Molecular Phylogenetics and Evolution 127 (2018) 502–512

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Employing hypothesis testing and data from multiple genomic compartments to resolve recalcitrant backbone nodes in *Goodenia* s.l. (Goodeniaceae)

Rachel S. Jabaily[a,b,*], Kelly A. Shepherd[c], Pryce S. Michener[b], Caroline J. Bush[b], Rodrigo Rivero[d,e], Andrew G. Gardner[f], Emily B. Sessa[d,g]

[a] Department of Organismal Biology & Ecology, Colorado College, Colorado Springs, CO 80903, USA
[b] Department of Biology, Rhodes College, Memphis, TN 38112, USA
[c] Western Australian Herbarium, Department of Biodiversity, Conservation and Attractions, Kensington, WA 6151, Australia
[d] Department of Biology, University of Florida, Gainesville, FL 32607, USA
[e] Department of Natural Resources and Environmental Management, University of Hawaii– Mānoa, Honolulu, HI 96822, USA
[f] Department of Biological Sciences, California State University, Stanislaus, One University Circle, Turlock, CA 95382, USA
[g] Genetics Institute, University of Florida, Gainesville, FL 32607, USA

ARTICLE INFO

Keywords:
Chloroplast genome
Conserved ortholog set (COS)
Genome skimming
Goodeniaceae
Mitochondrial genome
Phylogeny
Rapid radiation
Topology testing

ABSTRACT

Goodeniaceae is a primarily Australian flowering plant family with a complex taxonomy and evolutionary history. Previous phylogenetic analyses have successfully resolved the backbone topology of the largest clade in the family, *Goodenia* s.l., but have failed to clarify relationships within the species-rich and enigmatic *Goodenia* clade C, a prerequisite for taxonomic revision of the group. We used genome skimming to retrieve sequences for chloroplast, mitochondrial, and nuclear markers for 24 taxa representing *Goodenia* s.l., with a particular focus on *Goodenia* clade C. We performed extensive hypothesis tests to explore incongruence in clade C and evaluate statistical support for clades within this group, using datasets from all three genomic compartments. The mitochondrial dataset is comparable to the chloroplast dataset in providing resolution within *Goodenia* clade C, though backbone support values within this clade remain low. The hypothesis tests provided an additional, complementary means of evaluating support for clades. We propose that the major subclades of *Goodenia* clade C (C1–C3 + *Verreauxia*) are the result of a rapid radiation, and each represents a distinct lineage.

1. Introduction

Goodeniaceae R.Br. is a family of primarily Australian angiosperms characterized by their extraordinary floral diversity and cup-like stylar indusia (Jabaily et al., 2014). The family includes 420+ species in 12 genera and is sister to the clade of Asteraceae plus Calyceraceae (Tank and Donoghue, 2010). The first family-wide chloroplast DNA (cpDNA)-based molecular phylogeny of Goodeniaceae (Jabaily et al., 2012) determined that a number of genera were monophyletic (e.g. *Anthotium* R.Br., *Brunonia* R.Br., *Dampiera* R.Br., *Lechenaultia* R.Br.). However, it was clear that the currently accepted generic and infrageneric concepts (*sensu* Carolin, 1992) of the remaining genera within the family (e.g. *Coopernookia* Carolin, *Scaevola* s.l. and *Goodenia* s.l. in 'Core Goodeniaceae') were not all supported. The largest genus in the family, *Goodenia* Sm., which includes approximately 220 species in two subgenera, four sections, five subsections, and two series (Carolin, 1992), was particularly taxonomically problematic as the genus was rendered paraphyletic in the Jabaily et al. (2012) analyses. Species in *Goodenia* resolved into three subclades (designated *Goodenia* clades A, B, and C) within the broader *Goodenia* s.l. clade, which also included multiple smaller genera such as *Coopernookia*, *Pentaptilon* E.Prtizel, *Selliera* Cav., *Velleia* Sm., and *Verreauxia* Benth., and the taxon *Scaevola collaris* F.Muell. Moreover, the backbone relationships between the major clades in this diverse *Goodenia* s.l. clade were poorly resolved, impeding potential taxonomic conclusions. Re-circumscribing *Goodenia* to encompass all of the observed molecular and morphological diversity in the large *Goodenia* s.l. clade was thought to be suboptimal because no clear morphological synapomorphies unite all members, and the smaller, embedded genera have apparently strong morphological support (Jabaily et al., 2012). We determined that more comprehensive

