Although we have used the polymerization of recMoPrP(89–230) into amyloid fibrils to generate prion infectivity, we hasten to add that other β-rich forms of recMoPrP(89–230) may also harbor infectivity. Preliminary results suggest that preparations of β-oligomers formed from recMoPrP(89–230) may also contain low levels of prion infectivity (*33*). Such findings emphasize the need to define optimal conditions for prion formation in vitro under which high levels of PrP^{Sc} can be generated. Moreover, previous difficulties in creating infectious prions in vitro from recPrPs enriched for β-structure may be due to the tendency of mammalian PrPs to fold into biologically irrelevant β-rich isoforms (*3*, *4*, *11*). In studies of fungal prions, the ease of assaying infectivity (*34*) and the ability to study millions of colonies made the creation of in vitro infectivity from recombinant proteins more tractable (*35*–*37*). Whereas yeast prions form within the cytoplasm (*38*), mammalian prions are thought to be produced on the cell surface in caveolae-like domains (*39*, *40*).

From Tg mouse studies, it is well established that templates improve the likelihood of forming an infectious β-rich isoform (*8*, *12*). In the studies reported here, we see evidence that seeded amyloid fibrils exhibit shorter incubation times than their unseeded progenitor (Fig. 1A). It remains to be determined whether this is due to the greater number of PrP^{Sc} molecules within seeded fibrils relative to unseeded fibrils, or whether this reflects strain differences.

Our results have important implications for human health. The formation of prions from recPrP demonstrates that PrP^{C} is sufficient for the spontaneous formation of prions; thus, no exogenous agent is required for prions to form in any mammal. Our findings provide an explanation for sporadic Creutzfeldt-Jakob disease for which the spontaneous formation of prions has been hypothesized. Understanding sporadic prion disease is particularly relevant to controlling the exposure of humans to bovine prions (*41*). That bovine prions are pathogenic for humans is well documented; more than 150 teenagers and young adults have already died from prion-tainted beef derived from cattle with bovine spongiform encephalopathy (*42*). Moreover, the sporadic forms of Alzheimer's and Parkinson's diseases as well as amyotrophic lateral sclerosis and the frontal temporal dementias are the most frequent forms of these age-dependent disorders, as is the case for the prion diseases. Important insights in the etiologic events that feature in these more common neurodegenerative disorders, all of which are caused by the aberrant processing of proteins in the nervous system, are likely to emerge as more is learned about the molecular pathogenesis of the sporadic prion diseases.

References and Notes

1. S. B. Prusiner, in *Prion Biology and Diseases,* S. B. Prusiner, Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2004), pp. 89–141.
2. D. Peretz *et al*., *Neuron* **34**, 921 (2002).
3. I. Mehlhorn *et al*., *Biochemistry* **35**, 5528 (1996).
4. A. F. Hill, M. Antoniou, J. Collinge, *J. Gen. Virol.* **80**, 11 (1999).
5. K. K. Hsiao *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9126 (1994).
6. K. Kaneko *et al*., *J. Mol. Biol.* **295**, 997 (2000).
7. P. Tremblay *et al*., *J. Virol.* **78**, 2088 (2004).
8. S. B. Prusiner *et al*., *Cell* **63**, 673 (1990).
9. S. B. Prusiner *et al*., *Cell* **35**, 349 (1983).
10. See supporting data on *Science* Online.
11. I. V. Baskakov, G. Legname, M. A. Baldwin, S. B. Prusiner, F. E. Cohen, *J. Biol. Chem.* **277**, 21140 (2002).
12. S. Supattapone *et al*., *J. Virol.* **75**, 1408 (2001).
13. S. B. Prusiner, M. R. Scott, S. J. DeArmond, G. Carlson, in *Prion Biology and Diseases,* S. B. Prusiner, Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2004), pp. 187–242.
14. G. C. Telling *et al*., *Genes Dev.* **10**, 1736 (1996).
15. G. Legname *et al*., data not shown.
16. K. Post, D. R. Brown, M. Groschup, H. A. Kretzschmar, D. Riesner, *Arch. Virol. Suppl.*, 265 (2000).
17. H. Wille, S. B. Prusiner, *Biophys. J.* **76**, 1048 (1999).
18. C. Govaerts, H. Wille, S. B. Prusiner, F. E. Cohen, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8342 (2004).
19. M. P. McKinley *et al*., *J. Virol.* **65**, 1340 (1991).
20. D. C. Gajdusek, *Br. Med. Bull.* **49**, 913 (1993).
21. J. T. Jarrett, P. T. Lansbury, Jr., *Cell* **73**, 1055 (1993).
22. H. Wille, G.-F. Zhang, M. A. Baldwin, F. E. Cohen, S. B. Prusiner, *J. Mol. Biol.* **259**, 608 (1996).
23. G. Forloni *et al*., *Nature* **362**, 543 (1993).
24. H. LeVine, *Protein Sci.* **2**, 404 (1993).
25. A. Taraboulos *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8262 (1990).
26. S. J. DeArmond *et al*., *Neuron* **19**, 1337 (1997).
27. P. Gambetti, P. Parchi, *N. Engl. J. Med.* **340**, 1675 (1999).
28. J. A. Mastrianni *et al*., *N. Engl. J. Med.* **340**, 1630 (1999).
29. J. Collinge, K. C. L. Sidle, J. Meads, J. Ironside, A. F. Hill, *Nature* **383**, 685 (1996).
30. R. A. Bessen, R. F. Marsh, *J. Virol.* **68**, 7859 (1994).
31. G. C. Telling *et al*., *Science* **274**, 2079 (1996).
32. S. B. Prusiner, *Annu. Rev. Microbiol.* **43**, 345 (1989).
33. G. Legname *et al*., unpublished data.
34. R. B. Wickner, *Science* **264**, 566 (1994).
35. M. L. Maddelein, S. Dos Reis, S. Duvezin-Caubet, B. Coulary-Salin, S. J. Saupe, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7402 (2002).
36. C. Y. King, R. Diaz-Avalos, *Nature* **428**, 319 (2004).
37. M. Tanaka, P. Chien, N. Naber, R. Cooke, J. S. Weissman, *Nature* **428**, 323 (2004).
38. M. M. Patino, J.-J. Liu, J. R. Glover, S. Lindquist, *Science* **273**, 622 (1996).
39. A. Gorodinsky, D. A. Harris, *J. Cell Biol.* **129**, 619 (1995).
40. P. J. Peters *et al*., *J. Cell Biol.* **162**, 703 (2003).
41. C. Casalone *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3065 (2004).
42. R. G. Will, M. P. Alpers, D. Dormont, L. B. Schonberger, in *Prion Biology and Diseases,* S. B. Prusiner, Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2004), pp. 629–671.
43. D. Peretz *et al*., *Nature* **412**, 739 (2001).

44. We thank P. Culhane, K. Giles, P. Lessard, M. Scott, S. Supattapone, and the staff at the Hunters Point Animal Facility. Supported by NIH grants AG02132, AG10770, and AG021601, the G. Harold and Leila Y. Mathers Charitable Foundation, the Dana Foundation, and the Sherman Fairchild Foundation. G.L., D.R., F.E.C., S.J.D., and S.B.P. have a financial interest in InPro Biotechnology, which applies research on prions and prion diseases to promote human health.

Supporting Online Material

www.sciencemag.org/cgi/content/full/305/5684/673/DC1
Materials and Methods
Figs. S1 to S3
Table S1
References

11 May 2004; accepted 1 July 2004

Host-to-Parasite Gene Transfer in Flowering Plants: Phylogenetic Evidence from Malpighiales

Charles C. Davis[1]* and Kenneth J. Wurdack[2]

Horizontal gene transfer (HGT) between sexually unrelated species has recently been documented for higher plants, but mechanistic explanations for HGTs have remained speculative. We show that a parasitic relationship may facilitate HGT between flowering plants. The endophytic parasites Rafflesiaceae are placed in the diverse order Malpighiales. Our multigene phylogenetic analyses of Malpighiales show that mitochrodrial (*matR*) and nuclear loci (18*S* ribosomal DNA and *PHYC*) place Rafflesiaceae in Malpighiales, perhaps near Ochnaceae/Clusiaceae. Mitochondrial *nad1B-C*, however, groups them within Vitaceae, near their obligate host *Tetrastigma*. These discordant phylogenetic hypotheses strongly suggest that part of the mitochondrial genome in Rafflesiaceae was acquired via HGT from their hosts.

Malpighiales are one of the most diverse clades of flowering plants uncovered in recent phylogenetic analyses. The order comprises 27 families (*1*) previously assigned to 13 different orders (*2*), including more than 16,000 species spanning tremendous morphological and ecological diversity (*3*). Recent surprising additions to Malpighiales are the endophytic holoparasites Rafflesiaceae (*4*), which lack leaves, stems, and roots, and rely entirely on their host plants, species of *Tetrastigma* (Vitaceae), for their nutrition. Despite their extreme vegetative reduction, they are unmistakable in flower, producing the largest flowers in the world, which mimic rotting flesh—an enticement to the carrion flies that pollinate them (*5*).

