ELECTRONIC APPENDIX

This is the Electronic Appendix to the article

Gene transfer from a parasitic flowering plant to a fern

by

Charles C. Davis, William R. Anderson & Kenneth J. Wurdack

Proc. R. Soc. B (doi:10.1098/rspb.2005.3226)

Supplementary on-line Material for **Gene transfer from a parasitic flowering plant to a fern** (C. C. Davis, W.R. Anderson, & K. J. Wurdack)

Materials and Methods

Gene Isolation and Sequencing. Twenty-one, 60, 73, 81, and four sequences for *atp1*, *atp6*, *nad1B-C*, *matR*, and *rbcL* were newly obtained for this study (GenBank numbers DQ110142–DQ110380: *atp1*, DQ110142–DQ110222; *nad1B-C*, DQ110223–DQ110295; *matR*, DQ110296–DQ110376; *rbcL*, DQ110377–DQ110380), respectively. Otherwise, sequences were obtained from GenBank. Total cellular DNA was prepared following Davis *et al.* (2002). Amplification and automated sequencing protocols for *matR* and *nad1B-C* followed Davis & Wurdack (2004). These protocols were similarly applied to *atp1*, *atp6*, and *rbcL* using locus-specific primer combinations atp6f (5'-CTTTGTTTATGCTRCTMACTCTCAG-3') and atp6r (5'-TTGATGGAGATTTATAGCATCATTC-3') for *atp6*; AF15 (5'-GTTGRAGAT GGRATTGCMCG-3') and AR18 (5'-C AAAAGCGGCHACTTCRCG-3') for *atp1*; and published primer sequences 26f and 1379r for *rbcL* (Little & Barrington 2003). Amplification of fern *matR* as a single fragment used matR5f-nsp (5'-CTATCACCCCTRCTRTGCAAYATYTA-3') and matR3r-nsp (5'-CTRGTAGTCGACAATCGYTCGGAC-3'); amplification as two halves used matR5f-nsp with R1 (5'-GAAGCWGTKAKATGGATACGGTGCT-3') and matR3r-nsp with F2 (5'-ATCCGRCCTGAACCTTKRGGTAGGCTC-3') or F3 (5'-GARTAAAGCACCGTATCCATCTCA-3'). Additional internal sequencing primers for

this region were F1 (5'-GACCCAGTAAATTGCCAACTCCAC-3') and R2 (5'-GTTGGCTTTACRAATGAAGGC-3'). The universal primers F2 and matR3r-nsp were used to co-amplify native and transgene *matR* in a single PCR reaction. Direct sequencing of PCR products from this primer combination (i.e., F2 and matR3r-nsp) yielded unambiguous chromatograms for all ferns sampled except *B. virginianum*. The latter had overlapping peaks indicative of multiple copies, which were cloned to verify native (i.e., fern) and transgenic (i.e., angiosperm/santalalean) *matR*. Using the same strategy as above for Santalales, we isolated only native angiosperm *matR* from these samples. We also sequenced four clones from four individuals of *B. viriginianum* for transgenic *matR* and *nad1B-C* and found very little sequence heterogeneity between clones, suggesting that there are no divergent copies of these transferred gene regions.

Functionality of transferred genes. Total RNA was extracted with the RNeasy Plant Mini Kit (Qiagen, Valencia, California) and treated with DNase I (DNA-free; Ambion, Austin, Texas) to degrade contaminating genomic DNA. Reverse transcription used Omniscript Reverse Transcriptase (Qiagen) and *matR* specific primers followed by amplfication and direct sequencing of the resulting PCR products. The authenticity of the cDNA sequences derived from RT-PCR was established by the lack of detectable genomic DNA contamination in minus-RT controls. No RNA edited sites were detected in the native (i.e., fern) transcribed copy of *matR* we sequenced.

Phylogenetic analyses. Nucleotide and, where appropriate, amino acid sequences were aligned by eye; the ends of sequences, as well as ambiguous internal regions composed

mostly of overlapping insertions, were trimmed from each data set to maintain complementary data between taxa. The aligned *atp1*, *atp6*, *nad1B-C*, *matR*, and *rbcL* data sets included 1264, 643, 1712, 1927, and 1206 base pairs (bp), across 145, 74, 100, 208, and 118 taxa, respectively. We replaced a 1525 bp region (2064 bp in *Botrychium* subgenus *Sceptridium*) in the *nad1B-C* alignment with ?'s in the non-angiosperm taxa because this region was particularly divergent relative to angiosperms.

The general time reversible model (GTR), or a submodel of the GTR, was selected for maximum likelihood (ML) optimization and implemented in PHYML *ver*. 2.4.4 (Guindon & Gascuel 2003). For the *atp1* data, the GTR+G+I model was selected with gamma distribution of rates among sites with alpha shape 0.6572, and proportion of invariant sites 0.2864. For the *atp6* data, the GTR+G model was selected with gamma distribution of rates among sites with alpha shape 0.5384. For the *matR* data, the K81uf+G model was selected with gamma distribution of rates among sites with alpha shape 1.2897. For the *nad1B-C* data, the K81uf+G model was selected with gamma distribution of rates among sites with alpha shape 0.9460. For the combined data set, the K81uf+G model was selected with gamma distribution of rates among sites with alpha shape 0.8053. For the *rbcL* data, the GTR+G+I model was selected with gamma distribution of rates among sites with alpha shape 0.9261, and proportion of invariant sites 0.4022. In all cases, the optimal rate of nucleotide substitution and transition/transversion ratios were estimated from the data during ML searches.

