Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution 46 (2008) 830–843
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Mitochondrial sequence data and Dipsacales phylogeny: Mixed models, partitioned Bayesian analyses, and model selection

Richard C. Winkworth [a,b,*], Charles D. Bell [a,c], Michael J. Donoghue [a]

[a] Department of Ecology and Evolutionary Biology, Yale University, P.O. Box 208106, New Haven, CT 06520, USA
[b] Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, Caixa Postal 11461, CEP 05422-970 São Paulo, SP, Brazil
[c] Department of Biological Sciences, University of New Orleans, 2000 Lakeshore Drive, New Orleans, LA 70148, USA

Received 21 November 2006; revised 20 October 2007; accepted 23 November 2007
Available online 5 December 2007

Abstract

Phylogenetic analyses of chloroplast DNA sequences, morphology, and combined data have provided consistent support for many of the major branches within the angiosperm clade Dipsacales. Here we use sequences from three mitochondrial loci to test the existing broad scale phylogeny and in an attempt to resolve several relationships that have remained uncertain. Parsimony, maximum likelihood, and Bayesian analyses of a combined mitochondrial data set recover trees broadly consistent with previous studies, although resolution and support are lower than in the largest chloroplast analyses. Combining chloroplast and mitochondrial data results in a generally well-resolved and very strongly supported topology but the previously recognized problem areas remain. To investigate why these relationships have been difficult to resolve we conducted a series of experiments using different data partitions and heterogeneous substitution models. Usually more complex modeling schemes are favored regardless of the partitions recognized but model choice had little effect on topology or support values. In contrast there are consistent but weakly supported differences in the topologies recovered from coding and non-coding matrices. These conflicts directly correspond to relationships that were poorly resolved in analyses of the full combined chloroplast–mitochondrial data set. We suggest incongruent signal has contributed to our inability to confidently resolve these problem areas.

Keywords: Bayesian inference; Caprifolieae; Dipsacales; *Heptacodium*; Linnaeeae; Mitochondrial DNA; Mixed models; Data partitions

1. Introduction

Dipsacales is a clade containing ca. 1050 species of flowering plants. Its members display a wide array of vegetative and reproductive morphologies. For example, they range from small herbaceous annuals to shrubs and trees, their flowers range from small and radially symmetrical to large and bilaterally symmetrical, and their fruits may be drupes, berries, capsules, or achenes (see Donoghue et al., 2003 and references therein). The group has a predominantly Northern Hemisphere distribution, occurring in temperate and boreal forests, the seasonally arid Mediterranean region, and the mountains of Asia, Europe, and North America. In the Southern Hemisphere species diversity is centered primarily on the mountains of Central America and Southeast Asia (see Donoghue et al., 2003 and references therein).

Over the past decade analyses of chloroplast DNA sequences, morphological characters, and combined data have done much to clarify the broad structure of Dipsacales phylogeny (see Donoghue et al., 2003 and references therein). Previous analyses have consistently recovered two major lineages, (i) a larger clade including Diervilleae, Caprifolieae, and Linnaeeae of the traditional Caprifoliaceae, as well as the Morinaceae, Valerianaceae, and Dipsacaceae, and (ii) a smaller clade containing *Viburnum* and *Sambucus* of the traditional Caprifoliaceae, plus

* Corresponding author. Present address: School of Biology, Chemistry, and Environmental Sciences, University of the South Pacific, Private Bag, Laucala Campus, Suva, Fiji. Fax: +55 203 432 3854.
E-mail address: rwinkworth@gmail.com (R.C. Winkworth).

1055-7903/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2007.11.021

Sinadoxa, *Tetradoxa*, and *Adoxa*. For the most part these studies also consistently resolve relationships within the two basal lineages. For example, in the smaller clade *Viburnum* always branches first with *Adoxa*, *Sinadoxa*, and *Tetradoxa* forming a clade sister to *Sambucus*. Clearly the traditional concept of Caprifoliaceae is not supported by phylogenetic analyses and several revised classifications have been proposed based upon this insight (e.g., Backlund and Pyck, 1998; Benko-Iseppon and Morawetz, 2000; Donoghue et al., 2001). Here we adopt the explicitly phylogenetic scheme of Donoghue et al. (2001) (Fig. 1). This has two key advantages over the remaining proposals. First it does not require name changes solely to adjust rank, clades corresponding to previously recognized groups retain their traditional names (e.g., Caprifolieae, Morinaceae, Dipsacaceae). Second, names are given to several clades not recognized in the other systems. Specifically, Donoghue et al. (2001) apply the name Caprifoliaceae to the larger of the two basal clades, while Valerina and Linnina refer to sub-clades within Caprifoliaceae (Fig. 1).

One might assume that since previous studies have consistently recovered the same, often well-supported topology there is little left to learn about broad scale relationships within Dipsacales. However, there remain several outstanding issues; specifically, the placement of *Heptacodium* and relationships among the exemplars of Caprifolieae and Linnaeeae are not yet consistently resolved. Although seemingly minor, resolution of these problems is critical for understanding character evolution in Dipsacales as a whole. For example, the current tree suggests a complex pattern of fruit evolution, but since *Heptacodium* has achenes and both drupes and berries occur in Caprifolieae we cannot be entirely certain of the evolutionary pattern (Donoghue et al., 2003). Generally it has been assumed that uncertainty in Dipsacales phylogeny resulted from limited signal and that simply adding data would solve the problem. However, comparison of two recent studies suggests that limited data alone is not a sufficient explanation. Specifically, the combined matrices compiled by Bell et al. (2001) and Zhang et al. (2003) contain 1819 and 331 parsimony informative positions, respectively. Despite the matrix containing more than five times amount of information, the analyses of Bell et al. (2001) do not recover substantially better support values for any of the three problem areas.

So why have previous analyses been unable to resolve the uncertainty? Partly the problem may reflect taxon sampling. Previous analyses of the broad relationships (see Donoghue et al., 2003 and references therein) have often included sin-