Barkman *et al*. (*4*) used mitochondrial (mt) *matR* sequences to place Rafflesiaceae firmly with Malpighiales [100%

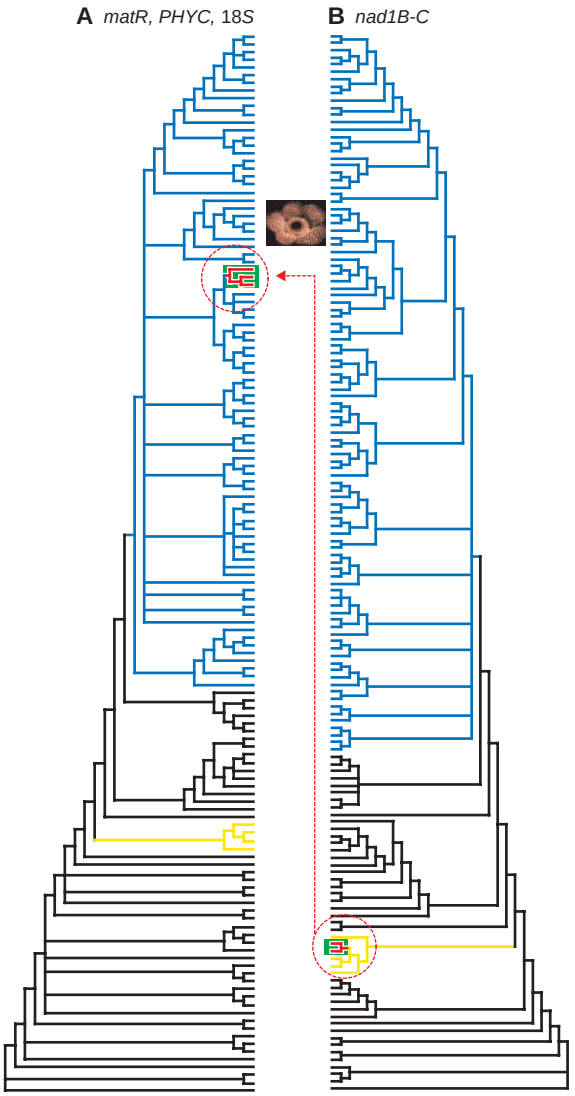


Fig. 1. Two conflicting hypotheses about the phylogenetic placement of Rafflesiaceae. (**A**) The strict consensus of 136 angiosperms for combined mt *matR* and nuclear (*PHYC* and ribosomal 18*S*) data showing a well-supported (100% BP) Malpighiales clade (in blue), which includes all members of the order sensu APG II (*1*) plus Rafflesiaceae (in red; *Rafflesia*, *Rhizanthes*, and *Sapria*). (**B**) The strict consensus of 147 angiosperms for mt *nad1B-C* (the *nad1* intron 2 and part of the adjacent exons b and c) showing a well-supported (100% BP) Malpighiales clade, which includes all members of the order except Rafflesiaceae. Rafflesiaceae (*Rafflesia* and *Sapria*) are strongly placed (100% BP) in the basal eudicot family Vitaceae (in yellow) near their host genus, *Tetrastigma*. The dashed line is the hypothesized host/parasite HGT.

bootstrap percentage (BP)]. Their use of a single mt gene was appropriate in a family that has resisted placement with standard genetic loci. To further examine this placement, we obtained sequences representing all families of Malpighiales, all genera of Rafflesiaceae, and numerous basal eudicots for four loci from the mt and nuclear genomes (*6*). Low-copy nuclear genes are an underused resource for resolving the placement of problematic taxa, and *phytochrome C* (*PHYC*), as used here, has been useful for revealing relationships within Malpighiales (*7*).

Our phylogenetic analyses are summarized in Fig. 1 (*8*). The tree created from the *matR* and nuclear loci firmly (100% BP) place Rafflesiaceae within Malpighiales. In contrast, the mt locus *nad1B-C* suggests that Rafflesiacecae are not members of Malpighiales but belong (100% BP) in Vitaceae near their host *Tetrastigma.* Each of these mutually exclusive hypotheses cannot be attributable to contamination (*9*), and each receives strong support from parsimony analyses and from alternative topology tests.

Which of these conflicting hypotheses reflect the true species affinities of Rafflesiaceae? Vitaceae possess several synapomorphies that are rare among angiosperms, including sieve-tube plastids with starch and protein inclusions, pearl glands, stamens opposite the petals, and seeds with a cordlike raphe. If Rafflesiaceae were embedded in Vitaceae, as suggested by *nad1B-C*, we would expect species to possess at least some of these characters, but they do not (*2*, *3*). A definitive malpighialean sister group for Rafflesiaceae is unclear, given our data. However, the closest relatives suggested in the combined analysis (*10*), Ochnaceae and Clusiaceae sensu lato, share tenuinucellate ovules (among mostly crassinucellate relatives) and staminal fusion with Rafflesiaceae (*2*, *3*).

The position of Rafflesiaceae based on *nad1B-C* provides a new example of horizontal gene transfer. If *nad1B-C* were vertically transmitted, as we believe to be the case for the other loci, we would expect Rafflesiaceae to group with Malpighiales. Instead, phylogenetic evidence from *nad1B-C* suggests that part of the mt genome in Rafflesiaceae originated from their hosts, *Tetrastigma* (either stem or crown group members), and was horizontally transferred to these obligate parasites. A similar horizontal gene transfer (HGT) of *nad1B-C* was recently reported (*11*) in seed plants, involving a transfer from an asterid to *Gnetum*. And Bergthorsson *et al*. (*12*) have documented several instances of mt HGT between distantly related angiosperm groups.

The underlying mechanism for HGT between sexually unrelated plants, however, has been elusive. Various pathogens have been suggested as primary vector agents (*11*, *12*). Our study documents a case in which there is no need to propose an intermediary vector for HGT. In these plants, the transfer appears to have been facilitated by the intimacy of the association between the host and the endophytic parasite, which lives its whole vegetative life as "an almost mycelial haustorial system," "ramifying and anastomosing throughout the [tissues of the] host" (*13*). This pattern may be an important mechanism by which parasites assemble their genetic architecture, and additional cases of HGT should be sought among other endophytic parasites and their hosts. It will also not be surprising if reciprocal genetic transfers are found to have occurred, from parasite to host.

[1]Ecology and Evolutionary Biology, University of Michigan Herbarium, 3600 Varsity Drive, Ann Arbor, MI 48108–2287, USA. [2]Department of Botany and Laboratories of Analytical Biology, Smithsonian Institution, 4210 Silver Hill Road, Suitland, MD 20746, USA.

*To whom correspondence should be addressed. E-mail: chdavis@umich.edu.

References and Notes

1. The Angiosperm Phylogeny Group, *Bot. J. Linn. Soc.* **141**, 399 (2003).
2. A. Cronquist, *An Integrated System of Classification of Flowering Plants* (Columbia Univ. Press, New York, 1981).
3. P. F. Stevens, *Angiosperm Phylogeny Website*, ed. 4 (2001 onward) (www.mobot.org/MOBOT/research/Apweb).
4. T. J. Barkman, S.-K. Lim, K. Mat Salleh, J. Nais, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 787 (2004).
5. R. S. Beaman, P. J. Decker, J. H. Beaman, *Am. J. Bot.* **75**, 1148 (1988).
6. For primer and GenBank information see the supporting online material (SOM) on *Science* Online.
7. C. C. Davis, M. W. Chase, *Am. J. Bot.* **91**, 262 (2004).
8. For detailed information on materials and methods, data sets, phylogenetic analyses, and complete annotated figures see the SOM. We were unable to sample *PHYC* from all families of Malpighiales because it is probably absent from some clades (*7*). Hence, we analyzed these data in combination with 18*S* to ensure that all families were represented for the nuclear genome. We also examined the placement of Rafflesiaceae with 18*S* in two ways. First, we excluded the most divergent domains, V2 and V4, from the analysis (*14*) (322 of 1813 base pairs)] across all taxa. Second, we treated these domains as missing data

only for Rafflesiaceae. Both approaches yielded congruent topologies, with the latter (shown here) being much better resolved. The spurious placement of Rafflesiaceae as sister to all angiosperms when 18*S* is used (*4*) may be attributable to high divergence in these small domains and suggests that 18*S* may not be as useless for placing problematic taxa as previously suggested (*4*).

9. We took several precautions to avoid and detect contamination. First, our results were independently corroborated in the laboratories of each author. Second, Rafflesiaceae data were acquired before starting any work on Vitaceae. Third, if our DNA were cross-contaminated we would not expect such strongly conflicting results regarding the placement of Rafflesiaceae, given the same DNA. Nor would we expect such a high degree of sequence divergence of *nad1B-C* between Rafflesiaceae and *Tetrastigma* (or other Vitaceae we sampled). If contamination occurred, we would expect sequences to be nearly identical to those of other sampled Vitaceace, especially given the relatively low amount of sequence divergence between all accessions of Vitaceae in *nad1B-C* (0.51 to 0.95% sequence divergence). Not only is *nad1B-C* divergent within Rafflesiaceae (6.2%), but it is also highly divergent from other phylogenetically diverse Vitaceae (*15*) included here (2.5 to 3.3%).
10. Taxa in the combined analysis of 18*S*, *PHYC*, and *matR* were included if they were sampled for *matR* plus at least one of the two nuclear loci. We believe this analysis to represent the best estimate of Malpighiales phylogeny. Similar approaches combining multiple genes have provided powerful insights into angiosperm phylogeny where single gene studies have failed (*16*).
11. H. Won, S. S. Renner, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 10824 (2003).
12. U. Bergthorsson, K. L. Adams, B. Thomason, J. D. Palmer, *Nature* **424**, 197 (2003).
13. J. Kuijt, *The Biology of Parasitic Flowering Plants* (Univ. of California Press, Berkeley, CA, 1969).
14. D. L. Nickrent, E. M. Starr, *J. Mol. Evol.* **39**, 62 (1994).
15. M. J. Ingrouille *et al.*, *Bot. J. Linn. Soc.* **138**, 421 (2002).
16. Y.-L. Qiu *et al.*, *Nature* **402**, 404 (1999).
17. We thank W. Anderson, D. Boufford, Y.-L. Qiu, and P. Stevens. C.C.D. was supported by grants from the University of Michigan, NSF AToL 0431266, and the Michigan Society of Fellows. This paper is dedicated to V. Morrison.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1100671/DC1
Materials and Methods
Figs. S1 to S5
References

25 May 2004; accepted 6 July 2004
Published online 15 July 2004; 10.1126/science.1100671
Include this information when citing this paper.