Topologies generated from *nad1B-C* and *matR* were congruent (Whitten *et al.* 2000; Reeves *et al.* 2001), and were analyzed in combination to better assess the placement of *B. virginianum* in Santalales. This data set contained 70 taxa (3821 bp)

spanning most major eudicot lineages and was rooted with the monocot *Cypripedium*. Both gene regions were sampled for all taxa with the exception of Cucurbitaceae, Fabaceae, and Rubiaceae in which cases two different genera from the same family served as placeholders for these clades. The branching pattern inferred for Santalales in this analysis agreed with all independent analyses in this study, and with other published topologies for the order (Nickrent & Malécot 2001; Malécot *et al.* 2004). A combined analysis using non-angiosperms as outgroups yielded topologically identical results for Santalales (results not shown).

The Kishino-Hasegawa (Kishino & Hasegawa 1989), Shimodaira-Hasegawa (Kadowaki *et al.* 1996), and parametric bootstrap tests (Huelsenbeck *et al.* 1996) were used to examine the robustness of the hypothesis of vertical versus horizontal gene transmission of *B. virginianum* for all four mt gene regions. For the parametric bootstrap the significance of the difference in likelihood scores between constrained and unconstrained trees was determined by comparing the distribution of 100 simulated data sets. To create these simulated data we used Seq-Gen *ver*. 1.2.7 (Rambaut & Grassly 1997) under the optimal models of sequence evolution and the constrained tree topology. One hundred new data matrices were sampled and analyzed with ML. In each replicate, the tree length of the most likely solution was compared to that of the constraint tree; the resulting distribution of these differences was then plotted as a histogram and compared to the observed difference.

Complete ML trees with parsimony and likelihood support values are shown below for *nad1B-C* (Fig. 1), *matR* (Fig. 2), *atp1* (Fig. 3), *atp6* (Fig. 4), combined *matR* and *nad1B-C* (Fig. 5), and *rbcL* (Fig. 6).

Biogeographic Analyses. Our data matrix used to assess ancestral areas using dispersal-vicariance analysis (DIVA) is shown in Fig. 7. The number of ancestral areas was restricted to two per node in DIVA analyses following the suggestions of Ronquist (1997), Donoghue *et al.* (2001), and Davis *et al.* (2002), but identical results were obtained for the node of interest with max areas set to six.

ML topologies from mt *atp1*, *atp6*, and plastid *rbcL* were nearly identical to the ones published by Hauk *et al.* (2003) for Ophioglossaceae. The only major topological difference between these analyses regards the placement of *Botrychium lanuginosum* (subgenus *Japanobotrychium*). The study by Hauk *et al*. (2003) did not include this taxon and our data suggested two alternate placements for this species. The *rbcL* gene (Fig. 6) placed it moderately (i.e., 65%/58% likelihood and parsimony bootstrap support, respectively) as sister to *Botrychium* subgenus *Sceptridium,* whereas *atp6* (Fig. 4) placed it strongly (93%/91% likelihood and parsimony bootstrap support, respectively) as sister to subgenera *Sceptridium* + *Botrychium*. Since the latter well-supported placement is also suggested on the basis of morphology (Hauk *et al.* 2003) we assessed ancestral areas in DIVA under this placement, rather than the moderately supported placement based on *rbcL*. In all cases ancestral area reconstructions using DIVA and MacClade suggest that *B. virginianum* diverged from its closest relatives in Asia (Fig. 7).

We used the ML topology shown in Fig. 6 to test for rate constancy among lineages under the assumption of a molecular clock. Fossil age constraints described in the main text of the paper are shown in Fig. 8.

REFERENCES

Davis, C. C., Bell, C. D., Fritsch, P. W. & Mathews, S. 2002 Phylogeny of *Acridocarpus-Brachylophon* (Malpighiaceae): implications for Tertiary tropical floras and Afroasian biogeography. *Evolution* **56**, 2395-2405.

Davis, C. C. & Wurdack, K. J. 2004 Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. *Science* **305**, 676-678.

Donoghue, M. J., Bell, C. D. & Li, J. 2001 Phylogenetic patterns in Northern Hemisphere plant geography. *Int. J. Pl. Sci.* **162**, S41-S52.

Guindon, S. & Gascuel, O. 2003 A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696-704.

Hauk, W. D., Parks, C. R. & Chase, M. W. 2003 Phylogenetic studies of Ophioglossaceae: evidence from *rbcL* and *trnL-F* plastid DNA sequences and morphology. *Mol. Phylog. Evol.* **28**, 131-151.

