Syst. Biol. 58(6):612–628, 2009

DOI:10.1093/sysbio/syp068
Advance Access publication on October 21, 2009

Reticulation, Data Combination, and Inferring Evolutionary History: An Example from Danthonioideae (Poaceae)

MICHAEL D. PIRIE[1,2,*], AELYS M.HUMPHREYS[1], NIGEL P. BARKER[3], AND H. PETER LINDER[1]

[1]*Institute for Systematic Botany, University of Zurich, CH-8008 Zurich, Switzerland;*
[2]*Present address: Department of Biochemistry, University of Stellenbosch, Private Bag X1, Matieland 7602, Stellenbosch, South Africa; and*
[3]*Molecular Ecology and Systematics Group, Department of Botany, Rhodes University, Grahamstown 6140, South Africa;*
**Correspondence to be sent to: Department of Biochemistry, University of Stellenbosch, Private Bag X1, Matieland 7602, Stellenbosch, South Africa; E-mail: mpirie@sun.ac.za.*

Abstract.—We explore the potential impact of conflicting gene trees on inferences of evolutionary history above the species level. When conflict between gene trees is discovered, it is common practice either to analyze the data separately or to combine the data having excluded the conflicting taxa or data partitions for those taxa (which are then recoded as missing). We demonstrate an alternative approach, which involves duplicating conflicting taxa in the matrix, such that each duplicate is represented by one partition only. This allows the combination of all available data in standard phylogenetic analyses, despite reticulations. We show how interpretation of contradictory gene trees can lead to conflicting inferences of both morphological evolution and biogeographic history, using the example of the pampas grasses, *Cortaderia*. The characteristic morphological syndrome of *Cortaderia* can be inferred as having arisen multiple times (chloroplast DNA [cpDNA]) or just once (nuclear ribosomal DNA [nrDNA]). The distributions of species of *Cortaderia* and related genera in Australia/New Guinea, New Zealand, and South America can be explained by few (nrDNA) or several (cpDNA) dispersals between the southern continents. These contradictions can be explained by past hybridization events, which have linked gains of complex morphologies with unrelated chloroplast lineages and have erased evidence of dispersals from the nuclear genome. Given the discrepancies between inferences based on the gene trees individually, we urge the use of approaches such as ours that take multiple gene trees into account. [Biogeography; *Cortaderia*; Danthonioideae; gynodioecy; hybridization; incongruence; molecular dating; phylogeny.]

When reconstructing phylogeny, some authors combine independent sources of data irrespective of conflict under the assumption that the combined analysis will maximize explanatory power of the phylogeny ("total evidence," Kluge 1989). This may be an appropriate approach for weakly supported conflict, which could be the result of sampling error (de Queiroz et al. 1995). However, differences between gene trees can be informative with respect to the evolutionary history of a group, for example, by revealing past hybridization events (Rieseberg and Brunsfield 1992; Rieseberg and Morefield 1995). Such differences cannot be represented in a single bifurcating topology (Rieseberg and Morefield 1995), and combined analysis of the data will, at best, discard information. Combined analysis is likely to lead to decreased, rather than increased, support for clades identified by one gene and contradicted by another (Bull et al. 1993; Lecointre and Deleporte 2005). In a worst-case scenario, combination may even result in the inference of spurious relationships supported by neither data partition individually (McDade 1992; Bull et al. 1993).

Most authors agree that data partitions with strongly supported conflicting phylogenetic signals should not be combined in phylogenetic analyses. However, when such conflict is encountered, it is treated in a number of different ways. The individual gene trees can be interpreted separately (e.g., Shaw 2002; Barber et al. 2007) or used to calculate a consensus tree or network (Huson et al. 2004; McBreen and Lockhart 2006). Phylogenetic networks are more precise representations of conflict than bifurcating consensus trees: A number of different network patterns could each be summarized as the same consensus tree in which reticulating lineages are collapsed to a common ancestral node. Neither approach reaps the benefit of the improved phylogenetic resolution that could result from combining the nonconflicting parts of the data (de Queiroz et al. 1995; Pisani and Wilkinson 2002). Where data are to be combined in a single analysis, conflicting taxa are usually excluded (Rodrigo et al. 1993; de Queiroz et al. 1995; Kellogg et al. 1996), or conflicting data partition(s) corresponding to such taxa are recoded as missing data (Lecointre and Deleporte 2005). The latter option might be chosen under the assumption that one or other partition is in some way misleading, for example, due to paralogy or due to an elevated substitution rate resulting in long branch attraction (Felsenstein 1978). When conflicting taxa have been removed from combined analyses, they can be reintroduced thereafter as reticulations (Funk 1985; Kellogg et al. 1996). Crucially, in both cases, part of the available data are excluded from the analyses. Combined analysis including all the taxa and data can be justified only if the method used does not assume a bifurcating tree (e.g., Dickerman 1998; Morrison 2005; McBreen and Lockhart 2006) or if it permits representation of taxa more than once in a tree. Such methods are not widely used in higher level systematics.

The importance of representing and correctly interpreting differences between gene trees is highlighted by the numerous phylogenetic studies, which have demonstrated gene tree conflict above the species level (e.g., Kellogg et al. 1996; Comes and Abbott 2001; Barber et al.

2007; Fehrer et al. 2007; Maureira-Butler et al. 2008; Pirie et al. 2008; reviewed for plants in Vriesendorp and Bakker 2005). Phylogenetic studies of plants often reveal conflict between chloroplast- and nuclear-encoded DNA markers, the latter frequently nuclear ribosomal internal transcribed spacer regions (ITS; e.g., Comes and Abbott 2001; Barber et al. 2007; Fehrer et al. 2007; Pirie et al. 2008). In some cases, one or other data partition may better reflect the taxonomy of a group (Barber et al. 2007) and may even be assumed to therefore represent the "true tree" (Fehrer et al. 2007). However, this attitude may be over simplistic in many cases, as we will demonstrate with an example from the grass subfamily Danthonioideae.

Danthonioideae comprises 282 currently recognized species distributed in temperate regions mostly limited to the southern continents. Increasingly, robust phylogenetic hypotheses for the subfamily have been proposed (Barker et al. 1999, 2000, 2003, 2007; Linder and Barker 2005; Pirie et al. 2008). In a recent study (Pirie et al. 2008), greatly increased sampling of DNA sequence data from the chloroplast genome (cpDNA) allowed the delimitation of a number of areas of well-supported conflict between relationships inferred using cpDNA and those inferred using sequences from ITS.

One area of conflict in the Danthonioideae phylogeny involves the pampas grasses, *Cortaderia*. The general appearance of species of *Cortaderia* is distinctive: They share a combination of characteristics, which appear to be functionally linked to life history traits such as robust longevity and gynodioecy (Barker et al. 2003). Although the former trait is common to different genera in Danthonioideae (Barker et al. 2003), gynodioecy is only found in *Cortaderia* and in one species of *Chionochloa* (*Chionochloa bromoides*; Connor 1990) and is uncommon in grasses as a whole (Connor 1973, 1981). Nevertheless, both cpDNA and nuclear ribosomal DNA (nrDNA) evidence suggest that *Cortaderia* is not monophyletic (Barker et al. 2003, 2007). The relationships between clades of *Cortaderia* and a number of related but morphologically dissimilar genera differ in part according to the 2 data partitions. These different hypotheses of relationship paint contrasting pictures of the evolution of the "*Cortaderia* morphological syndrome," ranging from paraphyly of *Cortaderia*, which might indicate a single gain, to polyphyly, which might indicate multiple gains. Moreover, the topological differences involve taxa distributed in Australia, New Guinea, New Zealand, and South America. Even superficial examination of the gene trees suggests that the biogeographical inferences made with these data might differ profoundly.

We explore different ways of extracting phylogenetic signal from DNA sequence data: We combine cpDNA and ITS sequence data using different methods, producing phylogenetic trees representing a greater or lesser proportion of the reticulation in the Danthonioideae phylogeny. We investigate the implications of the different resulting phylogenies using character optimizations (morphological and continental distribution) and molecular dating analyses, with *Cortaderia* and related genera as an example.

Materials and Methods

Taxon and Character Sampling

We selected a subset of the taxa and molecular markers used in a comprehensive study of the Danthonioideae phylogeny (Pirie et al. 2008). We kept taxon representation of the *Danthonia*/*Cortaderia* clade high but reduced sampling of the other Danthonioideae clades; retaining taxa representative of their major lineages and geographical distribution (see Appendix S1 available from http://www.sysbio.oxfordjournals.org./). This was done in order to reduce analysis time, allowing more detailed analyses of target clades than would otherwise have been possible. Three taxa representing the grass subfamily Chloridoideae (*Centropodia glauca*, *Merxmuellera papposa*, and *Merxmuellera rangei*) were selected as outgroups, following the results of Bouchenak-Khelladi et al. (2008), who showed Chloridoideae to be sister to Danthonioideae.

We used markers *trnL-F* (including the *trnL* intron and *trnL-F* spacer); the *rpl16* intron; and *rbcL*, *ndhF*, and *matK* genes (the latter including partial flanking spacer regions) from the chloroplast genome and the ITS regions of nrDNA. These regions were amplified and directly sequenced following protocols as described in Pirie et al. (2008) and previous studies from which part of the sequence data were obtained (Barker et al. 1995, 2000, 2003, 2007; Verboom et al. 2006; Galley and Linder 2007). Although there was no evidence of multiple copies, we cloned polymerase chain reaction (PCR) products from 5 samples representing clades with conflicting signals in order to test for the presence of minority copies of ITS. Products from 3 PCR reactions applying differing annealing temperatures (50 °C, 52 °C, and 54 °C) were pooled and cloned using TOPO TA cloning kit for sequencing (Invitrogen Corp., Carlsbad, CA). Up to 20 individual colonies per sample were sequenced in one direction using the PCR primers. Accessions details and Genbank accession numbers are presented in Appendix S1 and the matrix is available on TreeBASE (www.treebase.org; accession numbers S2474, M4713). The alignments of the different regions were adopted without change from Pirie et al. (2008), as were gap characters (indels), which had been coded using the simple indel coding method of Simmons and Ochoterena (2000) as implemented in the program SeqState (Müller 2006).