KIF1A Alternately Uses Two Loops to Bind Microtubules

Ryo Nitta,[1] Masahide Kikkawa,[1,2] Yasushi Okada,[1] Nobutaka Hirokawa[1]*

The motor protein kinesin moves along microtubules, driven by adenosine triphosphate (ATP) hydrolysis. However, it remains unclear how kinesin converts the chemical energy into mechanical movement. We report crystal structures of monomeric kinesin KIF1A with three transition-state analogs: adenylyl imidodiphosphate (AMP-PNP), adenosine diphosphate (ADP)–vanadate, and ADP-AlFx (aluminofluoride complexes). These structures, together with known structures of the ADP-bound state and the adenylyl-(β,γ-methylene) diphosphate (AMP-PCP)–bound state, show that kinesin uses two microtubule-binding loops in an alternating manner to change its interaction with microtubules during the ATP hydrolysis cycle; loop L11 is extended in the AMP-PNP structure, whereas loop L12 is extended in the ADP structure. ADP-vanadate displays an intermediate structure in which a conformational change in two switch regions causes both loops to be raised from the microtubule, thus actively detaching kinesin.

To move along microtubules, kinesins (*1*) must alternate rapidly between tightly bound and detached states. How both dimeric (*2*, *3*) and monomeric (*4*, *5*) kinesins achieve this remains unclear. Because the binding energy in the strong-binding state [10 to 20 k_BT (*3*, *4*), where k_B is the Boltzmann constant and T is absolute temperature] is too large for rapid spontaneous release, the energy for fast detachment of kinesin from the microtubule must come from a step of the ATP hydrolysis cycle. Large change(s) in free energy are expected to occur during four steps: ATP binding, hydrolysis, phosphate release, and ADP release. Both conventional kinesin and KIF1A bind tightly to microtubules in the nucleotide-free state and in the ATP-bound state. In the ADP-bound state, conventional kinesin is detached from microtubules, whereas KIF1A is partially detached and diffuses freely along the microtubule. This is because loose binding of ADP-bound KIF1A is supported by the KIF1 family–specific K-loop at loop L12. A mutant KIF1A that lacks the K-loop detaches from the microtubule in the ADP-bound state, and the dissociation constant markedly varies depending on the type of bound nucleotide, as is true for conventional kinesin (*4*). For historical reasons, the tightly bound state is called the strong-binding state, and the fully or partially detached state is called the weak-binding state. Recent work detected the phosphate release from a mutant kinesin, which stalls before the detached state (*6*, *7*). This means that detachment occurs just at or after the phosphate release. Thus, the active process to detach kinesin from the microtubule should occur at the transition from the strong-binding state to the weak-binding state.

The active detachment process can be detected in KIF1A because of its property of binding to the microtubule during adenosine triphosphatase (ATPase) cycling. The apparent dissociation constant of KIF1A in the presence of ATP is the weighted average of the equilibrium dissociation constant of various intermediate states during the ATPase turnover. Because the dissociation constant is not significantly different between two major intermediate states, the AMP-PNP–bound and ADP-bound states (Table 1) (fig. S1) (*8*), the apparent dissociation constant during the ATPase turnover was not expected to be fundamentally different from these values. However, the apparent dissociation constant in the presence of 2 mM ATP was twice the expected

Table 1. (Apparent) equilibrium dissociation constants (K_d) for microtubules. K_d values are reported as means ± SEM of at least three independent experiments. Conditions: 2 mM nucleotide or its analog, 50 mM imidazole, 5 mM Mg-acetate, 1 mM EGTA, and 50 mM K-acetate, pH 7.4 at 27°C (nd, not determined).

Nucleotide	K_d (nM)			
	Wild type	L12†	L11‡	L8§
AMP-PNP	4.2 ± 1.3	6.0 ± 1.4	20.2 ± 4.0	25.0 ± 6.0
ADP	6.8 ± 2.5	23.5 ± 8.4	12.3 ± 4.0	26.5 ± 5.0
ATP*	10.8 ± 1.8	40.5 ± 11.8	nd	nd
ADP-AlFx	5.9 ± 1.5	7.1 ± 1.7	nd	nd
ADP-Vi	21.4 ± 4.3	167 ± 66	nd	nd

*ATP regeneration system was used to maintain ATP/ADP level. †L12: CK1 (*4*). ‡L11: K261A/R264A/K266A. §L8: K161A/R167A/R169A/K183A.

[1]Department of Cell Biology and Anatomy, University of Tokyo, Graduate School of Medicine, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [2]Department of Cell Biology, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390, USA.

*To whom correspondence should be addressed. E-mail: hirokawa@m.u-tokyo.ac.jp

Supporting Online Material for **Host-to-parasite gene transfer in flowering plants** (C. C. Davis and K. J. Wurdack)

Materials and Methods

Taxon sampling. The familial- and ordinal-level circumscriptions for this study follow the APG II (*S1*) system. All families of Malpighiales (including alternative APG II circumscriptions), mostly representing multiple accessions per family, were sampled for *matR, nad1B-C,* and for the combined nuclear data set (*PHYC* plus ribosomal 18S; see below). We similarly sampled all three genera of Rafflesiaceae sensu stricto (cf. *S2*), several species from the closely related clades Celastrales and Oxalidales, members of most basal eudicot families, some monocots, and several "basal angiosperm" groups (*S3*). *Amborella*, *Cabomba*, *Illicium*, and *Nymphaea* represent the earliest diverging angiosperm lineages (*S4, S5*) and were used as outgroups. The *nad1B-C* data set did not include these "basal angiosperms," in which case, representatives of the earliest diverging eudicots (*S3*), Ranunculales, were used as outgroups.

Gene isolation and sequencing. One hundred one, 146, 24, and 69 sequences for *matR*, *nad1B-C*, *PHYC*, and 18S were newly obtained for this study (GenBank numbers AY674447–AY674785), respectively. Otherwise, sequences were obtained from GenBank. Total cellular DNA was prepared following Davis *et al.* (*S6*) or Davis and Chase (*S7*). Amplification and sequencing protocols for *nad1B-C* followed Freudenstein and Chase (*S8*), using primers B and C with new internal sequencing primers nad1-774F

(5'–CCGCCCGCCTTCATTTCGTGGA–3') and nad1-856R (5'–ATTCATCTATACTCGCTGCCAC–3'); 18S followed Soltis and Soltis (*S9*). *matR* followed those protocols above with amplification (26F: 5'–GACCGCTNACAGTAGTTCT–3'; 1858R 5'–TGCTTGTGGGCYRGGGTGAA–3') and internal primers (matR 879F: 5'–ACTAGTTATCAGGTCAGAGA–3'; matR 1002R 5'–CACCCACGATTCCCAGTAGT–3') provided by Zhiduan Chen (personal communication, The Chinese Academy of Sciences). *PHYC* sequences were obtained using previously detailed (*S4, S10*) PCR, cloning, and sequencing procedures.

Rafflesiaceae were strongly supported as monophyletic across all data sets. We sequenced 10 *PHYC* clones from *Sapria*, and found no evidence of a duplication event, which is consistent with previous findings using *PHYC* (*S4, S7*). The four minor sequence variants of these clones (GenBank numbers AY674464-AY674467) differed at a total of eight nucleotide sites. All variants formed a strongly supported monophyletic group and were reduced to a single placeholder in the analyses presented here.

Nucleotide and, where appropriate, amino acid sequences were aligned by eye; the ends of sequences, as well as ambiguous internal regions, were trimmed from each data set to maintain complementary data between taxa. We also examined the impact of excluding the V4 and V2 domains from 18S for all members of Rafflesiaceae (*S11*; see also main text). Nickrent and Starr (*S11*) found the V4 domain to be particularly divergent—it contains 26% of the sequence variation in *Rafflesia* despite its short length (227 of 1813 base pairs (bp); 232 of 1652 included characters in this study). Our analyses confirm the

exceptional divergence of V4 for Rafflesiaceae. Also, the much smaller V2 domain (90 bp in our study) displayed similarly high amounts of divergence. For example, uncorrected "p" values for Rafflesiaceae when compared with *Malpighia* and *Ochna* (respectively) were: entire 18S (11.1–13.2%, 11.6–13.4%), 18S excluding both V4 and V2 from Rafflesiaceae (8.2–8.9%, 8.5–9.2%), V4 only (23.3–31.8%, 23.2–31.2%), and V2 only (29.7–38.6%, 35.7–41.5%). We examined the placement of Rafflesiaceae with 18S by excluding V2 and V4 across all taxa and by treating these regions as missing data only for Rafflesiaceae. Both approaches resulted in congruent topologies, with the latter (i.e., the ones presented here) being much better resolved.

The *matR*, combined nuclear (*PHYC* and 18S), combined *matR* plus nuclear, and *nad1B-C* data sets included 2042, 2756, 4798, and 3884 bp, across 238, 156, 136, and 147 taxa, respectively. All taxa and voucher information, including GenBank accession numbers, and sequence alignments, for these analyses have been archived on GenBank or is available on TreeBASE (www.treebase.org).

Phylogenetic analysis. Individual and combined parsimony analyses were performed in PAUP* ver. 4.0b10 (*S12*). We restricted our analyses to parsimony partly due to the computational intensity of likelihood and Bayesian methods on such large datasets, but more importantly, because there is not yet a consensus on how to handle likelihood/Bayesian analyses when missing data are included. This is a potential problem given that *PHYC* is absent from some members of the order (*S7*). Combined analyses were performed when there were no strongly supported (≥ 85% bootstrap) incongruent clades between topologies generated from independent data sets (*S13, S14*). We did not analyze *nad1B-C* in

combination because topologies from this locus conflicted strongly regarding the placement of Rafflesiaceae (see also below). With the exception of Rafflesiaceae, however, all individual data sets showed congruent relationships within Malpighiales, including *nad1B-C*.

PHYC and 18S data were analyzed simultaneously as a single nuclear data set. The reason for combining these data sets, as noted in the main text, is because *PHYC* is believed to be absent from at least some families of Malpighiales including Salicaceae and some of their closest relatives (*S7*). 18S includes representatives of all families not sampled for *PHYC*, and by combining these data we ensured that all families of Malpighiales were sampled for the nuclear genome. Each data set independently produced topologically congruent results. With respect to Rafflesiaceae, 18S placed them weakly with Linaceae, and *PHYC* weakly with Ochnaceae. Taxa in the nuclear analysis were only included if both genes were sampled, except for family representatives not available for *PHYC*, in which case 18S was included. We similarly analyzed the combined mt *matR* and nuclear data.

