitivity to LBA of morphological data, a conflict between a molecular tree and a morphological tree or evidence does not separate the artifacts of LBA from biological realities that can cause a mismatch between gene trees and species trees (Pamilo and Nei, 1988; Doyle, 1997; Maddison, 1997; Nichols, 2001).

This is why separate partition analyses of molecular and morphological data are not by themselves a valid test of LBA, but, together with other tests like long-branch exclusion (below), they serve as an important exploratory tool for identifying the source of conflict and possible LBA artifacts. Even if a total evidence approach is preferable as a final analysis (Kluge, 1989, 1998; Grant and Kluge, 2003), testing molecules and morphology separately can inform us as to whether some specific relationship, involving long branches—advocated by the molecules but not by morphology—is carried through in the combined analysis.

In a similar manner, comparing the results from different genes evolving at different rates has also been applied in order to see if a fast gene might show signs of grouping long branches together, in contrast to slower evolving genes (Moreira et al., 2002). Although suffering from the same problem as those above, i.e., it does not consider that gene trees do not need to match species trees and thus not each other, it could be indicative given that the problem of LBA should preferentially occur at higher rates of evolution (Huelsenbeck and Lander, 2003).

A related approach instead compares trees derived from sites with differing rates, or with and without fast evolving sites (Lake, 1998; Brinkmann and Philippe, 1999; Dacks et al., 2002; Schwarz et al., 2004; see also Lopez et al., 1999). A matrix with fast sites versus one or several matrices with slower evolving sites, assigned through some threshold value, are analyzed separately and if conflicting topologies are observed, then LBA might be suspected from the fast evolving sites. Alternatively, preceding a full matrix analysis, fast evolving sites are excluded and the analysis re-run to look for differences in the outcome that can be attributed to LBA. For protein coding genes it is not uncommon to analyse first and second positions separately from the normally faster evolving third codon position (although it is more common that third positions are simply excluded). This approach is particularly interesting (e.g., Sanderson et al., 2000; Debry, 2003) because it actually circumvents the major disadvantage of the latter two tests; differing histories among codon positions within a gene are hardly possible (all third codon positions in a gene cannot be inherited independently of all the first and second positions in the same gene). The mitochondrial genome represents another opportunity where different genes can be tested against each other (see for example Cao et al., 1998 and Waddell et al., 1999), although the obligate maternal and non-recombining inheritance of the mitochondrial DNA molecule is being more and more questioned (Piganeau et al., 2004). If gene trees, or at least codon trees, cannot have independent histories, then differing topologies from two or more partitions can only be due to the inference method and artifacts or random chance. To separate

these two possibilities it is important to look for support values; if the conflicting topologies from the two (or more) partitions are in addition strongly supported, then chance alone is unlikely to be responsible. As for comparing morphology and molecular data, other tests need to be combined with partitioned analyses to infer LBA. A practical problem can be that the number of informative characters usually is far greater in third codon positions and the sometimes low number of informative slower sites neither give much resolution nor strong support. I am not convinced by the arguments of Grant and Kluge (2003) denying the heuristic nature of data exploration, as it applies to analysis of separate partitions like the three approaches above. They argue that for any such comparison of characters or character partitions to be truly heuristic “they must be based on the results of the total-evidence analysis” (Grant and Kluge, 2003, p. 409). Inconsistently, they judge the method of long-branch extraction (Siddall and Whiting, 1999) (see below) to be “strongly heuristic” (Grant and Kluge, 2003, p. 398) although this is exactly the same thing; evaluating and comparing the outcome from analyses of subsets of the total evidence data. In conclusion then, separate partition analyses are an informative and heuristic guide to the researcher in identifying data partitions that, based on other tests, are likely to be responsible for LBA artifacts, and can guide the researcher in future taxon and character sampling strategies.

Long-branch extraction

Siddall and Whiting (1999) and Pol and Siddall (2001) suggested a simple test for cases where LBA is suspected to be a problem. They noted the obvious fact that for a long branch to be able to attract or be attracted there needs to be another long branch simultaneously in the analysis. So, in a case where two long branched taxa are grouped together, removing one while keeping the other in and vice versa would allow them to find their correct position in two separate analyses. If the clade was correct, then the separate analyses would not alter the position of the long branches in the tree. If however, one branch “flies away” to another part of the tree then it would be suspected that the clade was an LBA artifact. This approach is particularly appealing since it experimentally tests the LBA hypothesis by removing the potential causative agent while minimizing the amount of data excluded for the test. In comparison with testing LBA with the inclusion of, and then removing, third codon positions, where one-third of the data is excluded for the test, here only a small fraction (depending on the taxonomic sampling) need to be excluded. This is a critical issue because the more data excluded for the test, the weaker the test. Although it is a strong test, there are a few problems; first it only indirectly tests the relation

of the long branches to each other. The optimal parsimony solution cannot be found with certainty unless all taxa and all characters are simultaneously in the analysis. Second, even if they do end up in the same place in the separate analyses, there is no answer as to whether they should in fact be sister groups or consecutive branchings separated by a single internode. In spite of these drawbacks the long-branch extraction test combines power with simplicity and is therefore very appealing, and I recommend the use of it whenever LBA is suspected. It has recently been applied by Hampl et al. (2004) and Bergsten and Miller (2004), and is used in the two examples below.

Outgroups and adding artificial LB sequences

Several studies have tested LBA by creating artificial taxa with random (long-branched) sequences (Sullivan and Swofford, 1997; Philippe and Forterre, 1999; Stiller and Hall, 1999; Qiu et al., 2001; Stiller et al., 2001; Graham et al., 2002), and this approach dates back to Wheeler (1990), who showed that a random sequence is expected to attach to a phylogeny on the longest branch. Wheeler’s concern was that the use of a too-distant outgroup will act as a random sequence and artificially root the ingroup on the longest branch. Since the very rooting procedure is done after the search for a most parsimonious unrooted network, where outgroup taxa are treated identically to the ingroup taxa, it is easy to see that this is exactly the same as LBA. Thus a distant outgroup works as an attractor of long branched ingroup taxa. The test of LBA functions as follows; create a number of (e.g., 100) random sequences of the same length as the original. Exchange the real outgroup with the artificial sequences one by one. Run a parsimony analysis for each exchange and compare them where the tree is rooted with the root of the original analysis using the real outgroup sequence. If the tree is rooted at the same place with the original outgroup as with a high percentage of the random sequences, then it is suspected that the rooting is not based on a phylogenetic signal but rather on LBA artifacts. Needless to say, the weakness of the test is that there is no convincing argument as to why a true phylogeny could not in fact have its root at the longest branch. On the contrary, assuming a molecular clock, every asymmetric topology should have basal taxa with longer branches which means that the central issue is the proportion of asymmetric topologies to expect in phylogenetic studies (that the molecular clock is frequently violated does not really change the expectation in any direction). Although the matter of asymmetric versus symmetric trees has been thoroughly dealt with, most studies predominantly focus on biological reasons (e.g., variation in speciation/extinction rates; Heard, 1992; Guyer and Slowinski, 1993; Rogers, 1993) or methodological

differences (cladograms versus phenograms; Colless, 1995; Farris and Källersjö, 1998), but do not adequately take into account the crucial issue of taxon sampling in phylogenetic analyses. Taxon sampling is normally neither complete nor random, but usually aims to represent major lineages and not insignificantly influenced by the rarity and availability of fresh material, in particular for molecular studies. Consequently whether we should expect many trees to be asymmetric and rooted on long branches or not depends on many more factors than just variation in speciation/extinction rates or methods. Nonetheless the mere presence of an asymmetric base of the tree has been taken as indicative of LBA artifacts (Philippe and Laurent, 1998; Moreira et al., 1999). In phylogenetic studies the combination of variation in branch lengths in the ingroup and using outgroups for rooting is very common, almost universal. Outgroups are by reasons of sampling almost always long branches as compared to the branches in the ingroup. This is a simple consequence of taxon sampling, the breaking up of long branches, which is normally much more extensive in the ingroup of interest than outside the ingroup (which is not of interest). For this statement possibly to be false, outgroup taxa need either be sampled as extensively as the ingroup, or to have a significantly lower rate of evolution. The latter could of course be true but is not more likely than the opposite, after all ingroup taxa and outgroup taxa are only defined in the context of a specific study. Thus outgroups will be a potential misleading attractor of relatively longer branches among the ingroup (Philippe and Laurent, 1998). Several very common causes can give rise to situations where the longest ingroup branch should not be most basal, including unequal taxonomic sampling across different groups, extinctions and unequal evolutionary rates. I believe that the matter of outgroup rooting and LBA is of major concern, especially in pure molecular studies, and theoretical as well as empirical work is warranted on the subject.

Recently, Huelsenbeck et al. (2002a) and Holland et al. (2003) performed some simulation studies related to the performance of different methods to root trees. Huelsenbeck et al. (2002a) concluded that the outgroup-criteria for rooting performed better than both a molecular clock and a non-reversible model of DNA substitution. This was based both on a model tree and the use of a “known” phylogeny. Unfortunately, it was not a very strong test for the performance of the outgroup method since the simulation tree had all equal branch lengths and the “known” rooting of the four-taxon tree was on the longest branch. Wheeler’s (1990) conclusions were however, confirmed in the simulation: when the outgroup branch increased (became more distant), rooting accuracy decreased. The simulation study of Holland et al. (2003) on the other hand tested in particular outgroup misplacement in a five-taxon tree

(4 + outgroup) with long external branches and short internal branches, the correct rooting in the model tree being on the short internal branch. For a detailed comparison of their results between methods, the reader should consult the original paper, but the most important results concerning parsimony are repeated here. Parsimony methods were as expected, inconsistent in a corner of the parameter space (the Felsenstein-zone), but also estimated four incorrect trees more often than the correct tree close to the border but outside the inconsistency-zone, up to sequence lengths of 5000. These incorrect trees were all a result of the outgroup attaching the ingroup tree at one of the long external branches rather than at the correct short internal branch. Holland et al. (2003) also analyzed all repetitions, both with (five-taxon) and without (four-taxon) the outgroup and could therefore divide up the incorrect topologies into: (1) those with the correct ingroup topology but incorrect outgroup attaching, (2) those with incorrect ingroup topology where the ingroup by itself resolved correctly and only with the addition of the outgroup ended up incorrect, and (3) those where the ingroup remained incorrect with the addition of the outgroup. The most common error with all the methods was that only the outgroup was erroneously placed; the LBA was between the outgroup and a terminal ingroup branch. However, in a not insignificant number of cases had the addition of the outgroup disrupted the ingroup to become incorrect. This was much more frequent than the opposite, i.e., that the addition of an outgroup corrected an incorrect ingroup-topology. For this reason the authors recommended that trees should be constructed both with and without outgroups, and if the outgroup changes the ingroup topology, it is likely that the ingroup-alone topology is more accurate.

The behavior of rooting in relation to LBA needs much further attention, for example whether it is always beneficial the more outgroups are sampled, and how outgroups should be sampled. In summary, the random outgroup approach is not sufficient to conclude LBA by itself; the true root can very well be on the longest branch in the ingroup. Although not a good enough test of LBA, as Wheeler (1990, p. 367) pointed out “the roots determined by distant [random] outgroups should be suspect.” (my addition in brackets). The suggestion to run all phylogenetic analyses both with and without the outgroup is simple, warranted and can be very informative, especially if combined with separate partition analyses. Holland et al.’s (2003) study provided the theoretical justification for preferring the unrooted ingroup topology if altered by the inclusion of the outgroup. Can separate partition analyses in addition show the outgroup rooted topology to be restricted to a certain molecular partition? If other partitions give the ingroup-only topology even with outgroups included, then the case is strong.

Parametric simulation

Huelsenbeck et al. (1996) and Huelsenbeck (1997, 1998) used parametric bootstrapping to address the question of whether two branches are long enough to attract, even though they are not each other's closest relatives. Also called parametric simulation, the method has been quite popular and has been applied in several other studies in order to investigate possible LBA artifacts (Maddison et al., 1999; Tang et al., 1999; Tourasse and Gouy, 1999; Sanderson et al., 2000; Wiens and Hollingsworth, 2000; Omilian and Taylor, 2001; Wilcox et al., 2004). The procedure to test if two sister taxa in a specific study with some data are long enough to artificially attract each other works as follows: (1) assume a model and a model tree where the long branches are not sister taxa, (2) estimate the model parameters (incl. branch lengths) from the real data, (3) simulate a number of replicated datasets of similar size with this model tree and parameters, and finally (4) analyze the replicated datasets with the method used in the original analysis. Conclude, if the two taxa group together in a high proportion of the replicated datasets (although they were apart in the model tree), that the branches are long enough to artificially attract with the applied method. Although seemingly elegant, the outcome is somewhat disappointing in its restricted conclusions; the answer is “yes” or “no” as to whether the branches are or are not, long enough to attract. Preferably an LBA detector excludes or at least renders unlikely other possibilities. The obvious fact that two “long enough branches to attract” could nevertheless be each other's true relatives is not made any less likely by the test. Accordingly, Huelsenbeck (1997) suggested that, following a positive outcome, a method (likelihood) less commonly affected by LBA should be shown to pull them apart (methodological discordance is dealt with below). Wiens and Hollingsworth (2000) added a third criterion to this multiple test combination: that evidence should be provided (from other datasets) that the long branches are not actually sister taxa. Showing that a molecular dataset groups two long branches together only with a method known to be sensitive to LBA—that even if apart in a model tree this method will, under the estimated model of evolution group them erroneously together—and that a morphological dataset strongly contradicts the grouping, is undoubtedly taken together a very powerful indication of LBA. The necessarily model-based and assumption-rich nature of the test have been criticized by Siddall and Whiting (1999) and Pol and Siddall (2001).

RASA

Lyons-Weiler and Hoelzer (1997) proposed that their method, called RASA (Relative Apparent Synapomor-

phy Analysis) and originally developed to measure phylogenetic signal (Lyons-Weiler et al., 1996) can be used to detect long branched and “problematic taxa”. According to them, the examination of the taxon-variance plot¹ where long branches should be detectable as outliers, and an increase in phylogenetic signal (measured by t_{RASA})¹ when removed from the matrix suffice to judge specific taxa as problematic. Case studies that have used RASA in this aspect are numerous (Stiller and Hall, 1999; Barkman et al., 2000; Belshaw et al., 2000; Bowe et al., 2000; Culligan et al., 2000; Reyes et al., 2000b; Teeling et al., 2000; Stiller et al., 2001; Dacks et al., 2002). Lyons-Weiler and Hoelzer (1997) further suggested that when or if detected, problematic long branches should be excluded or, if of particular interest, their effect mitigated in other ways, e.g., by eliminating third codon positions or by sampling more taxa to break up the long branches. These proposals have already been discussed (see above). A critique of RASA and the recommended procedures is severe, in particular as a measure of phylogenetic signal but also as a detector of problematic long branches (Faivovich, 2002; Farris, 2002; Simmons et al., 2002; Xiang et al., 2002) and in general as a scientific endeavor (Grant and Kluge, 2003). RASA both fails to identify long branches when present (Simmons et al., 2002; Xiang et al., 2002), and identifies problematic long branched taxa when they are not problematic (Faivovich, 2002), even when they are of length zero (Simmons et al., 2002). Grant and Kluge (2003) adequately summarized why RASA and its recommended procedures should not be utilized.

Split decomposition and spectral analysis

A few studies have used split decomposition (Bandelt and Dress, 1992) and related spectral analyses (Hendy and Penny, 1993) as a way of checking for long branch attraction (Flook and Rowell, 1997; Kennedy et al., 1999; Waddell et al., 1999; Lockhart and Cameron, 2001; Clements et al., 2003). Split decomposition and spectral analysis are methods for detecting conflicting signals in phylogenetic data and illustrate this in a way that bifurcating trees are unable to do. Since LBA is due to convergent transformations, overwhelming the true phylogenetic signal, there should be conflict in the dataset, and this is arguably the background to how spectral analysis and split decomposition can be used to detect LBA. The split graphs indicate where there is conflict in the data by displaying a box-like structure in the tree, but does nothing to test which of the alternatives for a certain branch is the artifactual signal. Grant

¹I refer to the original papers of Lyons-Weiler et al. (1996) and Lyons-Weiler and Hoelzer (1997) or to Faivovich (2002), Farris (2002) and Grant and Kluge (2003), critical papers in *Cladistics*, for explanations.

and Kluge (2003) discussed some of the merits of spectral analysis and concluded with regards to spectral analysis as a method of data exploration that, apart from the “data correction involved” the plotting procedure was heuristic, providing indication of conflict in the data and pointing towards suboptimal hypotheses worthy of further testing and consideration. Although not a direct LBA detector, this method can more or less serve the same purpose as the third criterion of Wiens and Hollingsworth (2000), but searches for conflict within instead of between datasets.