Huelsenbeck, J. P., Hillis, D. M. & Jones, R. 1996 Parametric bootstrapping in molecular phylogenetics: applications and performance. In *Molecular zoology: advances, strategies, and protocols* (eds. J. D. Ferraris & S. R. Palumbi), pp. 19–45. New York: Wiley-Liss.

Kadowaki, K.-i., Kubo, N., Ozawa, K. & Hirai, A. 1996 Targeting presequence acquisition after mitochondrial gene transfer to the nucleus occurs by duplication of existing targeting signals. *EMBO J.* **15**, 6652-6661.

Kishino, H. & Hasegawa, M. 1989 Evaluation of the maximum likelihood estimates of the evolutionary tree topologies from sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29**, 170-179.

Little, D. P. & Barrington, D. S. 2003 Major evolutionary events in the origin and diversification of the fern genus *Polystichum* (Dryopteridaceae). *Amer. J. Bot.* **90**, 508-514.

Malécot, V., Nickrent, D. L., Baas, P., Oever, L. v. d. & Lobreau-Callen, D. 2004 A morphological cladistic analysis of Olacaceae. *Syst. Bot.* **29**, 569-586.

Nickrent, D. L. & Malécot, V. 2001 A molecular phylogeny of Santalales. In *Proceedings of the 7th. International Parasitic Weed Symposium* (eds. A. Fer, P. Thalouarn, D. M. Joel, L. J. Musselman, C. Parker & J. A. C. Verkleij), pp. 69-74. Nantes, France: Faculté des Sciences, Université de Nantes.

Rambaut, A. & Grassly, N. C. 1997 Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* **13**, 235-238.

Reeves, G., Chase, M. W., Goldblatt, P., Rudall, P., Fay, M. F., Cox, A. V., Lejeune, B. *et al.* 2001 Molecular systematics of Iridaceae: evidence from four plastid DNA regions. *Amer. J. Bot.* **88**, 2074-2087.

Ronquist, F. 1997 Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Syst. Biol.* **46**, 195-203.

Rothwell, G. W. & Stockey, R. A. 1989 Fossil Ophioglossales in the Paleocene of Western North America. *Amer. J. Bot.* **76**, 637-644.

Sanderson, M. J. 2003 Molecular data from 27 proteins do not support a Precambrian origin of land plants. *Amer. J. Bot.* **90**, 954-956.

Schneider, H., Schuettpelz, E., Pryer, K. M., Cranfill, R., Magallón, S. & Lupia, R. 2004 Ferns diversified in the shadow of angiosperms. *Nature* **428**, 553-557.

Whitten, W. M., Williams, N. H. & Chase, M. W. 2000 Subtribal and generic relationships of Maxillarieae (Orchidaceae) with emphasis on Stanhopeinae: combined molecular evidence. *Amer. J. Bot.* **87**, 1842-1856.

Fig. 1. Full maximum likelihood topology of mt *nad1B-C* sequence data for main manuscript Fig. 1a. Santalales are depicted in red, Ophioglossaceae in blue. Transgenic copy of *nad1B-C* shown in Santalales in blue. Likelihood and parsimony bootstrap values, respectively, are given for those clades supported at >50%. Asterisks indicate <50% bootstrap support. Topology with branch lengths shown in inset. Scale bar subtending inset corresponds to 0.1 nucleotide substitutions per site.

Fig. 2. Full maximum likelihood topology of mt *matR* sequence data for main manuscript Fig. 1b. Santalales are depicted in red, Ophioglossaceae in blue. Transgenic copy of *matR* shown in Santalales in blue. Likelihood and parsimony bootstrap values, respectively, are given for those clades supported at >50%. Asterisks indicate <50% bootstrap support. Topology with branch lengths shown in inset. Scale bar subtending inset corresponds to 0.1 nucleotide substitutions per site.

Fig. 3. Full maximum likelihood topology of mt *atp1* sequence data for main manuscript Fig. 1c. Santalales are depicted in red, Ophioglossaceae in blue. Likelihood and parsimony bootstrap values, respectively, are given for those clades supported at >50%. Asterisks indicate <50% bootstrap support. Topology with branch lengths shown in inset. Scale bar subtending inset corresponds to 0.1 nucleotide substitutions per site.

Fig. 4. Full maximum likelihood topology of mt *atp6* sequence data for main manuscript Fig. 1d. Santalales are depicted in red, Ophioglossaceae in blue. Likelihood and parsimony bootstrap values, respectively, are given for those clades supported at >50%. Asterisks indicate <50% bootstrap support. Topology with branch lengths shown in inset. Scale bar subtending inset corresponds to 0.1 nucleotide substitutions per site.

Fig. 5. Full maximum likelihood topology of combined mt *matR* and *nad1B-C* sequence data for main manuscript Fig. 2. Santalales are depicted in red. Transgenic *B. virgininaum* shown in Santalales in blue. Likelihood and parsimony bootstrap values, respectively, are given for those clades supported at >50%. Asterisks indicate <50% bootstrap support. Topology with branch lengths shown in inset. Scale bar subtending inset corresponds to 0.1 nucleotide substitutions per site.