An initial heuristic search of 100 random taxon addition replicates (RAS) was conducted on each data set with tree-bisection-reconnection (TBR) branch swapping and MulTrees on, retaining only ten trees per replicate. The resulting consensus tree was then used as a backbone constraint to search for trees not consistent with the initial trees. These searches were conducted as above with 1,000 RAS. This strategy was employed due to the excessive number of trees generated for unconstrained heuristic searches, and should detect that there are no shorter trees (*S15*, see also *S16*). In all instances, this search strategy failed to

find any trees that were more parsimonious. The strict consensus trees of these most parsimonious trees are presented below (Figs S1–S4).

Parsimony bootstrap percentages (*S17*) for each clade were estimated as above with 1,000 RAS, TBR branch swapping, and saving no more than 10 trees per iteration.

To assess alternate topologies generated from single and combined data set analyses we employed the Templeton (TEMP; *S18, S19*), Kishino-Hasegawa (KH; *S20*), and Shimodaira-Hasegawa (SH; *S21*) tests. Parsimony searches were performed on the constraint tree of interest using the strategy presented above. For the SH test branch lengths were optimized onto competing trees under the preferred model of sequence evolution as determined by a series of hierarchical likelihood ratio tests (*S22, S23*) using Modeltest ver. 3.06 (*S24*). The selected optimal models were all submodels of the general time reversible (GTR) model (*S25*); (*matR* [TIM+Γ], nu [GTR+I+Γ], *matR* plus nuclear [GTR+I+Γ], and *nad1B-C* [SYM+Γ]). For more detailed model parameters see archived data sets.

Each test strongly favored the placement of Rafflesiaceae in the unconstrained trees over alternative topologies. The unconstrained *matR* trees were favored (TEMP, KH: $P = 0.0002$; SH: $P = 0.01$) over the trees grouping Rafflesiaceae with Vitaceae; the unconstrained nuclear trees were favored (TEMP: $P=0.0006$; KH: $P = 0.0007$) over the trees grouping Rafflesiaceae with Vitaceae; the unconstrained combined *matR* and nuclear trees were favored (TEMP, KH: $P < 0.0001$; SH: $P < 0.01$) over the trees grouping Rafflesiaceae with Vitaceae; and last, the unconstrained *nad1B-C* trees were favored (TEMP, KH: $P < 0.0001$; SH: $P < 0.05$) over the constrained trees grouping Rafflesiaceae with

Malpighiales. We were unable to obtain an SH value for the combined nuclear data due to missing *PHYC* sequence data. Less optimal trees were otherwise strongly rejected in all comparisons, including those using the nuclear data set.

Fig. S1. Strict consensus of 30 equally parsimonious trees based on mt *matR* sequence data. Bootstrap values are given for those clades supported at >50%. L = 4706; CI = 0.4533; RI = 0.7028. Malpighiales (sensu *S1*) in blue; Rafflesiaceae in red; Vitaceae in yellow.

Fig. S2. Strict consensus of 10 equally parsimonious trees based on combined nuclear *PHYC* and ribosomal 18S sequence data. Bootstrap values are given for those clades supported at >50%. L = 12871; CI = 0.1962; RI = 0.4727. Malpighiales (sensu *S1*) in blue; Rafflesiaceae in red; Vitaceae in yellow.

Fig. S3. Strict consensus of 60 equally parsimonious trees based on combined *matR* and nuclear sequence data. Bootstrap values are given for those clades supported at >50%. L = 12047; CI = 0.3202; RI = 0.5124. Malpighiales (sensu *S1*) in blue; Rafflesiaceae in red; Vitaceae in yellow. Tree summarized as Fig 1A in main text.

Fig. S4. Strict consensus of 10 equally parsimonious trees based on mt *nad1B-C* sequence data. Bootstrap values are given for those clades supported at >50%. L = 3724; CI = 0. 6856; RI = 0. 7112. Malpighiales (sensu *S1*) in blue; Rafflesiaceae in red; Vitaceae in yellow. Tree summarized as Fig 1B in main text.

Fig. S5. Phylograms for combined *matR* and nuclear sequence data (A) and mitochondrial *nad1B-C* (B). Each tree represents one of the most parsimonious trees from the pool of most parsimonious trees summarized in S3 and S4,

respectively. Malpighiales (sensu *S1*) in blue; Rafflesiaceae in red; Vitaceae in yellow.

References

S1. APGII, *Bot. J. Linn. Soc.* **141**, 399 (2003).

S2. T. J. Barkman, S.-K. Lim, K. Mat Salleh, J. Nais, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 787 (2004).

S3. D. E. Soltis *et al.*, *Amer. J. Bot.* **90**, 461 (2003).

S4. S. Mathews, M. J. Donoghue, *Science* **286**, 947 (1999).

S5. M. J. Zanis, D. E. Soltis, P. S. Soltis, S. Mathews, M. J. Donoghue, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6848 (2002).

S6. C. C. Davis, C. D. Bell, P. W. Fritsch, S. Mathews, *Evolution* **56**, 2395 (2002).

S7. C. C. Davis, M. W. Chase, *Amer. J. Bot.* **91**, 262 (2004).

S8. J. V. Freudenstein, M. W. Chase, *Syst. Bot.* **26**, 643 (2001).

S9. D. E. Soltis, P. S. Soltis, *Amer. J. Bot.* **84**, 504 (1997).

S10. C. C. Davis, *Amer. J. Bot.* **89**, 699 (2002).

S11. D. L. Nickrent, E. M. Starr, *J. Mol. Evol.* **39**, 62 (1994).

S12. D. L. Swofford, *Paup*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.* (Sinauer Associates, Sunderland, MA, 2003).

S13. W. M. Whitten, N. H. Williams, M. W. Chase, *Amer. J. Bot.* **87**, 1842 (2000).

S14. G. Reeves *et al.*, *Amer. J. Bot.* **88**, 2074 (2001).

S15. P. Catalán, E. A. Kellogg, R. G. Olmstead, *Mol. Phylog. Evol.* **8**, 150 (1997).

S16. C. C. Davis, W. R. Anderson, M. J. Donoghue, *Amer. J. Bot.* **88**, 1830 (2001).

S17. J. Felsenstein, *Evolution* **39**, 783 (1985).

S18. A. R. Templeton, *Evolution* **37**, 221 (1983).

S19. A. Larson, in *Molecular Ecology and Evolution: Approaches and Applications* B. Schierwater, B. Streit, G. P. Wagner, R. DeSalle, Eds. (Birkhauser Verlag, Basel, Switzerland, 1994) pp. 371-390.

S20. H. Kishino, M. Hasegawa, *J. Mol. Evol.* **29**, 170 (1989).

S21. K. Kadowaki, N. Kubo, K. Ozawa, A. Hirai, *EMBO J.* **15**, 6652 (1996).

S22. J. Felsenstein, *J. Mol. Evol.* **17**, 368 (1981).

S23. J. P. Huelsenbeck, B. Rannala, *Science* **276**, 227 (1997).

S24. D. Posada, K. A. Crandall, *Bioinformatics* **14**, 817 (1998).

S25. F. Rodríguez, J. F. Oliver, A. Marín, J. R. Medina, *J. Theor. Biol.* **142**, 485 (1990).