Methodological discordance

Comparing the results from different inference methods—such as parsimony versus likelihood versus Bayesian, neighbour-joining with distances calculated differently (Tourasse and Gouy, 1999), or a simpler model versus a more parameter-rich model using Maximum likelihood (VanDePeer et al., 1996)—is a very widespread protocol in phylogenetic studies and differences are sometimes taken as evidence of LBA (e.g., Schwarz et al., 2004; Wilcox et al., 2004), as suggested by Huelsenbeck (1997). Grant and Kluge (2003) criticized the common notion that increased support can be sought by inferring the same result with different methods, “methodological concordance”. The use of methodological discordance to infer methodological artifacts however, is not hit by the exact same lines of arguments. Accurate methods infer by definition the same true tree. Differences are thus caused by one or several methods not being accurate in a specific case. Imagine a case where parsimony groups two long branches together, but a likelihood analysis places them apart. Whether the difference are caused by parsimony producing LBA artifacts or likelihood failing to correctly place two long branches as sister groups is impossible to tell, the true tree being unknown. Most importantly, it is a circular test in that the outcome is predictable from the merits and assumptions of the methods. That long branches end up not being sister groups in the likelihood case is directly related to the use of a method which is deliberately designed to interpret many potential synapomorphies as independently derived. Consequently to infer LBA through methodological discordance one needs to fall back to simulation studies and theory claiming higher accuracy in recovering the true phylogeny involving long branches, with some methods over others. I summarized the result of simulation studies in the introduction above with four main conclusions. In particular, no method is immune to LBA, so methodological concordance does not rule out LBA. However as Pol and Siddall (2001) argued, parsimony suffers from a strong but topologically identifiable bias of LBA, and Swofford et al. (2001) confirmed this relatively stronger bias in parsimony

compared to likelihood. Accordingly, comparisons can have an indicative value of LBA, and methodological discordance is a valid approach but LBA suspicion should be further evaluated using other tests.

For the following two cases I exemplify the use of long-branch extraction, separate partition analysis, methodological discordance, and the adding data to detect and overcome LBA. In both examples, I find the result from the combined usage of several tests sufficient to state that LBA is the least refuted hypothesis. For an illustration of methods covered but not used below, I refer the reader to the examples cited under the descriptions. Both examples involve LBA where the outgroup is involved, which is not a coincidence since this was the most common situation in the literature scanned for this review.

Placental mammal example

The phylogeny of the placental mammals serves well to illustrate the problem of outgroups and long-branch attraction. The rapidly growing literature on mammalian phylogeny was recently reviewed by Springer et al. (2004). Molecular data have revolutionized the study of the higher level phylogeny of mammals and overthrown some classical ideas based on morphology. One of the most exciting findings was the discovery of an African clade (Springer et al., 1997; de Jong, 1998; Stanhope et al., 1998a; Stanhope et al., 1998b) which included not only the paenungulates (elephants, sirenians and hyraxes) but also the aardvark, elephants shrews, golden moles and tenrecs. However, the early history of the now emerging mammalian molecular phylogeny has also been bordered with radical news from mitochondrial genome data such as the classical “the guinea-pig is not a rodent” (D’Erchia et al., 1996), and that the hedgehog is the most basal placental mammal (Krettek et al., 1995). Both these hypotheses are related to the rooting of the eutherian tree; the hedgehog obviously being the edge where the marsupial outgroup attaches on the placental tree to make it the first offshoot, and the guinea-pig was not a rodent because the tree was rooted inside Rodentia on the mouse/rat (murid rodents) clade. Both the hedgehog and the murids have been shown to differ markedly in the inferred evolutionary rate as well as in base composition and always end up with long branches (Sullivan and Swofford, 1997; Waddell et al., 1999; Nikaido et al., 2001, 2003; Lin et al., 2002a,b).

“The guinea-pig is not a rodent”

“The guinea-pig is not a rodent” claim was forcefully disputed and shown to be susceptible to: (1) additional taxon-sampling (Philippe, 1997), and (2) taking rate-heterogeneity across sites into account (Sullivan and

Swofford, 1997). Sullivan and Swofford (1997) additionally showed that the opossum outgroup behaves like a random sequence and thus rooted the ingroup on the longest branches (within rodents) as expected (Wheeler, 1990). Nevertheless, Reyes et al. (1998, 2000a,b) stubbornly continued to argue for the non-monophyly of rodents. Reyes et al. (2000a) completely missed the point though, when based on a relative-rate test they argued that the basal position of the murid clade cannot be an LBA artifact since similar rates are found in the other rodents that branch second (nucleotides) or third (amino acids) from the ingroup root. They argue that the non-murid rodents in such cases should have been affected by the outgroup “in the same way” as the murids. However, in what way one or several branches are pulled down towards the outgroup by LBA cannot be easily predicted in other ways than maybe by parametric simulations (see above). Needless to say there might be long branches in a tree not affected by LBA at the same time as equally long branches are affected by LBA; a larger number of synapomorphies that overcomes possible artifacts might exist in one case, but not in the other. In addition, how the outgroup is supposed to affect two or more long branches simultaneously “in the same way”, being able to root only on one branch at a time, is to me a riddle. Reyes et al. (2000b) on the other hand, refute LBA as a cause of rodent non-monophyly by using RASA (Lyons-Weiler and Hoelzer, 1997), since no “problematic long-branched taxa” were found by that method. This needs no further comments than what has been said about RASA, above, and by Simmons et al. (2002), Faivovich (2002), Farris (2002) and Grant and Kluge (2003).

The hedgehog

When the mitochondrial genome of the hedgehog was first sequenced, this species ended up basal in the eutherian tree (Krettek et al., 1995), but even as more insectivore mitochondrial genomes became available, the hedgehog would not group with the mole or shrew but was always pulled down to the outgroup (Mouchaty et al., 2000a,b; Nikaido et al., 2001). In several mitochondrial eutherian studies the hedgehog sequence has simply been excluded a priori because it is argued to be too different, represent a long problematic branch, a rogue taxon, and violates assumptions inherent to phylogenetic methods (Sullivan and Swofford, 1997; Cao et al., 1998, 2000; Reyes et al., 2000a,b; Lin et al., 2002b). The impressive datasets of nuclear genes on the other hand firmly place the hedgehog with the other eulipotyphlans in the Laurasiatheria clade (Madsen et al., 2001; Murphy et al., 2001a). Combining these two latter studies, Murphy et al. (2001b) analyzed the resulting 16 397 bp dataset with Bayesian methods and arrived at a well supported phylogeny with monophyletic Eulipotyphla and Rodentia, as well as a basal

afrotherian clade, a clade with Xenarthra (sloths, armadillos and anteaters) and the two large sister clades Euarchontoglires and Laurasiatheria. This is where we stand today in placental phylogenetics (Springer et al., 2004). Finally, analyses of mitochondrial genomes also started to support a non-basal position of the hedgehog and eulipotyphlan monophyly (Nikaido et al., 2003), as well as the monophyly of rodents (Lin et al., 2002b) and came more and more into agreement with nuclear genes (Lin et al., 2002a; Reyes et al., 2004), with the notion that previous results were based on LBA artifacts (Waddell et al., 2001). The artifacts were mitigated by increasing taxon sampling and better likelihood models. Although remnants of support for the hedgehog-is-basal hypothesis can still be found (Arnason et al., 2002) the defense for “the guinea-pig is not a rodent” claim probably died out with Reyes et al. (2000b) denying the artifact using RASA.

Re-analyses

Since I consider the LBA artifacts from mitochondrial genomes concerning the rodents and hedgehog to be adequately analyzed and already demonstrated, I will instead show some effects of parsimony on the 16 kbp predominantly nuclear dataset of Murphy et al. (2001b). This is the largest nuclear (+3 mt genes) dataset available for placental mammals, and Springer et al. (1999, 2001) convincingly showed that nuclear sequences are better in retrieving deep-level benchmark mammalian clades than are similar-length mitochondrial sequences. Since Murphy et al. (2001b) never analyzed the dataset with parsimony, this will be informative. In addition, I have reanalyzed the dataset with Bayesian methods to investigate if allowing different parameters to different codon positions, mitochondrial- and untranslated regions in the dataset will alter the result, since Murphy et al. (2001b) forced a single GTR + Γ + I model across all genes, concatenated and treated as one. The dataset consists of 19 nuclear and three mitochondrial genes (see appendix Table 1) for 44 taxa representing all 18 presently recognized placental mammal orders and two marsupial outgroup taxa. Of the 16 397 base pairs in the matrix, 7785 represent parsimony informative sites

Bayesian analysis

First I reanalyzed the dataset with Bayesian inference methods but allowed the GTR + Γ + I model to be separate for five partitions defined as follows: mtRNA, untranslated nuclear regions, first positions in nuclear protein coding genes, second and third positions ditto. With so many different protein coding genes and with the possibility of assigning separate models to different genes and within genes to codon positions, etc., one is easily tempted to over-parameterize the analysis (i.e.,

Table 1

Jackknife support for different rooting alternatives of the full matrix analysis, and with three analyses with varying number of long branched taxa excluded

Analysis	Afrotheria	Tenrec	Xenarthra	Murids	Hyst./Cavi.	Macroscel.	Treeshrew
Full matrix		35	3	45		2	7
No tenrec	2		9	56		13	11
No tenrec, or murids	6		21		11	31	24
No tenrec, murids or elephant shrews	24		30		10		28

assign more parameters than the data could reasonably estimate). Over-parameterization can often be detected in Bayesian analyses, in that parameter estimations show a very high variance, and if plotted as a function of Markov chain generations do not show signs of stabilization. Two separate, 3 million Markov chain generations were run and sampled every 300 generations, starting from a random topology. A burn-in of the first 1.5 million generations was deleted in each run and the inference drawn from the remaining pooled sample of 2×5000 sampled generations, that had reached stability. Four simultaneous chains, one cold and three incrementally heated were run with priors and proposal settings set to their default values in MrBayes 3.0b4 (Ronquist and Huelsenbeck, 2003). The acceptance ratio of parameters and mixing of chains was checked, and the convergence of overall likelihood—as well as for each parameter—was controlled graphically within both runs and compared between them.

Parsimony analysis

Second I analyzed the data with parsimony in NONA (Goloboff, 1999) spawned from Winclada (Nixon, 2002). Search strategies were heuristic with the commands hold 10000, mult*100, hold/10, max*. Gaps were treated as missing data. I also used the parsimony ratchet (Nixon, 1999) with 200 iterations, hold/2 and 10% of the characters sampled to check that no shorter trees, or additional equally short trees, could be found, however, with a 44-taxa matrix the common heuristic search should suffice. Jackknife analyses used 1000 replicates (mult*5 and hold/1 per replication, no max*(TBR)).

Bayesian results

Results from the Bayesian analysis are shown in Figs 1 and 2. Estimated model parameters can be found in the Appendix table A1. Figure 1 shows the phylogram with estimated branch lengths that represents both the majority-rule consensus tree and the topology with highest posterior probability ($P = 0.266$). The distantly related outgroup (i.e., the long ingroup branch) is the first thing to strike the observer. Figure 2 is the same tree with clades named, and with the proportion of times different clades occurred in the sample indicated as

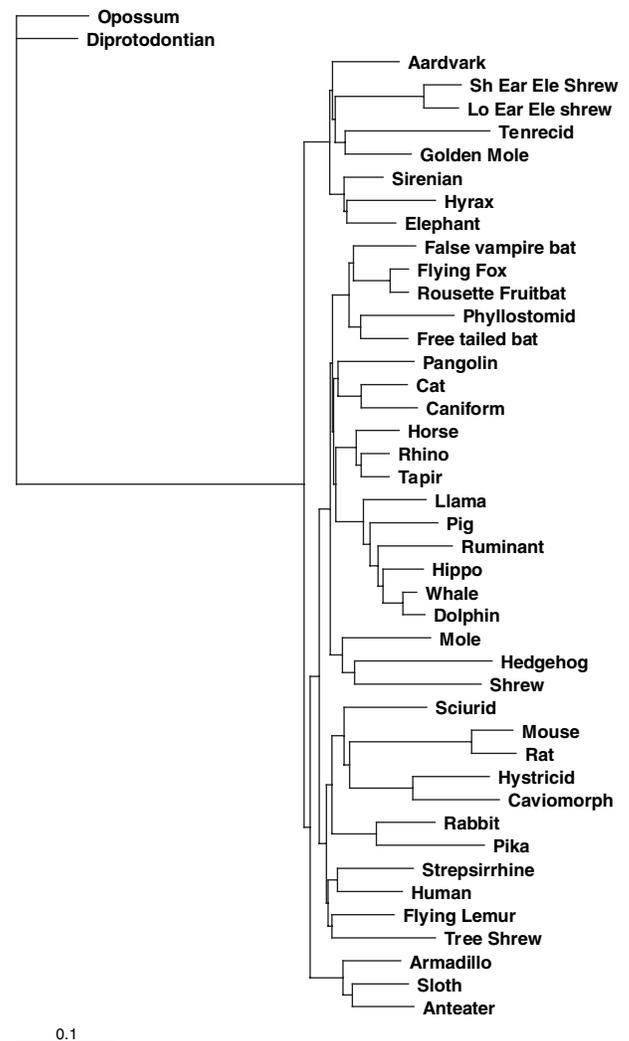


Fig. 1. Majority-rule consensus phylogram of placental mammals, as well as the topology with highest posterior probability ($P = 0.266$) from the Bayesian analysis using a separate GTR + Γ + I model for five partitions (mtRNA, untranslated regions, 1st, 2nd and 3rd codon positions of nuclear protein coding genes). Note the long ingroup branch from the outgroup, the short internodes of early divergences, and the long-terminal edges. The scale bar represents the number of expected substitutions per site.

support values. The four major supraordinal clades Afrotheria, Xenarthra, Euarchontoglires and Laurasiatheria are retrieved and rooted at the base of Afrotheria,

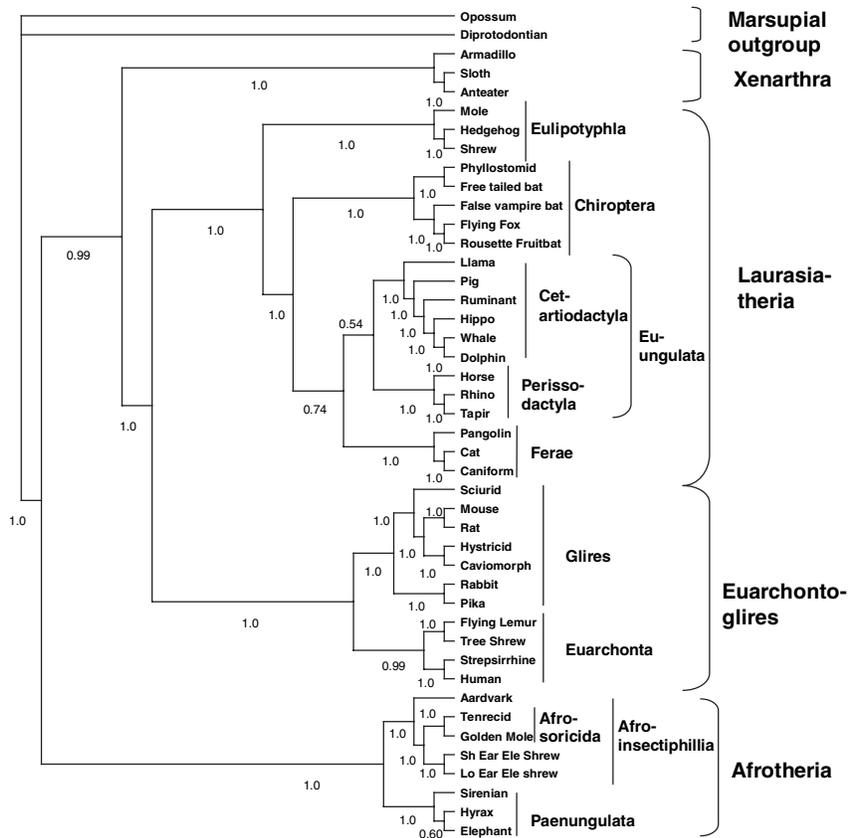


Fig. 2. Topology of placental mammals with highest posterior probability, $P = 0.266$, from Bayesian analysis using separate GTR + Γ + I models for five partitions (mtRNA, untranslated regions, 1st, 2nd and 3rd codon positions of nuclear protein coding genes). Numbers on internodes are the proportion of trees from the sampled chain in which the clade occurred after burn-in was removed.

all in agreement with Murphy et al.’s (2001b) result. The only difference compared to their tree can be found within the Laurasiatheria and Paenungulates; the odd-toed ungulates form Euungulata together with Cetartiodactyla, whereas they grouped with the carnivores + pangolin in Murphy et al. (2001b). Within paenungulates, hyrax and elephant are sister taxa whereas Murphy et al. got ((sirenia + hyrax) elephant). Both these two differences however, are those least supported in the tree, both here and in Murphy et al. (2001b). The four topologies with the highest posterior probability in the analysis contain these two plus two conflicting solutions and have a cumulative posterior probability of 0.71.