Fig. 6. Maximum likelihood topology of plastid *rbcL* sequence data. Likelihood and parsimony bootstrap values, respectively, are given for those clades supported at >50%. Asterisks indicate <50% bootstrap support. Topology with branch lengths shown in inset. Scale bar subtending inset corresponds to 0.1 nucleotide substitutions per site. This topology was also used for molecular divergence time estimation. Node of interest for DIVA and divergence time estimates indicated with a star (see Figs 7 and 8).

Fig. 7. DIVA analysis. Ancestral areas optimized onto the nodes of Ophioglossaceae

using the ML topology derived from plastid *rbcL* data. Scoring for extant distributions of Ophioglossaceae shown next to taxon names. Abbreviations for taxon names as follows: Bo=*Botrychium*; Op=*Ophioglossum;* He=*Helminthostachys*.

Fig. 8. Fossil constraints. Minimum fossil age constraints of major fern and land plant lineages, labeled nodes A2-A30, taken directly from Schneider *et al.* (2004). The two maximum age constraints of 380 mya and 450 mya for node A1 are taken from Schneider *et al.* (2004) and Sanderson (2003), respectively. The minimum age constraint for stem group *Botrychium* (node C1) is from Rothwell & Stockey (1989).

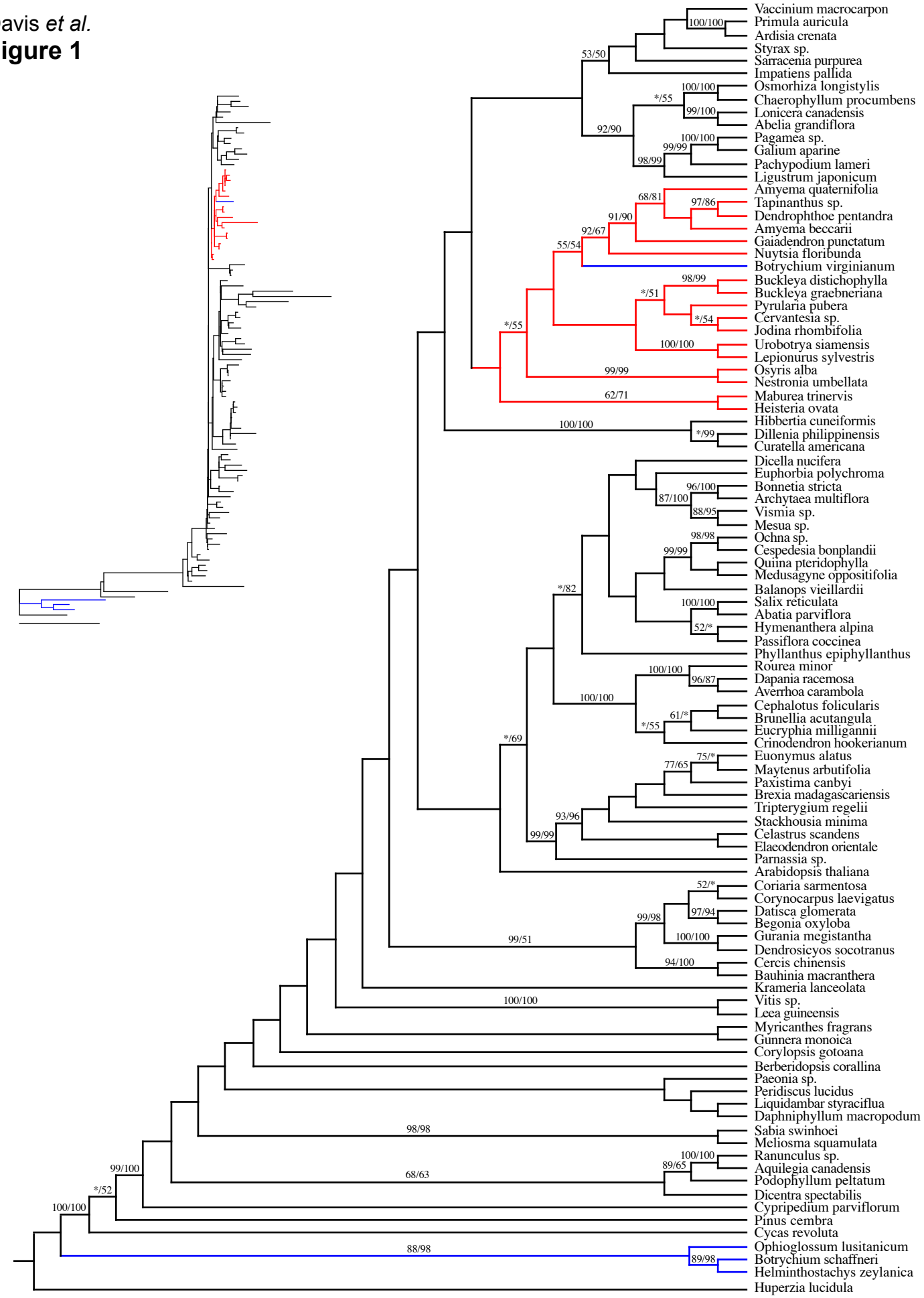
Table 1. Individuals of *Botrychium virginianum* sampled for transgenic mt regions *matR* and *nad1B-C*. Herbarium acronyms from *Index Herbariorum*, http://sciweb.nybg.org/science2/IndexHerbariorum.asp. USA= United States of America. GenBank accessions provided for sampled transgenic *matR* and *nad1B-C*.

