Partitioned Likelihood Support and the Evaluation of Data Set Conflict

M. S. Y. Lee1,2 and A. F. Hugall1,3

1Department of Environmental Biology, University of Adelaide, Adelaide, South Australia 5005, Australia; E-mail: lee.mike@sa.gov.au (M.S.Y.L.), ahugall@zen.uq.edu.au (A.F.H.)
2Department of Palaeontology, South Australian Museum, North Terrace, Adelaide, South Australia 5000, Australia
3Department of Zoology and Entomology, University of Queensland, Saint Lucia, Brisbane, Queensland 4072, Australia

Abstract.—In simultaneous analyses of multiple data partitions, the trees relevant when measuring support for a clade are the optimal tree, and the best tree lacking the clade (i.e., the most reasonable alternative). The parsimony-based method of partitioned branch support (PBS) forces each data set to arbitrate between the two relevant trees. This value is the amount each data set contributes to clade support in the combined analysis, and can be very different to support apparent in separate analyses. The approach used in PBS can also be employed in likelihood: a simultaneous analysis of all data retrieves the maximum likelihood tree, and the best tree without the clade of interest is also found. Each data set is fitted to the two trees and the log-likelihood difference calculated, giving “partitioned likelihood support” (PLS) for each data set. These calculations can be performed regardless of the complexity of the ML model adopted. The significance of PLS can be evaluated using a variety of resampling methods, such as the Kishino-Hasegawa test, the Shimodaira-Hasegawa test, or likelihood weights, although the appropriateness and assumptions of these tests remains debated. [Artiodactyls; cetaceans; Kishino-Hasegawa test; partitioned branch support; partitioned likelihood support; Shimodaira-Hasegawa test; Templeton test.]

Most current tests for data incongruence measure overall levels of disagreement between characters in different data sets and do not identify the particular clades that are in conflict (e.g., Wiens, 1998; Gatesy et al., 1999). Thus, when a “significant” level of incongruence between data sets is identified, it cannot be readily ascertained whether this conflict is localized to particular areas of the tree or spread throughout the entire tree. This is the case for the nonparametric test of Templeton (1983; Larson, 1994) and for the incongruence length difference test (Farris et al., 1994), also called the partition homogeneity test (Swofford, 2000). However, partitioned branch (or partitioned Bremer) support (Baker and DeSalle, 1997) allows determination of the level of support from each data set for each clade in a combined analysis and can thus identify the particular areas of a phylogeny that are the source of conflict between data sets. Partitioned branch support is being increasingly employed in simultaneous analyses of multiple data sets (DeSalle and Brower, 1997; Baker et al., 1998, 2001; O'Grady et al., 1998; Renssen and DeSalle, 1998; Gatesy et al., 1999; Gatesy and Arctander, 2000; Lambkin, 2000; Jamieson et al., 2002; Lambkin et al., 2002).

The measure proposed by Bremer (1988) and termed branch support (Bremer, 1994) is the number of steps it takes to collapse a clade, i.e., the length difference between the most-parsimonious tree (which has the clade in question) and the best tree lacking the clade (the constrained tree). Partitioned branch support (PBS) extends this measure to each data partition. To calculate PBS for a particular clade, each data partition is forced to arbitrate between the most-parsimonious tree and the best tree lacking the clade for the combined data. The data partitions are subjected to a simultaneous analysis; as with branch support, for each clade in the resultant most-parsimonious tree(s), the shortest tree(s) without the clade (the constrained tree) is identified. Each data partition is optimized onto these two trees; the PBS for a data partition is the extent (measured in number of steps) to which the data partition supports the most-parsimonious tree over the constrained tree. Positive PBS values indicate the data partition supports the clade in the context of a simultaneous analysis, and negative PBS values indicate the data partition contradicts the clade. PBS can identify particular regions of a tree that are the source of conflict between data partitions. Clades in such a region will have large negative values for at least some partitions. The sum of the PBS values for all the data partitions equals the overall branch support.