Identifying and Eliminating Conflict and Combining Data Partitions

Phylogenetic analyses followed a standard approach, first inferring gene trees based on individual markers, identifying conflict between different gene trees and addressing any such conflict before combining multiple markers in further analyses. The most widely used of the available methods for testing for incongruence

between data partitions, the incongruence length difference test (ILD; Mickevich and Farris 1981), can result in false positives for conflict (Sullivan 1996) and, in addition, only provides a measure of overall incongruence between partitions rather than identifying localized incongruence caused by specific taxa or clades. We therefore compared individual and combined gene topologies by hand, invoking conflict only when conflicting nodes were supported by at least 70% bootstrap support (BS) and 0.95 Bayesian posterior probability (PP). In contrast to the ILD, this method will only result in false identification of conflict in cases of systematic error and will pinpoint the taxa and clades causing the incongruence.

No conflict was observed between the individual cpDNA partitions (data not shown), and these were therefore combined in further parsimony and Bayesian analyses. Comparison of the cpDNA and ITS gene trees revealed a number of incidences of conflict. We addressed conflict between gene trees following 3 general principles, as follows (and summarized in Table 1):

1. The cpDNA and ITS data are in conflict and we therefore cannot justify combining them. The data are analyzed separately (Matrices 1 [only cpDNA] and 2 [only ITS]) and differences among gene trees resulting from these analyses can be summarized in a network.
2. Parts of the cpDNA and ITS data are in conflict: One or other part of the data are in some way misleading and must be removed before the data can be combined. This is achieved 1) by excluding taxa and clades with conflicting positions in the 2 gene trees (Matrix 3a) or 2) by excluding one or other of the data partitions (cpDNA or ITS) of conflicting taxa or clades by recoding it as missing data (Matrices 4 and 5, respectively).
3. Parts of the cpDNA and ITS data are in conflict, but all parts are informative because the conflicting elements represent different evolutionary histories that have been united in the same organisms through hybridization. Therefore, in principle, all the data should be included in analyses. This is achieved 1) by creating a matrix in which taxa with conflicting positions in the 2 gene trees are duplicated, with 1 taxon copy represented by the cpDNA sequences only (with the ITS partition coded as missing data), the other taxon copy represented only by the ITS sequences (with the cpDNA coded as missing data [Matrix 6]), resulting in a multilabeled phylogenetic tree (sensu; Huber et al. 2006) and 2) by combining trees resulting from Matrices 4 and 5, or summarizing the tree resulting from Matrix 6, in a network.

In the absence of contradictory evidence, we assume that the signals of the data partitions that are to be combined are congruent. Where 1 of the 2 partitions is less informative, as is the case in the present study, the test of congruence is weak. A further test can be applied under the assumption that the combination of congruent data partitions ought to result in improved resolution and support with respect to both individual trees. However, a number of the above methods of combining the data introduce missing data and increase the number of taxa in the analysis. Missing data can lead to loss of support when those taxa are sensitively placed (Wiens 2006) and an increased number of taxa will require more data overall to achieve the same levels of support (Bremer et al. 1999). It is therefore only possible to directly compare support values when the conflicting taxa are excluded. ITS and cpDNA matrices were therefore also analyzed separately with all conflicting taxa excluded (Matrices 3b and 3c, respectively) in order to compare support values with those of the combined analysis of Matrix 3a.

Phylogenetic Analysis

Parsimony.—Data were analyzed under parsimony using the software package PAUP* 4.0b10 (Swofford 2000), assuming unordered character state transformation (Fitch parsimony [FP]; Fitch 1971) and equal weights. Shortest trees were calculated using heuristic search of 1000 random addition sequence replicates (RAS), tree bisection and reconnection (TBR), saving a maximum of 10 trees each replicate. Support was estimated using bootstrap analyses of 1000 replicates with heuristic searches of 50 RAS, TBR, saving 10 trees each replicate. Lack of resolution between ITS sequences of South American *Cortaderia* has the potential to bias character optimizations in favor of gynodioecy and South American distribution at the node of the most recent common ancestor (MRCA) of *Cortaderia* and *Danthonia* (see below). In order to test the potential impact of such a bias on those results, we therefore repeated the parsimony analyses of Matrices 4 and 6, constraining

TABLE 1. Combinations of data partitions and taxa used for phylogenetic analyses

Matrix	Data	Treatment of conflicting taxa	Referred to as:
1	cpDNA	None	Gene trees
2	ITS	None	
3a	cpDNA + ITS	Excluded	Excluded tree
3b	cpDNA	Excluded	
3c	ITS	Excluded	
4	ITS (+cpDNA)	cpDNA recoded as missing data	Combined "gene" trees
5	cpDNA (+ITS)	ITS recoded as missing data	
6	ITS + cpDNA	Duplicated with (a) ITS and (b) cpDNA recoded as missing data	Duplicated tree

the ITS sequences of South American *Cortaderia* to form a monophyletic group; that is, a resolution of this uncertainty not contradicted by the supported nodes, which minimizes the chances of the South American *Cortaderia–Danthonia* node optimizing to these states.

Bayesian inference.—ModelTest 3.06 (Posada and Crandall 1998) was used to select the substitution model best fitting the each data partition under the Akaike Information Criterion, using an arbitrary most parsimonious tree topology. Matrices 1–3a and 4–6 were analyzed using Bayesian inference, as implemented in MrBayes version 3.1.2 (Huelsenbeck and Ronquist 2001). The data were partitioned according to the different markers used and both rates and substitution models were allowed to vary across the partitions. Priors for the number of parameters in the DNA substitution models were applied to each partition (as determined using ModelTest above). Prior probabilities for all topologies were assumed to be equal. Two or more independent Markov chain Monte Carlo (MCMC) analyses were set to run indefinitely with 4 simultaneous MCMC chains. One tree per 1000 generations was saved. We checked for convergence during the runs by comparing the mean likelihoods (minus burn-in) and estimated sample sizes (ESS) using Tracer v.1.3 (Rambaut and Drummond 2003) and by visualizing topological convergence within and between runs using "Are We There Yet" (Wilgenbusch et al. 2004). The runs were terminated when both had converged to the same mean likelihood with ESS $>$ 100, with consistent clade PPs. The burn-in values were determined empirically from the likelihood values and 50% majority rule consensus trees calculated together with approximations of the PP for the observed bipartitions.

Rate-smoothing and Coalescent Simulations

We performed rate-smoothing analyses 1) to compare age estimates derived from the different gene trees and 2) to obtain trees in which branch lengths are proportional to time for use in coalescent simulations. The latter were performed to test under which conditions differences between the gene trees could be explained by lineage sorting.

The sequence data and best fitting substitution model were used to calculate branch lengths over an arbitrarily chosen most parsimonious tree (the first saved in each parsimony analysis) from each of the analyses of Matrices 3–6 and the topology-constrained analyses of Matrices 4 and 6, using the maximum-likelihood (ML) criterion as implemented in PAUP*. The likelihood-ratio test resulted in rejection of the molecular clock ($P \leq 0.0001$); therefore, the trees were rate smoothed using penalized likelihood (Sanderson 2002a) as implemented in r8s (Sanderson 2002b) with the root node fixed to an age of 26.1 million years (Christin et al. 2008). Error margins were estimated using 100 topology-constrained trees with branch lengths derived from bootstrap resampled data (Wikström et al. 2001).

The rate-smoothed trees derived from analyses of Matrices 4 and 5 were used as competing hypothetical "species trees" to simulate samples of 1000 gene trees using the "Coalescence Contained within Current Tree" module of Mesquite version 2.5 (Maddison and Maddison 2006). We assumed a generation time of 1 year, panmixis, and constant effective population size (N_e), each of which assumptions are likely to be violated (see Discussion section). Higher N_e results in longer coalescence times and thus potentially larger differences between gene trees. Thus, for a range of values for N_e increasing from 10,000 to 1,000,000, we then compared the difference between the 2 species trees to the distribution of differences between each species tree and its corresponding simulated gene trees, as calculated using the partition metric (Penny and Hendy 1985; implemented in PAUP* as the symmetric distance). Following the approach of Maureira-Butler et al. (2008), we concluded that lineage sorting alone is unlikely to explain the difference between the 2 hypothetical species trees when this difference is higher than 95% of the difference simulated under coalescence.

Hybridization Network Construction

A large number of network methods are reviewed in Morrison (2005) and in McBreen and Lockhart (2006), but few are yet capable of building networks by directly combining DNA sequence data corresponding to different gene trees. RETICLAD (Rieseberg and Morefield 1995) was developed to identify hybrids based on the expectation that they will combine the characters of their parents. It, however, only tests reticulation events between terminal branches and is therefore inappropriate for a number of the examples presented here. Dickerman (1998) proposed a parsimony-based analysis method that seeks to minimize both homoplasy and reticulation in constructing networks. Heuristic search of the combined data scores networks according to a combined cost reflecting tree length and numbers of reticulations. There is, however, no objective way of setting the relative costs of these events, and thus, no way of determining the optimality criterion itself.

The most widely used methods for inferring phylogenetic networks take gene trees from phylogenetic analyses as input rather than working with the primary data itself (McBreen and Lockhart 2006). Networks can be constructed from a sample of trees (such as the shortest trees from a parsimony analysis), or from consensus trees, in which nodes with support below a certain level can be collapsed. Applying a threshold for support of nodes provides a natural means to distinguish incongruence attributable to hybridization from that resulting from stochastic error (McBreen and Lockhart 2006). We adopted this approach in order to represent as networks the differences between gene trees as revealed by the different phylogenetic approaches used here. Consensus trees resulting from analyses of 1) Matrices 1 and 2 (gene trees) and 2) Matrices 4 and 5 ("combined" gene trees) were edited to collapse nodes with <70% BS. These trees

were used as input for Splitstree 4.8 (Huson and Bryant 2006), and 2 hybridization networks were computed under method "RECOMB2005."