Population locality	**Voucher**	***matR***	***nad1B-C***
Asia			
China: Henan	D.E. Boufford et al. 26167 (A)	DQ110301	DQ110230
Japan: Aomori	K. Yonekura 6862 (A)	DQ110302	DQ110231
Japan: Miyagi, Sendai-shi	K. Yonekura 992 (A)	DQ110303	DQ110232
Japan: Miyagi, Kesennuma	D.E. Boufford et al. 25368 (A)	DQ110304	DQ110233
Japan: Nagasaki	K. Yonekura 96199 (A)	DQ110305	DQ110234
Japan: Shizuoka	T. Nakaike & M. Sakakibara 129 (M)		DQ110235
Central America			
Costa Rica: San José	A.M. Evans & D. B. Lellinger 229 (US)	DQ110306	DQ110236
Panama: Chiriquí	D.B. Lellinger 1904 (US)	DQ110307	DQ110237
Europe			
Austria: Styria	O. Angerer s.n. (M)	DQ110308	DQ110238
Germany: Bavaria	J. Pfadehauser s.n. (M)	DQ110309	
Switzerland: Prättigau	R. Göldi 7247 (Z)		DQ110239
Greater Antilles			
Dominican Republic: Pedernales	S.A. Thompson 7401 (NY)	DQ110310	DQ110240
North America			
Canada: Alberta	W.H. Wagner 83320 (MICH)	DQ110311	DQ110241
Canada: Ontario	W.H. Wagner 93016 (MICH)	DQ110312	DQ110242
Canada: Saskatchewan	V.L. & R.M. Harms 35604 (GH)		DQ110243
Mexico: Nayarit	J.I. Calzada et al. 19056 (GH)		DQ110244

USA: Alaska, Simeonof Island	S.S. Talbot 92 (MICH)	DQ110313	DQ110245
USA: Alaska, Simeonof Island	S.S. Talbot 363 (MICH)	DQ110314	DQ110246
USA: Idaho	J.F. Smith 3108 (MICH)	DQ110315	DQ110247
USA: Illinois	L.R. Phillipe & E. Ulaszek 14014 (MICH)	DQ110316	DQ110248
USA: Iowa	D.R. Farrar s.n. (MICH)	DQ110317	DQ110249
USA: Kentucky, Floyd County	W.H. Wagner 85062 (MICH)	DQ110318	DQ110250
USA: Kentucky, Rockcastle County	D.D. Taylor 3726 (M)	DQ110319	DQ110251
USA: Louisiana	W.H. Wagner 83203 (MICH)	DQ110320	DQ110252
USA: Michigan	C.C. Davis 08-03 (A)	DQ110321	DQ110253
USA: Montana	W.H. Wagner 80123 (MICH)	DQ110322	DQ110254
USA: New Jersey	R.K. Godfrey 62195 (MICH)	DQ110323	DQ110255
USA: New York	T.J. Rawinski 9197 (GH)	DQ110324	DQ110256
USA: North Carolina	C.C. Davis & R.E. Spangler 43 (A)		DQ110257
USA: Virginia	W.H. Wagner 62139 (MICH)	DQ110325	DQ110258
USA: Wisconsin	D.R. DeJoode 1583 (MICH)	DQ110326	DQ1102**59**
South America			
Bolivia: Santa Cruz	R.C. Moran 5914 (US)	DQ110327	DQ110260
Brazil: Minas Gerais	L.C.N. Melo 52 (NY)	DQ110328	DQ110261
Peru: Cusco	J.R.Grant & J.R.Rundell 01-3905 (NY)	DQ110329	DQ110262