Calculating PBS for a clade for each data partition in the context of a combined analysis is very different from and potentially more illuminating than calculating branch support for each clade in a separate analysis of each data partition (Baker and DeSalle, 1997; Gatesy et al., 1999; Gatesy and Arctander, 2000). For instance, a particular data partition analyzed in isolation might not retrieve a clade of interest found on the optimal combined tree and thus may appear to not support that clade. However, the data partition can bolster support for the clade in the context of a combined analysis with other data partitions. When forced to arbitrate between the most-parsimonious tree for the combined data (which has the clade) and the constrained tree for the combined data (best tree lacking the clade), the data partition might still favor the first tree over the second, resulting in an extreme case of hidden branch support (Gatesy et al., 1999). The PBS for that data partition will be positive, correctly indicating that in the context of a combined analysis this partition increases support for the clade in question. The observation that the optimal tree for the data partition is different from either tree is not particularly relevant in the context of a combined analysis, because it lies outside the two possible topologies evaluated when branch support is calculated for the combined analysis. Conversely, a data partition analyzed by itself might produce a tree (the separate analysis tree) with a particular clade also
found on the most-parsimonious combined tree, but rela-
tionships among the remaining taxa might be very differ-
ent. When the data partition is forced again to arbitrate
between the most-parsimonious combined tree (which
contains the clade of interest but conflicts extensively
with the rest of the separate analysis tree) and the best
constrained tree (which lacks the clade of interest but
is more congruent with the rest of the separate analysis
tree), the data partition might this time favor the second
tree over the first, resulting in an extreme case of hidden
conflict (Gatesy et al., 1999). The data partition that su-
perficially appears to support the clade of interest in fact
reduces support for that clade in the context of a com-
 pared analysis, as correctly reflected in the negative PBS
value. In most cases, however, hidden support (or hid-
den conflict) might not be so extreme. A clade might be
weakly supported (or weakly contradicted) in a separate
analysis of a data partition but more strongly supported
(or more strongly contradicted) by that data partition in
a combined analysis when that data partition is forced to
arbitrate between the optimal and best constrained tree
for the full data set.

Here, we discuss a simple way to evaluate the signif-
ificance of PBS values using the nonparametric sign-rank
test of Templeton (1983). We also extend the method to a
maximum likelihood (ML) framework. Partitioned like-
elihood support (PLS) identifies the likelihood support
(log-likelihood difference) that each data partition pro-
vides for each resolved clade in a simultaneous analysis,
and the significance of this support can be evaluated us-
ing the various resampling tests (Kishino and Hasegawa,
1989; Shimodaira and Hasagawa, 1999; Goldman et al.,
2000). We also provide a worked example for the new
methods. As with PBS, PLS allows identification (this
time in a likelihood framework) of hidden support and
hidden conflict between data partitions and identifica-
tion of particular regions of the tree that are the source
of the conflict. The example data set is taken from the
study by O’Leary (1999; table 3, alignment A) and con-
sists of 10 extant species of ungulate mammals coded for
three nuclear genes ($\beta$-casein, $\kappa$-casein, $\gamma$-fibrinogen),
a mitochondrial gene (cytochrome $b$), SINEs, general (non-
dental) morphology, and dentition. The morphological
data matrix (O’Leary and Geisler, 1999) has been divided
into dental and nondental partitions based on the ar-
guments of Naylor and Adams (2001) rather than into
the osteological and soft anatomical partitions adopted
by O’Leary (1999). The numerous fossil taxa included in
the original morphological study are not considered in
the example calculations. The file containing the matrix
and nexus commands used in this analysis is available
in TreeBASE and at the Systematic Biology website.

EVALUATING THE SIGNIFICANCE OF PARTITIONED
BRANCH SUPPORT

As with branch support (BS), there remains the prob-
lem of evaluating the significance of PBS: does a particu-
lar PBS value represent significant support for or conflict
with a particular clade? As with BS, there is no standard
significant level for PBS. The significance of a particu-
lar value of BS (or PBS) depends on the behavior of the
character evidence across the competing trees (e.g., num-
ber of informative characters, number supporting and
contradicting each tree, amount of homoplasy in each
character). As discussed elsewhere (Macey et al., 1999;
Whitlock and Baum, 1999; Lee, 2000), a simple way to
evaluate the significance of a particular BS value for a
clade is to employ the test of Templeton (1983). Instead
of measuring the length difference by counting the char-
acter steps for and against (BS), these characters are an-
alyzed using a nonparametric sign-rank test to evaluate
the null hypothesis that characters differentially favor-
ing a tree are equally likely to favor either topology.
AutoCladeS (Erickson, 2001) automatically calculates the
Templeton test significance for each clade on a tree. The
same method for evaluating the significance of BS can be
extended to PBS. To evaluate whether a particular
PBS value is significant, that data partition is optimized
onto the most-parsimonious and constrained trees found
in a combined analysis, and the characters supporting
one topology over the other are identified. However, in-
stead of counting character steps for and against (PBS),
Templeton’s test is used to evaluate the null hypothesis
that characters that differ in the number of steps on these
trees are equally likely to favor either topology.