Fig. S1. *matR*

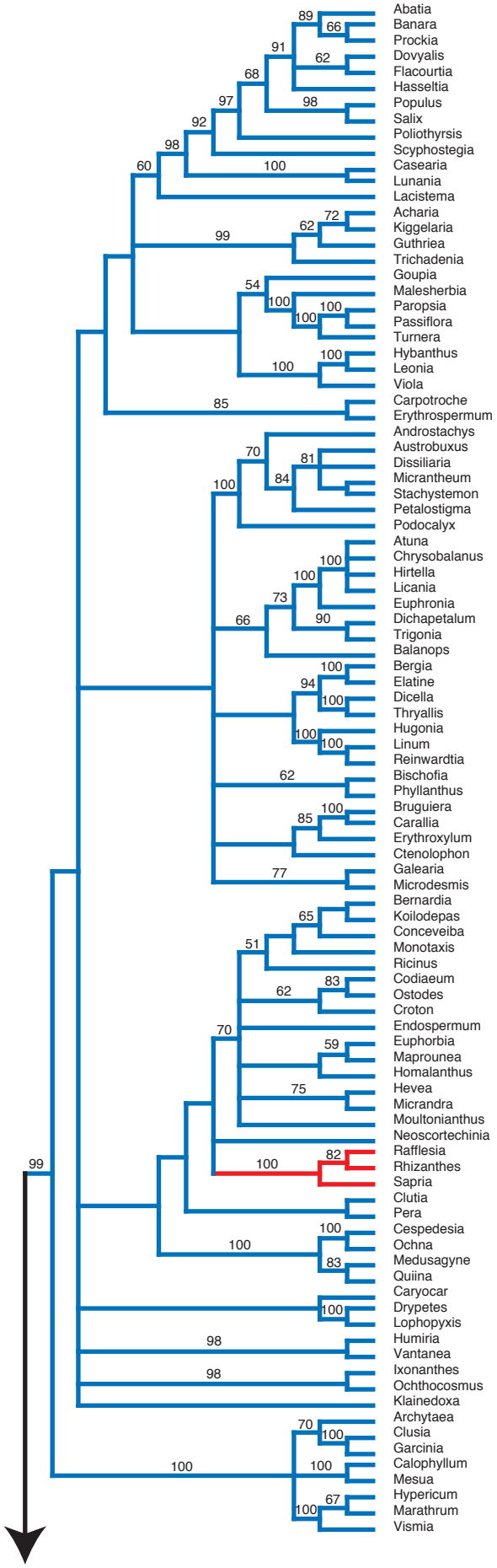

Abatia
Banara
Prockia
Dovyalis
Flacourtia
Hasseltia
Populus
Salix
Poliothyrsis
Scyphostegia
Casearia
Lunania
Lacistema
Acharia
Kiggelaria
Guthriea
Trichadenia
Goupia
Malesherbia
Paropsia
Passiflora
Turnera
Hybanthus
Leonia
Viola
Carpotroche
Erythrospermum
Androstachys
Austrobuxus
Dissiliaria
Micrantheum
Stachystemon
Petalostigma
Podocalyx
Atuna
Chrysobalanus
Hirtella
Licania
Euphronia
Dichapetalum
Trigonia
Balanops
Bergia
Elatine
Dicella
Thryallis
Hugonia
Linum
Reinwardtia
Bischofia
Phyllanthus
Bruguiera
Carallia
Erythroxylum
Ctenolophon
Galearia
Microdesmis
Bernardia
Koilodepas
Conceveiba
Monotaxis
Ricinus
Codiaeum
Ostodes
Croton
Endospermum
Euphorbia
Maprounea
Homalanthus
Hevea
Micrandra
Moultonianthus
Neoscortechinia
Rafflesia
Rhizanthes
Sapria
Clutia
Pera
Cespedesia
Ochna
Medusagyne
Quiina
Caryocar
Drypetes
Lophopyxis
Humiria
Vantanea
Ixonanthes
Ochthocosmus
Klainedoxa
Archytaea
Clusia
Garcinia
Calophyllum
Mesua
Hypericum
Marathrum
Vismia

Fig. S1. *matR* (cont. 2)

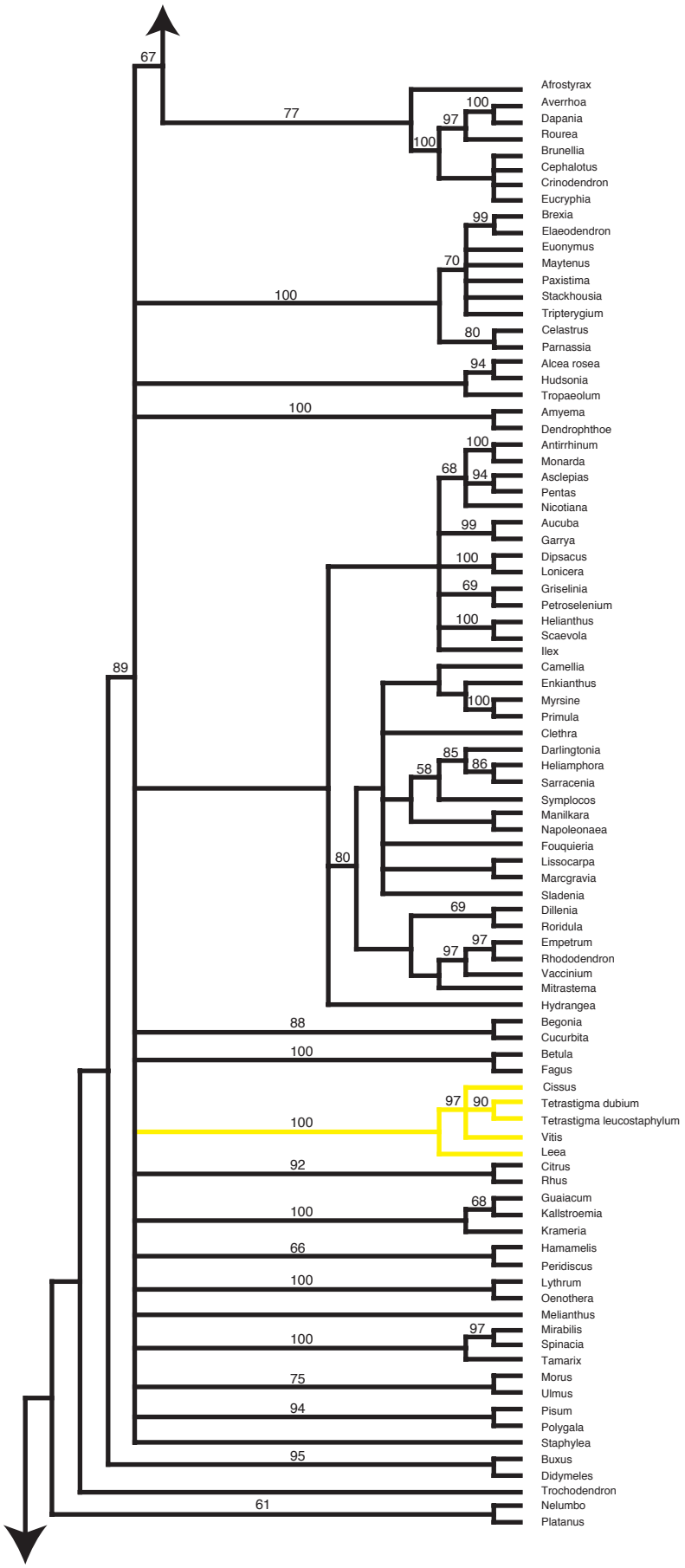


Fig. S1. *matR* (cont. 3)

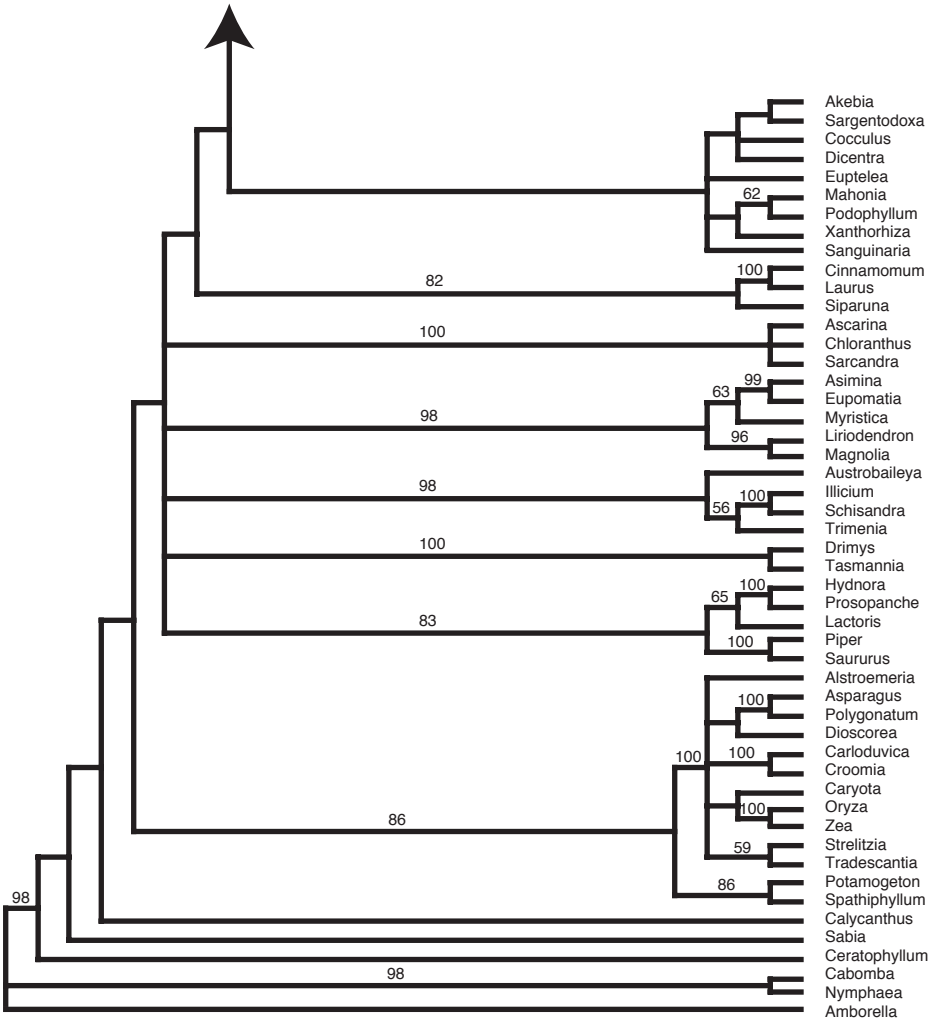


Fig. S2. nuclear

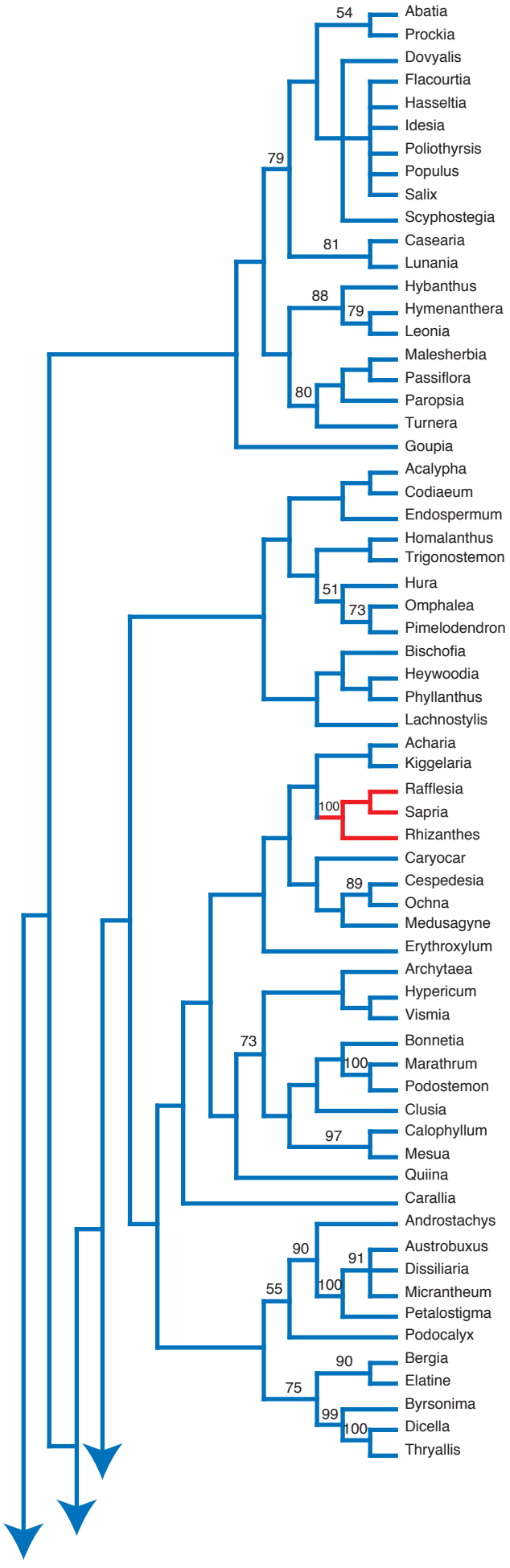


Fig. S2. nuclear (cont. 2)