Parsimony results

The parsimony analysis resulted in one most parsimonious tree (Fig. 3) with a tree length of 41 937 steps, ci of 0.40 and ri of 0.39. It is identical to the Bayesian analysis except at two points: (1) the Dermoptera + Scandentia clade is sister group with Glires instead of with the primates, and most notably (2) The Madagascan tenrec is pulled from a position as sister group to the golden moles within the African

clade, to the base, and sister to the remaining Eutheria.

The tenrec and long-branch extraction

Now before making a guinea-pig out of the tenrec, let us test whether the basal position could be due to LBA to the outgroup. Let us first predict the outcome if we remove the outgroup taxa; if the tenrec indeed has this peculiar position in the eutherian tree, then removing the outgroups and searching for the most parsimonious network would result in an unrooted tree with the tenrec in a trichotomy with the African clade and the remaining eutherians as the other two branches. If, on the other hand the tenrec was only pulled to the basal part of the tree by LBA artifacts then we predict that removing the outgroup would result in the tenrec re-entering the African clade as a sister group of the golden mole (forming afrosericidans) in the most parsimonious tree. The search resulted in the tree seen in Fig. 4 where the tenrec is indeed an afrotherian sister to the golden mole. (Note that Waddell et al. (2001) identified the tenrec and elephant in addition to the hedgehog and murid rodents as possible problematic taxa.) But where do we root this

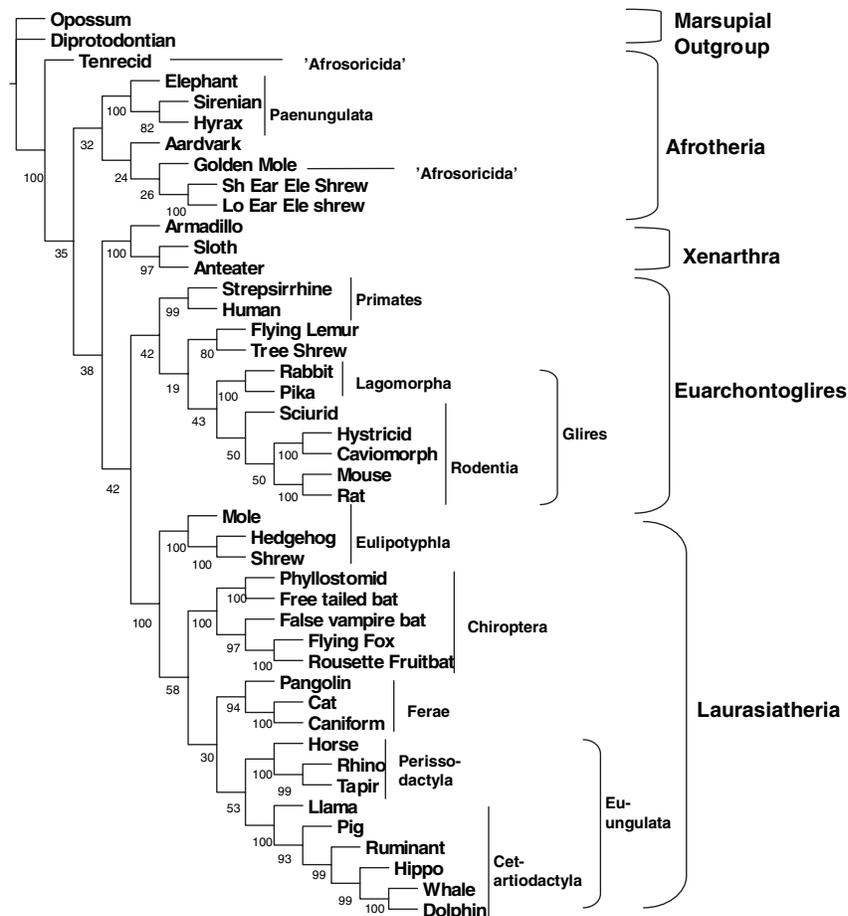


Fig. 3. Single most parsimonious tree of placental mammals from the 16 397 nucleotide dataset (22 genes) of Murphy et al. (2001b). Numbers below nodes are jackknife values from 1000 replications. $L = 41\ 937$, $ci = 0.40$, $ri = 0.39$.

tree? Let us then remove the tenrec from the matrix, put back the outgroups, and search again with the prediction that if the tenrec was the only problematic taxa we should get a more proper rooting. The outcome is surprising; the tree is now rooted instead among the rodents, with the mouse + rat clade the most basal (Fig. 5) just as in numerous analyses with complete mitochondrial genomes (D'Erchia et al., 1996; Pumo et al., 1998; Reyes et al., 1998, 2000a,b; Arnason et al., 2002).

Morphology

The most important argument against this hypothesis and criticism towards those who defend it I have not touched upon yet, and that is morphology. Considering the amazing amount of convergent or parallel morphological evolution the polyphyletic origin of rodents favored by, for example Reyes et al. (2000b, their Fig. 1) would imply (see Luckett and Hartenberger, 1993), it is surprising that the result has ever been seriously argued for. The modest recom-

mendation of Luckett and Hartenberger (1993, p. 143) is well worth repeating: "We also recommend that the accumulated wisdom from 300 years of assessing cranioskeletal and dental characters be considered when collecting and evaluating molecular or other biological data." Actually, adding to the matrix 21 dummy-characters representing the nine uncontroversial dental and cranioskeletal characters existing in support of rodent monophyly (Luckett and Hartenberger, 1993) and the 12 additional unambiguous dental, cranioskeletal and unique fetal membrane characters supporting a monophyletic Glires (rodents + Lagomorpha) (Luckett and Hartenberger, 1993), the monophyly of rodents and Glires is restored. Twenty-one morphological characters can change the result of a 16 397 bp (7785 parsimony informative) large nucleotide matrix! No doubt it is not a question of the 21 characters overwhelming the nucleotide dataset, but rather strengthening the underlying phylogenetic signal already present in it to overcome the artifactual rooting (remember the "hidden clade support" from the introduction). It certainly

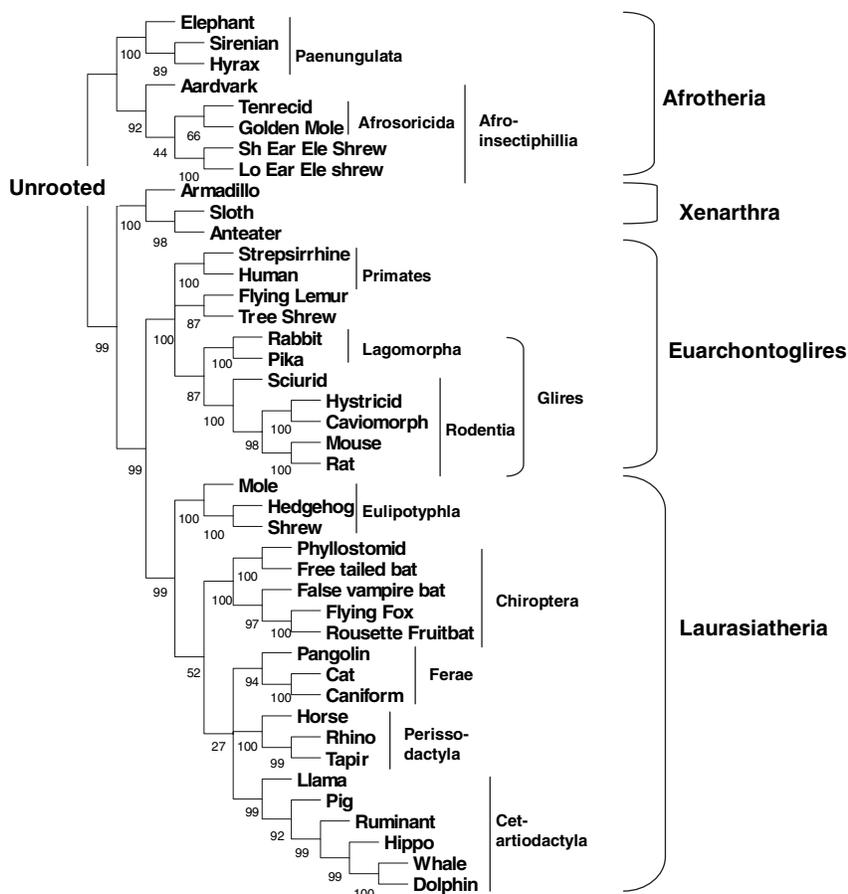


Fig. 4. Unrooted strict consensus of three most parsimonious trees ($L = 38\,510$, $ci = 0.44$, $ri = 0.48$) of placental mammals from the 16 397 nucleotide dataset (22 genes) of Murphy et al. (2001b) from which the two marsupial outgroup taxa were removed. Since unrooted, named clades only represent groups with a possibility of being monophyletic to illustrate the complete lack of conflict at this stage with the rooted Bayesian analysis. Note the possible sister group relationship between the tenrec and golden mole (Afrosoricida) compared to the rooted analysis (Fig. 2), where the tenrec is pulled out of the afrotherian group. Numbers below nodes are jackknife values from 1000 replications.

casts doubts over Scotland et al.’s (2003, p. 543) molecular point of view that “we disagree that morphology offers any hope for the future to resolve phylogeny at lower or higher taxonomic levels”. I disagree, in particular for the relevance of detecting LBA in molecular results I find morphological data very important. In addition I have yet to hear a good argument of why morphological data should not bear its evidence on the phylogenetic hypothesis but rather be interpreted a posteriori and ad hoc as in, for example, the mitochondrial study of Reyes et al. (2000a, p. 184) “the existence of two rodent clades is in great disagreement with morphological and paleontological data [refs.] suggesting that the degree of convergent and/or parallel evolution between murid and non-murid rodents may have been higher than thought...” With regard to Scotland et al. (2003) this paper has already been scrutinized, and criticized, and basically all their arguments proven unsupported or simply wrong by Jenner (2004) and Wiens (2004).

Rooting

The tree with rodents monophyletic is now rooted on the elephant shrews which is the same result as when the mouse and rat (and tenrec) are excluded and there are no dummy characters. The elephant shrews of course also represent very long branches (see Fig. 1). Also excluding the elephant shrews actually results in a single tree (TL 36 455, $ci = 0.46$, $ri = 0.53$) with the same rooting as in the Bayesian analysis, at the base of the African clade, which was also the result of Murphy et al. (2001b). There are two reasons to prefer this rooting as compared to all the previous ones. The first reason is that all previous rootings have either altered the ingroup topology after the inclusion of the outgroups, i.e., not only rooted the tree at one long edge but pulled the long branch to another position in the tree when included (conversely it jumps back when they are excluded), or in the case of the basal mouse/rat clade, have a truckload of morphological convergences or reversals to explain. The second

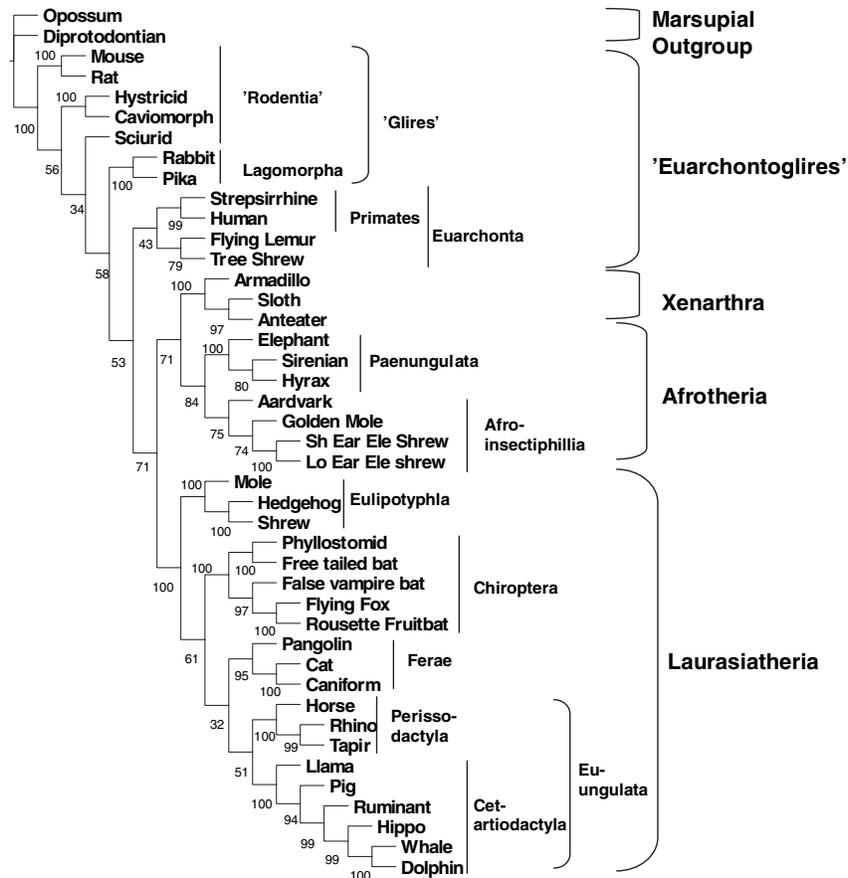


Fig. 5. Single most parsimonious tree of placental mammals from the 16 397 nucleotide dataset (22 genes) of Murphy et al. (2001b) when the tenrec is removed. Numbers below nodes are jackknife values from 1000 replications. $L = 40\,571$, $ci = 0.42$, $ri = 0.43$.

reason is that the first split between an African clade and the remaining South American/Northern Hemisphere taxa can make sense in a geographic and plate tectonical perspective, where molecular dating of the split between South America and Africa (see Murphy et al., 2001b; Springer et al., 2003, 2004). The other rootings above have severe problems in making similar sense, although, for example, a rooting with xenarthrans basal, as morphology and morphologists have long suggested, can make equal sense. However, neither the jackknife support for the afrotherian rooting nor the other alternatives is very well supported (Table 1). In the full matrix analysis the rooting on the tenrec branch is actually very weakly supported and in 1000 jackknife replicates the root is on the tenrec edge in 35% of the trees but on the mouse/rat clade in 45% of them. In the analysis with the matrix stripped of the tenrec, murids and elephant shrews, the root is on Afrotheria 24% and the xenarthrans 30%. As seen in Table 1 no rooting was ever supported with a jackknife value of more than 60% and only the murid rodents rooting with the tenrec removed was supported with more than 50%. It is also

interesting to note that the rooting preferred in the most parsimonious cladogram was not always that receiving the highest jackknife value.

Conclusions from placental mammal example

I conclude, as acknowledged previously (Philippe and Laurent, 1998; Waddell et al., 2001; Delsuc et al., 2002; Lin et al., 2002a,b; Springer et al., 2004), that the rooting is indeed the major problem in placental mammal phylogeny reconstruction, and of course excluding taxa is the wrong direction for further progress. This should only be used for exploratory purposes in relation to long-branch extraction tests, not for deriving a better result. Rather, improvements in the understanding of mammalian phylogeny during the molecular revolution has been achieved by: (1) constantly improving taxon sampling (breaking up long branches), (2) collecting more data, and especially switching focus to nuclear genes rather than focusing only on mitochondrial genomes, and (3) improvements in methods for analyzing the data. If molecular data continue to give the African clade

basal as the best alternative for the root (Murphy et al., 2001a,b; this analysis) the morphological data should be reconsidered and weighted in to evaluate how strong it challenges this root as opposed to xenarthrans basal.

Gall wasps example

LBA examples as that above are neither unique to parsimony nor confined to early attempts in molecular phylogenetics using likelihood, when molecular models were still poorly developed. Even with the most complex models and using the newest, hottest, Bayesian inference methods, results can be prone to LBA artifacts which I illustrate with a study taken from the recent literature. Nylander et al. (2004) recently presented a combined study on gall wasps (Hymenoptera: Cynipidae) based on morphology (164 morphological and two ecological characters from Liljeblad and Ronquist, 1998), nuclear 28S rDNA (1154 bp), nuclear EF1 α (367 bp) and LWRh (481 bp) and the mitochondrial gene CO1 (1078 bp) for 32 taxa (IG: 29 Cynipidae taxa: OG:1 Figitidae, 1 Liopteridae, 1 Ibalidae).