If this procedure is repeated multiple times for each
data partition (e.g., for the constrained tree found for
each clade), a sequential Bonferroni correction to the
significance level can be used (Holm, 1979; Rice, 1989);
results deemed significant using uncorrected and cor-
corrected tests are shown in Figure 1. Other factors make
the Templeton test overly conservative and thus bias the
results the other way (Lee, 2000); thus, not using the
Bonferroni correction might crudely allow these oppos-
ing biases to cancel one another out. However, use of the
correction has the advantage of ensuring that the bias
will be in a known direction. Also, it has been argued
that Templeton’s test is appropriate only for resolved
trees specified a priori and not for topologies identified
a posteriori (Goldman et al., 2000). A modified test more
apposite for trees chosen a posteriori might therefore
be used when available (Goldman et al., 2000). How-
ever, in the current procedure, the trees evaluated are not
strictly (i.e., fully) specified either a priori or a posteriori.
If the trees used for calculating PBS for each data parti-
tion were identified entirely on the basis of independent
information (e.g., the other data partitions only), they
might be considered to be fully specified a priori. If they
were identified entirely on the basis of analyses of the
data partition in question, they might be considered to
be fully specified a posteriori. Of course, the situation lies
somewhere between these two extremes, with the opti-
mal and constrained trees being identified on the basis of
the combined data set, which contains partly inde-
pendent information (other data partitions) but also the
data partition of interest. The trees cannot be considered
to be a priori, given that they are partly (if only slightly)
influenced by the data under consideration (Goldman,
pers. comm.), but they cannot be considered to be totally
FIGURE 1. Parsimony analyses of ungulate phylogeny. (a) Most-parsimonious tree based on combined sequence, SINE, and morphological data. Clades A–F are identical to those found in the likelihood analyses (Fig. 2a). The PBS, overall BS, and associated Templeton test P values are shown for each clade. Uncorrected P values <0.05 are underlined; values that remain significant at P = 0.05 after a sequential Bonferroni correction (Rice, 1989) are marked with an asterisk. (b) PBS for each clade. The majority of negative values, including all large significant ones, are from the morphological data partitions.

a posteriori either, given that they might be almost entirely determined by independent data, especially if the data set under consideration is only one of very many. For this reason, it is difficult to decide whether the a priori Templeton test or the a posteriori modification of this test (Goldman et al., 2000) would be more applicable. The likelihood versions of both tests have recently been suggested to be potentially biased in such intermediate situations (Shimodaira, 2002).

In the example data set, all but two clades in the combined analysis are significantly supported according to Templeton’s test (Fig. 1). However, there is widespread data conflict. Six of the seven clades have at least one negative PBS value, and three have a significant negative PBS value. All the sizeable negative values of PBS (>3), including the three significant ones, are generated by the morphological data sets (Fig. 1). This finding is consistent with the conclusion that the combined analysis is driven by the larger molecular data sets, with a cohesive signal from the molecular partitions that is contradicted by morphology (e.g., Gatesy et al., 1999; O’Leary, 1999).

As suggested by Naylor and Adams (2001), the dental partition is extremely incongruent with the combined (molecule driven) topology; six of the seven PBS values are negative. However, the nondental morphology is also incongruent with the combined topology, with four negative PBS values. The intuitive observation that, all things being equal, larger data sets are likely to generate larger BS and PBS values by chance is also supported; PBS values of up to 12 for the large cyt b data set were identified as not significant, whereas the same or smaller PBS values for the smaller y-fibrinogen and morphological data sets were identified as highly significant. The levels of hidden support in these molecular and morphological ungulate data sets, as identified by PBS analysis, have been discussed elsewhere (Gatesy et al., 1999) and will not be discussed here.