Character Optimization

In order to illustrate the differences in inferences of evolutionary history that might be made given different representations of conflicting gene trees, we used FP and ML methods to optimize 2 characters (continental distribution and gynodioecy) over the topologies resulting from analyses of Matrices 3–6 (those resulting from analyses of Matrices 1 and 2 were not used, as these were less resolved). Continental distribution was coded as a single multistate character as follows: 0 = Africa, 1 = Australia plus New Guinea, 2 = New Zealand, and 3 = the Americas and Europe. There are only 2 European species of Danthonioideae: *Danthonia alpina* and *Danthonia decumbens*. Both are nested within the otherwise exclusively New World *Danthonia* clade. We treated Europe and the Americas as one area because our focus here was on *Cortaderia* for which the distributions of these 2 species were largely irrelevant. According to the age estimation of Christin et al. (2008), the Danthonioideae evolved within the last 26.1 million years. We therefore interpreted changes in continental distribution as dispersals, never as vicariance, and treated ambiguity as ambiguity rather than as widespread distribution. We did not attempt to model extinction. Presence versus absence of gynodioecy was used as a marker for the *Cortaderia* morphological syndrome (discussed below).

FP optimization was performed using Mesquite version 2.5 (Maddison and Maddison 2006). Ancestral states for the 2 characters were summarized from optimizations over all the most parsimonious trees found for each of the 4 unconstrained analyses of the combined data (Matrices 3–6) plus those of the 2 topology constrained analyses (Matrices 4 and 6). ML optimizations were performed with the rate-smoothed phylograms (estimated using r8s, as above, based on Matrices 3–6 and the 2 topology constrained analyses of Matrices 4 and 6) using the ML (multistate) mode of Bayestraits (Pagel and Meade 2006). Likelihoods of optimizations at given nodes were considered significant when $\geqslant$0.95. The significance of differences in rates of character state change were assessed using likelihood-ratio tests, comparing twice the difference in log likelihood between unrestricted versus restricted models (estimated using the "restrict" command in Bayestraits) to χ^2 distribution with degrees of freedom equal to the difference in the number of parameters in the model.

RESULTS

Phylogeny Reconstruction: Parsimony

Separate analyses of the cpDNA markers resulted in congruent gene trees (results not shown) and combining these markers (Matrix 1: in total, ca. 5000 sequenced bases providing 883 parsimony informative characters) resulted in increased resolution and support. Analysis of the shorter ITS region (Matrix 2: ca. 700 sequenced bases; 153 parsimony informative characters) resulted in a less resolved tree including a number of nodes which conflicted with the Matrix 1 tree. BS for clades from analysis of Matrices 1 and 2 are represented in Figure 1, with conflicting nodes subject to $\geqslant$70% BS indicated. Of the 5 samples for which ITS was cloned, not all resulting sequences were identical, but with one exception they all exhibited the same phylogenetic signal. A phylogram with BS for nodes is presented in Appendix S2. The single exception concerns 3 of the clones of the sample of *D. alpina* (accession MP480), which rather than forming a clade with the only other European species, *D. decumbens*, as do the other ITS copies, instead follow the cpDNA result, forming a clade with *D. alpina* accession MP481. The presence of both copy types in a single individual could indicate recent hybridization with incomplete homogenization (as might be suspected from the intermediate morphology of this specimen and its cooccurrence with both putative parents) or the inheritance of multiple copies from a common ancestor of *D. alpina* and *D. decumbens*. We did not follow this up with further cloning of other accessions of *D. alpina* and *D. decumbens* as the distinction in this case would be of no consequence for our further inferences with regard to *Cortaderia*.

On deletion of conflicting taxa and combined analysis of the cpDNA and ITS partitions (Matrix 3a) BS support values for 7 nodes increased by $\geq$10 relative to the combined cpDNA results, whereas support for 4 nodes (of which 3 were within the *Rytidosperma* clade) decreased by $\geq$10 (see Fig. 2). The combined tree included numerous well-supported nodes that were consistent with the less resolved ITS tree but which were unsupported in the tree based on ITS alone. Support for 8 nodes subject to $\geq$50% BS (2 $\geq$70% BS) in the ITS analysis increased by $\geq$10 compared with only 2 which decreased by $\geq$10. Only 1 well-supported node in the ITS tree was not recovered in the combined analysis (indicated in Fig. 1).

Results of analyses of Matrices 4 and 5 (cpDNA and ITS partitions of conflicting taxa, respectively, recoded as missing data) are represented in the network in Figure 3 (see below). Results of analysis of Matrix 6 (duplication of the conflicting taxa, with the same recoding of conflicting partitions) are represented in the chronogram in Figure 4. The 2 different positions of most of the conflicting taxa received strong support. The single exception was in the relationships between accessions of European *D. alpina* and *D. decumbens*, support for which was no longer significant according to analyses of Matrices 4 and 6.

Phylogeny Reconstruction: Bayesian Inference

Results from Bayesian phylogenetic inference were consistent with those obtained using parsimony. PPs for clades are summarized on the trees in Figure 1 (Matrices 1 and 2), 2 (Matrices 3a–c), and 4 (Matrix 6)

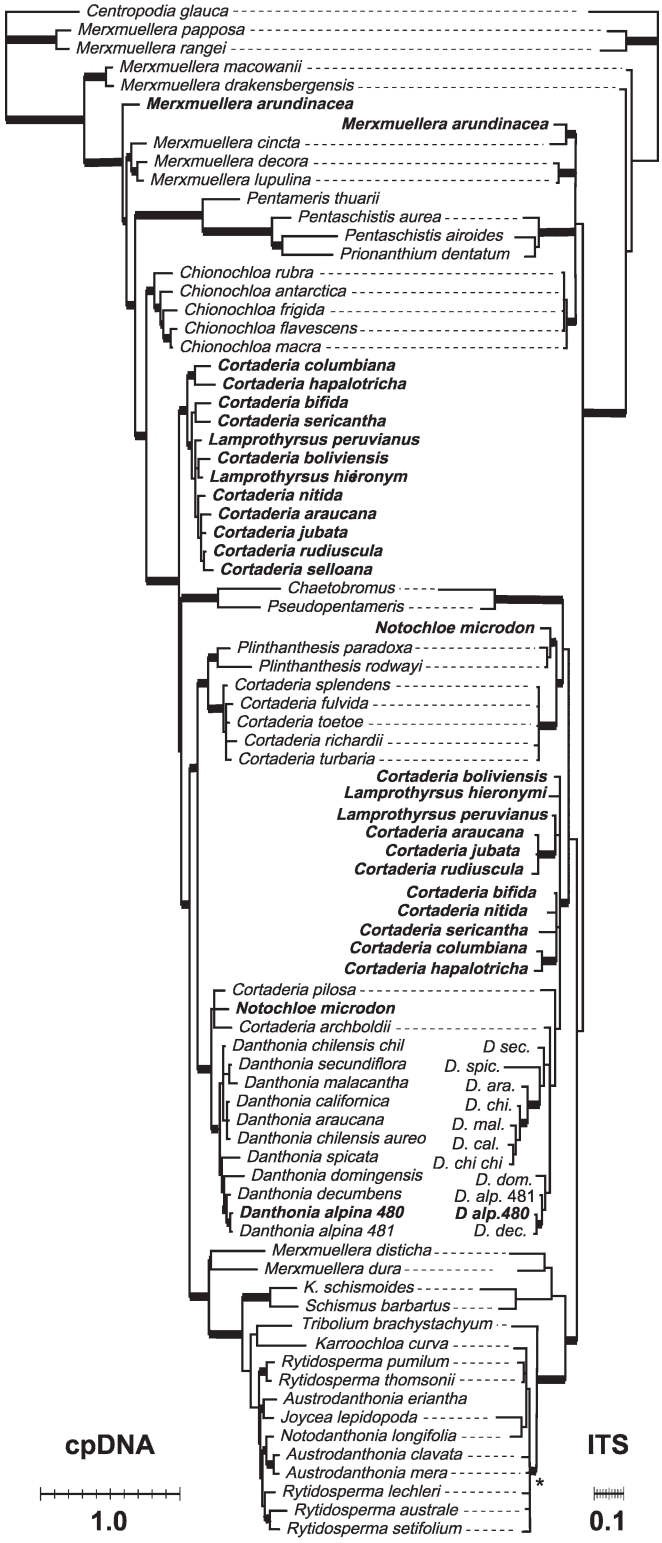

FIGURE 1. Fifty percent majority rule consensus trees with branch lengths from MrBayes (Huelsenbeck and Ronquist 2001) analyses of cpDNA (left) and of ITS (right) sequence data. Scale bars indicate expected substitutions per site. Nodes $\geq$70% BS (parsimony analysis) and 0.95 PPs (Bayesian analysis) are represented with thick lines. The names of taxa with conflicting positions are in bold type. One node subject to $\geqslant$70% BS (indicated by an asterisk) was not recovered in the combined analysis.

and presented in detail in Appendix S3 (Matrices 1–6). The different combination of the data in analyses 1–6 required differing lengths of analyses to reach convergence, with those including more data and greater proportions of missing data requiring generally longer runs with longer burn-in periods. Runs of Matrices 1 and 2 converged within 10 and 5 million generations, respectively, with burn-in of less than 0.5 million; those of Matrix 3 required up to 9.2 million generations and burn-in of up to 6.5 million; Matrices 4 and 5 required up to 14.6 million generations with burn-in of up to 4 million. Matrix 6, which included the greatest proportion of missing data, was particularly slow to converge, with some runs failing to converge within up to 40 million generations. Trees were finally sampled from 2 runs of 30 and 40 million generations with burn-in of up to 15 million.

Rate-Smoothing and Coalescent Simulations

In Table 2, relative age estimations with confidence limits for specified clades are reported, as derived from the r8s (Sanderson 2002b) analyses. Age estimations based on different data partitions are comparable only when the topologies are consistent. For example, the estimate for the stem node of *Notochloe* can be compared across all cpDNA topologies and all ITS topologies but not between the two. It may also be inappropriate to directly compare ITS and cpDNA estimations of the South American *Cortaderia* crown group, as the monophyly of the ITS sequences is not certain. Directly comparable nodes received mostly consistent age estimations, within the bounds of 95% error margins, with the exception of a number of estimates based on ITS only, compared with some based on the combined data (as indicated in Table 2).

The symmetric difference between "species" trees resulting from the analysis of Matrices 4 and 5 was 29, which was greater than 95% of the differences between gene trees simulated assuming an N_e of up to 500,000. At this value of N_e, the 95% distribution of the differences between the gene trees simulated for Matrix 4 was still lower than 29 (upper bound 27), but that of Matrix 5 was higher (upper bound 30). Thus, at $N_e = 500,000$ and above, the observed differences between the hypothetical species trees could be explained by coalescence alone. Detailed results are reported in Appendix S4 .