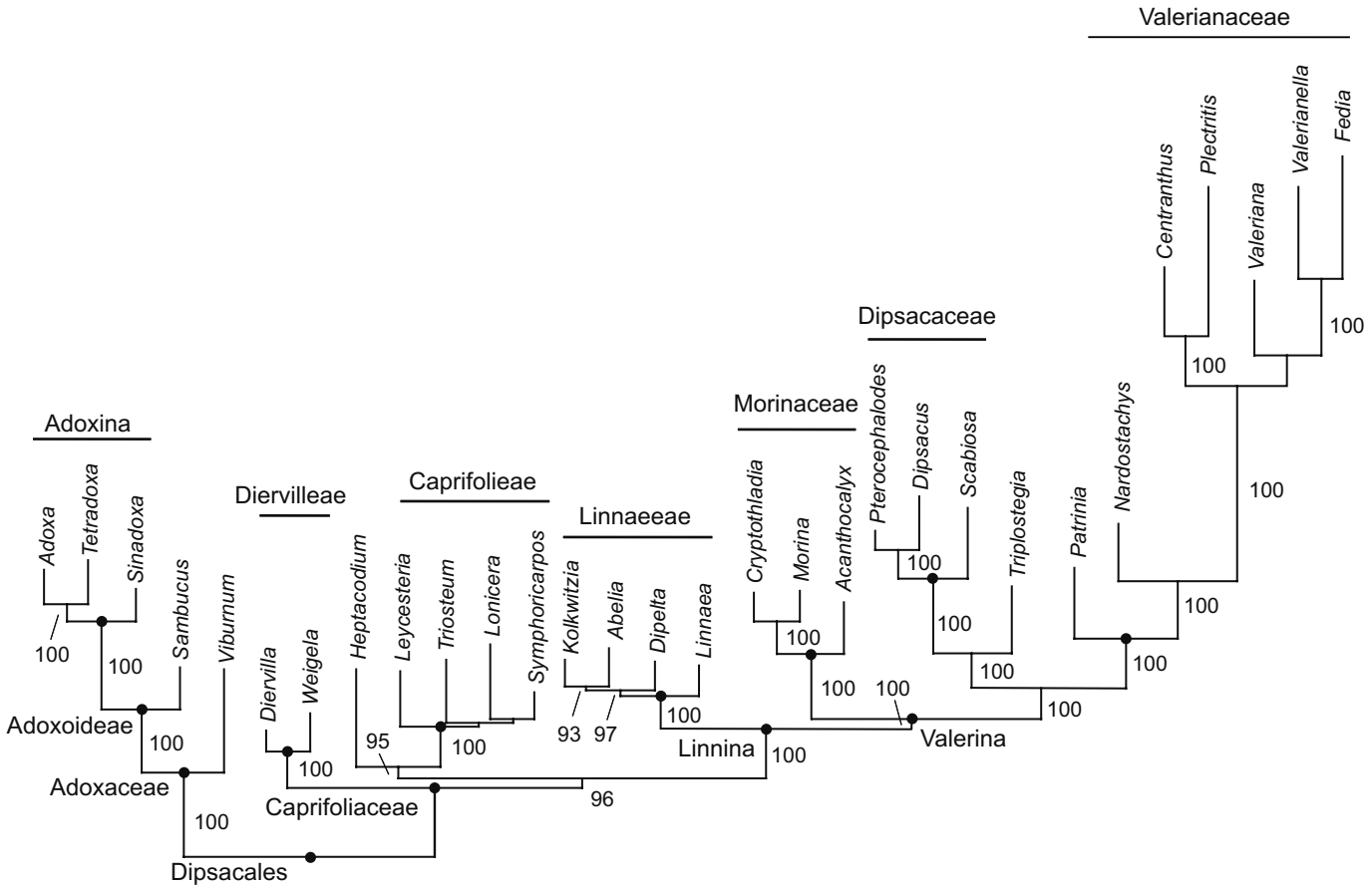

Fig. 1. The maximum likelihood topology from a combined analysis of ITS and chloroplast DNA sequence data (modified from Donoghue et al., 2003). Bootstrap support values are noted when greater than 90%. Clades named by Donoghue et al. (2001) are marked by black circles; names are given either at the node or above the corresponding clade.

gle representatives of key genera. Almost certainly this fails to adequately represent the variation within lineages. For example, including a single species from each of the four genera of Caprifolieae under represents *Lonicera* (ca. 180 species) relative to *Leycesteria*, *Symphoricarpos*, and *Triosteum* (ca. 30 species in total). Also potentially problematic is the almost exclusive use of chloroplast sequences. While sampling of markers has increased dramatically (e.g., compare Backlund and Bremer (1997) with Bell et al. (2001)) the limitations of relying on sequences from a single genome are widely recognized (e.g., Soltis and Soltis, 2004). It is therefore important to test the existing phylogeny against nuclear and mitochondrial sequences. The present study specifically addresses this issue by adding a large mitochondrial matrix. We also prepare combined chloroplast–mitochondrial matrices and use these to examine whether differences between data partitions may be influencing topology and clade support in broad analyses of Dipsacales phylogeny.

2. Materials and methods

2.1. Taxon sampling

One goal of the present study is to compare phylogenetic signal from different data partitions. In order to facilitate this comparison we, with one exception, maintained the sampling used by Bell et al. (2001). We included an additional species of *Valeriana*, *V. celtica*. This species has recently been shown to represent a basal lineage of Valerianaceae, rather than be closely related to the majority of *Valeriana* (see Hidalgo et al., 2004; Bell, 2004, Bell, 2007). Wherever possible we also used the same DNA accession as Bell et al. (2001), where this was not possible (i.e., DNA extraction exhausted) an alternative accession was included. For example, the chloroplast matrix of Bell et al. (2001) included *V. acerifolium* sequences but mitochondrial sequences are represented by *Viburnum dentatum*. Voucher information and GenBank numbers for the corresponding sequences are given in Appendix A.

2.2. DNA extraction, PCR amplification, and DNA sequencing

Genomic DNA was isolated from silica dried leaf tissue or herbarium samples using either a modified cetyltrimethylammonium bromide (CTAB) protocol (Doyle and Doyle, 1990), or the Qiagen DNeasy Plant Kit (La Jolla, CA).

PCR amplifications were performed in 25 μl reaction volumes, containing 1× PCR buffer (Applied Biosystems), 6.25 mM $MgCl_2$, 625 μM each dNTP (Invitrogen), 5% bovine serum albumin (v/v; New England Biolabs), 10 pM each amplification primer, 1 U AmpliTaq DNA polymerase (5 U/μl; Applied Biosystems) and 10–100 ng of total cellular DNA. We used the primers described by Duminil et al. (2002) for amplification of the cytochrome oxidase 1 (*cox1*) and cytochrome oxidase 3 (*cox3*) genes, and sequences from an intron in the NADH dehydrogenase I subunit 5 (*nad5*) gene. Thermocycling conditions for PCR were: initial denaturation at 98 °C for 3 min, 35–40 cycles of 1 min at 95 °C (denature), 1 min at 50–55 °C (annealing), and 2–3 min at 72 °C (extension), with a final 72 °C incubation for 5 min. Amplification products were purified using the QIAquick PCR Purification Kit (Qiagen). Automated sequencing reactions used the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction kit and were analyzed on either an ABI 3100 Genetic Analyzer or MJ Research BaseStation 51 Fragment Analyzer. PCR fragments were sequenced using the corresponding amplification primers; specific internal sequencing primers were also used for the *cox1* (DIPScox1F 5′-CTAGATACCCGA AGACCGG-3′ and DIPScox1R 5′-CTTCGGTCATCCA GAGGTG-3′) and *nad5*/4-5 (DIPSnad5F 5′-CGAGTG AAGTGCTTACGCC-3′ and DIPSnad5R 5′-GGCAGTG AAGGCTCGCGA-3′) regions.

2.3. Data sets

We prepared multiple sequence alignments for individual mitochondrial loci using ClustalX, with subsequent visual inspection. Incongruence between loci was evaluated using the incongruence length difference (ILD) tests as implemented in PAUP*4.0b10 (Swofford, 2002); these test used default parameters and 1000 replicate searches. We tested all pairwise comparisons and also performed a simultaneous test with all three markers. By concatenating individual alignments we constructed a single mitochondrial data matrix for phylogenetic analysis.

To further examine why specific relationships have remained poorly resolved we constructed a series of matrices that combine Bell et al.'s (2001) five-locus chloroplast matrix with our mitochondrial data set. We first constructed a full combined matrix by concatenating the two data sets; conflict between mitochondrial and chloroplast partitions was evaluated using the ILD test, with default parameters and 1000 replicate searches. Although genome-level differences are perhaps the most obvious source of conflict in the combined data set, functional differences between the sampled loci are another potential source. To examine conflict between functional classes we complied separate coding and non-coding matrices—the coding data set contained mitochondrial *cox1* and *cox3* sequences, along with chloroplast *matK*, *ndhF*, and *rbcL* genes, the non-coding matrix contained sequences from the mitochondrial *nad5*/4-5 intron, chloroplast *trnL-F* region, and chloroplast *atpB-rbcL* intergenic spacer. Differences between the two were evaluated using the ILD test as above. Our data matrices and trees are available from TreeBASE (study accession number S1963, matrix accession numbers M3619–3622).

2.4. Phylogenetic analyses

2.4.1. Maximum parsimony and maximum likelihood

Phylogenetic trees for mitochondrial, coding, non-coding, and full combined data matrices were inferred using

both maximum parsimony (MP) and maximum likelihood (ML) as implemented in PAUP*4.0b10 (Swofford, 2002). Heuristic MP searches used "tree-bisection-reconnection" (TBR) branch swapping, zero-length branches collapsed, and all characters equally weighted. Analyses were repeated 100 times with RANDOM ADDITION. For MP and all subsequent analyses gaps were treated as missing data. A best-fitting substitution model was determined for each matrix using a hierarchical series of likelihood ratio tests as implemented in PORN* (Bell, 2001). Heuristic ML tree searches used the most appropriate model (with parameters simultaneously estimated via ML), TBR branch swapping, collapsed zero-length branches, and were repeated 100 times with RANDOM ADDITION. MP and ML bootstrap analyses used 1000 replicates with NNI branch swapping; for the ML tests parameter values were estimated from the optimal ML tree.

2.4.2. Bayesian inference

Estimates of topology and support from combined data may be biased by the use of uniform models because parameters values are averaged across loci and may not adequately describe one or more of the partitions (Leaché and Reeder, 2002; Reeder, 2003). In contrast, heterogeneous substitution models allow parameter values to be assigned to data partitions individually—the expectation being that this leads to better data fit and improved phylogenetic estimates (Fishbein and Soltis, 2004; Brandley et al., 2005). For each of our data sets we identified model partitions based on differences in function (i.e., coding versus non-coding) and genome (e.g., chloroplast versus mitochondrial), as well as partitioning matrices by locus (i.e., *rbc*L, *ndh*F, *mat*K, etc.). Based on previous analyses and preliminary tests using the Akaike Information Criterion (AIC; Akaike, 1974) and Bayesian Information Criterion (BIC; Schwarz, 1974; Kass and Wasserman, 1995) we assigned GTR+I+G substitution models to each of the data partitions. For the coding matrix we also used a pair of site-specific models, either constraining the rate parameters of the GTR model across all sites (i.e., linked model, C_4) or allowing them to vary according to codon position (i.e., unlinked model, C_5). Table 1 lists the models used in Bayesian analyses.