Morphology: woody galls monophyletic

Nylander et al.’s (2004) analyses of the morphological data alone gave a monophyletic clade of woody galls formed by oak galls (*Plagiotrochus*, *Andricus*, *Biorhiza*, *Neuroterus*) + woody non-oak galls (*Diplolepis*, *Pediaspis*, *Eschatocerus*) with very strong bootstrap support (96.9: parsimony with implied weights, $k = 2$, Goloboff, 1993) and posterior probability (1.0: Bayesian analysis with Markov k model of Lewis, 2001). A parsimony analysis without implied weights of the morphological data alone results in the tree in Fig. 6. [Searches heuristic with the same commands as in the mammal example. Bremer support calculated stepwise in NONA, i.e., “bs1”, “bs2”, “bs3”, etc.]

DNA: Woody galls not monophyletic

In conflict with this result, their Bayesian analysis of the molecular (all four genes) or combined (molecular + morphology) data, analyzed with many different models from two partitions and a simple JC69 model to five partitions each with a separate GTR + Γ + I model (morphology with Markov k + Γ or not), all gave the

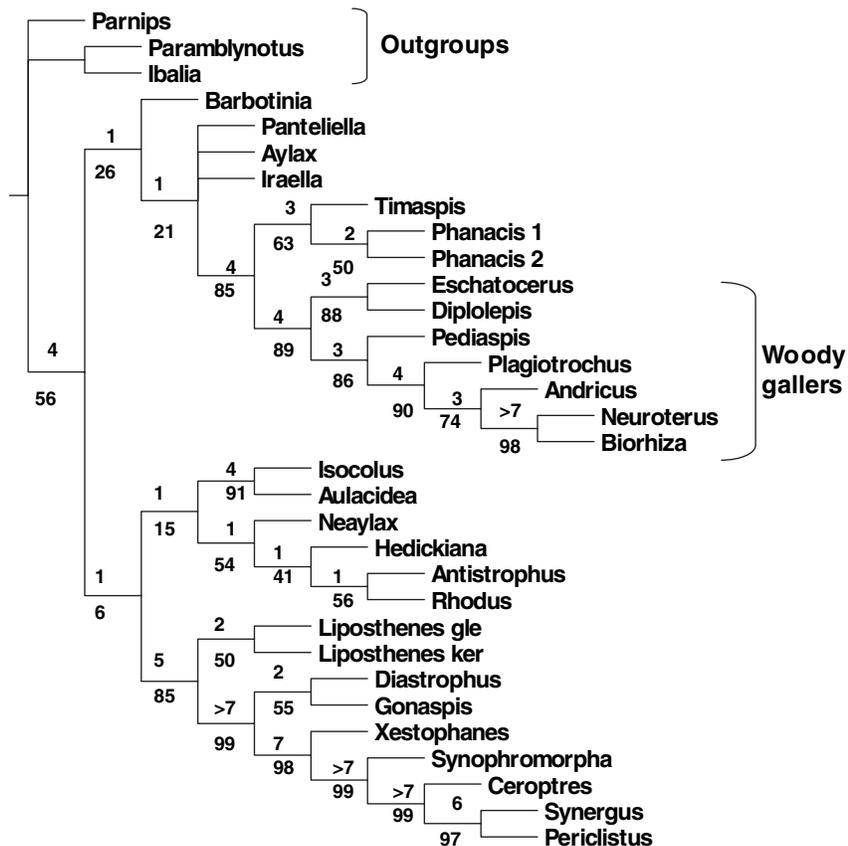


Fig. 6. Strict consensus of two most parsimonious trees (TL = 694, ci = 0.30, ri = 0.59) from only the morphological data of Nylander et al. (2004). Numbers below internodes refer to jackknife support values, and values above internodes refer to Bremer support values.

three woody non-oak galler taxa basal in the tree with conspicuously long branches. This result contrasts significantly with the inference of an ancestral herb galler by Ronquist and Liljeblad (2001), a conclusion stable even when taking uncertainty in the phylogeny into account (phylogeny estimation from more or less the same morphological data from Liljeblad and Ronquist, 1998). Nylander et al. (2004) leaned towards interpreting this result as “model imperfection rather than mismatch between morphological and molecular trees.” In the discussion they briefly provided two possible hypotheses of the former: morphological convergence grouping the gall inducers of woody hosts together in the morphological tree, or molecular process heterogeneity across the tree explaining why the three longest terminal branches appear basal in the molecular (and combined) tree. No further test was performed to assess whether the basal grouping of the woody non-oak gallers could be due to long-branch outgroup attraction.

Re-analysis

The method they used for the Bayesian inference of the combined morphological and molecular data, with the morphological data matrix “biased” to only include parsimony informative characters is not yet available (Ronquist & Huelsenbeck, in prep.), and I therefore could not exactly reproduce their analysis and test the basal position in their combined analyses with a LB extraction test. I ran the combined Bayesian analysis with MrBayes version 3.0b4 (Ronquist and Huelsenbeck, 2003) without taking this “bias” into account and received similar results, and I doubt the conclusions drawn from the LB extraction test below would be affected, had the bias been taken into account. Since some readers (including myself) might be skeptical towards analyzing morphological characters in a likelihood framework I also did a parsimony analysis on the combined data. Heuristic searches and jackknife analyses were done with the same commands as in the mammal example.

Bayesian analysis

In the Bayesian analysis I reproduced their most parameter rich model (45 parameters); five partitions, each gene with a separate GTR + Γ + I model and the morphological data with a Mk + Γ model. Since I am not convinced that a discrete gamma distribution approximates well the systematic difference of rates across codon position in protein coding genes (the faster rate in third codon positions in particular), I also ran the analysis with a slightly different model. Instead of simply dividing up the dataset into the four genes, I divided it into six partitions; 28S, CO1 third positions, CO1 first & second positions, EF1 & LWRh third

positions, EF1 & LWRh first & second positions and morphology. This was done in multiple single gene and combined analyses that gave insight into what parameters could be reasonably linked to avoid over-parameterization. For instance it became clear that all codon positions in all genes could not be treated separately, especially for the shorter segments of EF1 and LWRh. The six partitions were analyzed simultaneously with a mixed model setting; 28S was given a GTR + Γ + I model while the other four partitions were given a HKY85 + Γ + I model. Γ was linked across all nucleotide partitions but separate for morphology, I was separate for all partitions, state frequencies [A,C,T,G] were linked across the nuclear genes but separate for the mitochondrial gene, the six substitution parameters of GTR were (necessarily) unique for 28S, while the transition/transversion ratio was linked across third positions and first and second positions respectively. Finally, among-partition rate variation was allowed with a relative rate multiplier for each partition. Markov Chain settings: 3 million chain steps sampled every 100 generations. There were four simultaneous chains, one cold and three incrementally heated. Prior settings as well as all proposal mechanisms were left at their default values in MrBayes 3.0b4. Results were drawn from the 20 000 last sampled generations (i.e., burn-in = 10 000). Each run was repeated twice, starting with a random topology, to check for the convergence of results in topology, clade support and model parameter estimations. Apart from checking that the overall likelihood stabilized after a burn-in period, the single parameters were also graphically checked for stabilization.

Bayesian results

The result from the two Bayesian analyses resulted in the exact same preferred ingroup topology (Figs 7 and 8), with only minor differences in group support. This is also the exact same topology found by Nylander et al. (2004) with the same model, which certifies that not taking the “only parsimony-informative-sites” bias in the morphological matrix into account (however, their remedy was constructed; Ronquist & Huelsenbeck, in prep.) had no influence on the topology estimation. It was notable that the three long-branched (Fig. 7) woody non-oak gallers (*Eschatocerus*, *Pediaspis*, *Diplolepis*) group was basal with the outgroup, far away from the oak gallers (*Plagiotrochus*, *Andricus*, *Neuroterus*, *Biorhiza*). Note that at least two internodes support this division of the woody gallers with a posterior probability value of 1.0. This is, as noted above, in strong contrast to the morphology-alone result where woody gallers firmly form a monophyletic unit (Liljeblad and Ronquist, 1998; Ronquist and Liljeblad, 2001; Nylander et al., 2004).

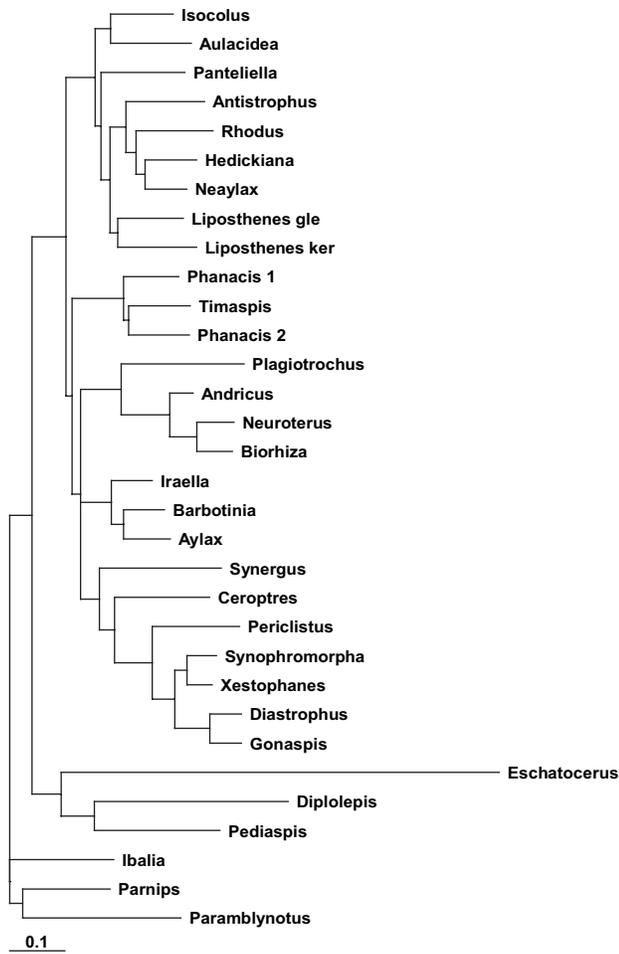


Fig. 7. Majority-rule consensus phylogram from a Bayesian analysis of the combined data of Nylander et al. (2004) using a separate GTR + Γ + I model for each of the four gene partitions and a Mk + Γ model for morphology (their 45-parameter analysis). The scale bar represents the number of expected substitutions per site.

Long branch extraction

The same analyses but with the matrix stripped of the outgroup again resulted in the exact same preferred unrooted ingroup tree in the two analyses (Fig. 9), with differences only in clade support values. The topology is quite different from the rooted analyses however, with the woody gallers now being potentially monophyletic and their split from the remaining taxa supported by the posterior probability of 1.0. Note that although unrooted, the splits in Fig. 9 and Fig. 8 are in conflict and cannot—despite the assertion by values of 1.0 in posterior probability clade supports—at the same time be true. The analysis with the outgroup taxa but without the three long-branched woody non-oak gallers results in a rooting at the same position had the *Eschatocerus–Pediaspis–Diplolepis* clade simply been pruned from the

tree in Fig. 8 (heavy arrow). Again, as in the mammalian example, the inclusion of the outgroups not only rooted the ingroup topology but simultaneously altered it and in this case pulled the *Eschatocerus–Pediaspis–Diplolepis* clade across the tree to a different position (thin arrows in Fig. 8).

Parsimony results

Parsimony analyses of the combined morphological and molecular data resulted in the topology shown in Fig. 10 with the woody gallers basal, paraphyletic, and the *Eschatocerus–Pediaspis–Diplolepis* clade most basal. Excluding the outgroups resulted in a topology where, as expected, woody gallers group together (Fig. 11), which is not in conflict with the rooted topology (Fig. 10). However the topological affinity between all woody gallers and the remaining ingroup taxa has changed, and basically instead of only pulling the *Eschatocerus–Pediaspis–Diplolepis* clade, this time all seven woody gallers have been moved (see thin arrows in Fig. 10). The rooting when the long branched woody non-oak gallers are excluded is indicated by the heavy arrow in Fig. 10, i.e. not the expected base of the oak galler clade if the rooting in Fig. 10 was correct. As seen from the bootstrap values though, the conflicting groupings between the rooted and unrooted trees are in this case not well supported. In fact, after collapsing internodes below 50% in bootstrap value (i.e., comparing the majority-rule consensus trees from 1000 bootstrap replicates instead) differences can only be found in degree of resolution.

A long branch extraction test with the molecular data alone and parsimony or Bayesian methods does not give the same easily interpreted LBA indication, certifying that the morphological data, being 16% of the informative characters, is influential on the above results.

Conclusions from gall wasp example

The conflict Nylander et al. detected (combined versus morphology) was, as seen from this re-analysis, first introduced when the outgroups were included to root the tree. Following the results and advice from Holland et al. (2003) it is more likely that the topological affinities between clades in the unrooted ingroup topology is correct rather than the altered rooted topology. This example could perhaps add to the use of morphology in a molecular millennium (Scotland et al., 2003; Jenner, 2004; Wiens, 2004; above) in that if molecular data conflicts with well supported groups based on morphology *only after rooting*, it is more likely that the rooting is artifactual rather than that the inclusion of the outgroups reveals true conflict between morphology and molecules. Morphology would here serve as a filter to screen against molecular LBA-to-

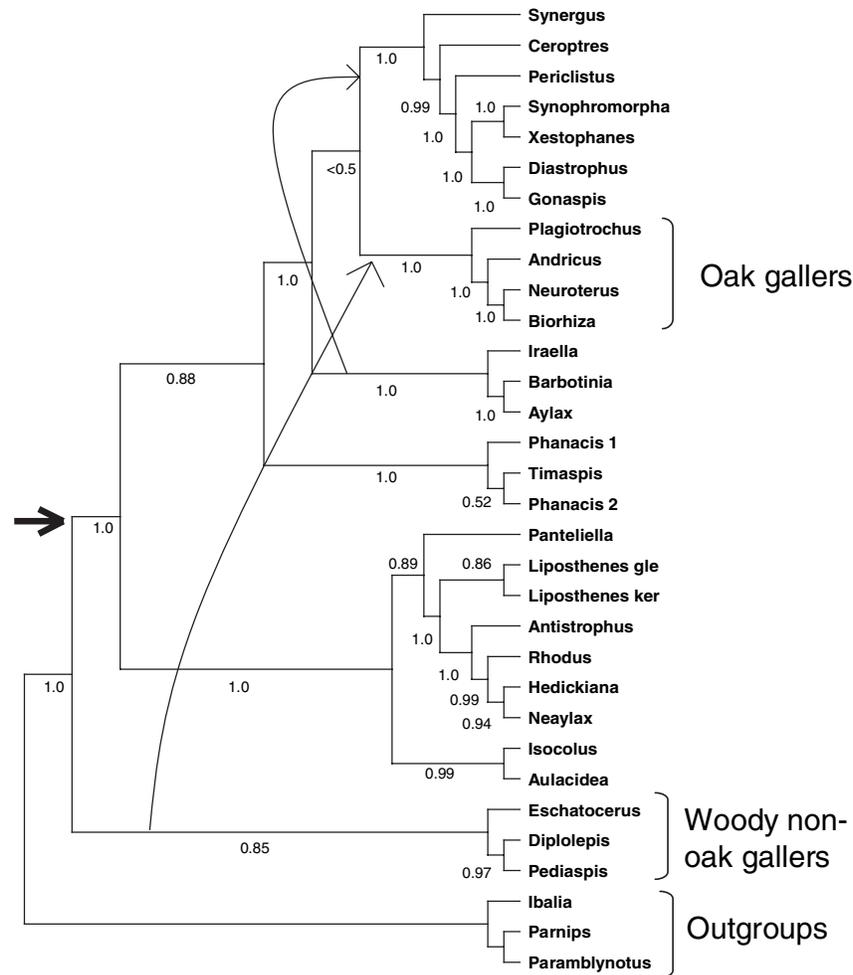


Fig. 8. Topology with highest posterior probability ($P = 0.079$) from a Bayesian analysis of the combined data of Nylander et al. (2004) using a separate GTR + Γ + I model for each of the four gene partitions and a Mk + Γ model for morphology (their 45-parameter analysis). Numbers on internodes are the proportion of trees from the sampled chain in which the clade occurred after burn-in was removed. Thin arrows indicate the different clade affinities compared to the unrooted ingroup topology result in Fig. 9. The thick arrow indicates the outgroup attachment in the search without the *Eschatocerus*–*Diplolepis*–*Pediaspis* clade.

outgroup rootings (Wheeler, 1990; Holland et al., 2003). Being only a tentative suggestion, the generality of this use needs to be much further evaluated by numerous empirical datasets. This is to my knowledge the first ever real example of long-branch attraction that has been inferred by a method, not only with high support values, but with support values indicating 100% certainty “that the clade is true” (Huelsenbeck et al., 2002b, p. 675). Moreover, it is demonstrated, using likelihood-based Bayesian inference with complex models of molecular (and morphological) evolution (partition specific GTR + Γ + I). Thus the problem of LBA is not universally solved simply by using likelihood-based methods, at least not with the models developed to date. The interpretation of the often very high posterior probability values of clades has been extensively debated in the recent literature (Huelsenbeck et al., 2002b;

Leaché and Reeder, 2002; Reed et al., 2002; Suzuki et al., 2002; Whittingham et al., 2002; Wilcox et al., 2002; Alfaro et al., 2003; Cummings et al., 2003; Douady et al., 2003; Erixon et al., 2003; Goloboff and Pol, 2004; Pickett et al., 2004; Simmons et al., 2004; Huelsenbeck and Rannala, 2004) and they should be taken with caution. Posterior probability values are certainly not directly comparable to bootstrap values, but whether they are more accurate or over-inflated is a matter of controversy in the cited references. Without getting involved in that debate, one can at least conclude from the gall wasp example that a clade or split supported by 1.0 in posterior probability is not a guarantee against alternative conflicting resolution—likewise supported with “certainty”—following increased taxon sampling (in this case outgroups used for rooting).