Davis *et al.*

Figure 1

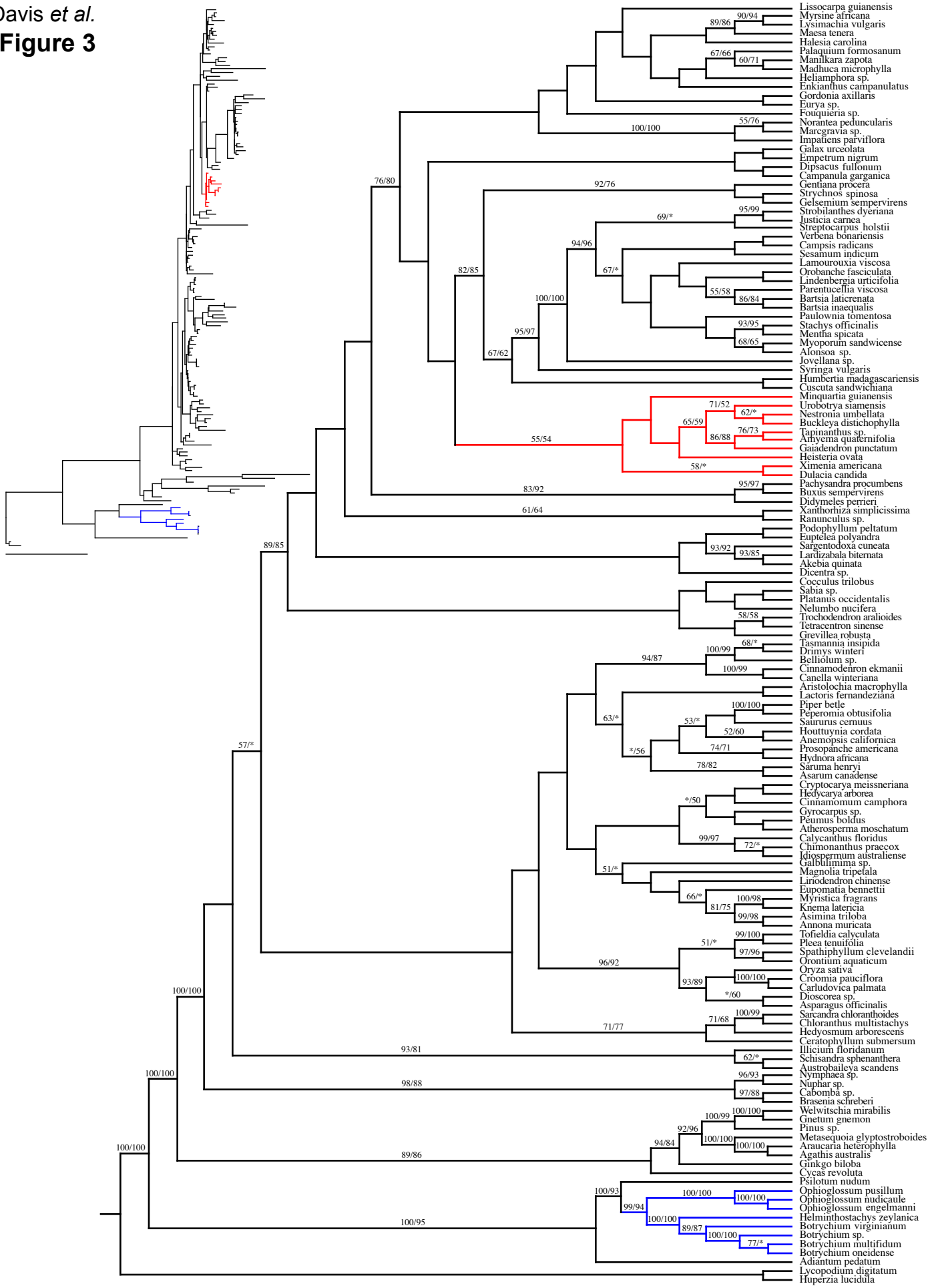


Davis *et al.*

Figure 2

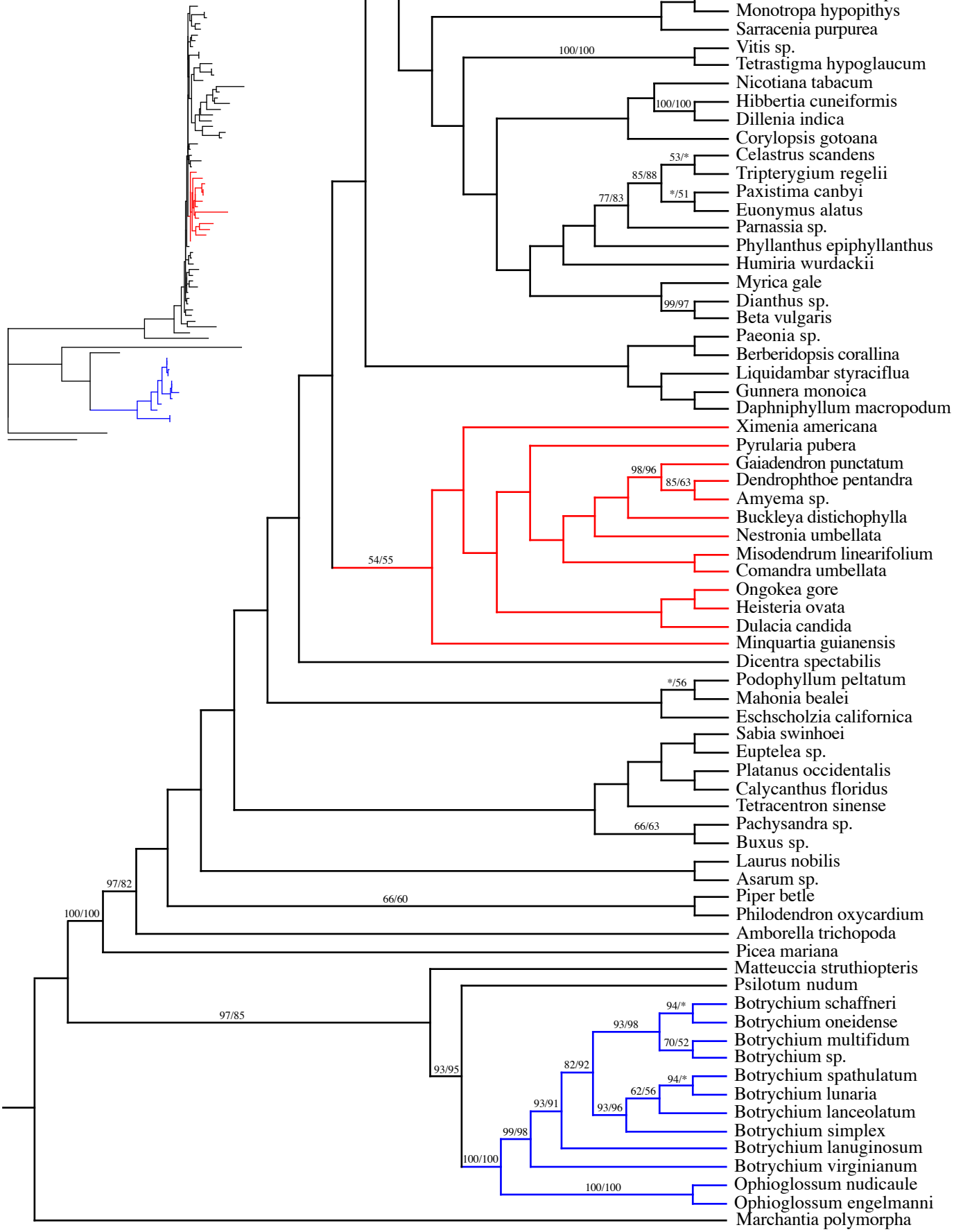