Bayesian analyses were performed using Metropolis-coupled Markov chain Monte Carlo as implemented in MrBayes ver. 3.1.2 (Ronquist and Huelsenbeck, 2003). Searches used default settings for an incremental heating scheme (i.e., three "heated" chains, and one "cold" chain) as well as defaults for the priors on the rate matrix (0–100), branch lengths (0–10), gamma shape parameter (0–10), and the proportion of invariable sites (0–1). A Dirichlet distribution was used for the base frequency parameters and an uninformative prior was used for the tree topology. Individual chains were initiated with a random starting tree and run for 5 million generations, sampling from the posterior distribution of trees every 100 generations (for a total of 50,000 samples). The "burn-in" was determined using convergence diagnostics and by plotting $-\ln L$ versus the number of generations; sample points collected prior to stationarity were eliminated and the remainder used to estimate posterior probabilities for clades. For each analysis four independent runs were performed to assess convergence.

2.4.3. Evaluating the fit of heterogeneous models

Using an appropriate or adequate model is a critical aspect of model-based phylogenetic inference (e.g., Posada, 2001; Posada and Crandall, 2001; Buckley, 2002; Posada and Buckley, 2004; Sullivan and Joyce, 2005). This is especially true for heterogeneous models where the number of estimated parameters can be large. The AIC, BIC, and Bayes factors are now commonly used to evaluate model fit. For each analysis AIC and BIC values were calculated using the best likelihood score from the post-burn-in sample with the best-fit model having the lowest score. Although commonly used, in a Bayesian context both these approaches have limitations as they ignore the influence of the prior and the analysis does not maximize the likelihood. Bayes factors, or more specifically the approximation provided by the harmonic mean of likelihoods from

Table 1
Statistics for the mitochondrial and three combined matrices, as well the corresponding parsimony and maximum likelihood analyses

	Mitochondrial	Mitochondrial and chloroplast		
		Full combined	Non-coding	Coding
Matrix length in nucleotides	3938	11,531	4526	7005
No. of varied nucleotides	246	3214	1359	1855
No. of MP informative positions	132	1972	844	1128
Maximum parsimony analysis				
No. of trees	66	1	1	4
No. of steps	334	2269	2466	3398
CI/RI	0.653/0.878	0.608/0.787	0.672/0.820	0.593/0.772
Maximum likelihood analysis				
Selected model (hLRT)	GTR + I + G	GTR + I + G	GTR + I + G	GTR + I + G
No. of trees	1	1	1	1
Likelihood score	7825.25735	50964.78297	20250.23606	30357.31307

the posterior sample, are an alternative (Ronquist and Huelsenbeck, 2003; Nylander et al., 2004). Brown and Lemmon (2007) have suggested that Bayes factors provide a robust means for choosing among partitioning strategies; although others have pointed out that the approach will tend to over-score complex models (Pagel and Meade, 2004; Lartillot and Philippe, 2006). Bayes factors were estimated for each analysis and the analyses compared in a pair-wise fashion for each data set. A positive value greater than two suggests positive evidence against the alternative hypothesis, while greater than 10 is considered very strong evidence (Kass and Raftery, 1995).

2.5. Comparing matrices and topologies

Trees based on analyses of coding and non-coding matrices suggest different relationships both within and among several of the major lineages of Caprifoliaceae. We tested the significance of these differences using the ILD test, significantly less parsimonious test (SLP test; Templeton, 1983), winning sites test (WS test; Sokal and Rohlf, 1995), and the Shimodaira–Hasegawa test (SH test; Shimodaira and Hasegawa, 1999) as implemented in PAUP*4.0b10.

We conducted ILD tests on a series of reduced matrices in order to evaluate the influence of specific differences between data sets. Reduced matrices were constructed by selectively removing taxa with conflicting placements in coding and non-coding topologies. All ILD tests used 1000 replicates and default search parameters. We used the optimal ML topologies for topology-based incongruence tests; for SH tests we used a GTR model (with model parameters estimated on the topologies of interest) and estimated the test distribution using 1000 RELL bootstrap replicates. As for ILD tests we conducted a series of analyses to evaluate the contribution of the different conflicts to the overall incongruence. For these we again used the ML topologies; rival trees were constructed by constraining the relationship of interest to reflect the topology suggested by the other analysis. We investigated the affect of differences in the placement of *Heptacodium*, Caprifolieae, and Linnaeeae, as well as relationships within the latter two.

3. Results

3.1. Data sets and phylogenetic analyses

3.1.1. Mitochondrial sequences

With one exception it was possible to amplify and sequence the three mitochondrial loci for each of the 31 taxa; *nad5*/4-5 could not be amplified from *Tetradoxa omeiensis*. Final alignments for the individual loci contained 1302, 658, and 1978 nucleotides for the *cox1*, *cox3*, and *nad5*/4-5 regions, respectively. In all pairwise combinations, and a comparison of all three markers, ILD tests suggest no significant conflict between partitions—*P* values ranged from 0.235 for *cox1–nad5*/4-5 to 0.012 for the three-marker comparison. Summary statistics for the mitochondrial matrix and subsequent phylogenetic analyses are given in Table 1. Trees from MP, ML, and Bayesian analyses are all highly similar to one another and are, for the most part, consistent with earlier chloroplast-based phylogenies. There is moderate to strong support from bootstrapping (both MP and ML) and Bayesian posterior probabilities for many of the relationships suggested by mitochondrial data. However, several parts of the tree are more poorly resolved and supported; specifically, (i) relationships within Caprifolieae and Linnaeae, (ii) the monophyly of Diervilleae, (iii) relationships of *Heptacodium*, and (iv) relationships among Linnaeae, Valerianaceae, Dipsacaceae, and Morinaceae (Fig. 2).

Sequence alignments for the *nad5*/4-5 locus contained gaps that were consistent with trees based on nucleotide substitutions in the combined mitochondrial data set. Inferred indels support Linnina (one deletion), Morinaceae (one deletion), Adoxoiodeae (one insertion, although *Tetradoxa* is not represented), and the pairing of *Cryptothladia* and *Morina* (three deletions). Consistent with the protein coding roles of *cox1* or *cox3* no indels were inferred in these alignments.

3.1.2. Analyses of combined chloroplast and mitochondrial sequences

Summary statistics for the three combined chloroplast–mitochondrial matrices and subsequent analyses are given in Tables 1 and 2.

Phylogenetic analyses of the full combined, coding, and non-coding matrices result in topologies that are highly similar to one another (Fig. 3A–C). Furthermore, clades corresponding to the major lineages (*sensu* Donoghue et al., 2001) often received bootstrap support of greater than 80% (both MP and ML) and Bayesian support values of 1.0 in these analyses. There are several differences between analyses of coding, non-coding, and combined data; these differences are recovered using MP, ML, and Bayesian analyses. Topological differences between the full combined and non-coding analyses are limited to relationships within Caprifolieae (Fig. 4A and C). The non-coding data also provided weaker support for relationships at the base of Valerianaceae. The positions of Caprifolieae, Linnaeeae, and *Heptacodium*, as well as within both Caprifolieae and Linnaeeae differ in coding analyses relative to the other data sets (compare Figs. 3A–C and 4A–C). However, these differences are only weakly supported by bootstrapping and posterior probabilities.