PARTITIONED LIKELIHOOD SUPPORT

The above approach can be extended to ML by partitioning the difference in negative log-likelihood between optimal and constrained trees in a combined analysis (the likelihood support [LS]) into components provided by each data partition (PLS). Also, for evaluating the significance of LS and PLS values, analogues of parsimony-based tests (Templeton, 1983; Goldman et al., 2000) exist in ML (Kishino and Hasegawa, 1989; Huelsenbeck et al., 1996b; Shimodaira and Hasagawa, 1999; Goldman et al., 2000).
First, a particular likelihood model must be adopted, e.g., based on sequentially adding parameters and evaluating whether the improvements in likelihood resulting from more complex models are significant. These evaluations are readily implemented using programs such as Modeltest (Posada and Crandall, 2001), although the caveat must be added that large numbers of variable sites are needed to avoid overspecifying the model (e.g., Goldman, 1993; Huelsenbeck et al., 1996a; Yang, 1997; Takahashi and Nei, 2000). Whether different data partitions require different parameter values or indeed different models also must be decided, perhaps based on whether improvements in likelihood generated by such mixed models are significant (e.g., Wilgenbusch and de Queiroz, 2000; Jamieson et al., 2002). These analyses are currently difficult to implement because no widely available phylogenetic program both allows mixed models and has good tree searching ability. For instance, PAUP* (Swofford, 2000) does not allow mixed models but has excellent tree searches, and PAML (Yang, 1999) allows mixed models but is relatively poor at searching for optimal trees. The example provided here employs a restricted general time reversible (GTR) model and constant parameter values across all partitions: however, the method will work with more complex analyses (e.g., different parameter values or even different models for each data partition).

Once a model is adopted, the ML tree for the combined data set is determined and the likelihood support for each clade is calculated. This process involves finding the ML tree(s) lacking a particular clade (constrained tree) using reverse constraint searches: in PAUP*, loading a constraint tree with a single resolved node (the clade under investigation) and then searching for the best tree incompatible with this constant. The difference in negative log-likelihood between the constrained tree and the ML tree corresponds to LS. As for BS in parsimony analyses, LS can be broken down into contributions from each data partition (by summing the site log-likelihoods for each partition calculated using the combined data). The difference in log-likelihood between the constrained and the ML trees for that data partition is the PLS. A positive PLS indicates that in the context of a combined analysis a data partition supports a clade, and a negative PLS indicates that a data partition contradicts a clade. For each clade, the sum of the PLS for all data sets equals the (total) LS in a combined analysis.

The significance of each PLS value can be evaluated using resampling tests, although the biases and appropriate context of each of the various methods are currently debated and beyond the scope of this paper (Kishino and Hasegawa, 1989; Shimodaira and Hasegawa, 1999; Goldman et al., 2000; Strimmer and Rambaut, 2001; Buckley, 2002; Shimodaira, 2002). We employed three widely used nonparametric tests: the Kishino-Hasegawa (KH), Shimodaria-Hasegawa (SH), and likelihood weight (LW) tests. The KH test is least biased (most appropriate) when the candidate trees have been fully specified a priori based on independent evidence, and there is usually a large likelihood difference between them (e.g., Goldman et al., 2000; Shimodaira, 2002; Fig. 2). The SH test is least biased (most appropriate) when the candidate trees have been selected based entirely on the data of interest, and are usually similar in likelihood (e.g., Goldman et al., 2000; Shimodaira, 2002; Fig. 2). The trees to be compared when computing PLS for a given data set are partly determined by the data at hand and partly determined by other independent data sets, so the situation is likely to be in the problematic intermediate area where all tests, including “appropriately unbiased” and weighted SH tests, are to some extent biased (Shimodaira, 2002; Fig. 2). For this reason, both KH and SH tests and the LW test (Strimmer and Rambaut, 2001) were used for the worked example, these being among the most commonly used possible tests. A sequential Bonferroni correction for multiple comparisons involving the same data set (e.g., across multiple constrained trees) might be used with the KH tests; the other tests compare multiple trees simultaneously.