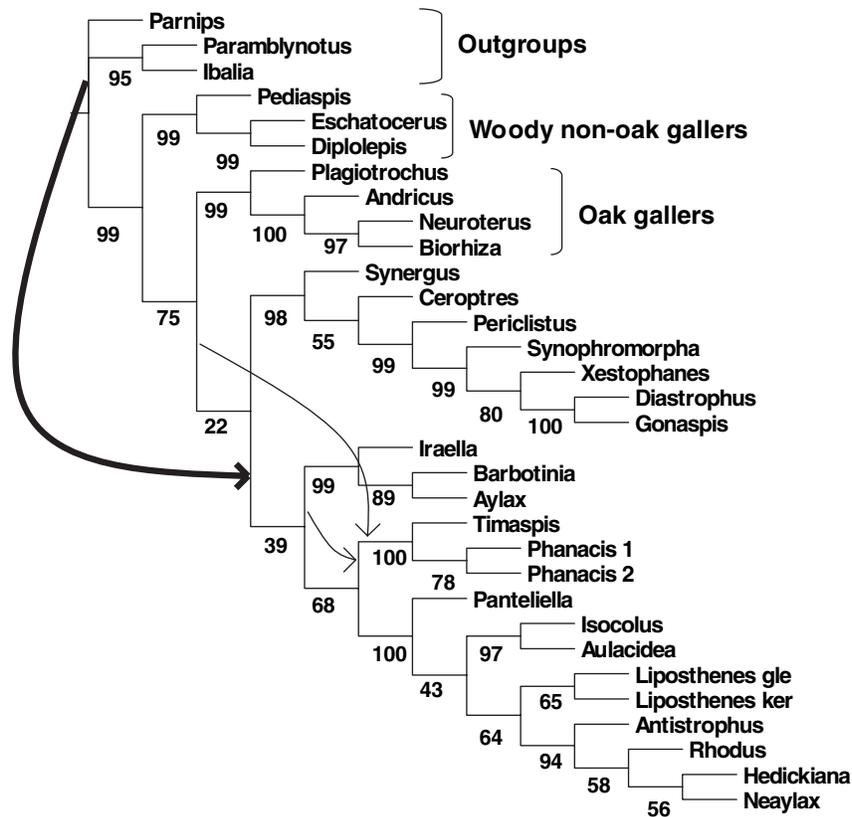


Fig. 10. Single most parsimonious tree from the analysis of the combined data of Nylander et al. (2004). Numbers on internodes refer to jackknife support from 1000 replicates. Thin arrows indicate the different clade affinities compared to the unrooted ingroup topology result in Fig. 11. The thick arrow indicates the outgroup attachment in the search without the *Eschatocerus–Diplolepis–Pediaspis* clade.

“...morphological characters that are less susceptible to long-branch attraction” non-arbitrary divisions of data exist with regard to sensitivity of LBA artifacts, which fulfill the heuristic merits of analyzing them separately as an exploratory tool for LBA. Second, that nucleotides within genes evolve independently has been found to be violated by several studies (Cummings et al., 1995; Naylor and Brown, 1998; Rokas et al., 2003), which, for an exploratory purpose, adds meaning to separate gene partition analyses. Although conflicts within and among data partitions could be detected by the *ci/r* exploration method suggested by Grant and Kluge (2003), it lacks the ability to point at possible LBA artifacts in a simple and illustrative way.

Apart from analyzing, for example, morphological versus DNA datasets separately I see a great potential in the analysis of partitions, which cannot have differing evolutionary histories, like codon positions (see Sanderson et al., 2000 and Debry, 2003) or the genes in the (normally) maternally inherited non-recombining mitochondrial and chloroplast genomes. Note that I am only arguing for separate partition analysis as an exploratory tool for LBA artifacts (as suggested by Wiens and Hollingsworth, 2000), not against the total evidence approach, with a simulta-

neous analysis of all available data being the severest test of competing phylogenetic hypothesis and identifying the hypothesis of greatest explanatory power (Grant and Kluge, 2003). On the contrary, Rokas et al.’s (2003) result highlights the importance of combining data from many unlinked genes to correctly infer the complete genome-tree.

In addition, Grant and Kluge (2003, p. 398) rapidly refute the comparison of results from parsimony and likelihood methods as neither a scientific nor heuristic exploratory tool of LBA artifacts by “...this procedure relies on the assumption that maximum likelihood is immune to long-branch attraction”, which admittedly it is not. However, although not *immune*, there is no doubt that likelihood based methods (including Bayesian) are *less sensitive* to LBA (Pol and Siddall, 2001; Swofford et al., 2001; see above), which is why comparing the outcome on long branches of these methods does have heuristic value. Comparing the outcome of parsimony and model-based inference methods, if LBA is suspected, is strongly heuristic in pointing towards taxa the position of which should be further tested, for example by adding taxa to break up long branches.

Finally, I agree with Grant and Kluge (2003) that the “long-branch extraction” method suggested by Siddall

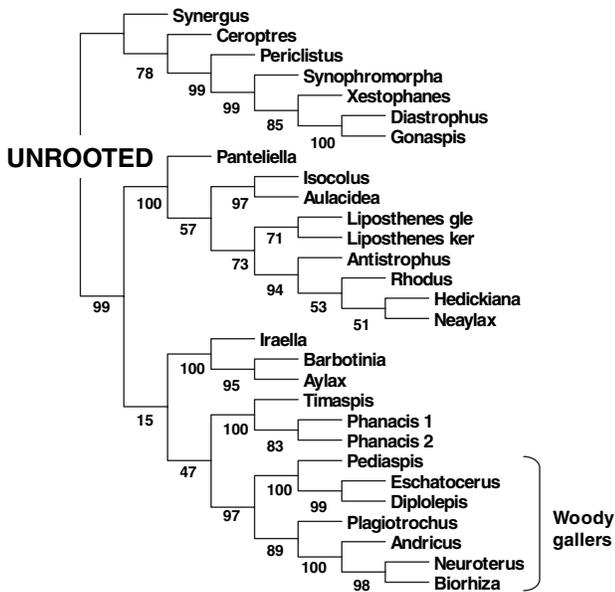


Fig. 11. Single most parsimonious unrooted ingroup topology from the analysis of the combined data of Nylander et al. (2004), excluding outgroups. Since unrooted, numbers on internodes are the proportion of trees from the 1000 jackknife replicates in which the split occurred, i.e. they do not indicate support for monophyletic groups. Neither do named clades indicate monophyletic groups, the grouping of “woody gallers” here is only done to illustrate that at this stage, before rooting, there is no conflict with morphology-alone results (see text).

and Whiting (1999) is one of the best heuristic methods for detecting possible LBA. The properties that makes this method especially appealing is its combination of simplicity with power, at the same time as minimizing the amount of data excluded for the test. Simplicity, because it is very quick to delete a taxon from a matrix and rerun the analysis, put it back, delete the other suspect and rerun the program. Power, because alternative explanations to LBA are hard to find if the two long branches end up in different parts of the tree when alone, but moved and pulled together when included simultaneously. Minimizing the amount of data that needs to be excluded for a test is essential since any conclusion drawn from the result will be weaker the fewer data are used to derive the result.

LBA to outgroups

I want in particular expose the fact that one or several outgroup taxa almost always form long, potentially problematic branches, unless the outgroup is sampled as extensively as the ingroup (which is rarely the case), and thus continuously represent a hazard towards misplacing long branched ingroup taxa. Note that this has nothing to do with unequal evolutionary rates, which, perhaps in the beginning, were thought to be a prerequisite of LBA. Numerous studies have shown that LBA can occur in trees following a molecular clock (Hendy

and Penny, 1989; Kim, 1996; Holland et al., 2003), and the conditions of branch length differences necessary to create LBA can easily be attained by unequal taxon sampling, without any need for unequal rates. The almost universal rooting with the outgroup taxa approach frequently creates such conditions. In this review the most common cause in studies suggesting LBA artifacts is related to ingroup taxa being pulled towards the long branched outgroup. As such, I would suggest that all phylogenetic analyses are run both with and without the outgroups to compare whether the outgroup only roots the ingroup tree, or if it simultaneously alters the ingroup topology. This has already been suggested by, for example Lin et al. (2002a) and Holland et al. (2003). Basically it is a variant of long-branch extraction where the outgroup is always one of the suspects. Although the arguments of Nixon and Carpenter (1993) for treating outgroup and ingroup terminals similarly in order to reach global parsimony as well as testing ingroup monophyly hold well, investigating the effects of the outgroup by also searching for the best unrooted ingroup topology serves as an important heuristic exploratory tool for identifying possible outgroup attraction artifacts. Certainly this has been of great value in identifying artifactual rootings, not only in placental mammals (Lin et al., 2002a,b; this analysis) but for example parabasalian protozoans (Hampl et al., 2004), monocotyledons (Graham et al., 2002) and birds (Garcia-Moreno et al., 2003). In both birds and mammals the investigation of unrooted ingroup topologies has also identified that severe, at first look “unsolvable”, conflicts between mitochondrial and nuclear datasets only pertain to outgroup rooted topologies—in the unrooted ingroup-only topologies there is basically no conflict (Lin et al., 2002a,b; Garcia-Moreno et al., 2003).

Few taxa, many characters: illusory safety

Another general pattern emerging from this survey was that spurious conclusions are often derived from an over-credibility of enormous numbers of nucleotide or amino acid characters (e.g., complete genomes) when combined with poor taxon sampling. “The guinea-pig is not a rodent” example (D’Erchia et al., 1996) is not unique. Passerine birds being most ancestral (Härlid and Arnason, 1999) as well as monocots being the basalmost angiosperm (Goremykin et al., 2003) follow the same pattern, based on complete mitochondrial or chloroplast genomes, high support values (independent of method used MP, ML, NJ, on amino acids or nucleotides) for a spurious result involving basal-most taxa (attracted to outgroups), and extremely poor taxon sampling (limited by available genomes), followed later by studies showing the sensitivity of these results to taxon-sampling (Philippe, 1997; van Tuinen et al., 2000; Soltis and

Soltis, 2004). Intuitively, the feeling is understandable that a dataset of 9255 or 30 017 aligned nucleotides should be able to correctly infer the relationship between four birds (Härlid and Arnason, 1999) or 13 plants (Goremykin et al., 2003), respectively. But the feeling is illusory, which becomes clear when simulation studies on methods and inconsistency are consulted. The very definition of inconsistency clearly states that adding more data will only strengthen the wrong tree (Sullivan and Swofford, 1997). Thus a bootstrap, jackknife or posterior probability support of 100% is no guarantee. However, the important point already stated earlier is that simulation studies on inconsistency assume that adding more data means adding more data *of the same kind*, which, when mitochondrial or chloroplast genomes are sequenced, might be at least partly fulfilled. The inconsistency prediction does not necessarily hold if the data to be added in practice is of another kind, e.g., adding morphological, behavioral, chemical datasets, sequences from other genomes, or genes with different properties, e.g., protein-coding versus ribosomal genes, slower evolving genes, etc. Numerous recent studies have concluded that success comes from combining different datasets to overcome artifacts from single-gene analysis (Qiu et al., 1999; Gatesy et al., 2003; Rokas et al., 2003; Wahlberg and Nylin, 2003; Bergsten and Miller, 2004; Gontcharov et al., 2004; Springer et al., 2004). In addition, the illusion of safety in studies with few taxa and many characters lacks empirical and theoretical support, which instead overwhelmingly stresses the importance of taxon sampling for accurate phylogenetic inference (Hillis, 1996, 1998; Graybeal, 1998; Poe, 1998; Rannala et al., 1998; Pollock et al., 2002; Zwickl and Hillis, 2002). To cite Soltis and Soltis (2004, p. 1000) “As exciting as genomic data are, extensive character sampling cannot compensate for inadequate taxon sampling.” I fully agree.

Suggestion

Summarizing my suggestions on what to do if LBA is suspected, I envision a hypothetical example:

1. Imagine a total evidence parsimony analysis of a dataset involving one gene and morphology that produces a surprising result, tree Z, where an unexpected taxon X of the monophyletic ingroup ends up most basal in the tree, making the group Y where it was formally classified polyphyletic. You suspect LBA. Test it:

2. Exclude the outgroup (whether one or several taxa) from the matrix and re-run the parsimony analysis. Does taxon X now end up in a position on the unrooted tree that makes the monophyly of clade Y possible?

3. Insert the outgroup again but remove taxon X from the matrix and re-run the analysis. Compare the rooting with tree Z and with results of step 4 and 5.

Include all taxa in the matrix again.

4. Analyze the gene and the morphological data separately by parsimony. Does morphology make group Y monophyletic, while the gene places taxon X basally, as in the total evidence analysis?

5. Analyze the gene with a method taking branch lengths into account, for example Bayesian phylogenetics with an appropriate model. Is clade Y now monophyletic in the result?

6. Are the estimated branches of taxon X and the outgroup among the longest in the tree?

If you have answered yes to all the questions above you have made LBA the least refuted hypothesis of the outcome in step 1, based on the combined tests of long branch extraction, separate partition analysis and methodological discordance. The outcome of the comparison in step 3 is not crucial for the conclusion, but can often provide additional information. This was admittedly a stereotypic example, but perhaps not uncommon. A “no” answer to one of the questions above opens up other explanations. A “no” answer on all, within the specifics of the results, has probably made an alternative hypothesis more likely and rejects LBA as an explanation (except if the results are inverse in step 4, and LBA of the morphological data is thus still a possibility). The gray zone in between urges for refinements of combined tests of LBA.

Optional

Test if the outgroup behaves like a random sequence.

Test if the outgroup branch and the branch of taxon X is long enough to attract with parsimony, even when the model trees have them apart with parametric simulation.

Continue to explore the conflict in the dataset by split graphs.

Remedy for the total evidence parsimony analysis:

1. If possible, increase taxon sampling such that the branch between taxon X and the outgroup are broken up.

2. Sequence another, unlinked, and more slowly evolving gene for the group and include it in the total evidence data.

The future

The evidence is now overwhelming that LBA occurs in real datasets. This urges us to develop methods or techniques to deal with the problem, overcome it, or at least identify it. To completely overcome the problem seems far away since, as shown here, even the promising new Bayesian application to phylogenetics and the constantly increasing complexity of new models are still susceptible to LBA. This should, more than today, push instead for further development of new methods and refinements of currently available ones to detect when LBA is present in a specific study. For example, when outgroups attach the longest ingroup branch but do not

alter the ingroup topology when analyzed alone. How do we separate the two very plausible explanations that: (1) the rooting is correct (that the basal-most branch is the longest is perfectly expectable), from (2) the rooting is wrong and due to the outgroup being too distant and random, artificially rooting on the longest branch (also perfectly expectable)? Thus far, long-branch extraction (including running the analyses with and without outgroups), separate partition analyses, methodological discordance, parametric simulation (including random outgroups) and split graphs are the available tools for the detection of LBA, none of which alone provides sufficient evidence, but in several combinations can make LBA the least falsified hypothesis. Theoretical work is badly needed to explore the relationship between outgroup rooting and LBA, especially how increasing different outgroup taxon sampling strategies affect the probability of LBA. Sampling more taxa to break up long branches and adding/combining more diverse data, especially morphological with molecular data and unlinked genes, are probably the best methods for overcoming the artifact, as well as using inference methods less sensitive to the phenomenon for a certain kind of data. However, whether one wants to use parsimony, likelihood or Bayesian methods for the final inference of the best phylogenetic hypothesis depends on many other considerations other than sensitivity to LBA, the exploration of which was never the intention of this paper.