3.2. Evaluating the fit of heterogeneous models

Test scores for the AIC and BIC are summarized in Table 2; Bayes factors in Table 3. Most often these selection criteria favored the use of more complex, parameter rich models. The AIC favors the "by locus" model for each of the three matrices; the BIC also prefers this model for the coding data but the less complex "by genome" models for the non-coding and full combined data sets. However, notice that the "by locus" model has a very similar score

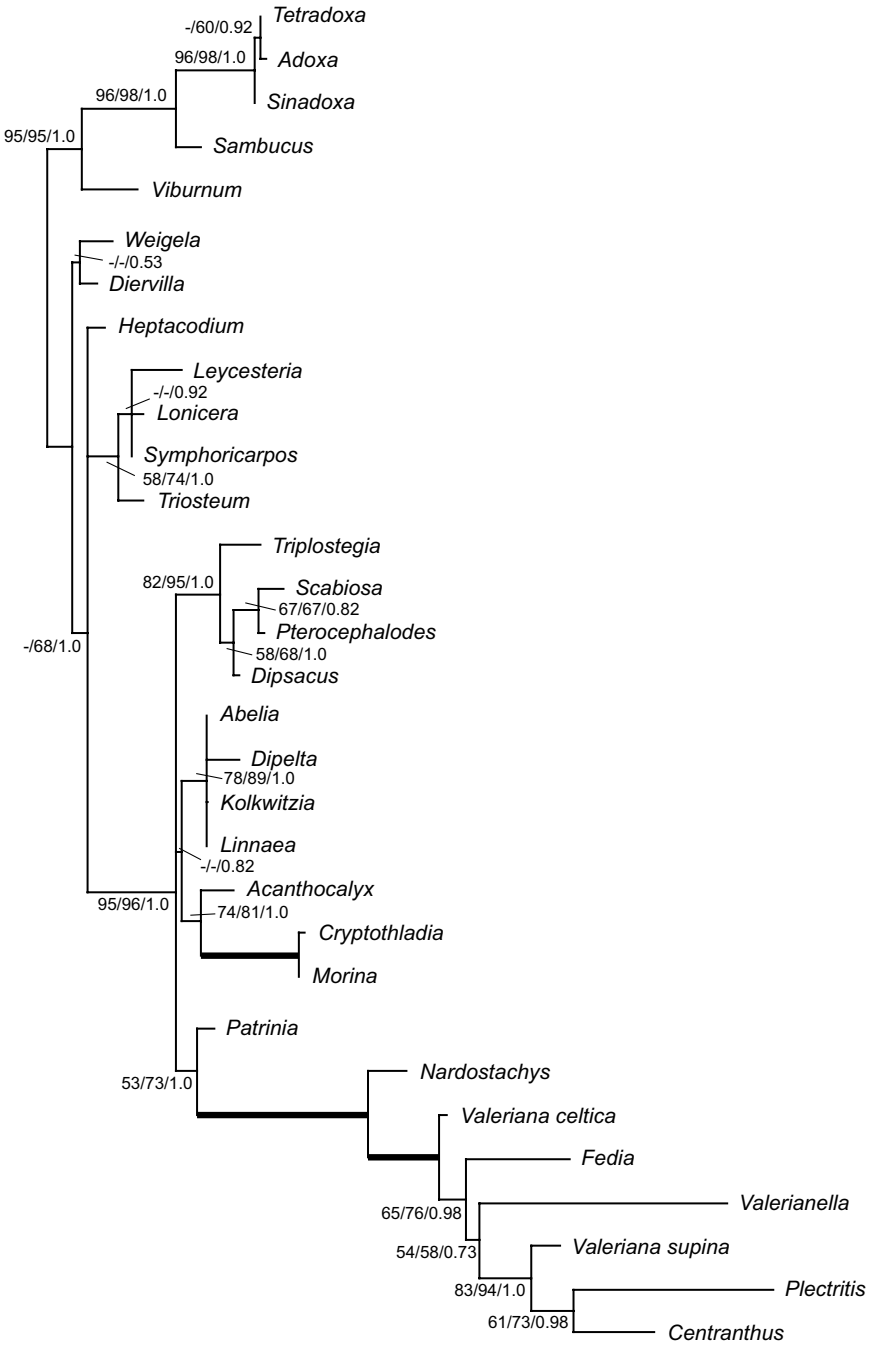

Fig. 2. Bayesian probability estimate of Dipsacales phylogeny based on combined mitochondrial DNA sequences and using substitution model MT_3. Branch lengths are calculated from means of the posterior probability density. Thickened edges receive bootstrap support of 100% (both MP and ML) and posterior probabilities of 1.00. For nodes where support values are lower individual values are given; the first number is the parsimony bootstrap value, followed by likelihood bootstrap and posterior probability. A dash indicates that support was less than 50% (for bootstrapping) or 0.50 (for Bayesian support values); for that node.

in both cases. More complex models also tend to be preferred by Bayes factors.

3.3. Comparing the topologies based on different functional data partitions

Visual comparison of trees indicates that several differences are consistently recovered. Although these differences are not strongly supported we further examined incongruence between data partitions using the ILD, SLP, and SH tests.

Results of ILD tests are summarized in Table 4. Consistent with the visual differences between trees a test of the full matrix indicates substantial conflict between the coding and non-coding partitions ($P = 0.108$). Tests on the reduced matrices suggest that the placement of *Heptacodium* contributes little to incongruence between data matri-

Table 2
Details of models used in Bayesian analyses of mitochondrial and combined data, as well as estimates of AIC and BIC

Data set	Model	Partitions	Free parameters	ln *L*	Harmonic mean	AIC	BIC
Mitochondrial	MT_1	1—uniform	10	−8046.68	−8094.64	16113.36	16129.31
	MT_2	2—by functional class	20	−7872.34	−7921.98	15784.68	15816.59
	MT_3	3—by locus	30	−7863.99	−7919.07	15787.98	15835.84
Full combined	FC_1	1—uniform	10	−50979.98	−51020.45	101979.96	102053.49
	FC_2	2—by functional class	20	−50918.66	−50962.85	101877.32	102024.38
	FC_3	2—by genome	20	−50303.42	−50511.89	100646.83	100793.88
	FC_4	4—by function and genome	40	−50197.61	−50395.29	100475.21	100769.32
	FC_5	9—by locus	90	−49972.25	−50210.13	100124.51	100786.26
Coding	C_1	1—uniform	10	−30373.63	−30412.25	60767.26	60835.80
	C_2	2—by genome	20	−30068.77	−30165.98	60177.54	60314.63
	C_3	5—by locus	50	−29915.43	−30046.90	59930.86	60273.58
	C_4	3—by codon position (linked)	11	−31217.33	−31333.87	62456.66	62532.06
	C_5	3—by codon position (unlinked)	27	−31166.03	−31245.87	62386.06	62571.13
Non-coding	NC_1	1—uniform	10	−20258.16	−20296.77	40536.32	40600.49
	NC_2	2—by genome	20	−19832.28	−20078.43	39704.56	39832.92
	NC_3	4—by locus	40	−19798.38	−19985.76	39676.75	39933.45

ces, whereas that associated with Caprifolieae and Linnaeeae makes a more substantial contribution. Results of topology-based tests are summarized in Table 5. Comparisons using the original topologies indicate substantial differences between partitions; tests targeting specific relationship are also consistent with differences.

4. Discussion

Although our understanding of Dipsacales phylogeny has benefited greatly from analyses of chloroplast sequences, several areas of uncertain resolution still remain. Here we examined a mitochondrial data set both as a test of the existing framework and in the hope these data would resolve the remaining uncertainty. In the following sections we consider the implications of our analyses for understanding Dipsacales phylogeny.

4.1. Support for the broad patterns of Dipsacales phylogeny

Gene trees from analyses of the mitochondrial matrix are all very similar to one another and to those from other recent molecular analyses (e.g., Bell et al., 2001; Donoghue et al., 2001, 2003; Zhang et al., 2003). In particular, mitochondrial data provide further support for the split between Adoxaceae and Caprifoliaceae as well as clades corresponding to previously recognized groups (sensu Donoghue et al., 2001; Fig. 2). Although consistent with earlier trees, analyses of mitochondrial data are more poorly resolved and supported than those based on the most comprehensive chloroplast matrix (i.e., Bell et al., 2001). This likely reflects a combination of a smaller matrix and the slower rates of sequence evolution commonly reported for the plant mitochondrial genome (e.g., Wolfe et al., 1987; Palmer et al., 2000). Not surprisingly trees from combined chloroplast–mitochondrial analyses are better resolved and supported than those based on mitochondrial data alone (compare Fig. 2 and Fig. 3). Indeed analyses of the full combined data set provide very strong support for the majority of relationships both within and among previously recognized lineages (Fig. 3A). Although many of the same relationships were recovered in analyses of the large chloroplast matrix of Bell et al. (2001), our full combined data further increases support for several nodes. In particular, likelihood bootstrap values are higher for the Valerina, the Linnina–Caprifolieae–*Heptacodium*, and the Caprifolieae–*Heptacodium* clades. Support for this latter relationship is particularly interesting since although previous analyses have often recovered this arrangement support values have tended to be much weaker (e.g., Pyck et al., 1999; Pyck and Smets, 2000; Bell et al., 2001; Donoghue et al., 2001; Zhang et al., 2003).