* Corresponding author at: Department of Organismal Biology & Ecology, Colorado College, Colorado Springs, CO 80903, USA.
E-mail addresses: rjabaily@coloradocollege.edu (R.S. Jabaily), kelly.shepherd@dbca.wa.gov.au (K.A. Shepherd), prycemichener@gmail.com (P.S. Michener), buscj-19@rhodes.edu (C.J. Bush), rodrigo25@ufl.edu (R. Rivero), agardner1@csustan.edu (A.G. Gardner), emilysessa@ufl.edu (E.B. Sessa).

https://doi.org/10.1016/j.ympev.2018.05.005
Received 22 November 2017; Received in revised form 22 March 2018; Accepted 7 May 2018

taxon and molecular sampling was required to fully explore the composition of clades within *Goodenia* s.l. before any taxonomic conclusions could be drawn.

Our research group has continued to work towards resolving *Goodenia* s.l., with the ultimate goal of naming monophyletic clades supported by molecular and morphological data. This necessitates having the strongest possible phylogenetic support for individual clades, as well as clear resolution of the relationships between those clades. Our efforts at improving the backbone phylogeny of *Goodenia* s.l. by maximizing both resolution and support have focused on broadening taxon and molecular sampling beyond the cpDNA-only dataset of Jabaily et al. (2012). To this end, we turned to a genome skimming approach with next-generation sequencing for 24 species representing most, but not all, of the major clades of *Goodenia* s.l., plus species from the two other clades within the broader Core Goodeniaceae (*Scaevola* s.l. and *Brunonia*) (Gardner et al., 2016a). We estimated a phylogeny based on complete plastome coding sequences (CDS), which did much to resolve our understanding of *Goodenia* s.l. (Gardner et al. 2016a). We also expanded taxon sampling in *Goodenia* s.l. via Sanger sequencing of targeted plastid loci, and when added to the 24-taxon CDS dataset, the combined dataset yielded a phylogeny with maximum support at all deepest nodes within *Goodenia* s.l., with *Coopernookia* placed sister to clade A plus clades B and C (the latter two sister to one another) (Gardner et al., 2016a).

Despite the overall success of this approach, new issues were also raised (Gardner et al., 2016a). In multiple cases, the taxonomic complexity within the clades (A, B, C) of *Goodenia* s.l. increased with expanded sampling. For example, in Jabaily et al. (2012), all (except one) species in *Goodenia* clade B were members of sect. *Goodenia* subsect. *Ebracteolatae* K.Krause. However, Gardner et al. (2016a) found that members of sect. *Porphyranthus* G.Don and sect. *Borealis* Carolin also fell within clade B and were in part sister to the remaining members of subsect. *Ebracteolatae*. Similarly, adding additional taxa from clade C, the most enigmatic of all the major groups recovered in *Goodenia* s.l., revealed a remarkably complex evolutionary history that frustrated our attempts at drawing taxonomic conclusions (Gardner et al., 2016a). This clade has remained the largest barrier to achieving our goal of a revised taxonomy for Goodeniaceae that reflects both morphological and molecular data.