To separate the LS in the combined analysis into components from each data partition requires that all the parameters used in that analysis, including branch lengths and site rate categories, remain constant during the optimization of each data partition onto the ML and constrained trees. Thus, if the same parameter values are used across all data partitions, as in most analyses, these values must be held constant when calculating PLS values. Often, these values cannot be held constant in widely used phylogenetic programs such as PAUP* (Swofford, 2000). Similarly, the KH and SH tests in PAUP* do not allow complete fixing of parameters for analyzing partitions of the whole combined data. A procedure for manually calculating PLS for each data partition by summing individual site log-likelihoods and for evaluating the significance of this value by performing certain versions of the KH and SH tests is given in the Appendix. If mixed models are used where different data partitions are permitted different models or parameter values, then the models and parameter values (in particular, branch lengths) adopted for each data partition in the combined analysis must also be used when calculating the PLS of each data partition.

Parametric bootstrapping (Huelsenbeck et al., 1996a, 1996b) could also be used to evaluate the significance of the PLS. In this case, the constrained tree (topology and branch lengths) found in the combined analysis should be used as the null hypothesis to generate replicate sets of each data partition evolving under the exact model and parameter values used in the combined analysis. For each replicate of each data partition, the relevant test statistic would be the difference between the likelihoods of the constrained and best trees found in the combined analysis. The actual PLS value could then be compared with the distribution of this test statistic to evaluate the $P$ value. However, parametric bootstrapping has recently been shown to have a much higher type 1 error than nonparametric methods, especially when the models are underspecified (Buckley, 2002; Shimodaira, 2002). The almost universal “significant” results have previously been commented upon as cause for
FIGURE 2. ML analyses of ungulate phylogeny. (a) ML tree based on combined sequence data. Clades A–F are identical to those found in parsimony analyses. The PLS, overall LS, and associated KH, SH, and LW P values are shown for each clade. Uncorrected P values <0.05 are underlined. For the KH test, values that remain significant at P = 0.05 after a sequential Bonferroni correction (Rice, 1989) are marked with an asterisk; the other tests perform multiple comparisons simultaneously. (b) PLS for each clade. There are no large negative values.

advantages of partitioned likelihood support

Many advocates of partitioned support have tended to favor parsimony over likelihood methods of phylogenetic reconstruction, and as a result, partitioned support has been applied exclusively in parsimony analyses. However, the technique and advantages of PBS can be extended into an ML framework. In particular, PLS reveals not only the amount of conflict between data partitions but also its distribution and significance across the various clades in the tree. Also, it allows detection of synergism among data sets at certain clades. For instance, under the KH test, clades B and C were not significantly supported by any data partition, but the combined data significantly supported both clades. Second, different data sets resolve different clades. For instance, under the SH test, clade F was significantly supported by only one data partition (γ-casein), whereas clade D was significantly supported by only γ-fibrinogen. The significance of the combined likelihood support of both these clades is largely contributed by a single data partition. Clade E was similarly almost entirely supported by cytochrome b, but this result is an artifact of missing data because no other gene was sequenced for both members of this clade.

advantages of partitioned likelihood support

Many advocates of partitioned support have tended to favor parsimony over likelihood methods of phylogenetic reconstruction, and as a result, partitioned support has been applied exclusively in parsimony analyses. However, the technique and advantages of PBS can be extended into an ML framework. In particular, PLS reveals not only the amount of conflict between data partitions but also its distribution and significance across the various clades in the tree. Also, it allows detection...
of hidden support and hidden conflict, which are secondary signals in each data partition that are revealed only in the context of combined analyses and the interaction of all data partitions (see Gatesy et al., 1999; Lambkin et al., 2002). Such signals cannot be identified using the common method of comparing the results of separate likelihood analyses of each data partition. For instance, in a separate analysis a particular gene might not retrieve a clade of interest found on the combined tree, but this gene can bolster support for that clade in a combined multigene analysis if it favors the ML tree over the best constrained tree. Although separate analysis of that gene alone would lead one to conclude that it does not support the clade, the positive PLS for that gene will correctly show that it increases support for the clade in a combined analysis. Conversely, a gene analyzed alone might produce a tree with a particular clade also found on the ML tree for the combined data, but relationships among the remaining taxa might be very different. When the gene is forced to arbitrate between the ML combined tree (which contains the clade of interest but conflicts extensively with the rest of the single-gene tree) and the best constrained tree (which lacks the clade of interest but is more congruent elsewhere with the single-gene tree) the gene might favor the second over the first, resulting in an extreme case of hidden conflict. Although a separate analysis of the gene might lead one to conclude that it supports the clade, the negative PLS value will show correctly that it reduces support for the clade in the context of a combined analysis.