Network Construction

In Figure 3, the 70% BS consensus trees resulting from parsimony analysis of 1) Matrices 1 and 2 and 2) Matrices 4 and 5 are represented in hybridization networks. In both cases, unsupported clades have been collapsed, thus reticulations in the networks represent conflict rather than uncertainty, and these are interpreted explicitly as hybridizations (Huson and Bryant 2006). The placement of *Notochloe microdon* differs according to the 2 networks. This is due to the way in

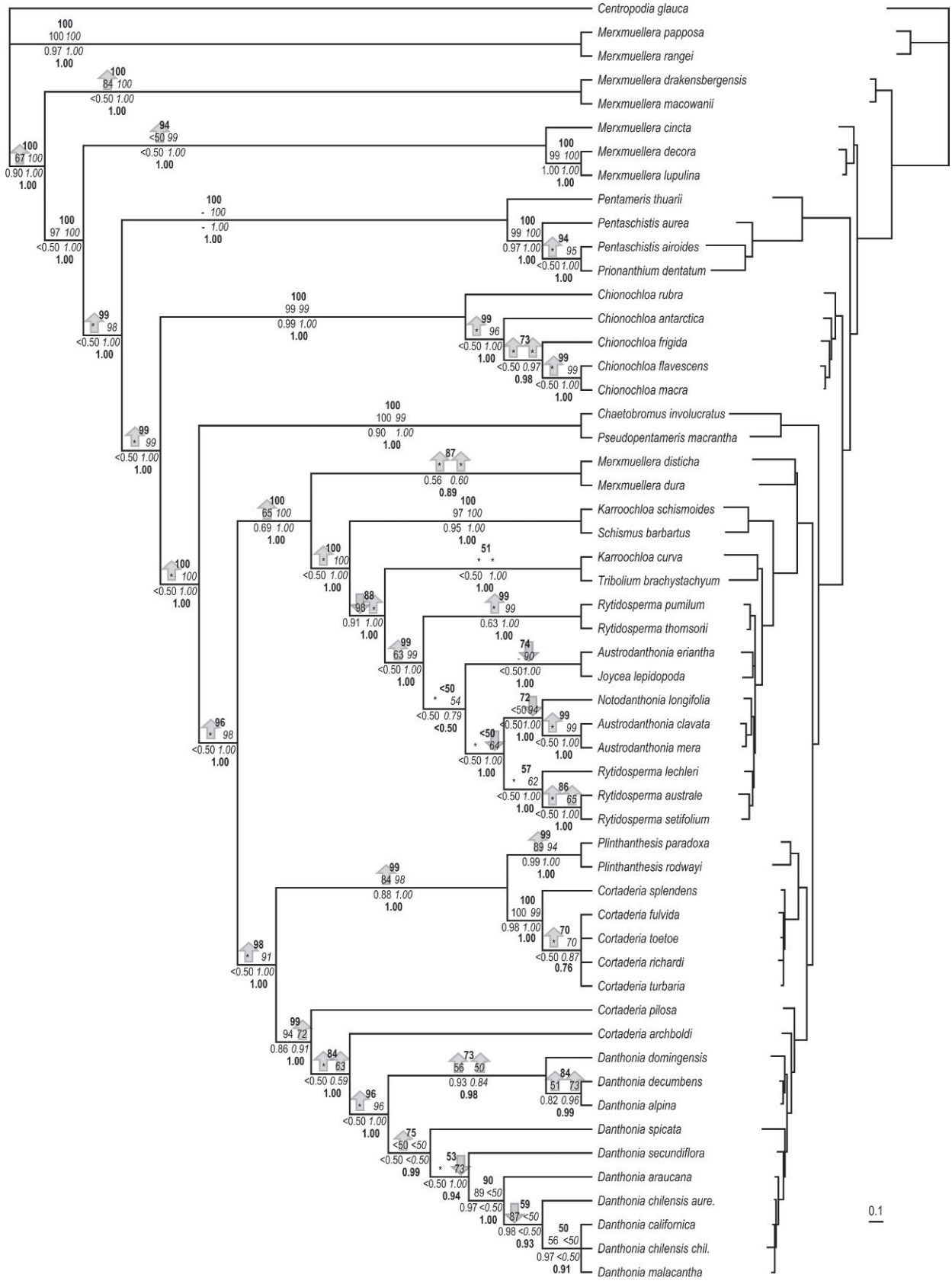

FIGURE 2. (Continued)

which the method treats the greater uncertainty in the gene trees compared with the corresponding combined "gene" trees (i.e., comparing Matrix 1 with Matrix 4 and Matrix 2 with Matrix 5). Only by combining the data do we obtain support for the sister group relationship between the chloroplast lineage of *N. microdon* and *Cortaderia pilosa*, resulting from the increase in support for relationships between related clades. Where there are polytomies in either tree, the hybridization network represents clades found in one and not contradicted by the other (equivalent to a semistrict consensus). This also means that the South American *Cortaderia* clade is represented as monophyletic, although its monophyly is only supported by cpDNA.

Character Optimizations

Ancestral state reconstructions (geographical distribution and gynodioecy, respectively) for a number of focal nodes according to topology (Matrices 3a, 4–6) and method are summarized in Appendix S5 and represented in Figures 5 and 6 as 4 diagrammatic trees showing *Cortaderia* and related genera. We focus on dispersals between Australia/New Guinea and the Americas and on gains and losses of gynodioecy, as inferred changes between these character states differed according to the different topologies used. Inferred numbers of state changes (parsimony optimization) are reported with corresponding estimations of rates of change (ML optimization). A single additional Australia/New Guinea to the Americas dispersal (within the *Rytidosperma* clade) is not represented in the diagrammatic trees in Figure 5 but is counted in the parsimony events in order to retain comparability with ML rates.

In general, the optimization results for Matrix 3 (conflicting taxa excluded) were largely ambiguous and those for Matrices 4 (conflicting taxa placed according to ITS), 6 (conflicting taxa duplicated), and especially 5 (conflicting taxa placed according to cpDNA) were more decisive. Where monophyly of the ITS South American *Cortaderia* clade was imposed in Matrices 4 and 6, the results for Matrix 4 remained the same (both parsimony and ML optimizations of both characters), but for Matrix 6, greater ambiguity appeared in most of the optimizations for the directly subtending node (indicated with an asterisk in Figs. 5 and 6). For a given topology, ML optimizations were largely congruent with parsimony optimizations. A number of inconsistencies were observed across different topologies, associated with nodes subtending the *Danthonia* clade, the ITS copies of South American *Cortaderia*, and cpDNA of *Notochloe*.

Parsimony optimization of the Matrix 5 topology suggests that the number of dispersals from the Americas to Australia/New Guinea (3 ➔ 1) and numbers of losses of gynodioecy (1 ➔ 0) are zero. In contrast, the results for Matrices 4 and 6 suggest roughly equal numbers of dispersals between the Americas and Australia/New Guinea and gains and losses of gynodioecy. Differences according to the different topologies were also apparent in the rates of these character state changes as estimated using ML. However, according to the results of likelihood-ratio tests, the difference in overall likelihood between models in which rates were restricted to be equal and those in which rates were unrestricted was not significant ($P > 0.05$) in any of the cases.

DISCUSSION

Character Conflict and Phylogeny Reconstruction

Gene trees based on DNA sequences can be compared to identify hybrid taxa and their parents (Rieseberg and Brunsfield 1992) in circumstances where morphology may display intermediate states (Seehausen 2004) and thus be harder to interpret (McDade 1992; Bull et al. 1993). However, if DNA sequence data containing conflicting signals are combined, misleading patterns can be observed. With the benefit of a robust chloroplast phylogeny of Danthonioideae, we can recognize that in previous studies, some taxa and clades have been placed in combined analyses either according to the signal present in cpDNA (e.g., *Notochloe* sister to *Danthonia*; Fig. 3 in Barker et al. 2000) or according to the signal present in ITS (*Notochloe* sister to *Plinthanthesis*; Barker et al. 2003) or in a unique position not supported by either data set individually (*Notochloe* sister to New Zealand *Cortaderia* and *Plinthanthesis*; Linder and Barker 2005; Barker et al. 2007). Identification and removal of conflicting taxa and subsequent simultaneous analysis of the remaining data here resulted in improved overall resolution and support. A drawback of this approach, however, is that it does not address the conflicting elements of the data.

Our first consideration should be whether conflicting signals of different markers are biologically meaningful. Paralogous copies of both nuclear- and plasmid-encoded markers can mislead phylogenetic inference (Alvarez and Wendel 2003; Pirie et al. 2007). Such problems are more likely within groups such as *Cortaderia*, which include polyploids with chromosome counts of $2n = 90$ (the New Zealand species) and $2n = 36$, through $2n = 72$ to $2n = 108$ (the South American species; Barker

FIGURE 2. Fifty percent majority rule consensus tree (left) and phylogram (right) from MrBayes analyses of combined cpDNA and ITS data with conflicting taxa excluded. The scale bar indicates expected substitutions per site. Values above the branches represent bootstrap proportions and values below the branches are Bayesian PPs: in bold for the combined analysis; to the left for ITS only, to the right (in italics) for cpDNA only. Where combination of the data resulted in the increase or decrease of BS values for clades by $\geq$10, these values are indicated by upward- or downward-pointing arrows, respectively. Nodes that break down in the parsimony strict consensus are represented by an asterisk instead of a bootstrap proportion.