Previous studies of morphology and chloroplast sequences have consistently recovered the same pattern of broad relationships within Dipsacales. Our analyses of mitochondrial data provide further confirmation of this structure. It is extremely comforting to find sequences from a second genome support basically the same topology as that suggested earlier. However, although analyses of the mitochondrial and full combined chloroplast–mitochondrial matrices increase confidence in the broad topology, these data do not resolve all the remaining uncertainty (Fig. 4A). Specifically, although we have added a new data set we have not clarified relationships within either Caprifolieae or Linnaeeae.

4.2. Insights from data partitioning

Previous studies have generally assumed that the remaining uncertainty in Dipsacales phylogeny resulted a lack of signal or inadequate taxon sampling. However, our analyses suggest incongruence between functional data partitions may be a more important source of uncertainty. Specifically, in our analyses there is a close correspondence

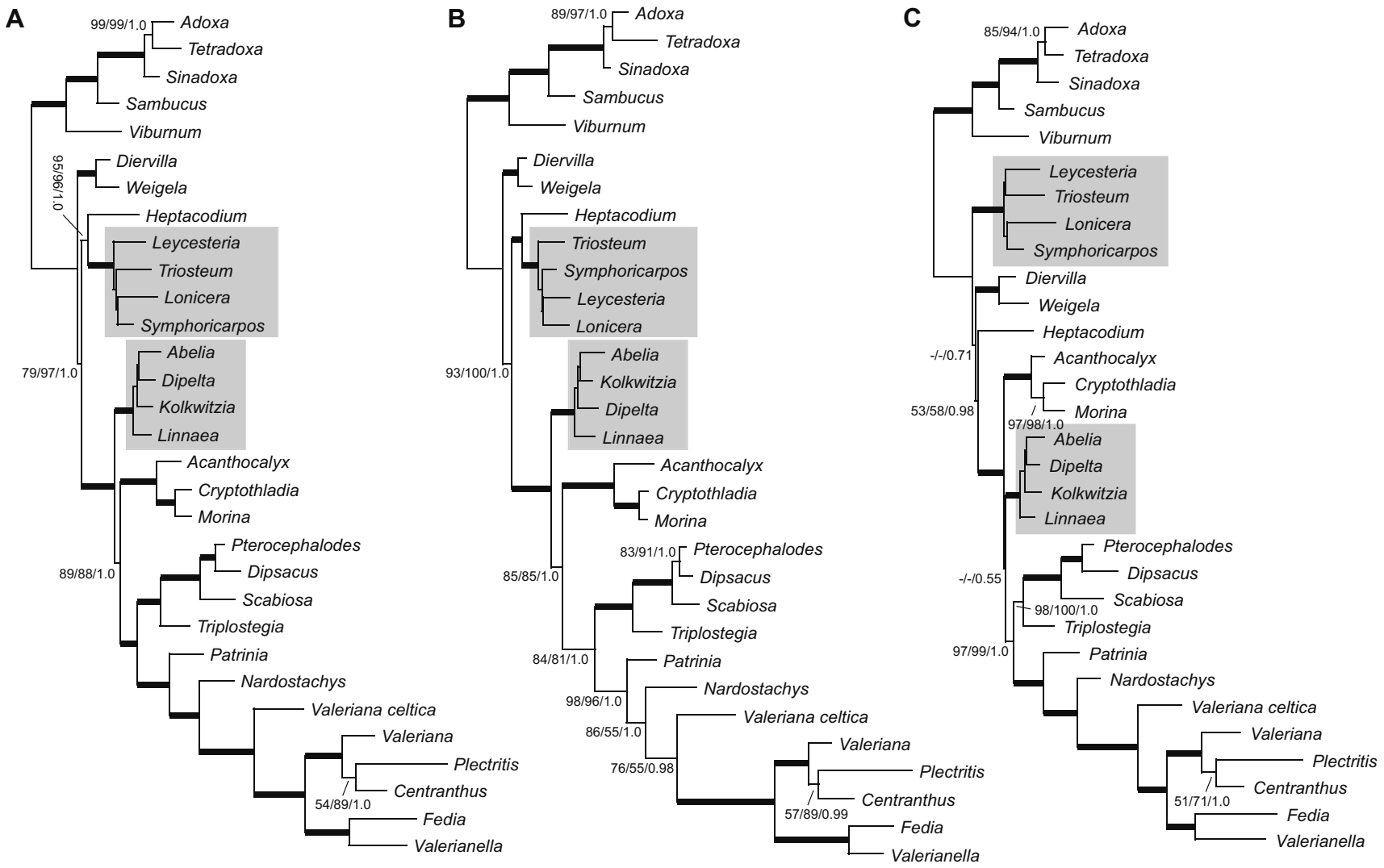

Fig. 3. Single optimal maximum likelihood estimates of Dipsacales phylogeny based on combined analyses of mitochondrial and chloroplast sequences. (A) Full combined matrix. (B) Non-coding data set. (C) Coding matrix. Support values are indicated as described in Fig. 2.

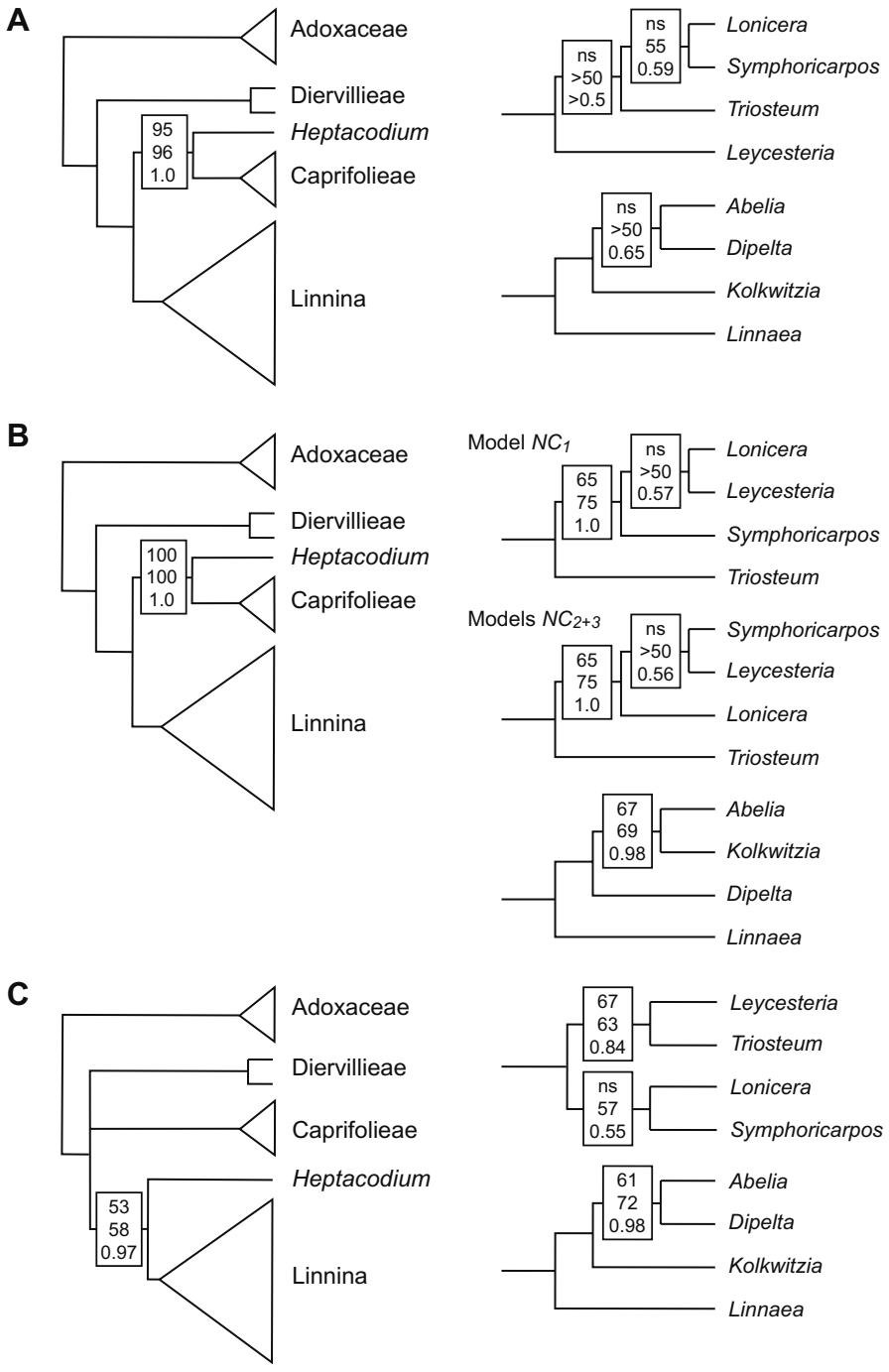

Fig. 4. Schematic diagrams indicating differences in the relationships suggested by analyses of different data sets and substitution models. (A) Topology based on full combined data. On the left is schematic of whole tree (Adoxaceae, Caprifolieae, and Linnina clades represented as triangles) and on the right are details of the relationships within Caprifolieae (above) and Linnaeeae (below). Support values (MP bootstrap, ML bootstrap, and Bayesian posterior probabilities in that order top to bottom) are associated with relevant edges; the minimum value of the posterior probability is given and "ns" indicates the edge was not present in the corresponding topology. (B) Topologies based on non-coding data. The upper two topologies on the right correspond to relationships in Caprifolieae recovered under different heterogeneous substitution models. (C) Topology based on coding data. Given the very weak support for differences in placement of Caprifolieae and Diervilleae when using standard and codon-specific models we represent this relationship as a polytomy.