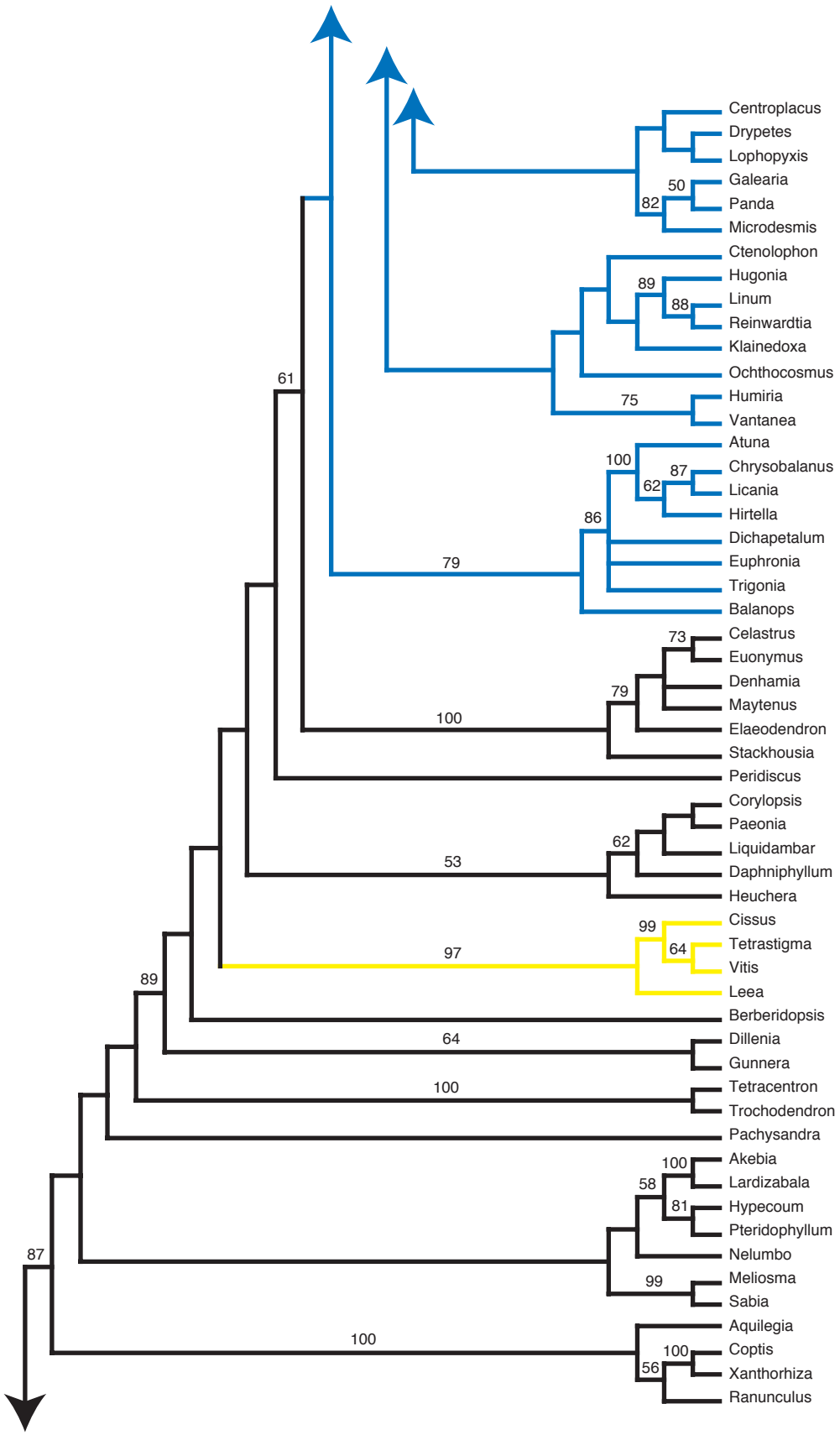


Fig. S2. nuclear (cont. 3)

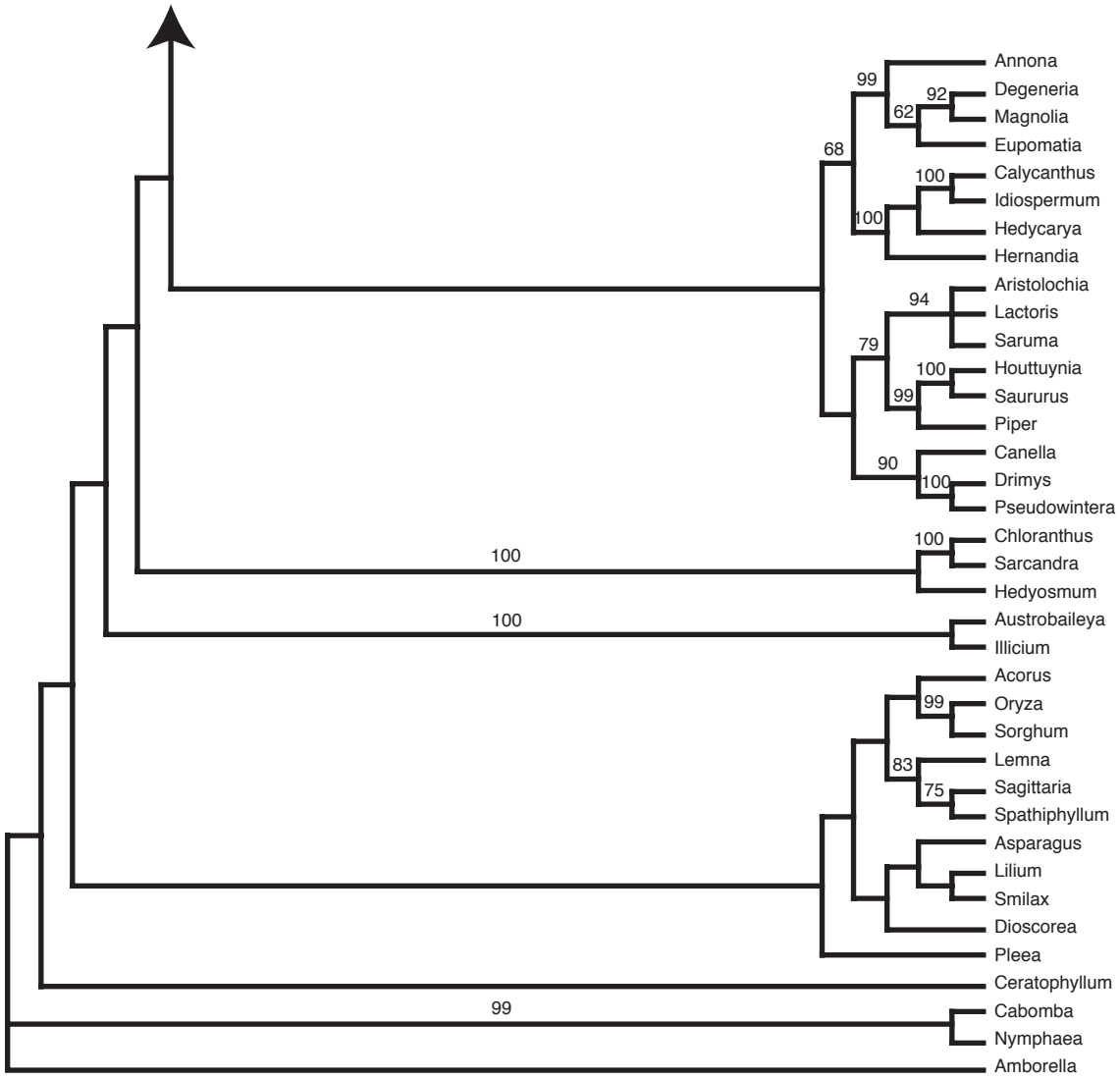


Fig. S3. *matR* + nuclear

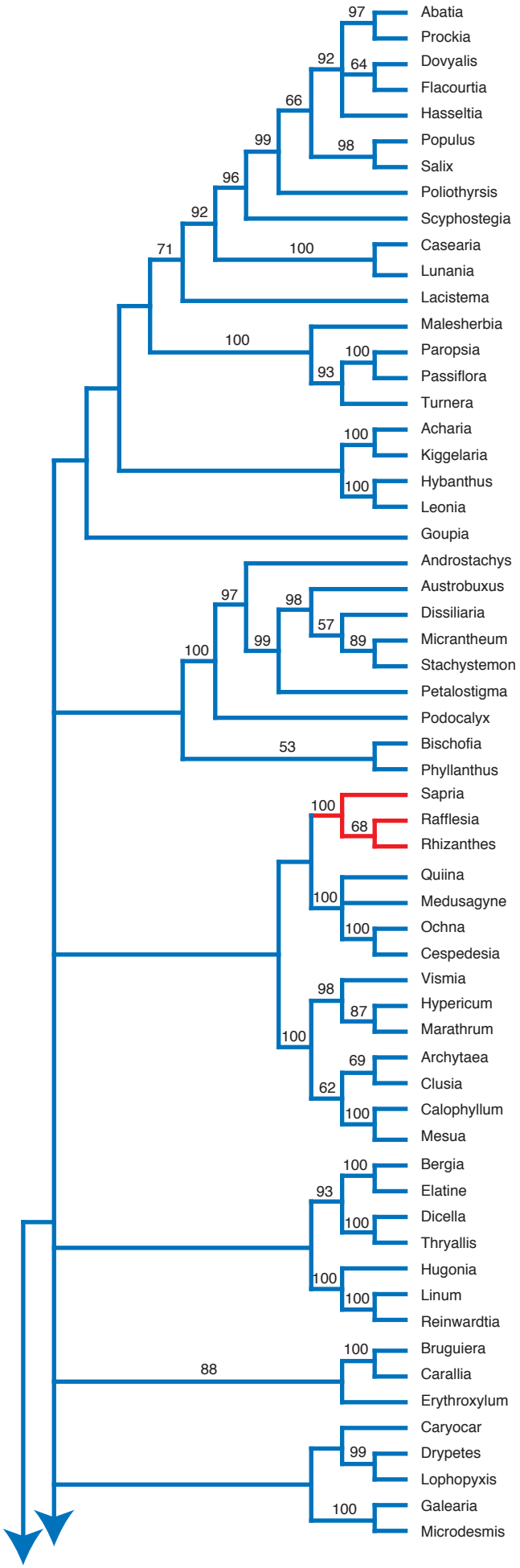


Fig. S3. *matR* + nuclear (cont. 2)