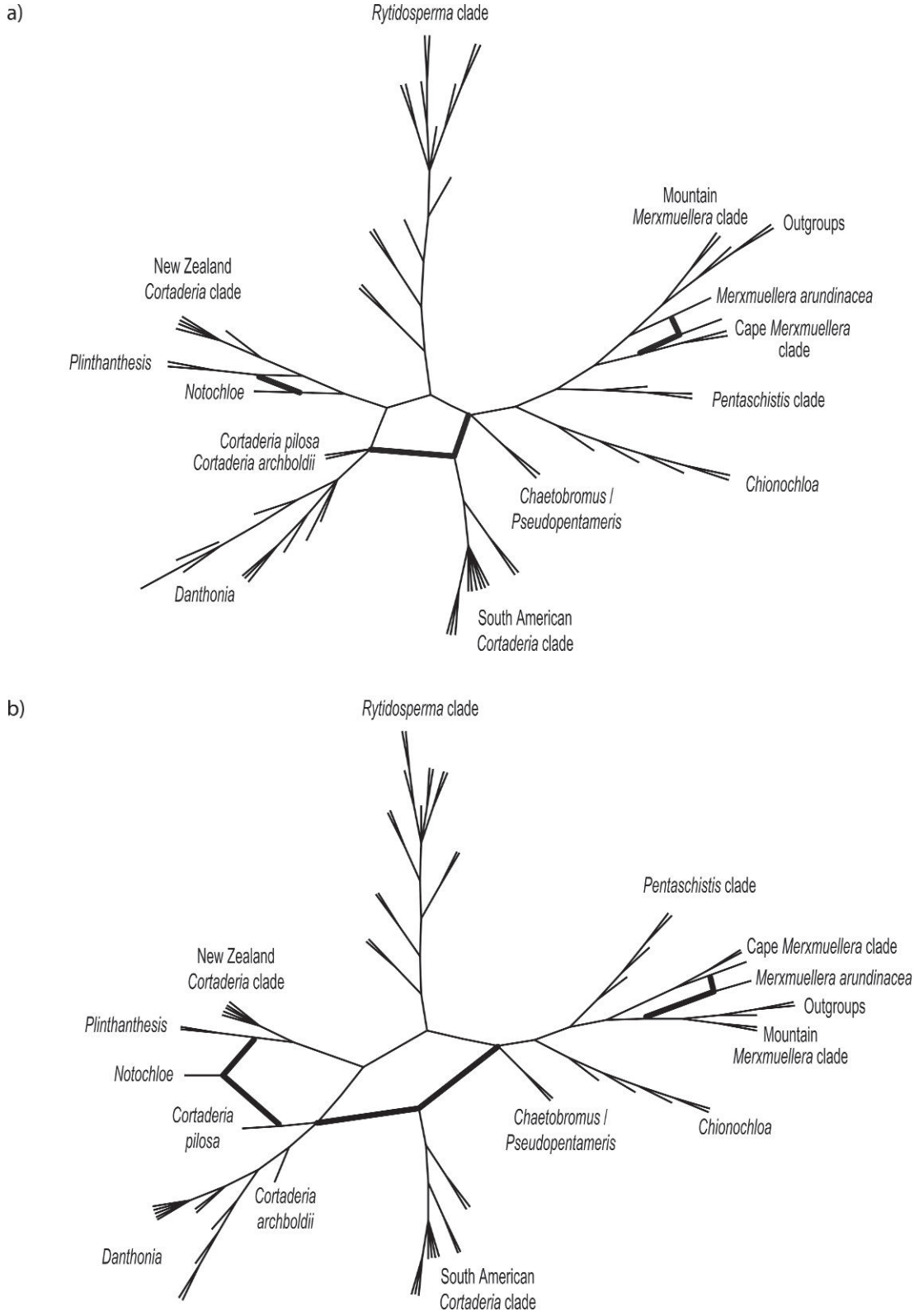

FIGURE 3. Hybridization networks generated from 70% bootstrap consensus trees of parsimony analyses of a) Matrices 1 and 2 and b) Matrices 4 and 5 using Splitstree version 4.8 (Huson et al. 2004). Hybridizations are subtended by thicker branches. Note the apparent monophyly of the South American *Cortaderia* clade, which is supported by cpDNA only.

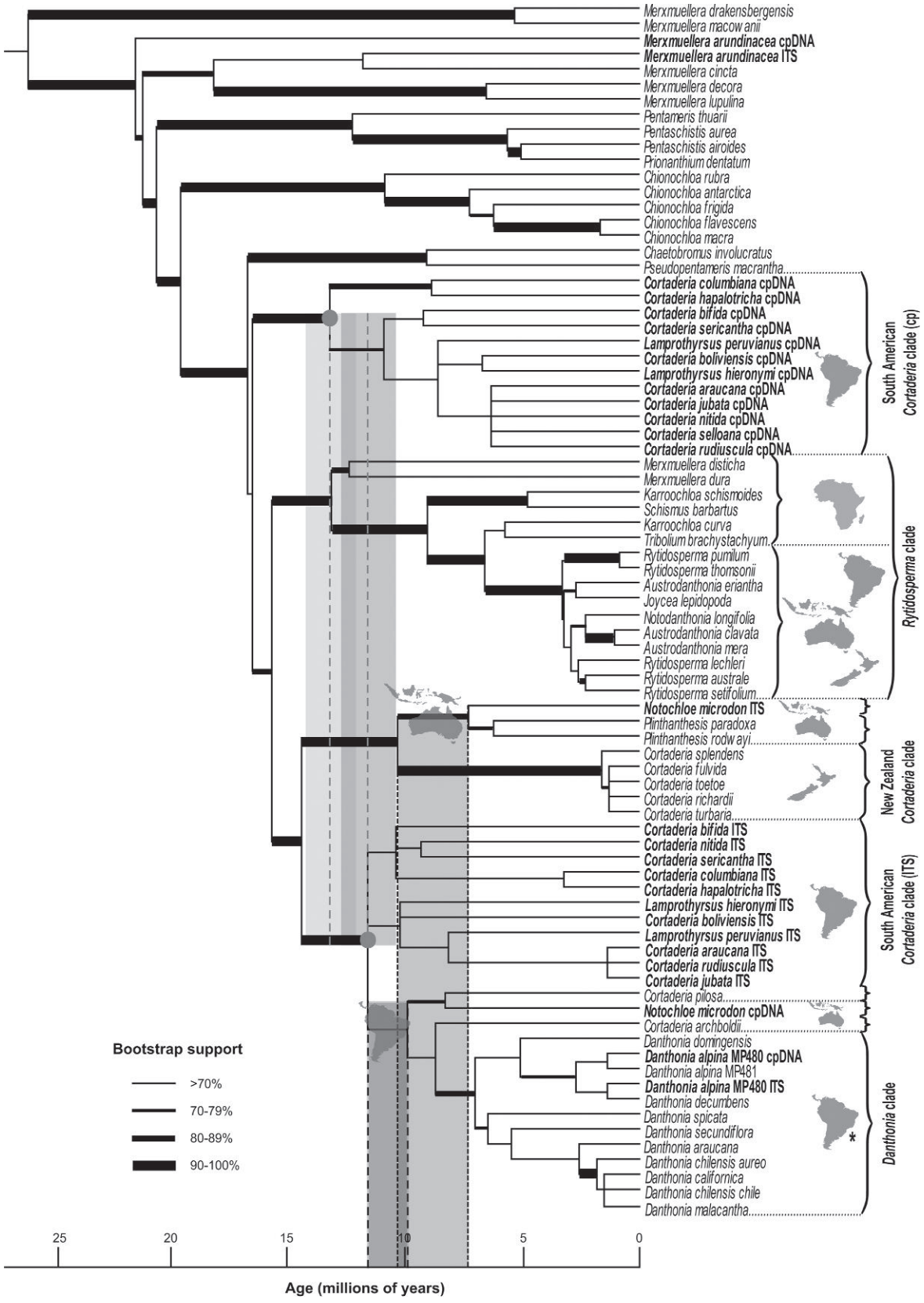

FIGURE 4. (Continued)

et al. 2003). Systematic error (e.g., long branch attraction; Felsenstein 1978) can result in strongly supported, but equally uninformative conflict. If a given marker contains misleading phylogenetic signal with respect to specific taxa, then the misleading signal can be selectively excluded from analyses by recoding the marker as missing data for the taxa in question (Matrices 4 and 5; Lecointre and Deleporte 2005). In the case presented here, however, we have no reason to distrust either data partition. Results obtained from ITS sequences were confirmed by Pirie et al. (2008), who sequenced multiple accessions of conflicting taxa and part of the 26S gene, and, in all but the case of European *Danthonia*, by the results of cloning experiments performed for this study. All the chloroplast markers supported the same conflicting topology, and both gene trees were retrieved consistently irrespective of analysis method.

Most of the examples of incongruence presented here involve branches that are distant in the tree, thus the pattern could be explained as resulting either from ancient hybridization or from deep lineage sorting (coalescent stochasticity). These 2 processes are difficult to discern (Comes and Abbott 2001), as illustrated by the simulation experiments performed here. We are able to infer conditions under which the observed differences between gene trees could have resulted from coalescent stochasticity (constant effective population size of 500,000 or higher, given a generation time of 1 year and panmixis), but these conditions are both difficult to test and are likely to have been violated. In the case of *Cortaderia*, such factors appear to be intimately entwined in the history of the group: Populations are likely to have undergone bottlenecks, in particular following long distance dispersal events, and the evolution of the *Cortaderia* syndrome will have influenced both generation time (robust longevity) and gene flow (gynodioecy). Although it is difficult to rule out lineage sorting in such cases (Comes and Abbott 2001; Maureira-Butler et al. 2008), hybridization is a common phenomenon in diverse plant and animal groups (Grant and Grant 1992; Ellstrand et al. 1996) and is particularly common in grasses (e.g. Kellogg et al. 1996; Connor 2004). Hybridization between native New Zealand *Cortaderia* species and between some South American species is known to produce viable progeny (Connor 2004) and crosses between New Zealand and South American species of *Cortaderia* are also possible, though producing sterile F1 hybrids (Connor 2004).

For cases of hybridization, the advantages of the taxon duplication method are considerable: It allows simultaneous analysis of all relevant data and results in a bifurcating tree, which can be used as the basis for inferences using standard techniques (such as ancestral area optimizations and molecular dating; see below). Our example is one of conflict between 2 genomes, but in principle, the method could be applied to incidences of significant incongruence between multiple data partitions, irrespective of their position in the genome, as long as incongruence is due to recombination between, not within, partitions, and is not the result of paralogy, or systematic error. The concept of duplicating hybrid taxa as a means of representing them in a bifurcating tree is not new (see page 87, figs. 7 and 8 in Nelson and Platnick 1980; Huber et al. 2006), but this method of analysis does not appear to have been used prior to our work here and in Pirie et al. (2008) to represent conflict between different markers. This may be due to the reticence with which, until recently, systematists have introduced missing entries into data matrices. However, Wiens (2003) and Wiens et al. (2005) demonstrated that missing data are not necessarily problematic for phylogenetic inference and that adding taxa represented by a small proportion of the data can in fact improve phylogenetic accuracy. We found that parsimony analyses of matrices with large proportions of missing data were unproblematic but that long runs were necessary to achieve convergence using the Bayesian method. This conclusion was also drawn by Wiens et al. (2005). However, Wiens et al. (2005) and Flynn et al. (2005) also reported that Bayesian analysis provided greater resolution and more robust support than parsimony in the presence of missing data.