Davis *et al.*

Figure 3



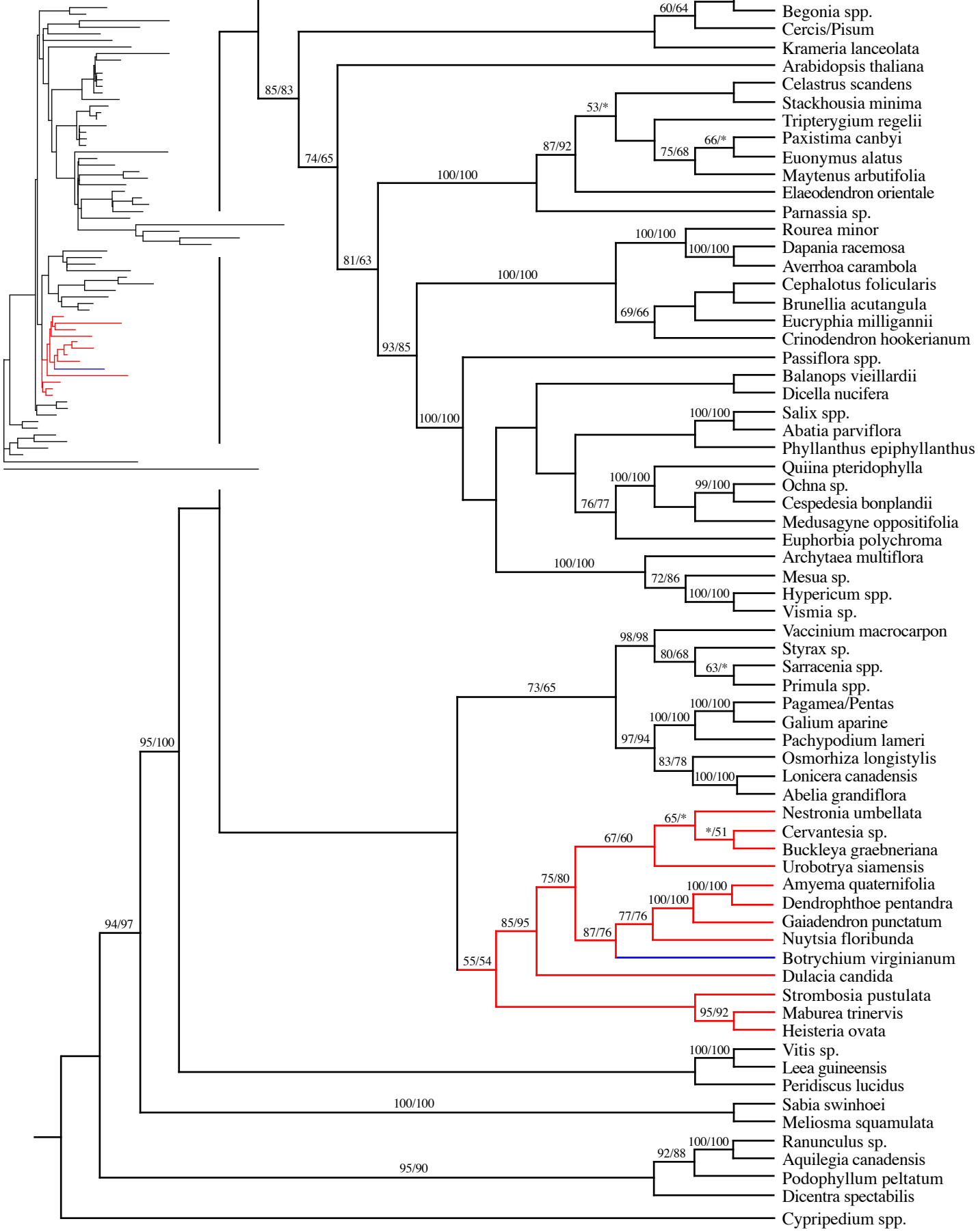
Davis *et al.*

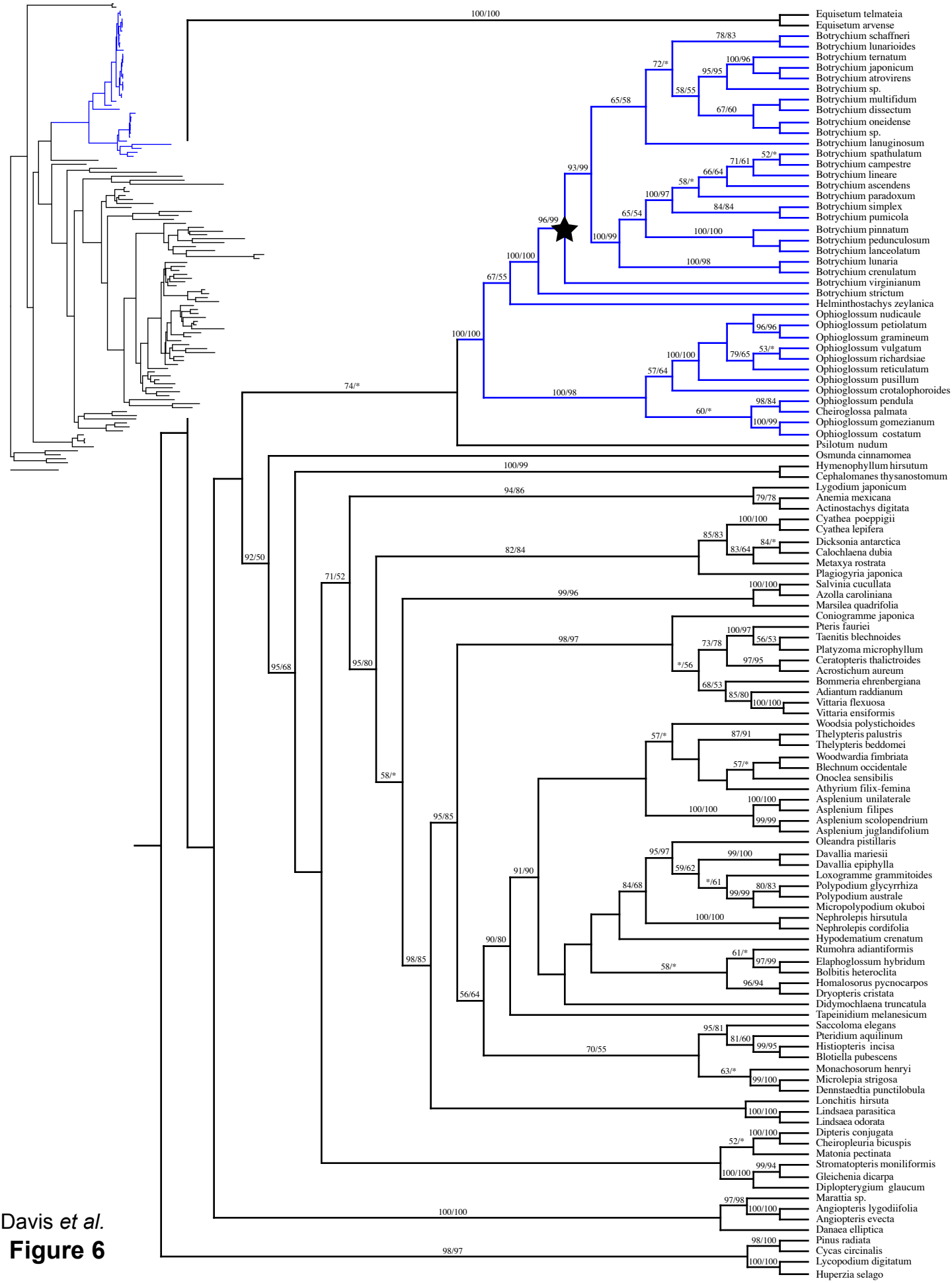
Figure 4



Davis *et al.*

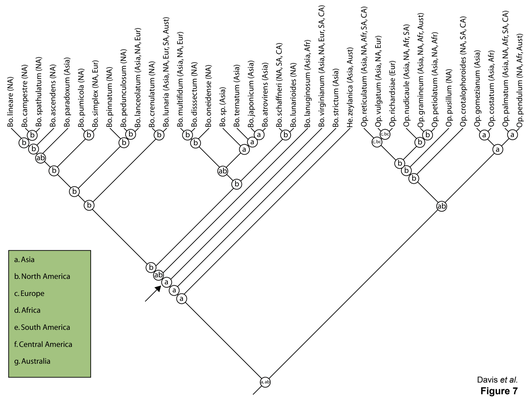
Figure 5





Davis *et al.*

Figure 6



Davis *et al.*
Figure 7

Davis *et al.*

Figure 8