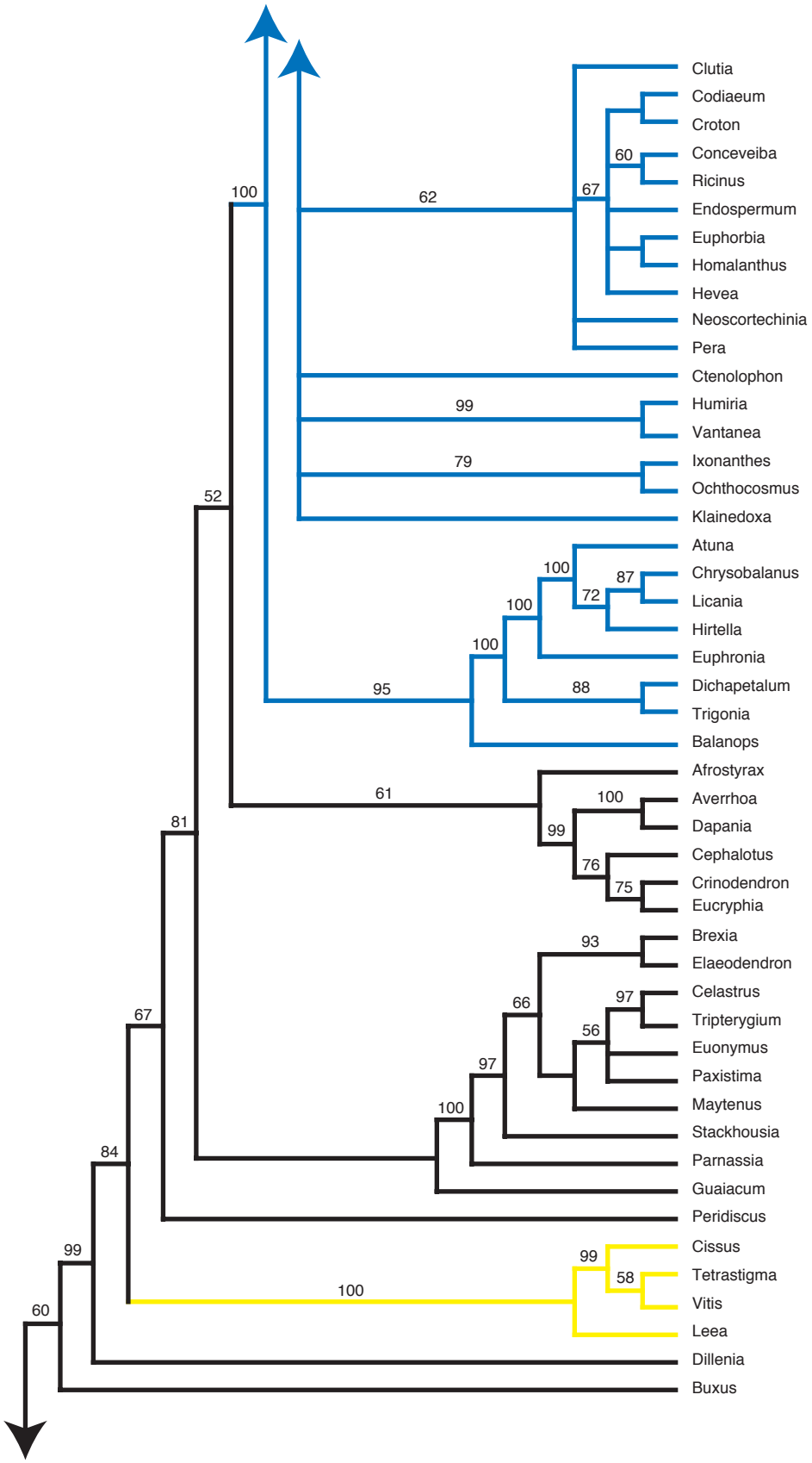


Fig. S3. *matR* + nuclear (cont. 3)