between relationships that are poorly resolved in the full combined topology and those that are incongruently resolved in coding and non-coding trees. Relationships that are confidently resolved in full combined analyses are also recovered with strong support in separate analyses of coding and non-coding data partitions (Fig. 3); furthermore many of these same relationships had been found in previous studies. In contrast, relationships with conflicting resolutions in coding and non-coding trees are those that have not been confidently resolved in earlier studies and were more poorly supported in our full combined analyses (Figs. 3 and 4). Despite the close correspondence between con-

Table 3
Evaluation of model adequacy using −2ln of the Bayes factor

Data matrix	Initial model				Alternative model
			MT_2	MT_3	
Mitochondrial			345.32	**351.14**	MT_1
				5.82	MT_2
	FC_2	FC_3	FC_4	FC_5	
Full combined	115.20	1017.12	1250.32	**1620.64**	FC_1
		901.92	1135.12	**1505.44**	FC_2
			233.20	**603.52**	FC_3
				370.32	FC_4
	C_2	C_3	C_4	C_5	
Coding	492.54	**730.70**	−1843.24	−835.62	C_1
		238.16	−2335.78	−2159.78	C_2
			−2573.94	−2397.94	C_3
				176.00	C_4
			NC_2	NC_3	
Non-coding			436.68	**622.02**	NC_1
				185.34	NC_2

Bold values are comparisons used to determine the optimal partitioning strategy.

Table 4
Incongruence length difference test scores from comparisons of coding and non-coding data partitions in a series of Dipsacales matrices

	N	*P* value
All Caprifolieae, all Linnaeeae, *Heptacodium*	1	0.108
All Caprifolieae, all Linnaeeae, no *Heptacodium*	1	0.140
All Caprifolieae, no Linnaeeae, *Heptacodium*	1	0.140
All Linnaeeae, no Caprifolieae, *Heptacodium*	1	0.520
All Caprifolieae, single Linnaeeae, no *Heptacodium*	4	0.269–0.548
All Linnaeeae, single Caprifolieae, no *Heptacodium*	4	0.050–0.484
Heptacodium, single Caprifolieae, single Linnaeeae	16	0.014–0.891
All Caprifolieae, no Linnaeeae, no *Heptacodium*	1	0.305
All Linnaeeae, no Caprifolieae, no *Heptacodium*	1	0.583
Single Caprifolieae, no Linnaeeae, no *Heptacodium*	4	0.288–0.878
Single Linnaeeae, no Caprifolieae, no *Heptacodium*	4	0.751–1.000
Heptacodium, no Caprifolieae, no Linnaeeae	1	1.000
No Caprifolieae, no Linnaeeae, no *Heptacodium*	1	1.000

flicting topologies in separate analyses and uncertainty in the combined results it could be argued that very weak support for the coding resolution suggests these relationships are simply artifacts. However, this interpretation seems less convincing since we recovered essentially the same topology in MP, ML, and each of the Bayesian analyses. These methods each have different properties we are not inclined to expect all analyses to recover the same artifactual topology. Instead we suggest that limited support for the coding resolution reflects conflict within the data set. In most cases winning sites tests indicate mixed signal in the coding matrix—some sites contradict (i.e., they require more steps) while some are better explained (i.e., they require fewer steps) by the rival, non-coding topology (Table 5). In contrast, there is little support for the coding resolution in reciprocal tests. It therefore seems likely that data incongruence, both within and between partitions, has contributed to our failure to fully resolve with confidence the broad phylogenetic structure of Dipsacales.

Although data conflict appears to be an important factor in ongoing uncertainty, it is more difficult to convincing explain the underlying causes of observed incongruence. Parallel substitutions at silent third positions or convergent protein evolution are possible explanations. However, contrary to our expectations conflicting sites are distributed more or less randomly with respect to both coding positions and genes. This suggests that alone neither of these mechanisms is sufficient to explain the observed conflicts. Instead we suspect that the explanation is more complex, probably involving a different combination of factors in each case. Several possibilities are outlined below.

4.2.1. Hybridization

While the full combined analyses provide solid support for the placement of *Heptacodium*, analyses of partitioned data and incongruence testing suggest that the issue is not yet completely settled. Indeed, the conflicting placements of *Heptacodium* in our molecular analyses become all the more intriguing when the morphological affinities of this lineage are considered. Specifically, morphological studies

Table 5
SLP, WS, and SH test scores from comparisons of coding and non-coding data sets with various rival constraints

Test data and tree	Rival constraint	SLP test		WS test		SH test
		N	*P* value	Counts	*P* value	*P* value
Coding	Non-coding	38	0.0094	27, −11	0.0150	0.033
	Caprifolieae—non-coding[1]	18	0.0593	13, −5	0.0963	0.043
	Caprifolieae—non-coding[2]	17	0.029	13, −4	0.0490	0.033
	Heptacodium—non-coding	26	0.2393	16, −10	0.3269	0.257
	Linnaeeae—non-coding	3	0.0833	3, 0	0.250	0.100
Non-coding[1]	Coding	35	<0.0001	31, −4	<0.0001	0.002
	Caprifolieae—coding	7	0.0082	7, 0	0.0156	0.063
	Heptacodium—coding	11	0.0009	11, 0	0.001	0.019
	Linnaeeae—coding	3	0.0833	3, 0	0.250	0.106

For the non-coding data topologies corresponding to both those recovered in Bayesian analysis were tested (see Fig. 4).
[1] Topology recovered using the uniform substitution model.
[2] Topology recovered using the heterogeneous by genome and by locus models.

have highlighted similar patterns of conflict (e.g., Golubkova, 1965; Pyck and Smets, 2000; Zhang et al., 2002). For example, *Heptacodium* shares a highly unusual form of ovary and fruit development with Linnina (Hara, 1983; Tang and Li, 1994) but is very similar to Caprifolieae with respect to inflorescence structure (Rehder, 1916; Airy Shaw, 1952; Weberling, 1966). One possible explanation for these conflicting morphological affinities is hybridization. Based on chromosome analyses Zhang et al. (2002) recently suggested that *Heptacodium* might have arisen as a hybrid between ancestors from Caprifolieae and Linnaeeae. Although conflict between genome partitions has often been used to infer hybridization (e.g., Sang et al., 2007) it is less obvious that differences between functional partitions in uniparentally inherited chloroplast or mitochondrial genomes could also be explained by hybridization. Clearly a complex mechanism would be required. One possibility is that if hybridization resulted in replacement of the maternal nuclear-encoded organelle-expressed genes by paternal sequence type then this might require compensatory changes in organellar genes; such changes could make the maternal gene more like the paternal copy.

4.2.2. Rapid diversification

Inspection of trees from coding, non-coding, and full combined data suggests that in comparison to other the major lineages the pattern of diversification within Caprifolieae and Linnaeeae may have been different. Specifically, compared to the other major lineages (e.g., Adoxaceae, Morina) internal edges in the Caprifolieae and Linnaeeae clades appear to be shorter. This suggests that perhaps these two groups diverged from one another over a shorter period of time. Bell and Donoghue (2005) provide absolute divergence time based on chloroplast sequences and relaxed-clock methods. In this analysis representatives of Caprifolieae and Linnaeeae diverged within 3–4 million years (my) and 6–9 my, respectively, depending on the clock method applied. For other major lineages diversification of the main groups occurs over time periods of approximately 15–60 my, again dependent on the method applied. Obviously, detailed studies are needed in order to confirm these observed differences. However, if the main Caprifolieae and Linnaeeae lineages did origin over shorter periods of time it is perhaps not surprising these relationships have been difficult to resolve. As Whitfield and Lockhart (2007) point out, we would expect little data to have supported the original pattern of diversification and, since it has been some 40 my (Bell and Donoghue, 2005) since these divergence events occurred, for the original signal to be further reduced by the accumulation of conflicting signal.