Acknowledgments

I thank Mark Springer for sending me their dataset of placental mammals and Johan Nylander et al. (2004) for depositing their gall wasp datasets in the freely, online accessible, TreeBase database. I also thank Anders N. Nilsson and Mats Wedin for comments on an earlier draft of this manuscript and two anonymous referees for significantly improving the manuscript through comments, questioning and criticism. This work was financially supported by Helge Ax:son Johnsons Stiftelse.

References

- Aguinaldo, A.M.A., Turbeville, J.M., Linford, L.S., Rivera, M.C., Garey, J.R., Raff, R.A., Lake, J.A., 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature*, 387, 489–493.
- Alfaro, M.E., Zoller, S., Lutzoni, F., 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Mol. Biol. Evol.* 20, 255–266.
- Andersson, F.E., Swofford, D.L., 2004. Should we be worried about long-branch attraction in real data sets? Investigations using metazoan 18S rDNA. *Mol. Phyl. Evol.* 33, 440–451.
- Archibald, J.M., O’Kelly, C.J., Doolittle, W.F., 2002. The chaperonin genes of jakobid and jakobid-like flagellates: implications for eukaryotic evolution. *Mol. Biol. Evol.* 19, 422–431.
- Arnason, U., Adegoke, J.A., Bodin, K., Born, E.W., Esa, Y.B., Gullberg, A., Nilsson, M., Short, R.V., Xu, X.F., Janke, A., 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. *Proc. Natl Acad. Sci. USA*, 99, 8151–8156.
- Baker, R.H., Yu, X., DeSalle, R., 1998. Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. *Mol. Phyl. Evol.* 9, 427–436.
- Bandelt, H.-J., Dress, A.W.M., 1992. Split decomposition: a new and useful approach to phylogenetic analysis of distance data. *Mol. Phyl. Evol.* 1, 242–252.
- Barkman, T.J., Chenery, G., McNeal, J.R., Lyons-Weiler, J., Ellisens, W.J., Moore, G., Wolfe, A.D., dePamphilis, C.W., 2000. Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proc. Natl Acad. Sci. USA*, 97, 13166–13171.
- Belshaw, R., Dowton, M., Quicke, D.L.J., Austin, A.D., 2000. Estimating ancestral geographical distributions: a Gondwanan origin for aphid parasitoids? *Proc. R. Soc. London (B), Biol. Sci.* 267, 491–496.
- Bergsten, J., Miller, K.B., 2004. *Acilius* phylogeny (Coleoptera: Dytiscidae), problems with long-branch attraction and morphological intersexual coevolution. *Cladistics*, 20, 76–77 [Meeting Abstract].
- Bocchetta, M., Gribaldo, S., Sanangelantoni, A., Cammarano, P., 2000. Phylogenetic depth of the bacterial genera *Aquifex* and *Thermotoga* inferred from analysis of ribosomal protein, elongation factor, and RNA polymerase subunit sequences. *J. Mol. Evol.* 50, 366–380.
- Bowe, L.M., Coat, G., dePamphilis, C.W., 2000. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales’ closest relatives are conifers. *Proc. Natl Acad. Sci. USA*, 97, 4092–4097.
- Brinkmann, H., Philippe, H., 1999. Archaea sister group of bacteria? Indications from tree reconstruction artefacts in ancient phylogenies. *Mol. Biol. Evol.* 16, 817–825.
- Bull, J.J., Cunningham, C.W., Molineux, I.J., Badgett, M.R., Hillis, D.M., 1993. Experimental molecular evolution of bacteriophage-T7. *Evolution*, 47, 993–1007.
- Cao, Y., Fujiwara, M., Nikaido, M., Okada, N., Hasegawa, M., 2000. Interordinal relationships and timescale of eutherian evolution as inferred from mitochondrial genome data. *Gene*, 259, 149–158.
- Cao, Y., Janke, A., Waddell, P.J., Westerman, M., Takenaka, O., Murata, S., Okada, N., Paabo, S., Hasegawa, M., 1998. Conflict among individual mitochondrial proteins in resolving the phylogeny of eutherian orders. *J. Mol. Evol.* 47, 307–322.
- Carnean, D., Crespi, B.J., 1995. Do long branches attract flies? *Nature*, 373, 666–666.
- Chang, J.T., 1996. Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. *Math. Biosci.* 134, 189–215.
- Chen, J.M., Cutler, C., Jacques, C., Boeuf, G., Denamur, E., Lecoindre, G., Mercier, B., Cramb, G., Ferec, C., 2001. A combined analysis of the cystic fibrosis transmembrane conductance regulator: implications for structure and disease models. *Mol. Biol. Evol.* 18, 1771–1788.
- Clements, K.D., Gray, R.D., Choat, J.H., 2003. Rapid evolutionary divergences in reef fishes of the family Acanthuridae (Perciformes: Teleostei). *Mol. Phyl. Evol.* 26, 190–201.
- Colless, D.H., 1995. Relative symmetry of cladograms and phenograms – an experimental study. *Syst. Biol.* 44, 102–108.
- Culligan, K.M., Meyer-Gauen, G., Lyons-Weiler, J., Hays, J.B., 2000. Evolutionary origin, diversification and specialization of eukaryotic MutS homolog mismatch repair proteins. *Nucl. Acids Res.* 28, 463–471.

- Cummings, M.P., Handley, S.A., Myers, D.S., Reed, D.L., Rokas, A., Winka, K., 2003. Comparing bootstrap and posterior probability values in the four-taxon case. *Syst. Biol.* 52, 477–487.
- Cummings, M.P., Otto, S.P., Wakeley, J., 1995. Sampling properties of DNA-sequence data in phylogenetic analysis. *Mol. Biol. Evol.* 12, 814–822.
- D'Erchia, A.M., Gissi, C., Pesole, G., Saccone, C., Arnason, U., 1996. The guinea-pig is not a rodent. *Nature*, 381, 597–600.
- Dacks, J.B., Marinets, A., Doolittle, W.F., Cavalier-Smith, T., Logsdon, J.M., 2002. Analyses of RNA polymerase II genes from free-living protists: phylogeny, long branch attraction, and the eukaryotic big bang. *Mol. Biol. Evol.* 19, 830–840.
- Debry, R.W., 2003. Identifying conflicting signal in a multigene analysis reveals a highly resolved tree: the phylogeny of Rodentia (Mammalia). *Syst. Biol.* 52, 604–617.
- Delsuc, F., Scally, M., Madsen, O., Stanhope, M.J., de Jong, W.W., Catzeflis, F.M., Springer, M.S., Douzery, E.J.P., 2002. Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. *Mol. Biol. Evol.* 19, 1656–1671.
- Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F., Douzery, E.J.P., 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Mol. Biol. Evol.* 20, 248–254.
- Doyle, J.J., 1997. Trees within trees: genes and species, molecules and morphology. *Syst. Biol.* 46, 537–553.
- Erixon, P., Svennblad, B., Britton, T., Oxelman, B., 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst. Biol.* 52, 665–673.
- Faivovich, J., 2002. On RASA. *Cladistics*, 18, 324–333.
- Farias, I.P., Orti, G., Sampaio, I., Schneider, H., Meyer, A., 2001. The cytochrome b gene as a phylogenetic marker: the limits of resolution for analyzing relationships among cichlid fishes. *J. Mol. Evol.* 53, 89–103.
- Farris, J.S., 1999. Likelihood and inconsistency. *Cladistics*, 15, 199–204.
- Farris, J.S., 2002. RASA attributes highly significant structure to randomized data. *Cladistics*, 18, 334–353.
- Farris, J.S., Källersjö, M., 1998. Asymmetry and explanations. *Cladistics*, 14, 159–166.
- Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27, 401–410.
- Felsenstein, J., 2003. *Inferring Phylogenies*. Sinauer Associates, Sunderland, MA.
- Flook, P.K., Rowell, C.H.F., 1997. The effectiveness of mitochondrial rRNA gene sequences for the reconstruction of the phylogeny of an insect order (Orthoptera). *Mol. Phyl. Evol.* 8, 177–192.
- Garcia-Moreno, J., Sorenson, M.D., Mindell, D.P., 2003. Congruent avian phylogenies inferred from mitochondrial and nuclear DNA sequences. *J. Mol. Evol.* 57, 27–37.
- Gatesy, J., Amato, G., Norell, M., DeSalle, R., Hayashi, C., 2003. Combined support for wholesale taxic atavism in gavialine crocodylians. *Syst. Biol.* 52, 403–422.
- Gaut, B.S., Lewis, P.O., 1995. Success of maximum-likelihood phylogeny inference in the 4-taxon case. *Mol. Biol. Evol.* 12, 152–162.
- Goloboff, P.A., 1993. Estimating character weights during tree-search. *Cladistics*, 9, 83–91.
- Goloboff, P., 1999. NONA (NO NAME), Version 2.0. Published by the Author, Tucuman, Argentina.
- Goloboff, P.A., Pol, D., 2004. Cases in which Bayesian phylogenetic analysis will be positively misleading. *Cladistics*, 20, 83–84. [Meeting Abstract].
- Gontcharov, A.A., Marin, B., Melkonian, M., 2004. Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and rbcL sequence comparisons in the Zygnemato-phyceae (Streptophyta). *Mol. Biol. Evol.* 21, 612–624.
- Goremykin, V.V., Hirsch-Ernst, K.I., Wolff, S., Hellwig, F.H., 2003. Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Mol. Biol. Evol.* 20, 1499–1505.
- Graham, S.W., Olmstead, R.G., Barrett, S.C.H., 2002. Rooting phylogenetic trees with distant outgroups: a case study from the commelinoid monocots. *Mol. Biol. Evol.* 19, 1769–1781.
- Grant, T., Kluge, A.G., 2003. Data exploration in phylogenetic inference: scientific, heuristic, or neither. *Cladistics*, 19, 379–418.
- Graybeal, A., 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* 47, 9–17.
- Gribaldo, S., Philippe, H., 2002. Ancient phylogenetic relationships. *Theor. Pop. Biol.* 61, 391–408.
- Guyer, C., Slowinski, J.B., 1993. Adaptive radiation and the topology of large phylogenies. *Evolution*, 47, 253–263.
- Halanych, K.M., 1998. Lagomorphs misplaced by more characters and fewer taxa. *Syst. Biol.* 47, 138–146.
- Hapl, V., Cepicka, I., Flegr, J., Tachezy, J., Kulda, J., 2004. Critical analysis of the topology and rooting of the parabasalian 16S rRNA tree. *Mol. Phyl. Evol.* 32, 711–723.
- Hanelt, B., VanSchyndel, D., Adema, C.M., Lewis, L.A., Loker, E.S., 1996. The phylogenetic position of *Rhopalura ophiocoma* (Ortho-nectida) based on 18S ribosomal DNA sequence analysis. *Mol. Biol. Evol.* 13, 1187–1191.
- Härlid, A., Arnason, U., 1999. Analyses of mitochondrial DNA nest ratite birds within the Neognathae: supporting a neotenus origin of ratite morphological characters. *Proc. R. Soc. London Series B, Biol. Sci.* 266, 305–309.
- Heard, S.B., 1992. Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees. *Evolution*, 46, 1818–1826.
- Hendy, M.D., Penny, D., 1989. A framework for the quantitative study of evolutionary trees. *Syst. Zool.* 38, 297–309.
- Hendy, M.D., Penny, D., 1993. Spectral analysis of phylogenetic data. *J. Classification*, 10, 5–24.
- Hillis, D.M., 1996. Inferring complex phylogenies. *Nature*, 383, 130–131.
- Hillis, D.M., 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Syst. Biol.* 47, 3–8.
- Hillis, D.M., Bull, J.J., White, M.E., Badgett, M.R., Molineux, I.J., 1992. Experimental phylogenetics – generation of a known phylogeny. *Science*, 255, 589–592.
- Hillis, D.M., Huelsenbeck, J.P., Cunningham, C.W., 1994a. Application and accuracy of molecular phylogenies. *Science*, 264, 671–677.
- Hillis, D.M., Huelsenbeck, J.P., Swofford, D.L., 1994b. Hobgoblin of phylogenetics. *Nature*, 369, 363–364.
- Holland, B.R., Penny, D., Hendy, M.D., 2003. Outgroup misplacement and phylogenetic inaccuracy under a molecular clock – A simulation study. *Syst. Biol.* 52, 229–238.
- Huelsenbeck, J.P., 1995. Performance of phylogenetic methods in simulation. *Syst. Biol.* 44, 17–48.
- Huelsenbeck, J.P., 1997. Is the Felsenstein zone a fly trap? *Syst. Biol.* 46, 69–74.
- Huelsenbeck, J.P., 1998. Systematic bias in phylogenetic analysis: is the Strepsiptera problem solved? *Syst. Biol.* 47, 519–537.
- Huelsenbeck, J.P., Bollback, J.P., Levine, A.M., 2002a. Inferring the root of a phylogenetic tree. *Syst. Biol.* 51, 32–43.
- Huelsenbeck, J.P., Hillis, D.M., 1993. Success of phylogenetic methods in the 4-taxon case. *Syst. Biol.* 42, 247–264.
- Huelsenbeck, J.P., Hillis, D.M., Jones, R., 1996. Parametric bootstrapping in molecular phylogenetics: application and performance. In: Ferraris, J.D., Palumbi, S.R. (Eds.), *Parametric Bootstrapping in Molecular Phylogenetics: Application and Performance*. John Wiley & Sons, New York.
- Huelsenbeck, J.P., Lander, K.M., 2003. Frequent inconsistency of parsimony under a simple model of cladogenesis. *Syst. Biol.* 52, 641–648.