Clade C is the most morphologically diverse group in *Goodenia* s.l. (Fig. 1), and it currently encompasses the genera *Velleia* (21 spp.), *Verreauxia* (3 spp.), and *Pentaptilon* (1 sp.) as well as *Goodenia* subgenus *Monochila* (G.Don) Carolin (9 spp.; 3 subspp., excluding *G. viscida* R.Br. see Gardner et al., 2016a), subsections *Coeruleae* (Benth.) Carolin (11 spp.) and *Scaevolina* Carolin (10 spp.) of section *Coeruleae*, and a number of species from the typical subsection of *Goodenia* (9 spp.; 1 subsp.; 2 var.) (Gardner et al., 2016a,b). In Gardner et al.'s (2016a) analyses of plastid loci, support values at several key nodes within clade C were low, with the exception that subg. *Monochila* and subsect. *Coeruleae* were each strongly supported as monophyletic. The low support values for the remaining relationships in clade C may have been impacted by sampling, as the 24-taxon CDS dataset only included *G. hassallii* F.Muell. (subsect. *Coeruleae*), *Verreauxia reinwardtii* (de Vriese) Benth., and four species from subg. *Monochila*. Gardner et al. (2016a) also performed individual analyses of two nuclear loci (NRR and G3PDH), which yielded topologies that were similar to one another, but which both conflicted significantly with the cpDNA phylogeny at backbone nodes within clade C and throughout *Goodenia* s.l. Given that the current taxonomic divisions within *Goodenia* s.l. are clearly non-monophyletic based on the phylogenies of Jabaily et al. (2012) and Gardner et al. (2016a), we are faced with two alternatives. First, there is support for most of the major clades and relationships among them, which could support the splitting of *Goodenia* s.l. into several genera corresponding to *Goodenia* A, *Goodenia* B, and the respective subclades within *Goodenia* C (Gardner et al., 2016a). Alternatively, we could expand the circumscription of *Goodenia* to encompass all the lineages of *Goodenia* s.l. (except *Coopernookia*), with the recovered clades representing infrageneric groups rather than segregate genera. Before we can commit to either approach, however, we feel it is essential to make every effort to explore relationships within and between the major clades, particularly within the morphologically diverse and, to date, poorly resolved *Goodenia* clade C.

The genome skimming data generated by Gardner et al. (2016a) represented a substantial investment of resources, and only ~3.5% of the data generated was used to produce the plastome CDS and NRR datasets. In order to maximize our investment in the genome skimming approach, we returned to those data in the current study to explore the phylogenetic utility of another high-copy genetic element that can be relatively easily assembled, at least in part, from such data – the mitochondrial genome. We also investigated another approach to deriving nuclear loci from genome skimming data, by assembling sequences belonging to the conserved ortholog set (COS) of loci (Mandel et al., 2014). Mitochondrial loci have typically been used much less frequently in phylogenetic analyses of plants than of animals because, in general, plant mitochondrial genomes are highly conserved at the sequence level but are divergent in structure and gene order even between close relatives, making them less than ideal for analyses at the family level and below (Wolfe et al., 1987; Palmer and Herbon, 1988; Drouin et al., 2008). Mitochondrial DNA (mtDNA) generally has a lower rate of synonymous substitutions when compared with chloroplast and nuclear sequences across diverse plant lineages (Drouin et al., 2008; Wang and Wang, 2014). However, recent comparative analyses of mtDNA sequences across angiosperms have revealed a much more dynamic and variable picture of mitochondrial sequence evolution. Some species and sets of close relatives have been found to have very high rates of synonymous substitution in mitochondrial genes (Parkinson et al., 2005; Mower et al., 2007; Zhu et al., 2014). Cho et al. (2004), in a study of mtDNA substitution rates in *Plantago* L. and representatives of other eudicots, found increased rates of molecular evolution in select taxa and highlighted *Goodenia ovata* Sm. as an outlier, with a 43-fold higher rate of synonymous mutation in surveyed mitochondrial genes compared to its closest included relative. Similarly, Qiu et al. (2010) identified Goodeniaceae (based on two accessions outside of *Goodenia* s.l.) as one of several groups with significant acceleration of mitochondrial mutation rates. Additional comparisons of diverse lineages of angiosperms may yield more exceptions to the rule of low mitochondrial substitution rates in plants, and these previous studies suggest that rates in Goodeniaceae in particular may be more variable and phylogenetically useful than most.

Our goal in the current work is to further leverage our existing genome skimming data from exemplar taxa to supplement previous efforts and produce a robust phylogenetic reconstruction that will support taxonomic revisions. We focus on *Goodenia* clade C in particular, and use mtDNA and a more extensive dataset of nuclear DNA (from COS loci) to reconstruct relationships within this clade and across *Goodenia* s.l. We investigate whether mtDNA is sufficiently variable across *Goodenia* s.l. to provide phylogenetic resolution at the generic and subgeneric levels, and we use an explicit hypothesis-testing framework to ask whether mtDNA sequence data corroborate or conflict with conclusions from cpDNA and nuclear loci. We follow Straub et al. (2014), who tested the ability of genome skimming data to resolve backbone phylogenies with short nodes, and who recommended performing extensive hypothesis testing of all possible topologies for the nodes in question, in addition to standard bootstrapping analyses. We likewise implement approximate unbiased (AU) tests (Shimodaira, 2002) in order to explore incongruence between topologies from different datasets, and to determine the statistical support for all possible relationships of the major lineages within *Goodenia* clade C.