Implications for Inferring Evolutionary History

We use the example of *Cortaderia* to illustrate the importance of using all the information available in differing gene trees when inferring evolutionary history. We found significant discrepancies between inferences based on phylogenies representing different proportions of the reticulations in the Danthonioideae phylogeny.

Implications for biogeography.—Sanmartin et al. (2007) used phylogenies of multiple plant groups distributed across the Southern Hemisphere to test the "West Wind Drift" (WWD) hypothesis: that floristic similarities between Australia, New Zealand, and South America are the result of concerted long-distance dispersals, the direction of which have been constrained by prevailing westerly winds or ocean currents. In other words, that dispersal from Australia/New Guinea to the Americas is more frequent than the other way round. The

FIGURE 4 (previous page). A rate-smoothed tree from parsimony analysis of Matrix 6 (conflicting taxa duplicated in the matrix). Categories of BS are indicated by branch thickness. The names of duplicated terminals are in bold type suffixed with either "cpDNA" or "ITS." The ages of the cpDNA and ITS crown nodes of the South American *Cortaderia* clade are indicated by dashed lines with confidence intervals represented by shaded boxes. Furthermore, shaded boxes represent the time frames within which an ancestor of *Notochloe* dispersed from South America to Australia. Clades represented in Figures 5 and 6 are indicated with parentheses, with distributions of the taxa indicated with maps (the asterisk indicates that the *Danthonia* clade is distributed into North America as well as South America, with 2 species also in Europe).

TABLE 2. Age estimations (in millions of years) for selected nodes with standard deviation calculated using the bootstrapping procedure described in the text, based on cpDNA, ITS, and 3 variants of the combined data (Matrices 1–2 and 4–6, as described in Table 1)

Node	Data	Matrix 2	Matrix 4	Matrix 6	Matrix 5	Matrix 1
A	ITS	16.4 ± 1.8*	11.0 ± 0.8	11.5 ± 2.3	—	—
	cpDNA	—	—	13.3 ± 2.1	12.0 ± 2.6	11.5 ± 2.3
B	ITS	19.6 ± 1.3*	13.8 ± 0.8	14.4 ± 1.6	—	—
	cpDNA	—	—	16.7 ± 1.3	15.1 ± 1.0	14.4 ± 1.3
C	ITS	10.4 ± 3.9	7.8 ± 1.3	7.3 ± 1.8	—	—
	cpDNA	—	—	8.4 ± 2.1	5.5 ± 1.0	4.7 ± 1.0

Notes: Node A, South American *Cortaderia* crown; node B, South American *Cortaderia* stem; and node C, *Notochloe* stem. Asterisks indicate inconsistent age estimations.

phylogeny of *Cortaderia* and related genera potentially provides further data for testing the WWD hypothesis, but the trees from Matrices 4 and 5 would lead us to contradictory conclusions with respect to both the numbers and the directions of dispersals (Fig. 5). The Matrix 5 topology (conflicting taxa placed according to cpDNA) would appear to lend support to the WWD hypothesis but that of Matrix 4 (conflicting taxa placed according to ITS) does not: No difference between dispersal rates in westerly and easterly directions is apparent. This latter conclusion is also drawn when the information contained in Matrices 4 and 5 is pooled (duplicated topology, Matrix 6). The ancestral areas optimized on the duplicated topology are less decisive than those optimized on either combined "gene" tree individually. They appear to agree with those derived from the Matrix 4 topology but reveal a larger number of dispersal events.

"Dispersal," as mapped in molecular phylogenies, represents dispersal plus successful establishment, survival, and sampling of genes. As such, it is an underestimation of actual dispersal rates. Extinction can remove evidence of successful dispersal. As demonstrated here, hybridization (followed by genetic drift) can have the same effect. For example, using the duplicated topology, we can infer that an ancestor of *N. microdon* dispersed from the Americas and hybridized with an ancestor of *Plinthanthesis* in Australia. In this case, use of a single gene/genome phylogeny would only reveal part of the species history and could lead to misleading ancestral area reconstructions. The possibility of these kinds of problems should not be overlooked even when multiple gene trees are available, as studies that sample a small number of gene trees are likely to miss those representing the minority of the genome (Rieseberg and Morefield 1995).

Effectively modeling rare events such as long-distance dispersal requires a great deal of data (i.e., from different groups of plants; Sanmartin et al. 2007). Our interest in recovering the evolutionary history of individual groups notwithstanding, differences in inferences made according to differing gene trees might better be considered as stochastic effects, best interpreted in the context of a larger body of information. It is, however, not unthinkable that similar biases could arise as a result of nonrandom processes across multiple independent groups. Cook and Crisp (2005) considered the possible misleading effects of strong directional asymmetry on inferences of dispersals. Our study might provide a further example of how such asymmetry could go undetected. When dispersal is followed by homoploid hybridization, the genes of the invading genome are likely to be greatly outnumbered by those of the native population (chromosome numbers of *N. microdon* and *Plinthanthesis*, and thus the mode of hybrid speciation in *Notochloe*, are unknown). Transfer of genetic adaptations through hybridization may facilitate the colonization of new habitats (Rieseberg and Brunsfield 1992) and potentially elevate the rates of response to selection, predisposing colonizing populations to rapid adaptive diversification under disruptive or divergent selection (Seehausen 2004). However, the chances of persistence of chloroplast and/or nuclear genomes/genes, which invade an area already inhabited by a close relative might be comparatively low. This effect could represent a bias against the discovery of concerted multiple dispersals of closely related organisms.

Implications for molecular dating.—Ancient hybridizations provide additional information relevant to the ages of nodes and of evolutionary events that would not be available if gene trees were congruent. For example, where hybridization appears to have followed dispersal, the different age estimations of the stem nodes of the duplicated taxon can be used to infer minimum and maximum ages of the dispersal event. The stem node of the duplicated *N. microdon* is represented by its ITS lineage in Matrix 4, by its cpDNA lineage in Matrix 5, and by both in Matrix 6 (Fig. 4). It follows from the ancestral area optimizations that the ancestor of *Notochloe* could not have dispersed prior to the age of its cpDNA MRCA with *C. pilosa* (in the Americas) but must have done so prior to the age of its ITS MRCA with *Plinthanthesis* (in Australia; see Fig. 4). This further illustrates the strength of the taxon duplication approach for biogeographic reconstruction, even if, in the example used here, the age estimations for the nodes in question are not sufficiently precise to be informative.

Implications for the evolution of complex characters.—Connor (1981, p. 67) considered gynodioecism "[. . .] so rare in the family [Poaceae] as to not merit much attention as an optimal breeding system." It might seem

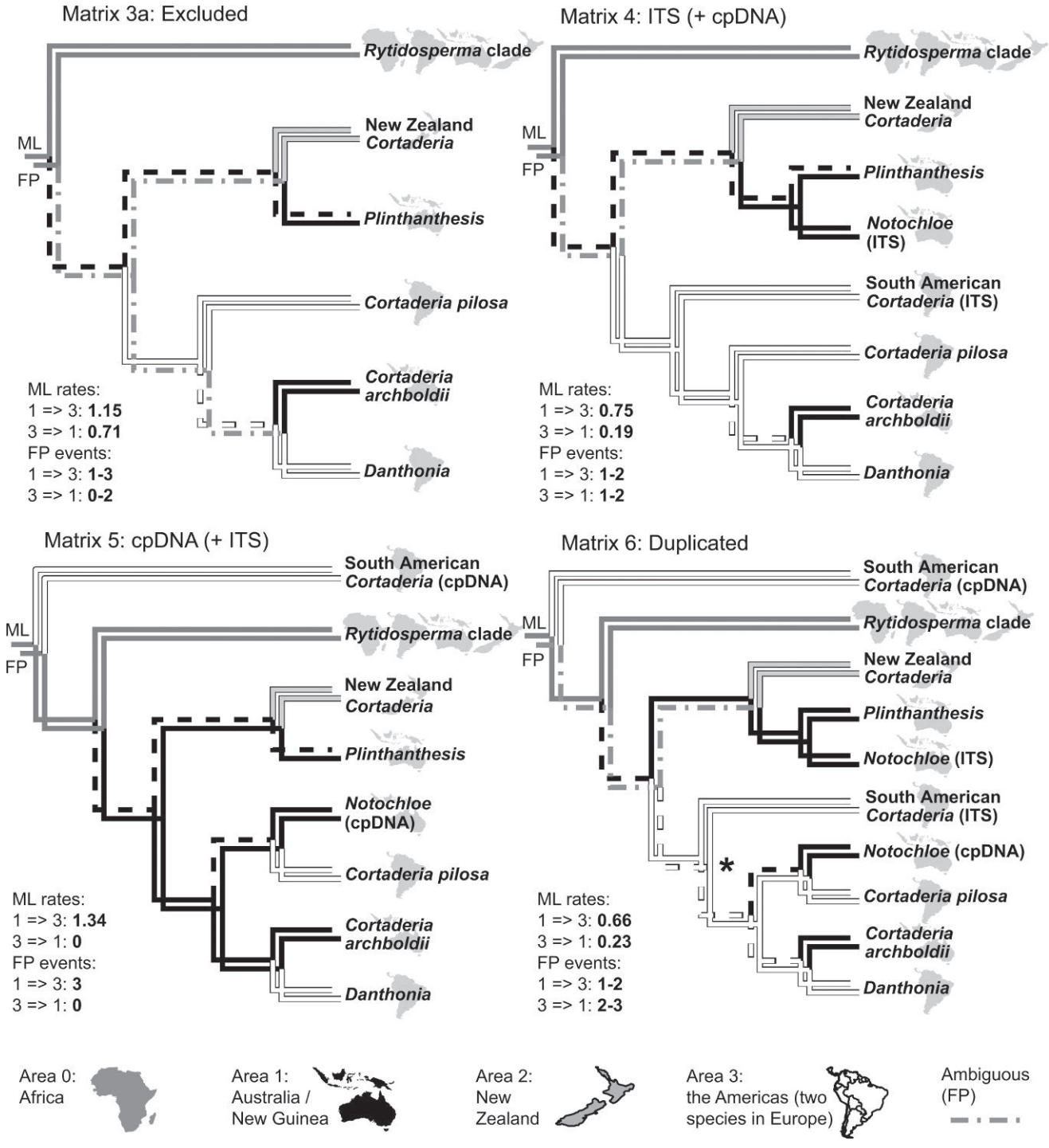

FIGURE 5. Ancestral state reconstructions of geographic distribution represented on 4 trees summarized from analyses 3–6. Branches subtending the reconstructed nodes are shaded by state as indicated, with ML represented on branches above and FP below. Ambiguous parsimony reconstructions are dashed, and where likelihood reconstructions for nodes scored less than 0.95, branches are dashed in the shade of the best-scoring reconstruction. Rates of change (ML) and numbers of dispersal events (FP) between Australia (Area 1) and the Americas and Europe (Area 3) are indicated. The parsimony optimization of the node indicated with an asterisk is ambiguous when monophyly of the South American *Cortaderia* clade is constrained, and this is therefore represented with a dashed line.