4.2.3. Other lineages

Focusing on the uncertain relationships when looking for underlying mechanisms is normal. However, an alternative is that at least partial the explanations lie with other members of Dipsacales. While obviously conflicting signals might arise between any pair of lineages we suspect that in this case Valerianaceae may be having a particularly strong affect. In most molecular analyses the branch lengths reported for this lineage are longer than in other parts of the tree (Fig. 3). This suggests that evolutionary rates have been faster in Valerianaceae. If so, then perhaps this group is more likely to give rise to conflicting signals. Since a generally well-supported phylogenetic structure is recover this is clearly not a pervasive problem. However, even if mutations in Valerianaceae lead to a few conflicting positions that are randomly distributed with respect to the remaining Dipsacales we might still expect the outcomes to be lineage-specific. Specifically, for relationships that are supported by numerous characters one or two conflicting sites may have little effect on the overall result. In contrast, these same one or two sites could have significant impacts on both resolution and support if there were few characters supporting the original relationships. This certainly appears possible for both Caprifolieae and Linnaeeae.

4.2.4. Limited taxon sampling

It is well known that limited taxon sampling may affect our ability to resolve relationships and obviously this is a potential problem for broad scale studies of any large group—sequencing all members of even a moderately diverse group remains impractical. In this case we have included 2–6 representatives of the previously recognized major lineages. The current sampling clearly under represents the diversity of almost all of the major lineages and not just those clades that remain poorly resolved. However, it does seem likely that limited sampling would have the greatest effect on poorly resolved relationships. The underlying causes of uncertainty in these groups are likely to be accentuated by sampling effects.

4.3. The influence of substitution models

In contrast to data partitioning we find that model selection has little effect on topology and support. Consistent with the expectation that more complex models provide better data fit, our tests indicate that models differ in adequacy and that more complex parameter-rich models are generally favored (Tables 2 and 3). However, these differences in model fit do not result in substantial differences between analyses with respect to resolution and support. For example, for the full combined matrix BIC values suggest that the "uniform" and "by locus" models are substantially worse than the others (Table 2) but all five analyses recover the same topology and very similar posterior probabilities. Since results for a given data set were all highly similar it appears that for these data the different models all offer broadly equivalent descriptions of the underlying substitution process.

Applying different models to a given data set does result in minor differences in topology and support. Interestingly these are all associated with Caprifolieae. One possibility is that differences between analyses reflect genuine differences in the appropriateness of the models. For example, our

tests indicate that the two locus-specific models are more similar to one another than either is to the uniform model (Tables 2 and 3); likewise the locus-specific models suggest the same relationships within Caprifolieae while for the uniform model a different resolution is recovered (Fig. 4B). In this case, since the uniform model is a worse fit we would favor the locus-specific topology. However, given that previous analyses have failed to confidently resolve the relationships within Caprifolieae it is perhaps not surprising that model selection appears to have an impact on relationships within this clade.

4.4. Furthering our understanding

The similarity of topologies from various analyses, including those presented here, provide good reason to be confident in the broad structure of Dipsacales phylogeny. However, it is clear from our study that resolving several outstanding issues poses a significant challenge.

As is often the case resolution will require further sampling of both genes and taxa. Based on our study it seems clear that simply adding markers to increase the size of the matrix is unlikely to provide a meaningful result. However, addition of data and careful analysis of partitions will be a critical step if we are to further improve our understanding of Dipsacales phylogeny. In particular, analyses of nuclear data are now very important. Obviously, such data would provide a further test of the broad structure of the phylogeny, but perhaps more importantly nuclear makers would allow us to further examine the reasons for conflicting data signals. More accurately representing lineage diversity will also be important and this has several potential benefits. Certainly adding taxa will be critical for improving resolution in the remaining problem areas. However, trees with increased sampling would also be useful when attempting to reconstruct broad patterns of morphological character evolution since diversity within groups could be better represented. In this respect a better understanding of the characters themselves, based on detailed macro-morphological and developmental comparisons, is also needed. Such studies will be critical for untangling the complex morphological affinities of several taxa.

Acknowledgments

We thank David Swofford and Jeff Thorne for use of computational resources as well as Johan Nylander and Paul Lewis for very helpful discussions on model selection. This research was supported in part by an NSF ATOL award to MJD. RCW was partially supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (Grant 04/09666-2).

Appendix A. Accession details for sampled taxa

Taxon	Collection/Voucher	GenBank Accession No.		
		cox1	*cox3*	*nad5*/4-5 intron
Abelia chinensis R. Br.	Cult. Arnold Arboretum 1033-85A; *Donoghue & Winkworth 23* (YU, A)	EF121453	EF121484	EF121515
Acanthocalyx alba (Hand.-Mazz.) M. Cannon	*Boufford et al. 28401* (A)	EF121454	EF121485	EF121516
Adoxa moschatellina L.	*Boufford et al. 28906* (A)	EF121455	EF121486	EF121517
Centranthus ruber L. DC.	*Bell 203* (YU)	EF121456	EF121487	EF121518
Cryptothladia chinensis (Pai) M. Cannon	*Boufford et al. 27782* (A)	EF121457	EF121488	EF121519
Diervilla sessilifolia Buckley	Cult. Arnold Arboretum; *Donoghue & Winkworth 29* (YU, A)	EF121458	EF121489	EF121520
Dipelta floribunda Maxim.	Cult. Arnold Arboretum 14514B; *Buckland and Kelly 32* (A)	EF121459	EF121490	EF121521
Dipsacus mitis D. Don	*Boufford et al. 27724* (A)	EF121460	EF121491	EF121522
Fedia cornucopiae (L.) Gaertner	Cult. Bergius Bot. Gard., Sweden; *Eriksson, s.n. 2 Nov. 1999* (SBT)	EF121461	EF121492	EF121523
Heptacodium miconioides Rehder	Cult. Arnold Arboretum 1549-80A; *Donoghue & Winkworth 7* (YU, A)	EF121462	EF121493	EF121524
Kolkwitzia amabilis Graebn.	Cult. Arnold Arboretum 20447-B; *Elsik, Michener, and Bailey 844* (A)	EF121463	EF121494	EF121525
Leycesteria sp.	*Boufford et al. 44597* (A)	EF121464	EF121495	EF121526
Linnaea borealis L.	Door County, WI; *Donoghue, 1990*, no voucher.	EF121465	EF121496	EF121527
Lonicera heteroloba	Cult. Arnold Arboretum 838-76-A; *Elsik, Dumaine, and Groves 1668* (A)	EF121466	EF121497	EF121528
Morina longifolia Wallich ex DC.	Cult. Bergius Bot. Gard., Sweden; *Eriksson s.n., 2 Nov. 1999* (SBT)	EF121467	EF121498	EF121529
Nardostachys jatamansi (D. Don) DC.	*Boufford et al. 28099* (A)	EF121468	EF121499	EF121530
Patrinia triloba Miq.	Cult. Bergius Bot. Gard., Sweden; *Eriksson 807* (SBT)	EF121469	EF121500	EF121531
Plectritis congesta (Lindl.) DC.	USA, Oregon, Benton Co.; *Shenk #308* (YU)	EF121470	EF121501	EF121532
Pterocephalodes hookeri (C.B.Clarke) V.Mayer & Ehrend.	*Boufford et al. 28691* (A)	EF121471	EF121502	EF121533
Sambucus canadensis L.	Cult. Marsh Botanical Garden, Yale Univ.; *Donoghue & Winkworth 37* (YU, A)	EF121472	EF121503	EF121534

(*continued on next page*)

Appendix A (*continued*)

Taxon	Collection/Voucher	GenBank Accession No.		
		cox1	*cox3*	*nad5*/4-5 intron
Scabiosa columbaria L.	*Bell 199* (YU)	EF121473	EF121504	EF121535
Sinadoxa corydalifolia C. Y. Wu, Z. L. Wu & R. F. Huang	*Boufford et al. 26555* (A)	EF121474	EF121505	EF121536
Symphoricarpos sp.	Cult. Arnold Arboretum; *Donoghue & Winkworth 28* (YU, A)	EF121475	EF121506	EF121537
Tetradoxa omeiensis (H. Hara) C. Y. Wu	*Donoghue et al. 4000* (A)	EF121476	EF121507	EF121538
Triosteum aurantiacum Bickn.	*Walters et al. s.n.* (BHO)	EF121477	EF121508	–
Triplostegia glandulifera Wall. ex DC.	*Boufford et al. 28440* (A)	EF121478	EF121509	EF121539
Valeriana celtica L.	*Bell SWITZ002* (YU)	EF121479	EF121510	EF121540
Valeriana supina Ard.	*Nyffeler 1076* (YU)	EF121480	EF121511	EF121541
Valerianella locusta (L.) Latarrade	*Patterson 2001* (SFSU)	EF121481	EF121512	EF121542
Viburnum dentatum L.	Cult. Arnold Arboretum 5070-1-A; *Donoghue & Winkworth 33* (YU, A)	EF121482	EF121513	EF121543
Weigela hortensia (Seib. & Zuck.) C. A. Mey.	Cult. Arnold Arboretum 1897-77-A; *Kelly and Buckland 28* (A)	EF121483	EF121514	EF121544

References

Airy Shaw, H.K., 1952. A second species of the genus *Heptacodium* Rehd. (Caprifoliaceae). Kew Bull. 7, 245–246.