- Huelsenbeck, J.P., Larget, B., Miller, R.E., Ronquist, F., 2002b. Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst. Biol.* 51, 673–688.
- Huelsenbeck, J.P., Rannala, B., 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex models. *Syst. Biol.* 53, 904–913.
- Inagaki, Y., Susko, E., Fast, N.M., Roger, A.J., 2004. Covariation shifts cause a long-branch attraction artefact that unites microsporidia and archaeobacteria in EF-1 alpha phylogenies. *Mol. Biol. Evol.* 21, 1340–1349.
- Jenkins, C., Fuerst, J.A., 2001. Phylogenetic analysis of evolutionary relationships of the planctomycete division of the domain bacteria based on amino acid sequences of elongation factor Tu. *J. Mol. Evol.* 52, 405–418.
- Jenner, R.A., 2004. Accepting partnership by submission? Morphological phylogenetics in a molecular millennium. *Syst. Biol.* 53, 333–342.
- de Jong, W.W., 1998. Molecules remodel the mammalian tree. *TREE*, 13, 270–275.
- Källersjö, M., Albert, V.A., Farris, J.S., 1999. Homoplasy increases phylogenetic structure. *Cladistics*, 15, 91–93.
- Kennedy, M., Paterson, A.M., Morales, J.C., Parsons, S., Winnington, A.P., Spencer, H.G., 1999. The long and short of it: branch lengths and the problem of placing the New Zealand short-tailed bat, *Mystacina*. *Mol. Phyl. Evol.* 13, 405–416.
- Kim, J.H., 1996. General inconsistency conditions for maximum parsimony: effects of branch lengths and increasing numbers of taxa. *Syst. Biol.* 45, 363–374.
- Kim, J.H., Kim, W., Cunningham, C.W., 1999. A new perspective on lower metazoan relationships from 18S rDNA sequences. *Mol. Biol. Evol.* 16, 423–427.
- 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., 1998. Total evidence or taxonomic congruence: cladistics or consensus classification. *Cladistics*, 14, 151–158.
- Kolaczowski, B., Thornton, J.W., 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature*, 431, 980–984.
- Krettek, A., Gullberg, A., Arnason, U., 1995. Sequence analysis of the complete mitochondrial DNA molecule of the hedgehog, *Erinaceus europaeus*, and the phylogenetic position of the Lipotyphla. *J. Mol. Evol.* 41, 952–957.
- Lake, J.A., 1987. A rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony. *Mol. Biol. Evol.* 4, 167–191.
- Lake, J.A., 1998. Optimally recovering rate variation information from genomes and sequences: pattern filtering. *Mol. Biol. Evol.* 15, 1224–1231.
- Leaché, A.D., Reeder, T.W., 2002. Molecular systematics of the Eastern Fence Lizard *Sceloporus undulatus*, A comparison of parsimony, likelihood, and Bayesian approaches. *Syst. Biol.* 51, 44–68.
- Lewis, P.O., 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* 50, 913–925.
- Liljeblad, J., Ronquist, F., 1998. A phylogenetic analysis of higher-level gall wasp relationships (Hymenoptera: Cynipidae). *Syst. Entomol.* 23, 229–252.
- Lin, Y.H., McLenachan, P.A., Gore, A.R., Phillips, M.J., Ota, R., Hendy, M.D., Penny, D., 2002a. Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. *Mol. Biol. Evol.* 19, 2060–2070.
- Lin, Y.H., Penny, D., 2001. Implications for bat evolution from two new complete mitochondrial genomes. *Mol. Biol. Evol.* 18, 684–688.
- Lin, Y.H., Waddell, P.J., Penny, D., 2002b. Pika and vole mitochondrial genomes increase support for both rodent monophyly and glires. *Gene*, 294, 119–129.
- Lockhart, P.J., Cameron, S.A., 2001. Trees for bees. *TREE*, 16, 84–88.
- Lockhart, P.J., Larkum, A.W.D., Steel, M.A., Waddell, P.J., Penny, D., 1996. Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. *Proc. Natl Acad. Sci. USA*, 93, 1930–1934.
- Lopez, P., Forterre, P., Philippe, H., 1999. The root of the tree of life in the light of the covariation model. *J. Mol. Evol.* 49, 496–508.
- Luckett, W.P., Hartenberger, J.-L., 1993. Monophyly or polyphyly of the order Rodentia: possible conflict between morphological and molecular interpretations. *J. Mamm. Evol.* 1, 127–147.
- Lyons-Weiler, J., Hoelzer, G.A., 1997. Escaping from the Felsenstein zone by detecting long branches in phylogenetic data. *Mol. Phyl. Evol.* 8, 375–384.
- Lyons-Weiler, J., Hoelzer, G.A., Tausch, R.J., 1996. Relative apparent synapomorphy analysis (RASA) I: the statistical measurement of phylogenetic signal. *Mol. Biol. Evol.* 13, 749–757.
- Maddison, W.P., 1997. Gene trees in species trees. *Syst. Biol.* 46, 523–536.
- Maddison, D.R., Baker, M.D., Ober, K.A., 1999. Phylogeny of carabid beetles as inferred from 18S ribosomal DNA (Coleoptera: Carabidae). *Syst. Entomol.* 24, 103–138.
- Madsen, O., Scally, M., Douady, C.J., Kao, D.J., DeBry, R.W., Adkins, R., Amrine, H.M., Stanhope, M.J., de Jong, W.W., Springer, M.S., 2001. Parallel adaptive radiations in two major clades of placental mammals. *Nature*, 409, 610–614.
- Moreira, D., Kervestin, S., Jean-Jean, O., Philippe, H., 2002. Evolution of eukaryotic translation elongation and termination factors: variations of evolutionary rate and genetic code deviations. *Mol. Biol. Evol.* 19, 189–200.
- Moreira, D., Le Guyader, H., Philippe, H., 1999. Unusually high evolutionary rate of the elongation factor lar genes from the ciliophora and its impact on the phylogeny of eukaryotes. *Mol. Biol. Evol.* 16, 234–245.
- Moreira, D., Le Guyader, H., Philippe, H., 2000. The origin of red algae and the evolution of chloroplasts. *Nature*, 405, 69–72.
- Mouchaty, S.K., Gullberg, A., Janke, A., Arnason, U., 2000a. The phylogenetic position of the Talpidae within Eutheria based on analysis of complete mitochondrial sequences. *Mol. Biol. Evol.* 17, 60–67.
- Mouchaty, S.K., Gullberg, A., Janke, A., Arnason, U., 2000b. Phylogenetic position of the Tenrecs (Mammalia: Tenrecidae) of Madagascar based on analysis of the complete mitochondrial genome sequence of *Echinops telfairi*. *Zool. Scripta*, 29, 307–317.
- Murphy, W.J., Eizirik, E., Johnson, W.E., Zhang, Y.P., Ryder, O.A., O'Brien, S.J., 2001a. Molecular phylogenetics and the origins of placental mammals. *Nature*, 409, 614–618.
- Murphy, W.J., Eizirik, E., O'Brien, S.J., Madsen, O., Scally, M., Douady, C.J., Teeling, E., Ryder, O.A., Stanhope, M.J., de Jong, W.W., Springer, M.S., 2001b. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science*, 294, 2348–2351.
- Naylor, G.J.P., Brown, W.M., 1998. Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. *Syst. Biol.* 47, 61–76.
- Nichols, R., 2001. Gene trees and species trees are not the same. *TREE*, 16, 358–364.
- Nikaido, M., Cao, Y., Harada, M., Okada, N., Hasegawa, M., 2003. Mitochondrial phylogeny of hedgehogs and monophyly of Eulipotyphla. *Mol. Phyl. Evol.* 28, 276–284.
- Nikaido, M., Kawai, K., Cao, Y., Harada, M., Tomita, S., Okada, N., Hasegawa, M., 2001. Maximum likelihood analysis of the complete mitochondrial genomes of eutherians and a reevaluation of the phylogeny of bats and insectivores. *J. Mol. Evol.* 53, 508–516.

- Nixon, C., 1999. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics*, 15, 407–414.
- Nixon, K., 2002. WinClada, Version 1.00.08. Published by the Author, Ithaca, NY.
- Nixon, K.C., Carpenter, J.M., 1993. On outgroups. *Cladistics*, 9, 413–426.
- Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. *Syst. Biol.* 53, 47–67.
- Omilian, A.R., Taylor, D.J., 2001. Rate acceleration and long-branch attraction in a conserved gene of cryptic daphniid (Crustacea) species. *Mol. Biol. Evol.* 18, 2201–2212.
- Page, R.D.M., Holmes, E.C., 1998. *Molecular Evolution, a Phylogenetic Approach*. Blackwell Science, Oxford.
- Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. *Mol. Biol. Evol.* 5, 568–583.
- Philippe, H., 1997. Rodent monophyly: pitfalls of molecular phylogenies. *J. Mol. Evol.* 45, 712–715.
- Philippe, H., Forterre, P., 1999. The rooting of the universal tree of life is not reliable. *J. Mol. Evol.* 49, 509–523.
- Philippe, H., Germot, A., 2000. Phylogeny of eukaryotes based on ribosomal RNA: long-branch attraction and models of sequence evolution. *Mol. Biol. Evol.* 17, 830–834.
- Philippe, H., Laurent, J., 1998. How good are deep phylogenetic trees? *Curr. Opin. General Dev.* 8, 616–623.
- Pickett, K.M., Simmons, M.P., Randle, C.P., 2004. Do Bayesian support values reflect probability of the truth? *Cladistics*, 20, 92–93 [Meeting Abstract].
- Piganeau, G., Gardner, M., Eyre-Walker, A., 2004. A broad survey of recombination in animal mitochondria. *Mol. Biol. Evol.* 21, 2319–2325.
- Poe, S., 1998. The effect of taxonomic sampling on accuracy of phylogeny estimation: test case of a known phylogeny. *Mol. Biol. Evol.* 15, 1086–1090.
- Poe, S., Swofford, D.L., 1999. Taxon sampling revisited. *Nature*, 398, 299–300.
- Pol, D., Siddall, M.E., 2001. Biases in maximum likelihood and parsimony: a simulation approach to a 10-taxon case. *Cladistics*, 17, 266–281.
- Pollock, D.D., Zwickl, D.J., McGuire, J.A., Hillis, D.M., 2002. Increased taxon sampling is advantageous for phylogenetic inference. *Syst. Biol.* 51, 664–671.
- Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics*, 14, 817–818.
- Pumo, D.E., Finamore, P.S., Franek, W.R., Phillips, C.J., Tarzami, S., Balzarano, D., 1998. Complete mitochondrial genome of a neotropical fruit bat, *Artibeus jamaicensis*, and a new hypothesis of the relationships of bats to other eutherian mammals. *J. Mol. Evol.* 47, 709–717.
- Qiu, Y.-L., Lee, J., Bernasconi-Quadroni, F., Soltis, D.E., Soltis, P.S., Zanis, M., Zimmer, E.A., Chen, Z., Savolainen, V., Chase, M.W., 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature*, 402, 404–407.
- Qiu, Y.L., Lee, J., Whitlock, B.A., Bernasconi-Quadroni, F., Dombrowska, O., 2001. Was the ANITA rooting of the angiosperm phylogeny affected by long-branch attraction? *Mol. Biol. Evol.* 18, 1745–1753.
- Rannala, B., Huelsenbeck, J.P., Yang, Z.H., Nielsen, R., 1998. Taxon sampling and the accuracy of large phylogenies. *Syst. Biol.* 47, 702–710.
- Reed, D.L., Carpenter, K.E., deGravelle, M.J., 2002. Molecular systematics of the Jacks (Perciformes, Carangidae) based on mitochondrial cytochrome b sequences using parsimony, likelihood, and Bayesian approaches. *Mol. Phyl. Evol.* 23, 513–524.
- Rexova, K., Frynta, D., Zrzavy, J., 2003. Cladistic analysis of languages: Indo-European classification based on lexicostatistical data. *Cladistics*, 19, 120–127.
- Reyes, A., Gissi, C., Catzeflis, F., Nevo, E., Pesole, G., Saccone, C., 2004. Congruent mammalian trees from mitochondrial and nuclear genes using Bayesian methods. *Mol. Biol. Evol.* 21, 397–403.
- Reyes, A., Gissi, C., Pesole, G., Catzeflis, F.M., Saccone, C., 2000a. Where do rodents fit? Evidence from the complete mitochondrial genome of *Sciurus vulgaris*. *Mol. Biol. Evol.* 17, 979–983.
- Reyes, A., Pesole, G., Saccone, C., 1998. Complete mitochondrial DNA sequence of the fat dormouse, *Glis glis*: further evidence of rodent paraphyly. *Mol. Biol. Evol.* 15, 499–505.
- Reyes, A., Pesole, G., Saccone, C., 2000b. Long-branch attraction phenomenon and the impact of among-site rate variation on rodent phylogeny. *Gene*, 259, 177–187.
- Rogers, J., S., 1993. Response of Colless tree imbalance to number of terminal taxa. *Syst. Biol.* 42, 102–105.
- Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, 425, 798–804.
- Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19, 1572–1574.
- Ronquist, F., Liljebäck, J., 2001. Evolution of the gall wasp–host plant association. *Evolution*, 55, 2503–2522.
- Sanderson, M.J., Wojciechowski, M.F., Hu, J.M., Khan, T.S., Brady, S.G., 2000. Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. *Mol. Biol. Evol.* 17, 782–797.
- Schuh, R.T., 2000. *Biological Systematics Principles and Applications*. Cornell University Press, Ithaca, NY.
- Schulmeister, S., 2004. Inconsistency of parsimony revisited. *Syst. Biol.* 53, 521–528.
- Schwarz, M.P., Tierney, S.M., Cooper, S.J.B., Bull, N.J., 2004. Molecular phylogenetics of the allodapine bee genus *Braunsapis*: A-T bias and heterogeneous substitution parameters. *Mol. Phyl. Evol.* 32, 110–122.
- Scotland, R.W., Olmstead, R.G., Bennett, J.R., 2003. Phylogeny reconstruction: the role of morphology. *Syst. Biol.* 52, 539–548.
- Siddall, M.E., 1998. Success of parsimony in the four-taxon case: long-branch repulsion by likelihood in the Farris Zone. *Cladistics*, 14, 209–220.
- 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.
- Simmons, M.P., Pickett, K.M., Miya, M., 2004. How meaningful are Bayesian support values? *Mol. Biol. Evol.* 21, 188–199.
- Simmons, M.P., Randle, C.P., Freudenstein, J.V., Wenzel, J.W., 2002. Limitations of relative apparent synapomorphy analysis (RASA) for measuring phylogenetic signal. *Mol. Biol. Evol.* 19, 14–23.
- Soltis, D.E., Soltis, P.S., 2004. *Amborella* not a ‘basal angiosperm’? Not so fast (vol. 91, pg. 997, 2004). *Am. J. Bot.* 91, 1199–1199.
- Springer, M.S., Amrine, H.M., Burk, A., Stanhope, M.J., 1999. Additional support for Afrotheria and Paenungulata, the performance of mitochondrial versus nuclear genes, and the impact of data partitions with heterogeneous base composition. *Syst. Biol.* 48, 65–75.
- Springer, M.S., Cleven, G.C., Madsen, O., de Jong, W.W., Waddell, V.G., Amrine, H.M., Stanhope, M.J., 1997. Endemic African mammals shake the phylogenetic tree. *Nature*, 388, 61–64.
- Springer, M.S., DeBry, R.W., Douady, C., Amrine, H.M., Madsen, O., de Jong, W.W., Stanhope, M.J., 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Mol. Biol. Evol.* 18, 132–143.
- Springer, M.S., Murphy, W.J., Eizirik, E., O’Brien, S.J., 2003. Placental mammal diversification and the Cretaceous–Tertiary boundary. *Proc. Natl Acad. Sci. USA*, 100, 1056–1061.

- Springer, M.S., Stanhope, M.J., Madsen, O., de Jong, W.W., 2004. Molecules consolidate the placental mammal tree. *TREE*, 19, 430–438.
- Stanhope, M.J., Madsen, O., Waddell, V.G., Cleven, G.C., de Jong, W.W., Springer, M.S., 1998a. Highly congruent molecular support for a diverse superordinal clade of endemic African mammals. *Mol. Phyl. Evol.* 9, 501–508.
- Stanhope, M.J., Waddell, V.G., Madsen, O., de Jong, W.W., Hedges, S.B., Cleven, G.C., Kao, D., Springer, M.S., 1998b. Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *Proc. Natl Acad. Sci. USA*, 95, 9967–9972.
- Steel, M., Huson, D., Lockhart, P.J., 2000. Invariable sites models and their use in phylogeny reconstruction. *Syst. Biol.* 49, 225–232.
- Stiller, J.W., Hall, B.D., 1999. Long-branch attraction and the rDNA model of early eukaryotic evolution. *Mol. Biol. Evol.* 16, 1270–1279.
- Stiller, J.W., Riley, J., Hall, B.D., 2001. Are red algae plants? A critical evaluation of three key molecular data sets. *J. Mol. Evol.* 52, 527–539.
- Sullivan, J., Swofford, D.L., 1997. Are Guinea pigs rodents? The importance of adequate models in molecular phylogenetics. *J. Mamm. Evol.* 4, 77–86.
- Suzuki, Y., Glazko, G.V., Nei, M., 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proc. Natl Acad. Sci. USA*, 99, 16138–16143.
- 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.), *Phylogenetic Inference*. Sinauer Associates, Sunderland, MA, USA, pp. 407–514.
- Swofford, D.L., Waddell, P.J., Huelsenbeck, J.P., Foster, P.G., Lewis, P.O., Rogers, J.S., 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Syst. Biol.* 50, 525–539.
- Tang, K.L., Berendzen, P.B., Wiley, E.O., Morrissey, J.F., Winterbottom, R. & Johnson, G.D., 1999. The phylogenetic relationships of the suborder Acanthuroidei (Teleostei: Perciformes) based on molecular and morphological evidence. *Mol. Phyl. Evol.* 11, 415–425.
- Teeling, E.C., Scally, M., Kao, D.J., Romagnoli, M.L., Springer, M.S., Stanhope, M.J., 2000. Molecular evidence regarding the origin of echolocation and flight in bats. *Nature*, 403, 188–192.
- Tourasse, N.J., Gouy, M., 1999. Accounting for evolutionary rate variation among sequence sites consistently changes universal phylogenies deduced from rRNA and protein-coding genes. *Mol. Phyl. Evol.* 13, 159–168.
- van Tuinen, M., Sibley, C.G., Hedges, S.B., 2000. The early history of modern birds inferred from DNA sequences of nuclear and mitochondrial ribosomal genes. *Mol. Biol. Evol.* 17, 451–457.
- VanDePeer, Y., Rensing, S.A., Maier, U.G., DeWachter, R., 1996. Substitution rate calibration of small subunit ribosomal RNA identifies chlorarachniophyte endosymbionts as remnants of green algae. *Proc. Natl Acad. Sci. USA*, 93, 7732–7736.
- Waddell, P.J., Cao, Y., Hauf, J., Hasegawa, M., 1999. Using novel phylogenetic methods to evaluate mammalian mtDNA, including amino acid invariant sites LogDet plus site stripping, to detect internal conflicts in the data, with special reference to the positions of hedgehog, armadillo, and elephant. *Syst. Biol.* 48, 31–53.
- Waddell, P.J., Kishino, H., Ota, R., 2001. A phylogenetic foundation for comparative mammalian genomics. *Genome Informatics*, 12, 141–154.
- Wahlberg, N., Nylin, S., 2003. Morphology versus molecules: resolution of the positions of *Nymphalis*, *Polygonia*, and related genera (Lepidoptera: Nymphalidae). *Cladistics*, 19, 213–223.
- Wheeler, W.C., 1990. Nucleic-acid sequence phylogeny and random outgroups. *Cladistics*, 6, 363–367.
- Wheeler, W.C., Whiting, M., Wheeler, Q.D., Carpenter, J.M., 2001. The phylogeny of the extant hexapod orders. *Cladistics*, 17, 113–169.
- Whiting, M.F., 1998. Long-branch distraction and the Strepsiptera. *Syst. Biol.* 47, 134–138.
- Whiting, M.F., Carpenter, J.C., Wheeler, Q.D., Wheeler, W.C., 1997. The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Syst. Biol.* 46, 1–68.
- Whittingham, L.A., Slikas, B., Winkler, D.W., Sheldon, F.H., 2002. Phylogeny of the tree swallow genus, *Tachycineta* (Aves, Hirundinidae), by Bayesian analysis of mitochondrial DNA sequences. *Mol. Phyl. Evol.* 22, 430–441.
- Wiens, J.J., 1998. The accuracy of methods for coding and sampling higher-level taxa for phylogenetic analysis: a simulation study. *Syst. Biol.* 47, 397–413.
- Wiens, J.J., 2004. The role of morphological data in phylogeny reconstruction. *Syst. Biol.* 53, 653–661.
- Wiens, J.J., Hollingsworth, B.D., 2000. War of the iguanas: conflicting molecular and morphological phylogenies and long-branch attraction in iguanid lizards. *Syst. Biol.* 49, 143–159.
- Wilcox, T.P., García de León, F.J., Hendrickson, D.A., Hillis, D.M., 2004. Convergence among cave catfishes: long-branch attraction and a Bayesian relative rates test. *Mol. Phyl. Evol.* 31, 1101–1113.
- Wilcox, T.P., Zwickl, D.J., Heath, T.A., Hillis, D.M., 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Mol. Phyl. Evol.* 25, 361–371.
- Willson, S.J., 1999. A higher order parsimony method to reduce long-branch attraction. *Mol. Biol. Evol.* 16, 694–705.
- Xiang, Q.Y., Moody, M.L., Soltis, D.E., Fan, C.Z., Soltis, P.S., 2002. Relationships within Cornales and circumscription of Cornaceae – matK and rbcL sequence data and effects of outgroups and long branches. *Mol. Phyl. Evol.* 24, 35–57.
- Yang, Z.H., 1997. How often do wrong models produce better phylogenies? *Mol. Biol. Evol.* 14, 105–108.
- Yeates, D.K., 1995. Groundplans and exemplars: paths to the tree of life. *Cladistics*, 11, 343–357.
- Zwickl, D.J., Hillis, D.M., 2002. Increased taxon sampling greatly reduces phylogenetic error. *Syst. Biol.* 51, 588–598.