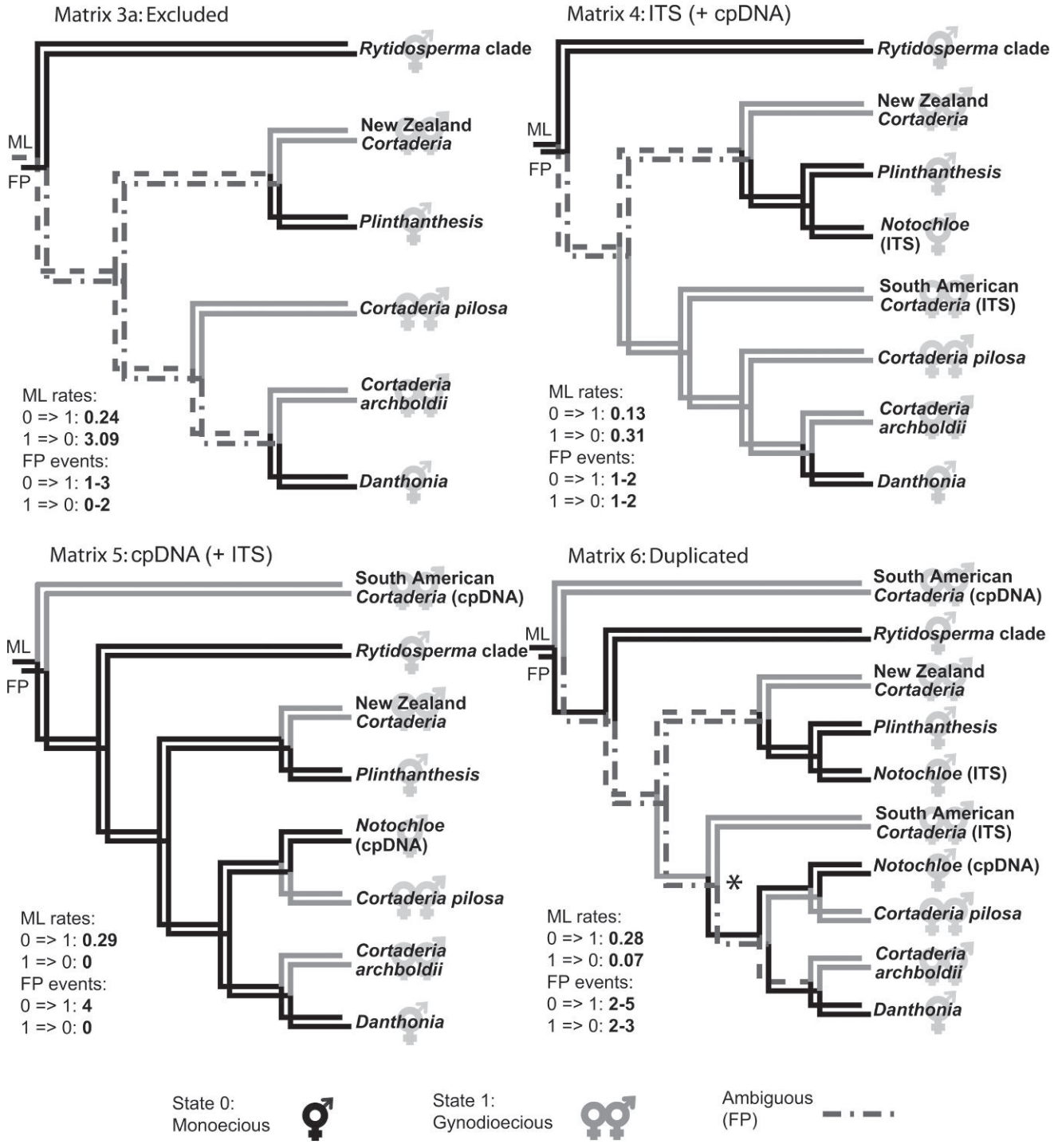

FIGURE 6. Ancestral state reconstructions of gynodioecy represented on 4 trees summarized from analyses 3–6. Branches subtending the reconstructed nodes are shaded by state, with ML represented on branches above and FP below. Gynodioecy (State 1) = light gray, dioecy (State 0) = black, and ambiguous = mid gray, dashed. Rates of change (ML) and numbers of changes (FP) are indicated. The parsimony optimization of the node indicated with an asterisk is ambiguous when monophyly of the cpDNA South American *Cortaderia* clade is constrained, and this is therefore represented with a dashed line.

curious then that optimization of gynodioecy (which here also represents a suite of characters that make up the *Cortaderia* morphological syndrome) over the Matrix 5 tree suggests a scenario of multiple independent gains in Danthonioideae, that is, convergence. In contrast, 1 of the 2 most parsimonious optimizations onto the Matrix 4 tree would lead us instead to infer a single gain followed by 2 losses (the other 2 gains followed by a single loss). The Matrix 4 scenario also appears more parsimonious when regarding similarities in the leaf anatomy of *Notochloe* and *Plinthanthesis* (Linder and Verboom 1996), which given the Matrix 5 topology would also have arisen twice independently rather than just once.

The term "chloroplast capture" has been used in situations where the chloroplast genome is all that remains of one of the parents in a hybrid lineage (Rieseberg and Soltis 1991). This affords a tempting explanation for the patterns of morphological variation observed in Danthonioideae. The "capture" of a complex morphology can be likened to the capture of a chloroplast genome: Although complex characters may be disrupted by chromosomal rearrangements following hybridization, both should be more likely than a de novo gain. A realistic model of such reticulating systems might thus include not only differences between the likelihoods of gains and losses of a character but also between gains and captures, as different modes of origin. Such a model may not be straightforward to apply. In contrast to the example of biogeographic reconstruction above, using the duplicated tree serves to highlight rather than to address the problem. One possible solution might be to estimate likelihoods of character changes using a topology in which the position of conflicting taxa implies the lowest numbers of secondary gains and losses. In our case, this would correspond to the Matrix 4 topology. This approach would, however, be at the cost of making assumptions about the process under study that are likely to influence heavily the outcome of the analyses. An alternative solution would be to optimize the character over a hybridization network rather than a bifurcating tree.

Conclusions

Quoting from Maureira-Butler et al. (2008): ". . . the ability to use trees to understand character evolution, to time speciation events, and to make predictions about taxa from their nearest relatives is confounded by a reticulate history." Hybridization has played an important role in shaping the evolutionary history of the Danthonioideae, apparently allowing the transfer of complex morphologies between lineages and masking evidence of long-distance dispersal. The implications for studies addressing evolutionary questions are clear: Individual gene trees can indeed be misleading. With respect to morphological evolution, we should certainly still use such gene trees to target the reassessment of characters that appear to have originated multiple times. However, the process of reciprocal illumination should also lead us to test the phylogenetic hypothesis with independent data. For cases in which the species phylogeny includes more than 1 significantly differing gene tree, we have demonstrated a simple method, whereby all available data can be combined to represent reticulations in a single bifurcating tree. This should be particularly useful to improve resolution where different gene trees are largely congruent and the areas of conflict can be clearly defined. We therefore challenge, at least in part, the conclusion of Maureira-Butler et al. (2008). When analyzed in this manner, conflicting gene trees provide additional information that can be used, for example, to both identify dispersal events and place logical constraints on the time frame within which they could have taken place.

Supplementary Material

Supplementary material can be found at http://www.sysbio.oxfordjournals.org/.

Funding

The project under which a major part of this research was conducted was funded by a Swiss National Science Foundation grant to H.P.L. (3100A0-107927) and additional support for fieldwork provided by a Swiss Academy of Sciences travel grant and a Claraz donation (M.D.P. and A.M.H.).

Acknowledgements

The authors would like to thank Alessia Guggisberg for helpful discussions of hybridization and network methods, Shuqing Xu for assistance with cloning, members of the Linder lab at the Institute for Systematic Botany, Zurich, for discussion of drafts of the manuscript, and Roberta Mason-Gamer, Jack Sullivan, Thorsten Lumbsch, and 2 anonymous reviewers for constructive criticism. Part of this work was carried out by using the resources of the Computational Biology Service Unit from Cornell University that is partially funded by Microsoft Corporation.

References

Alvarez I., Wendel J.F. 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. 29:417.

Barber J.C., Finch C.C., Fransisco-Ortega J., Santos-Guerra A., Jansen R.K. 2007. Hybridisation in Macaronesian Sideritis (Lamiaceae): evidence from incongruence of multiple independent nuclear and chloroplast sequence datasets. Taxon. 56:74–88.

Barker N.P., Galley C., Verboom G.A., Mafa P., Gilbert M., Linder H.P. 2007. The phylogeny of the austral grass subfamily Danthonioideae: evidence from multiple data sets. Plant Syst. Evol. 264:135–156.

Barker N.P., Linder H.P., Harley E.H. 1995. Polyphyly of Arundinoideae (Poaceae): evidence from *rbc*L sequence data. Syst. Bot. 20:423–435.

Barker N.P., Linder H.P., Harley E.H. 1999. Sequences of the grass-specific insert in the chloroplast *rpo*C2 gene elucidate generic relationships of the Arundinoideae (Poaceae). Syst. Bot. 23: 327–350.