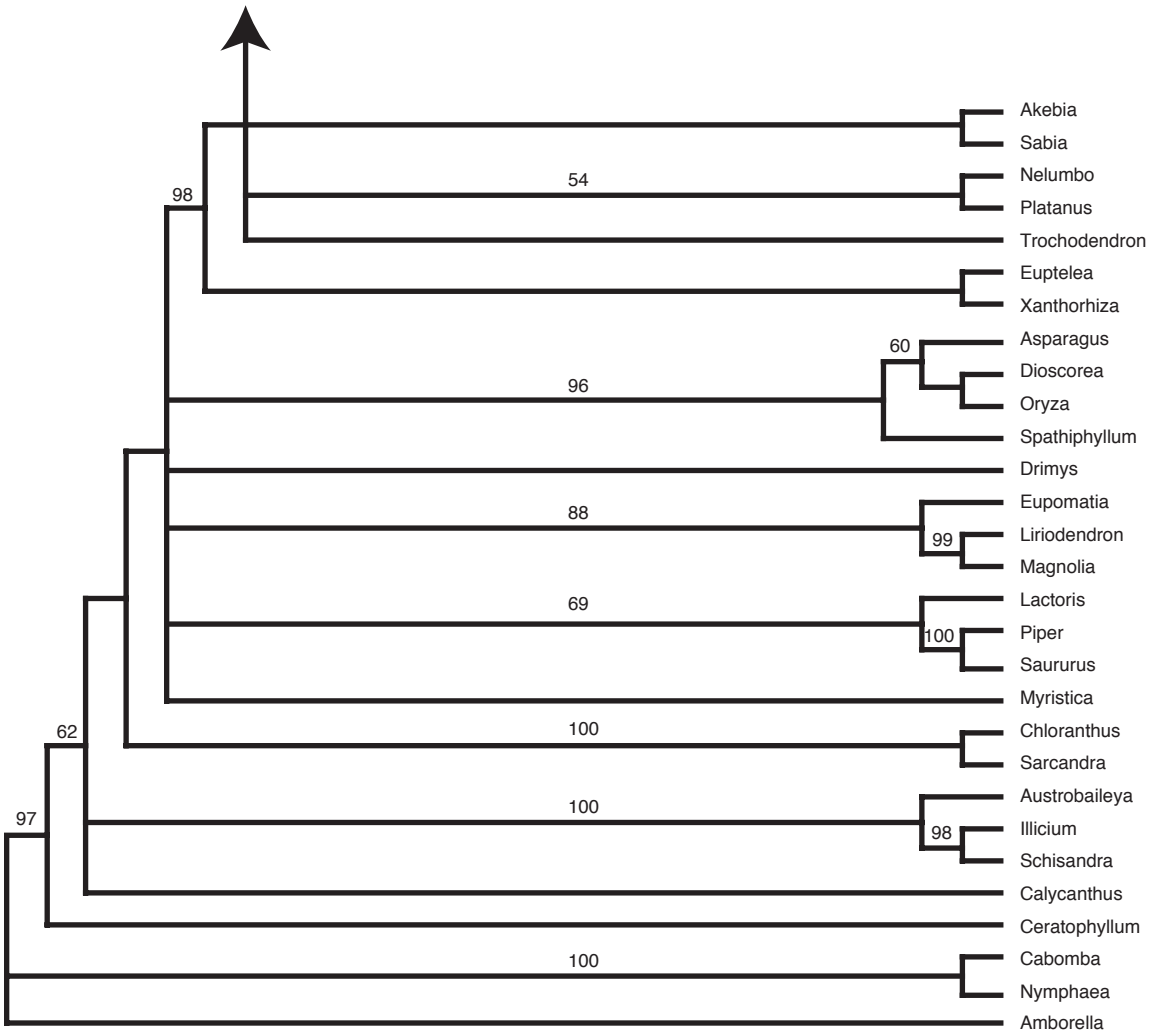


Fig. S4. *nad1B-C*

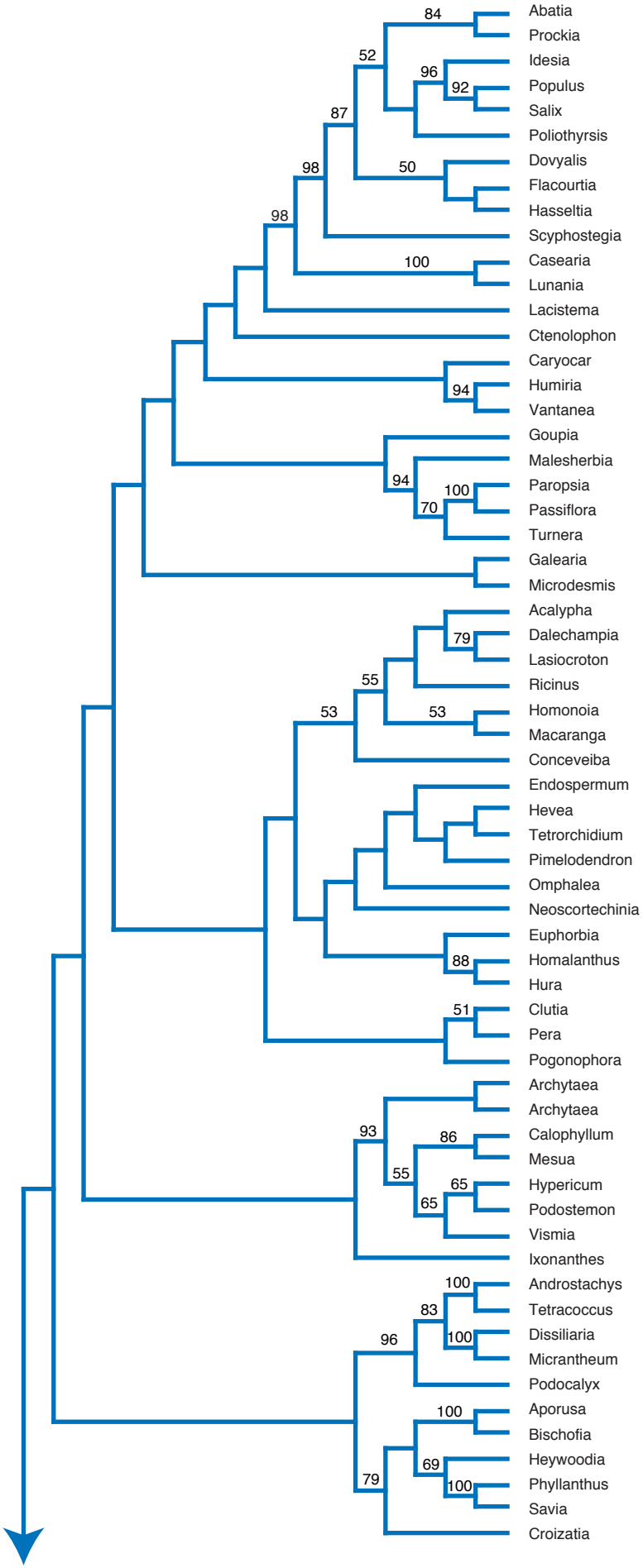


Fig. S4. *nad1B-C* (cont. 2)

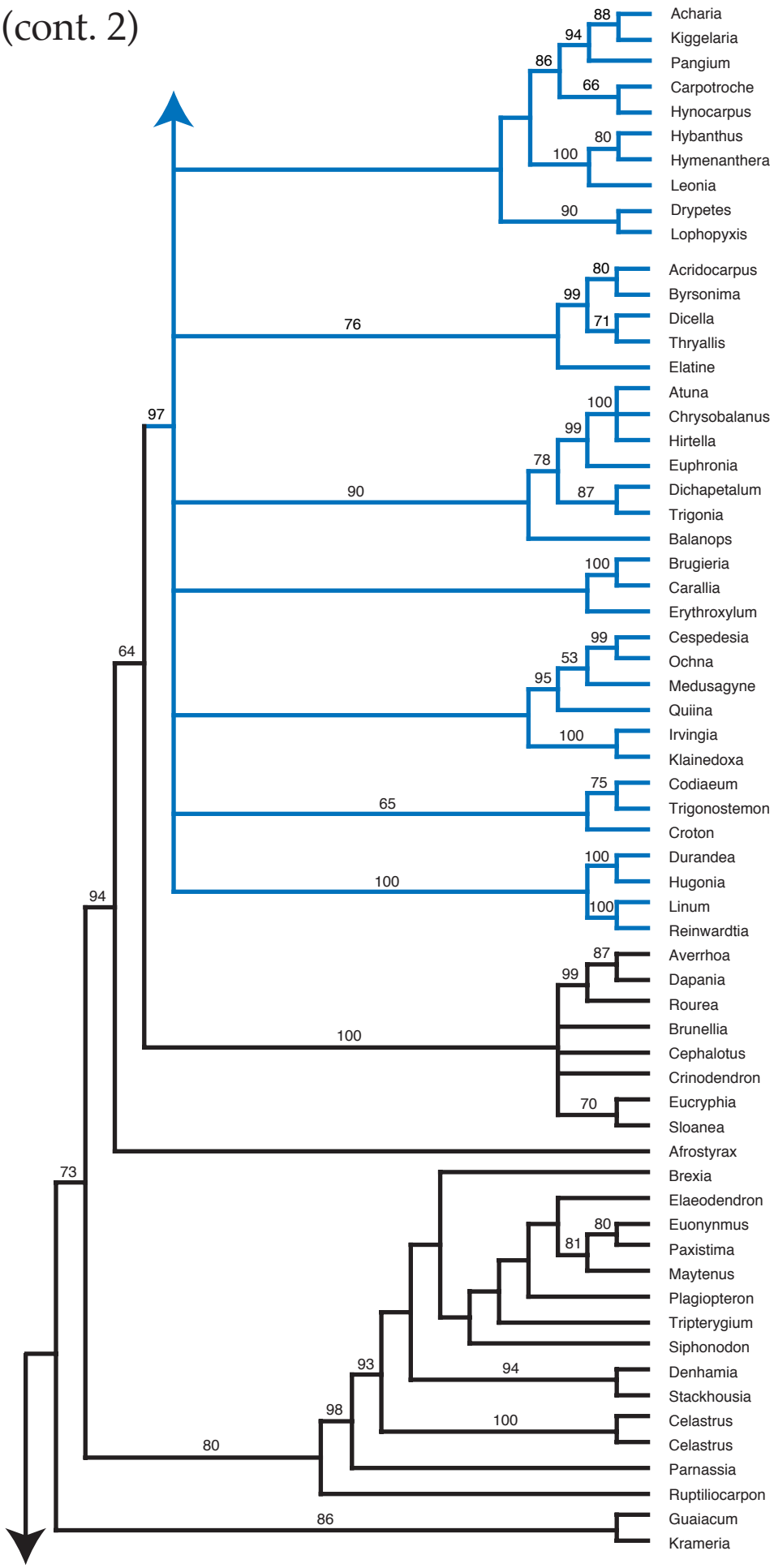
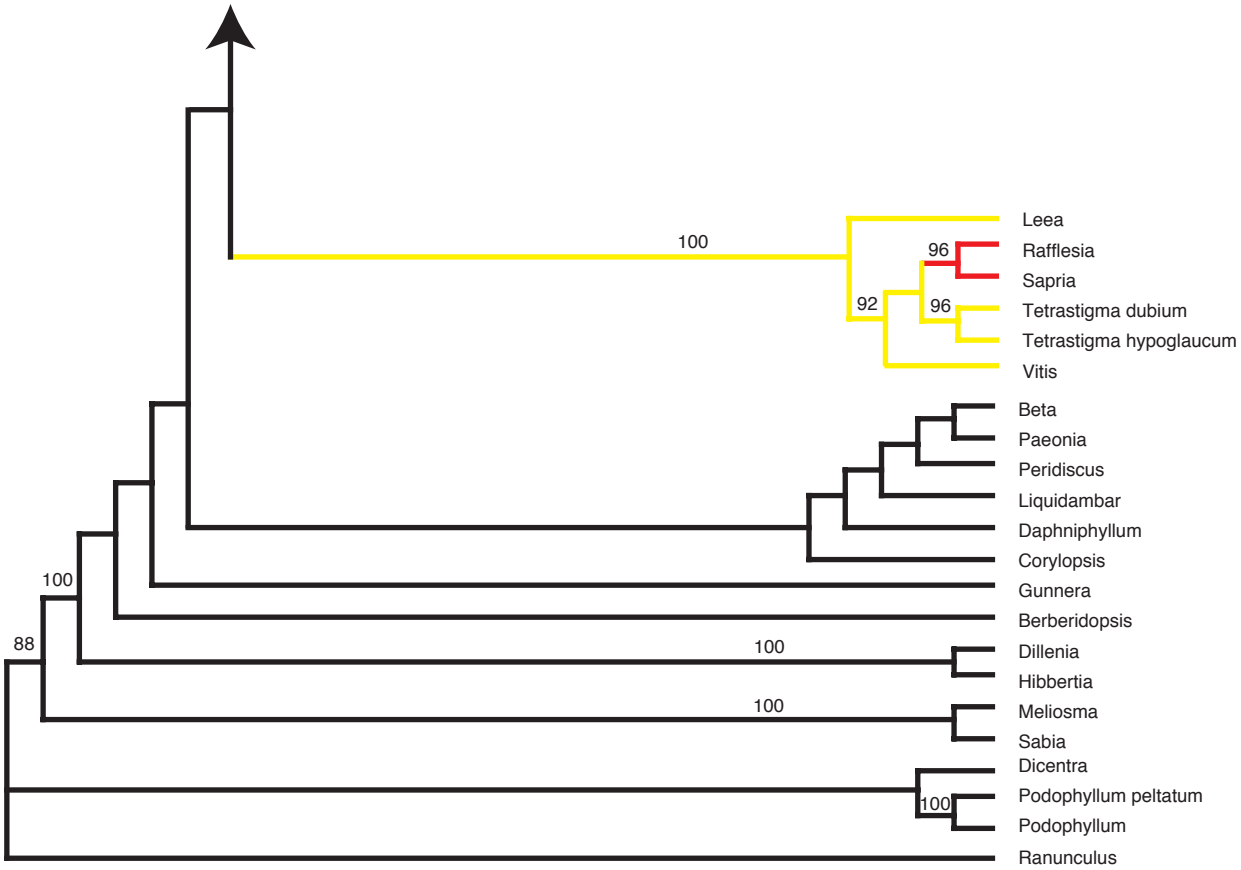


Fig. S4. *nad1B-C* (cont. 3)




A
50 changes
B
10 changes