Appendix 1

Table A1

Model parameter estimations from the Bayesian analysis (separate GTR + Γ + I for five partitions) of the placental mammal data of Murphy et al. (2001b). Partitions referred to by their numbers after the parameter name are as follows: 1 = mtRNA, 2 = untranslated nuclear regions, 3 = first codon positions of nuclear protein coding genes, 4 = second ditto, 5 = third ditto. TL = total tree length, $r(G \leftrightarrow T)$ = relative rate of G-to-T substitution, $\pi(A)$ = stationary base frequency of adenine, α = shape parameter of the gamma distribution of rate heterogeneity across sites, pinvar = proportion of invariable sites, m = relative rate multiplier for rate heterogeneity across partitions. Untranslated regions: APP, BMI1, CREM, PLCB4. Nuclear translated genes: ADORA3, ADRB2, ATP7A, BDNF, CNR1, EDG1, PNOG, RAG1, RAG2, TYR, ZFX, VWF, BRCA1, IRBP, A2AB. Mitochondrial DNA: 12S-tRNAval-16S. The 95% credibility interval is the Bayesian analogue to the confidence interval and defines the range within which the true value lies with a probability at 0.95

Parameter	Mean	Variance	95% Cred. interval	
			Lower	Upper
TL{all}	3.837485	0.001896	3.753000	3.925000
$r(G \leftrightarrow T)\{1\}$	1.000000	0.000000	1.000000	1.000000
$r(C \leftrightarrow T)\{1\}$	33.63619	44.641049	23.17955	49.871127
$r(C \leftrightarrow G)\{1\}$	0.693104	0.075659	0.276198	1.353993
$r(A \leftrightarrow T)\{1\}$	3.115709	0.453023	2.046610	4.699832
$r(A \leftrightarrow G)\{1\}$	14.06500	7.615208	9.700388	20.712920
$r(A \leftrightarrow C)\{1\}$	6.684784	2.051386	4.442389	10.006942
$r(G \leftrightarrow T)\{2\}$	1.000000	0.000000	1.000000	1.000000
$r(C \leftrightarrow T)\{2\}$	3.010491	0.076104	2.533615	3.625134
$r(C \leftrightarrow G)\{2\}$	0.912374	0.012014	0.714621	1.137358
$r(A \leftrightarrow T)\{2\}$	0.458235	0.002530	0.367969	0.562422
$r(A \leftrightarrow G)\{2\}$	3.538014	0.073964	3.025878	4.111975
$r(A \leftrightarrow C)\{2\}$	1.547945	0.027308	1.247209	1.890612
$r(G \leftrightarrow T)\{3\}$	1.000000	0.000000	1.000000	1.000000
$r(C \leftrightarrow T)\{3\}$	4.395056	0.077119	3.862167	4.979766
$r(C \leftrightarrow G)\{3\}$	1.096939	0.006875	0.940801	1.258942
$r(A \leftrightarrow T)\{3\}$	0.811495	0.004528	0.689257	0.949550
$r(A \leftrightarrow G)\{3\}$	3.971286	0.065228	3.515585	4.496017
$r(A \leftrightarrow C)\{3\}$	1.547882	0.013132	1.342237	1.776181
$r(G \leftrightarrow T)\{4\}$	1.000000	0.000000	1.000000	1.000000
$r(C \leftrightarrow T)\{4\}$	4.139286	0.086333	3.588733	4.731208
$r(C \leftrightarrow G)\{4\}$	2.023846	0.027377	1.717634	2.379036
$r(A \leftrightarrow T)\{4\}$	0.806164	0.004956	0.676358	0.951087
$r(A \leftrightarrow G)\{4\}$	6.610073	0.194760	5.757365	7.483312
$r(A \leftrightarrow C)\{4\}$	1.417865	0.014300	1.193859	1.662327
$r(G \leftrightarrow T)\{5\}$	1.000000	0.000000	1.000000	1.000000
$r(C \leftrightarrow T)\{5\}$	5.617898	0.056248	5.171903	6.088510
$r(C \leftrightarrow G)\{5\}$	0.926058	0.002403	0.831769	1.022499
$r(A \leftrightarrow T)\{5\}$	1.339881	0.006119	1.190691	1.491481
$r(A \leftrightarrow G)\{5\}$	5.934169	0.068960	5.450564	6.456804
$r(A \leftrightarrow C)\{5\}$	1.217481	0.005132	1.083843	1.361540
$\pi(A)\{1\}$	0.358751	0.000110	0.337873	0.379558
$\pi(C)\{1\}$	0.182235	0.000061	0.166616	0.197561
$\pi(G)\{1\}$	0.201584	0.000079	0.184604	0.218941
$\pi(T)\{1\}$	0.257430	0.000073	0.240894	0.273945
$\pi(A)\{2\}$	0.261830	0.000072	0.245897	0.278874
$\pi(C)\{2\}$	0.222490	0.000073	0.205835	0.239326
$\pi(G)\{2\}$	0.204249	0.000062	0.189067	0.220305
$\pi(T)\{2\}$	0.311431	0.000084	0.293225	0.329844
$\pi(A)\{3\}$	0.295051	0.000028	0.284478	0.305947
$\pi(C)\{3\}$	0.236021	0.000025	0.226160	0.245961
$\pi(G)\{3\}$	0.291376	0.000030	0.280722	0.301689
$\pi(T)\{3\}$	0.177552	0.000023	0.168105	0.187214

Table A1
Continued

Parameter	Mean	Variance	95% Cred. interval	
			Lower	Upper
$\pi(A)\{4\}$	0.296139	0.000034	0.284951	0.307244
$\pi(C)\{4\}$	0.239711	0.000029	0.229651	0.250030
$\pi(G)\{4\}$	0.185245	0.000023	0.175798	0.194632
$\pi(T)\{4\}$	0.278905	0.000034	0.267981	0.290155
$\pi(A)\{5\}$	0.174955	0.000014	0.167087	0.182593
$\pi(C)\{5\}$	0.328757	0.000019	0.320611	0.337677
$\pi(G)\{5\}$	0.274017	0.000020	0.265084	0.282586
$\pi(T)\{5\}$	0.222271	0.000016	0.214558	0.229531
$\alpha\{1\}$	0.508946	0.001647	0.437220	0.596225
$\alpha\{2\}$	0.912780	0.011072	0.740079	1.144125
$\alpha\{3\}$	1.162474	0.009012	0.982992	1.354583
$\alpha\{4\}$	0.958277	0.007148	0.803264	1.131110
$\alpha\{5\}$	1.600986	0.002901	1.501675	1.710638
$\text{pinvar}\{1\}$	0.420817	0.000503	0.375881	0.463973
$\text{pinvar}\{2\}$	0.069645	0.001172	0.007355	0.137037
$\text{pinvar}\{3\}$	0.316994	0.000300	0.280143	0.348986
$\text{pinvar}\{4\}$	0.376332	0.000389	0.335564	0.412984
$\text{pinvar}\{5\}$	0.005052	0.000013	0.000248	0.013376
$m\{1\}$	1.285717	0.004281	1.165483	1.415491
$m\{2\}$	0.681692	0.000420	0.640260	0.720464
$m\{3\}$	0.634405	0.000173	0.609615	0.660903
$m\{4\}$	0.520472	0.000147	0.496854	0.544825
$m\{5\}$	1.865779	0.000435	1.824598	1.907524

Table A2

Model parameter estimations from full matrix Bayesian analysis 1 (GTR + Γ + I for four partitions corresponding to genes and Mk + Γ for morphology) of the combined data of Nylander et al. (2004). Partitions referred to by their numbers after the parameter name are as follows: 1 = CO1, 2 = 28S, 3 = EF1, 4 = LWRh and 5 = morphology. Parameter abbreviations are the same as in Table A1

Parameter	Mean	Variance	95% Cred. interval	
			Lower	Upper
TL{all}	5.932432	0.111090	5.320000	6.600000
r(G↔T){1}	1.000000	0.000000	1.000000	1.000000
r(C↔T){1}	30.85950	128.65445	14.968297	58.574882
r(C↔G){1}	26.87936	116.26891	12.07196	53.336706
r(A↔T){1}	1.258682	0.188651	0.630405	2.279847
r(A↔G){1}	13.275670	17.20058	7.291687	22.925466
r(A↔C){1}	2.158422	0.843342	0.838977	4.372804
r(G↔T){2}	1.000000	0.000000	1.000000	1.000000
r(C↔T){2}	7.327556	1.852243	5.152499	10.468116
r(C↔G){2}	0.139967	0.005813	0.025018	0.311531
r(A↔T){2}	3.439390	0.592590	2.250840	5.208430
r(A↔G){2}	4.718721	0.909727	3.198567	6.958109
r(A↔C){2}	0.805343	0.055987	0.424446	1.344202
r(G↔T){3}	1.000000	0.000000	1.000000	1.000000
r(C↔T){3}	9.118440	12.419975	4.527899	17.659280
r(C↔G){3}	1.405417	0.593459	0.414780	3.307705
r(A↔T){3}	1.816427	0.566481	0.840680	3.717347
r(A↔G){3}	7.602743	7.728747	3.888020	14.428883
r(A↔C){3}	1.509906	0.481911	0.612542	3.295309
r(G↔T){4}	1.000000	0.000000	1.000000	1.000000
r(C↔T){4}	26.55366	214.754411	10.04240	67.352286
r(C↔G){4}	5.720286	12.678707	1.870012	15.639675
r(A↔T){4}	5.262656	9.642797	1.787063	13.848850
r(A↔G){4}	19.66847	122.80073	7.218202	49.695928
r(A↔C){4}	7.020082	16.441214	2.429796	18.174649
pi(A){1}	0.416537	0.000105	0.397228	0.436575
pi(C){1}	0.048102	0.000006	0.043294	0.053007
pi(G){1}	0.067279	0.000011	0.060817	0.073827
pi(T){1}	0.468082	0.000110	0.447298	0.488220
pi(A){2}	0.257164	0.000144	0.233751	0.280805
pi(C){2}	0.212588	0.000118	0.191874	0.234185
pi(G){2}	0.268663	0.000152	0.244760	0.293645
pi(T){2}	0.261585	0.000138	0.238581	0.284918
pi(A){3}	0.338148	0.000476	0.296040	0.381616
pi(C){3}	0.173255	0.000284	0.141650	0.207833
pi(G){3}	0.189780	0.000325	0.155298	0.226365
pi(T){3}	0.298818	0.000428	0.259517	0.340125
pi(A){4}	0.276855	0.000295	0.243543	0.311149
pi(C){4}	0.196282	0.000196	0.170440	0.225165
pi(G){4}	0.201157	0.000260	0.170776	0.234113
pi(T){4}	0.325706	0.000342	0.290491	0.362628
alpha{1}	0.354073	0.000273	0.323517	0.388571
alpha{2}	0.664228	0.030205	0.389813	1.072946
alpha{3}	3.528246	21.318493	0.287858	14.561565
alpha{4}	3.534917	13.312140	0.908467	11.963294
alpha{5}	1.262591	0.049745	0.833516	1.722314
pinvar{1}	0.333344	0.000419	0.293192	0.373511
pinvar{2}	0.537947	0.002677	0.423904	0.624560
pinvar{3}	0.570727	0.011153	0.172370	0.669874
pinvar{4}	0.488610	0.003968	0.352743	0.567252
m{1}	2.348544	0.001982	2.256495	2.430360
m{2}	0.192991	0.000292	0.162663	0.228490
m{3}	0.180667	0.000450	0.143967	0.224957
m{4}	0.272725	0.000654	0.227138	0.326558
m{5}	1.771523	0.034390	1.437307	2.172104

Table A3

Model parameter estimations from full matrix Bayesian analysis 2 (mixed model for molecular data and Mk + Γ for morphology) of the combined data of Nylander et al. (2004). Partitions referred to by their numbers after the parameter name are as follows: 1 = CO1 first and second codon positions, 2 = CO1 third codon positions, 3 = EF1 & LWRh first and second codon positions, 4 = EF1 & LWRh third codon positions, 5 = 28S, and 6 = morphology. Parameter abbreviations are the same as in Table A1 except kappa = the transition/transversion ratio

Parameter	Mean	Variance	95% Cred. interval	
			Lower	Upper
TL{all}	4.333285	0.053189	3.921000	4.827000
kappa{1,3}	4.181606	0.101737	3.606215	4.847860
kappa{2,4}	7.892526	0.318095	6.866367	9.068239
r(G↔T){5}	1.000000	0.000000	1.000000	1.000000
r(C↔T){5}	6.549434	1.277573	4.641110	9.052910
r(C↔G){5}	0.175557	0.006597	0.049474	0.363166
r(A↔T){5}	2.713458	0.299644	1.814161	3.945611
r(A↔G){5}	4.445831	0.621914	3.113099	6.193281
r(A↔C){5}	0.721531	0.038271	0.402343	1.162708
pi(A){1,2}	0.395691	0.000090	0.377435	0.414679
pi(C){1,2}	0.066631	0.000009	0.060901	0.072562
pi(G){1,2}	0.074015	0.000010	0.067940	0.080172
pi(T){1,2}	0.463663	0.000094	0.444152	0.481752
pi(A){3,4,5}	0.284297	0.000073	0.267865	0.300953
pi(C){3,4,5}	0.205820	0.000055	0.191067	0.220549
pi(G){3,4,5}	0.222374	0.000058	0.207727	0.237363
pi(T){3,4,5}	0.287509	0.000072	0.271630	0.304612
alpha{1–5}	0.731855	0.001900	0.648365	0.819855
alpha{6}	1.260026	0.048148	0.846888	1.712983
pinvar{1}	0.543395	0.000660	0.492203	0.592679
pinvar{2}	0.034449	0.000314	0.003667	0.071069
pinvar{3}	0.741027	0.001007	0.673368	0.798817
pinvar{4}	0.006603	0.000042	0.000165	0.024046
pinvar{5}	0.560604	0.000575	0.512118	0.606495
m{1}	0.402353	0.001284	0.336036	0.479842
m{2}	5.500854	0.036883	5.135265	5.898329
m{3}	0.094868	0.000193	0.070153	0.124876
m{4}	0.758108	0.005196	0.626929	0.913555
m{5}	0.258377	0.000416	0.219560	0.298764
m{6}	2.468843	0.062349	2.023286	2.999949