Barker N.P., Linder H.P., Morton C.M., Lyle M. 2003. The paraphyly of *Cortaderia* (Danthonioideae; Poaceae): evidence from morphology and chloroplast and nuclear DNA sequence data. Ann. Mo. Bot. Gard. 90:1–24.

Barker N.P., Morton C.M., Linder H.P. 2000. The Danthonieae: generic composition and relationships. In: Jacobs S.W.L., Everett J., editors. Grasses: systematics and evolution. Melbourne (Australia): CSIRO. p. 221–230.

Bouchenak-Khelladi Y., Salamin N., Savolainen V., Forest F., van der Bank M., Chase M.W., Hodkinson T.R. 2008. Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. Mol. Phylogenet. Evol. 47:488–505.

Bremer B., Jansen R.K., Oxelman B., Backlund M., Lantz H., Kim K.-J. 1999. More characters or more taxa for a robust phylogeny—case study from the coffee family. Syst. Biol. 48:413–435.

Bull J.J., Huelsenbeck J.P., Cunningham C.W., Swofford D.L., Waddell P.J. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42:384–397.

Christin P.-A., Besnard G., Samartitani E., Duvall M.R., Hodkinson T.R., Savolainen V., Salamin N. 2008. Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. Curr. Biol. 18:37–43.

Comes H.P., Abbott R.J. 2001. Molecular phylogeography, reticulation, and lineage sorting in Mediterranean *Senecio* Sect. *Senecio* (Asteraceae). Evolution. 55:1943–1962.

Connor H.E. 1973. Breeding systems in Cortaderia (Gramineae). Evolution. 27:663–678.

Connor H.E. 1981. Evolution of reproductive systems in the Gramineae. Ann. Mo. Bot. Gard. 68:48.

Connor H.E. 1990. Breeding systems in New Zealand grasses XI. Gynodioecism in *Chionochloa bromoides*. N.Z. J. Bot. 28:59–65.

Connor H.E. 2004. Flora of New Zealand—Gramineae supplement I: Danthonioideae. N.Z. J. Bot. 42:771–795.

Cook L.G., Crisp M.D. 2005. Directional asymmetry of long-distance dispersal and colonization could mislead reconstructions of biogeography. J. Biogeogr. 32:741–754.

de Queiroz A., Donoghue M.J., Kim J. 1995. Separate versus combined analysis of phylogenetic evidence. Annu. Rev. Ecol. Syst. 26:657–681.

Dickerman A.W. 1998. Generalizing phylogenetic parsimony from the tree to the forest. Syst. Biol. 47:414–426.

Ellstrand N.C., Whitkus R., Rieseberg L.H. 1996. Distribution of spontaneous plant hybrids. Proc. Natl. Acad. Sci. USA. 93:5090–5093.

Fehrer J., Gemeinholzer B., Chrtek J.J., Brautigam S. 2007. Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in Pilosella hawkweeds (Hieracium, Cichorieae, Asteraceae). Mol. Phylogenet. Evol. 42:347.

Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27:401–410.

Fitch W.M. 1971. Toward defining the course of evolution: minimum change for a specified tree topology. Syst. Zool. 20:406–416.

Flynn J., Finarelli J., Zehr S., Hsu J., Nedbal M. 2005. Molecular phylogeny of the Carnivora (Mammalia): assessing the impact of increased sampling on resolving enigmatic relationships. Syst. Biol. 54:317.

Funk V.A. 1985. Phylogenetic patterns and hybridization. Ann. Mo. Bot. Gard. 72:681–715.

Galley C., Linder H.P. 2007. The phylogeny of the Pentaschistis clade (Danthonioideae, Poaceae) based on chloroplast DNA, and the evolution and loss of complex characters. Evolution. 61: 864–884.

Grant P.R., Grant B.R. 1992. Hybridization of bird species. Science. 256:193–197.

Huber K.T., Oxelman B., Lott M., Moulton V. 2006. Reconstructing the evolutionary history of polyploids from multilabeled trees. Mol. Biol. Evol. 23:1784–1791.

Huelsenbeck J.P., Ronquist F. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics. 17:754–755.

Huson D.H., Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254–267.

Huson D.H., Dezulian T., Klopper T., Steel M.A. 2004. Phylogenetic super-networks from partial trees. IEEE/ACM Trans. Comput. Biol. Bioinform. 1:151.

Kellogg E.A., Appels R., Mason-Gamer R.J. 1996. When genes tell different stories: the diploid genera of Triticeae (Gramineae). Syst. Bot. 21:321–347.

Kluge A.G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes). Syst. Zool. 38:7–25.

Lecointre G., Deleporte P. 2005. Total evidence requires exclusion of phylogenetically misleading data. Zool. Scr. 34:101–117.

Linder H.P., Barker N.P. 2005. From Nees to now—changing questions in the systematics of the grass subfamily Danthonioideae. Nova Acta Leopold. 92:29–44.

Linder H.P., Verboom G.A. 1996. Generic limits in the Rytidosperma (Danthonieae, Poaceae) complex. Telopea. 6:597–627.

Maddison W.P., Maddison D.R. 2006. Mesquite: a modular system for evolutionary analysis. Version 2.5. Available from: http://mesquiteproject.org.

Maureira-Butler I.J., Pfeil B.E., Muangprom A., Osborn T.C., Doyle J.J. 2008. The reticulate history of *Medicago* (Fabaceae). Syst. Biol. 57:466–482.

McBreen K., Lockhart P.J. 2006. Reconstructing reticulate evolutionary histories of plants. Trends Plant Sci. 11:398.

McDade L.A. 1992. Hybrids and phylogenetic systematics II. The impact of hybrids on cladistic analysis. Evolution. 46:1329.

Mickevich M.F., Farris J.S. 1981. The implications of congruence in Menidia. Syst. Zool. 30:351–370.

Morrison D.A. 2005. Networks in phylogenetic analysis: new tools for population biology. Int. J. Parasitol. 35:567–582.

Müller K. 2006. SeqState. Version 1.32. Available from: http://www.botanik.uni-bonn.de/system/downloads/SeqStatus.

Nelson G., Platnick N.I. 1980. Multiple branching in cladograms: two interpretations. Syst. Zool. 29:86.

Pagel M., Meade A. 2006. BayesTraits. Available from: http://www.evolution.rdg.ac.uk.

Penny D., Hendy M.D. 1985. The use of tree comparison metrics. Syst. Zool. 34:75–82.

Pirie M.D., Humphreys A.M., Galley C., Barker N.P., Verboom G.A., Orlovich D., Draffin S.J., Lloyd K., Baeza C.M., Negritto M., Ruiz E., Cota Sanchez J.H., Reimer E., Linder H.P. 2008. A novel supermatrix approach improves resolution of phylogenetic relationships in a comprehensive sample of danthonioid grasses. Mol. Phylogenet. Evol. 48:1106–1119.

Pirie M.D., Vargas M.P.B., Botermans M., Bakker F.T., Chatrou L.W. 2007. Ancient paralogy in the cpDNA trnL-F region in Annonaceae: implications for plant molecular systematics. Am. J. Bot. 94:1003–1016.

Pisani D., Wilkinson M. 2002. Matrix representation with parsimony, taxonomic congruence, and total evidence. Syst. Biol. 51: 151–155.

Posada D., Crandall K.A. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics. 14:817–818.

Rambaut A., Drummond A.J. 2003. Tracer. Version 1.3. Available from: http://beast.bio.ed.ac.uk/Tracer.

Rieseberg L.H., Brunsfield S.J. 1992. Molecular evidence and plant introgression. In: Soltis P.S., Soltis D.E., Doyle J.J., editors. Molecular systematics of plants. London, UK: Chapman & Hall. p. 151–176.

Rieseberg L.H., Morefield J.D. 1995. Character expression, phylogenetic reconstruction, and the detection of reticulate evolution. In: Hoch P.C., Stephenson A.G., editors. Experimental and molecular approaches to plant systematics. p. 333–353 (Monographs in Systematic Botany from the Missouri Botanical Garden).

Rieseberg L.H., Soltis D.E. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evol. Trend Plant. 5:65–84.

Rodrigo A.G., Kelly-Borges M., Bergquist P.R., Bergquist P.L. 1993. A randomisation test of the null hypothesis that two cladograms are sample estimates of a parametric phylogenetic tree. N.Z. J. Bot. 31:257–268.

Sanderson M.J. 2002a. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19:101–109.

Sanderson M.J. 2002b. r8s. Version 1.50. Distributed by the author. Davis (CA): Section of Evolution and Ecology, University of California.

Sanmartin I., Wanntorp L., Winkworth R.C. 2007. West wind drift revisited: testing for directional dispersal in the Southern hemisphere using event-based tree fitting. J. Biogeogr. 34:398–416.
Seehausen O. 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19:198–207.
Shaw K.L. 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proc. Natl. Acad. Sci. USA. 99:16122–16127.
Simmons M.P., Ochoterena. H. 2000. Gaps as characters in sequence-based phylogenetic analysis. Syst. Biol. 49:369–381.
Sullivan J. 1996. Combining data with different distributions of among-site rate variation. Syst. Biol. 45:375–380.
Swofford D.L. 2000. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Sunderland (MA): Sinauer Associates.
Verboom G.A., Ntsohi R., Barker N.P. 2006. Molecular phylogeny of African Rytidosperma-affiliated danthonioid grasses reveals generic polyphyly and convergent evolution in spikelet morphology. Taxon. 55:337–348.
Vriesendorp B., Bakker F.T. 2005. Reconstructing patterns of reticulate evolution in angiosperms: what can we do? Taxon. 54:593–604.
Wiens J.J. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. Syst. Biol. 52:528–538.
Wiens J.J. 2006. Missing data and the design of phylogenetic analysis. J. Biomed. Inform. 39:34–42.
Wiens J.J., Parkinson C.L., Reeder T.W. 2005. Hylid frog phylogeny and sampling strategies for speciose clades. Syst. Biol. 54:719–748.
Wikström N., Savolainen V., Chase M.W. 2001. Evolution of the angiosperms: calibrating the family tree. Proc. R. Soc. Lond. Biol. 1482:2211–2220.
Wilgenbusch J.C., Warren D.L., Swofford D.L. 2004. AWTY: a system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. Available from: http://www.ceb.csit.fsu.edu/awty.

Received 8 February 2008; reviews returned 13 June 2008; accepted 8 September 2009
Associate Editor: Roberta J. Mason-Gamer