In most cases, however, hidden LS (or conflict) will not be so extreme: a clade might be weakly supported (or weakly contradicted) in a separate analysis of a gene but more strongly supported (or more strongly contradicted) by that data partition in the context of a combined multigene analysis, when that gene is forced to arbitrate between the ML and best constrained tree for all data sets. A good example in this analysis is the support from cyt $b$ for the hippo clade. In a separate analysis of this gene alone, the LS for this clade is 41.7, but in a combined analysis the gene contributes nearly twice as much support to this clade (PLS $= 80.7$). Some of this increase is no doubt due to parameters being optimized to cyt $b$ alone in the former analysis but to all the data in the latter analysis. However, some of the increase is due to hidden support. In the separate analyses, the ML tree for cyt $b$ is identical to the ML tree for the combined data, and the best constrained tree lacking the hippo clade has a highly heterodox topology in which hippos are paraphyletic with respect to the camel (Fig. 3a). However, that constrained tree is soundly rejected by the other molecular data sets (and morphological data); in the combined analysis, the best tree that lacks the hippo clade is a more plausible one where the two hippos form a polytomy with cetaceans (Fig. 3b). When cyt $b$ is included in a combined analysis and forced to arbitrate between the ML tree and the most reasonable (based on all the data) constrained tree that lacks the hippo clade, the cyt $b$ data are more decisive. Unfortunately, PAUP* does not yet permit “freezing” all parameters of the combined analysis and using them in the partitioned analysis, which is the approach needed to calculate the exact magnitude of this hidden support.

Although much of the above discussion has been couched in terms of data divided according to genes, PLS can be applied to any type of molecular data partition, e.g., by codon position, by intron-exon, by rate category, or by secondary structure. PLS can also be applied using any commonly used model of molecular evolution. PLS shows both the overall level of data set conflict and its distribution across the tree. In this way, one can identify particular nodes that are in conflict. The method involving separate analyses of each partition and using reciprocal KH or SH tests to see whether the optimal tree for one data set is rejected by the other data set (e.g., Wilgenbusch and de Queiroz, 2000) does not explicitly reveal the distribution of incongruence throughout the trees. One has to guess which areas are problematic by examination of conflicting clades among the trees. Most importantly, PLS divides the total LS for each clade in the combined analysis into the precise contributions (for and against) from each data partition, and these values can be evaluated for statistical significance. These contributions in combined analyses, which result from the complex interaction of primary and hidden signals among data sets, cannot be
identified with the more commonly used methods that involve separate analyses of each data partition.

ACKNOWLEDGMENTS

We thank Maureen O’Leary for providing a copy of the ungulate data set, Nick Goldman, Sarah Smith, Remko Leis, John Scanlon, and an anonymous referee for comments and suggestions, and the University of Queensland and Australian Research Council for funding.

REFERENCES


First submitted 3 September 2001; reviews returned 24 January 2002; final acceptance 30 September 2002

Associate Editor: Nick Goldman

APPENDIX:

CALCULATING AND EVALUATING THE SIGNIFICANCE OF PLS

Most likelihood analyses currently employ a single model and parameter set for all data partitions in a combined analysis; this same model and same parameter set must be used when optimizing each
data partition onto the optimal and constrained trees to obtain PLS values. Current versions of PAUP* cause problems in such situations when attempting to calculate PLS values for individual data partitions because they do not readily allow one to fix the trees and parameters to the values used in the combined analysis. For instance, when several rate categories are used and each data partition is optimized onto the optimal and constrained trees to obtain the likelihood difference (PLS) and its significance (KH and SH tests), it is impossible to fix the exact values of these rate categories to those employed in the combined analysis. One can only make the gamma parameter and number of rate categories the same, but during PLS calculations the precise values of these categories will be automatically (and inappropriately) optimized for each data partition. Even if a gamma parameter of 0.5 with four rate categories is used, the precise values assigned to these four rate categories when the cyt b data are optimized onto a tree will be faster than the corresponding values assigned to the four rate categories when the β-casein is optimized onto that same tree. These extra degrees of freedom mean that the sum of the log-likelihoods for the data partitions optimized separately onto a tree will not equal the log-likelihood of the data partitions optimized simultaneously, unless the precise values of the rate categories are fixed. Even if a single rate is assumed for all sites, this rate must be specified a priori, otherwise it will again change when each data partition is optimized. Some beta versions of PAUP* have problems optimizing data partitions onto trees with prespecified branch lengths, tending to crash if the branch length smoothing feature is turned off. Thus, in current version of PAUP*, parameters can be fixed to the same values in the combined analysis and PLS optimizations only if the very simple (and arguably overly simplistic) models are used, i.e., those employing a single fully specified rate for all sites. However, this problem can be circumvented by using individual site log-likelihoods from the combined analysis.

To calculate a gene’s contribution in the context of the combined analysis, one must use the exact tree topologies, branch lengths, models, and model parameters employed in the combined analysis when evaluating each data partition. This can be done in PAUP* by obtaining the individual site log-likelihoods for the reverse constraint and ML trees in the combined analysis. For each tree, the sum of sites for each data partition (e.g., gene) is that partition’s log-likelihood. The difference in a partition’s log-likelihood across the two trees is the PLS. The individual site log-likelihoods output by PAUP* can be exported to Microsoft Excel to facilitate these calculations. The SUM and INDEX functions in Excel are used to sum the site log-likelihoods for each partition for each tree and to calculate the difference in these sums (PLS).

To estimate the significance of PLS values, we used the KH and SH tests and the expected LW confidence set. The two trees evaluated were the ML tree and the constrained tree found in the analysis of the combined data set. Each data set was mapped onto these two trees using identical parameters employed in the combined analysis, the difference in log-likelihood was calculated and evaluated for significance using the preferred test. (This approach is different from testing the support in partitioned analyses, where the ML and constrained trees for each data partition would be evaluated and parameters would be optimized for each data set.) These tests are all based on a nonparametric bootstrap resampling estimate of the variation in likelihood of a set of trees, which is then transformed into the KH and SH test statistics as summarized by Goldman et al. (2000) and the expected LW (c) confidence set of Strimmer and Rambaut (2001). The exact version of these tests requires bootstrapping the data and recalculating the log-likelihoods. This process is awkward, requiring separate resamplings for each partition, and the KH and SH tests cannot be exactly implemented directly in PAUP*. Bootstrapping the site log-likelihoods from the original combined data analysis overcomes these complications, greatly simplifying the procedure and is considered an acceptable approximation. A further simplification estimates variance from the variance of the observed site log-likelihoods. In the examples above, the RELL approximation was used: for each data partition, the original site log-likelihoods were randomly resampled with replacement (bootstrapping), and the PLS was recalculated. Site log-likelihoods were bootstrapped in Microsoft Excel by using the RANDOM function to randomly select a cell within the range that covers the partition under investigation. One hundred pseudoreplicates were generated. PLS was then calculated for each bootstrapped data set using the above procedure. From these PLS values, the SH, KH, and LW (c) test statistics were calculated, again using the Excel spreadsheet program. Different selections of trees will influence the values obtained for the SH test, and the trees included in the SH test should include all trees that can possibly be entertained as the true topology. The set of trees included in the SH test should be an a priori decision, and the power of the test is then appropriate for the given set of trees. However, a major problem is that no precise criterion has yet been proposed to identify this set of trees (Buckley, 2002), and the safe option of including every possible tree (Goldman et al., 2000) will usually be computationally impractical and may result in a less sensitive test, enlarging the confidence interval of acceptable trees (Strimmer and Rambaut, 2001; Buckley, 2002; Shimodaira, 2002). In this study, the set of included trees comprises the ML tree and the optimal constrained tree for each clade. This set will contain the best and the second best trees and an efficient sampling of the best of all others, and might be a rational way to select a workable subset of trees from what could be an insuperable number of topologies.