Akaike, H., 1974. A new look at statistical model identification. IEEE Trans. Autom. Control. 19, 716–723.

Backlund, A., Bremer, B., 1997. Phylogeny of the Asteridae *s. str.* based on *rbcL* sequences, with particular reference to the Dipsacales. Plant Syst. Evol. 207, 225–254.

Backlund, A., Pyck, N., 1998. Diervillaceae and Linnaeaceae, two new families of caprifolioids. Taxon 47, 657–661.

Bell, C.D., 2001. PORN*: A Hierarchical Likelihood Ratio Calculator for LINUX. Yale University, New Haven, CT. Available from: <http://www.phylodiversity.net/cbell/pornstar>.

Bell, C.D., 2004. Preliminary phylogeny of Valerianaceae inferred from nuclear and chloroplast sequences DNA sequence data. Mol. Phylogenet. Evol. 31, 340–350.

Bell, C.D., 2007. Phylogenetic placement and biogeography of the North American species of *Valerianella* (Valerianaceae: Dipsacales) based on chloroplast and nuclear DNA. Mol. Phylogenet. Evol. 44, 929–941.

Bell, C.D., Donoghue, M.J., 2005a. Dating the Dipsacales: comparing models, genes, and evolutionary implications. Am. J. Bot. 92, 284–296.

Bell, C.D., Donoghue, M.J., 2005b. Phylogeny and biogeography of Valerianaceae (Dipsacales) with special reference to the South American valerians. Org. Evol. Div. 5, 147–159.

Bell, C.D., Edwards, E.J., Kim, S.-T., Donoghue, M.J., 2001. Dipsacales phylogeny based on chloroplast DNA sequences. Harv. Pap. Bot. 6, 481–499.

Benko-Iseppon, A.M., Morawetz, W., 2000. Viburnales: cytological features and a new circumscription. Taxon 49, 5–16.

Brandley, M.C., Schmitz, A., Reeder, T.W., 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of Scincid lizards. Syst. Biol. 54, 373–390.

Brown, J.M., Lemmon, A.R., 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Syst. Biol. 56, 643–655.

Buckley, T.R., 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. Syst. Biol. 52, 509–523.

Donoghue, M.J., Bell, C.D., Winkworth, R.C., 2003. The evolution of reproductive characters in Dipsacales. Int. J. Plant Sci. 164 (5 Suppl), S453–S464.

Donoghue, M.J., Eriksson, T., Reeves, P.A., Olmstead, R.G., 2001. Phylogeny and phylogenetic taxonomy of Dipsacales, with special reference to *Sinadoxa* and *Tetradoxa* (Adoxaceae). Harv. Pap. Bot. 6, 459–479.

Doyle, J.J., Doyle, J.L., 1990. Isolation of plant DNA from fresh tissue. Focus 12, 13–15.

Duminil, J., Pemonge, M.-H., Petit, R.J., 2002. A set of 35 consensus primer pairs amplifying genes and introns of plant mitochondrial DNA. Mol. Ecol. Notes 2, 428–430.

Fishbein, M., Soltis, D.E., 2004. Further resolution of the rapid radiation of Saxifragales (angiosperms, eudicots) supported by mixed-model Bayesian analysis. Syst. Bot. 29, 883–891.

Golubkova, V., 1965. De genere *Heptacodium* Rehd. e familia Caprifoliaceae. Nov. Syst. Pl. Vas. 2, 230–236.

Hara, H., 1983. A Revision of the Caprifoliaceae of Japan with Reference to Allied Plants in Other Districts and the Adoxaceae. Academia Scientific Books, Tokyo.

Hidalgo, O., Garnatje, T., Susanna, A., Mathez, J.L., 2004. Phylogeny of Valerianaceae based on *matK* and ITS markers, with reference to *matK* individual polymorphism. Ann. Bot. 93, 283–293.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

Kass, R.E., Wasserman, L., 1995. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. J. Am. Stat. Assoc. 90, 928–934.

Lartillot, N., Philippe, H., 2006. Computing Bayes factors using thermodynamic integration. Syst. Biol. 55, 195–207.

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.

Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldry, J.L., 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53, 47–67.

Pagel, M., Meade, A., 2004. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. Syst. Biol. 53, 561–581.

Palmer, J.D., Adams, K.L., Cho, Y., Parkinson, C.L., Qiu, Y.L., Song, K., 2000. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. Proc. Natl. Acad. Sci. USA 97, 6960–6966.

Posada, D., 2001. The effect of branch length variation on the selection of models of molecular evolution. J. Mol. Evol. 52, 434–444.

Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53, 793–808.

Posada, D., Crandall, K.A., 2001. Selecting the best-fit model of nucleotide substitution. Syst. Biol. 50, 580–601.

Pyck, N., Roels, P., Smets, E., 1999. Tribal relationships in Caprifoliaceae: evidence from a cladistic analysis using *ndhF* sequences. Syst. Geogr. Pl. 69, 145–159.

Pyck, N., Smets, E., 2000. A search for the phylogenetic position of the seven-son flower (*Heptacodium*, Dipsacales): combining molecular and morphological evidence. Plant Syst. Evol. 225, 185–199.

Reeder, T.W., 2003. A phylogeny of the Australian *Sphenomorphus* group (Scincidae: Squamata) and the phylogenetic placement of the crocodile skinks (Tribolonotus): Bayesian approaches to assessing congruence and obtaining confidence in maximum likelihood inferred relationships. Mol. Phylogenet. Evol. 27, 384–397.

Rehder, A., 1916. Caprifoliaceae. In: Sargent, C.S. (Ed.), Plantae Wilsonianae 2. Arnold Arboretum, Boston, pp. 617–619.

Ronquist, F., Huelsenbeck, J.P., 2003. Mr.Bayes3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Schwarz, G., 1974. Estimating the dimensions of a model. Ann. Stat. 6, 461–464.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of loglikelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Sang, T., Crawford, D.J., Stuessy, T.F., 2007. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). Am. J. Bot. 84, 1120–1136.

Sokal, R.R., Rohlf, F.J., 1995. Biometry: The Principles and Practice of Statistics in Biological Research, third ed. W.H. Freeman and Co., New York, NY.

Soltis, D.E., Soltis, P.S., 2004. *Amborella* not a 'basal angiosperm'? Not so fast. Am. J. Bot. 91, 997–1001.

Sullivan, J., Joyce, P., 2005. Model selection in phylogenetics. Ann. Rev. Ecol. Evol. Syst. 36, 445–466.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis using Parsimony (* and Other Methods), version 4. Sinauer Associates, Sunderland, MA.

Tang, Y.-C., Li, L.-Q., 1994. The phytogeography of Caprifoliaceae s.str. with its implications for understanding eastern asiatic flora. Acta Phytotax. Sinica 32, 197–218.

Templeton, A.R., 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution 37, 221–244.

Weberling, F., 1966. Zur Systematischen Stellung der Gattung *Heptacodium* Rehd. Bot. Jahrb. 85, 253–258.

Whitfield, J.B., Lockhart, P.L., 2007. Deciphering ancient rapid radiations. TREE 22, 258–266.

Wolfe, K.H., Li, W.-H., Sharp, P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA 84, 9054–9058.

Zhang, W.-H., Chen, Z.-D., Li, J., Chen, H.-B., Tang, Y.-C., 2003. Phylogeny of the Dipsacales s.l. based on chloroplast *trn*L-F and *ndh*F sequences. Mol. Phylogenet. Evol. 26, 176–189.

Zhang, Z.-Y., Zhou, Z.-K., Li, J., Gu, Z.-J., 2002. Karyomorphology of *Heptacodium* (Caprifoliaceae s.str.) and its phylogenetic implications. Taxon 51, 499–505.