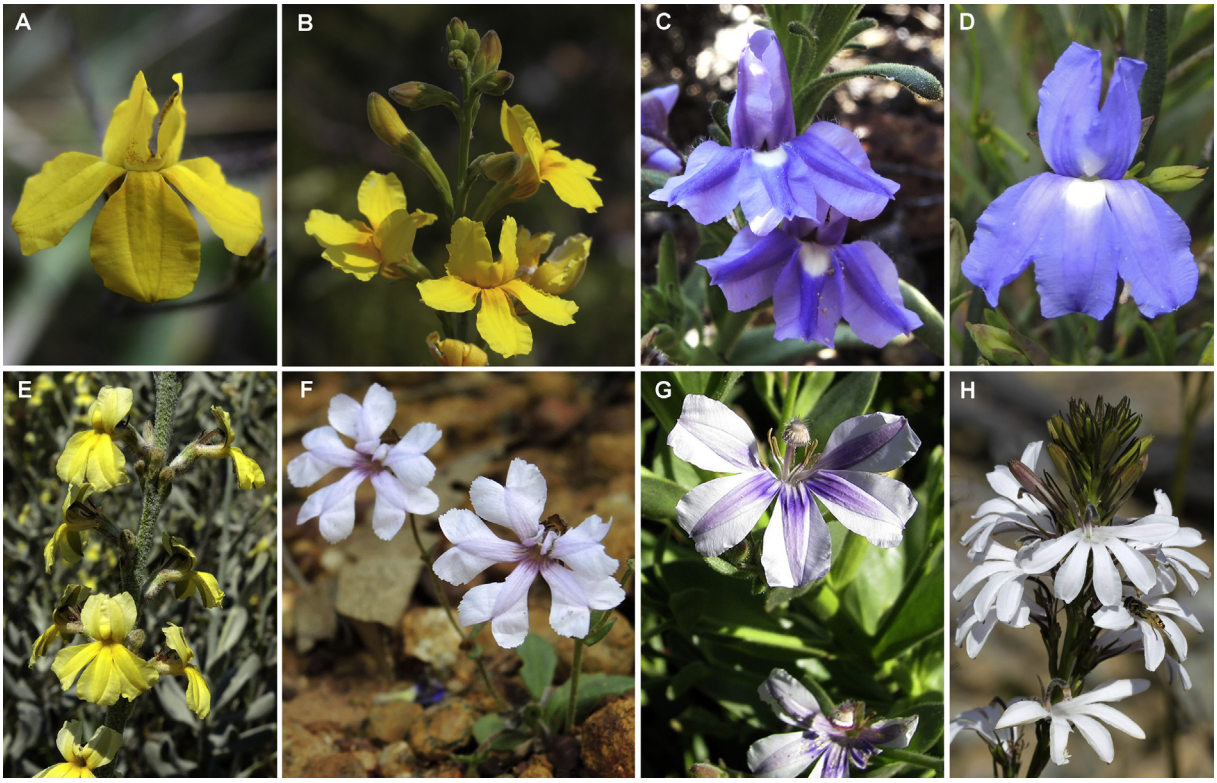

Fig. 1. Representative taxa of *Goodenia* clade C. (A) *G. dimorpha* var. *angustifolia* (subsect. *Goodenia*); (B) *G. quadrilocularis* (subsect. *Goodenia*); (C) *G. xanthotricha* (subsect. *Goodenia*); (D) *G. hassallii* (subsect. *Coeruleae*); (E) *Verreaxia reinwardtii*; (F) *Velleia rosea* (sect. Menoceras); (G) *G. stobbsiana* (subsect. *Scaevolina*); and (H) *G. decursiva* (subgen. Monochila). Vouchers: *K.A. Shepherd & J.A. Wege* KS1541(A); *K.A. Shepherd & J.A. Wege* KS1580 (B); *K.A. Shepherd & S.R. Willis* KS1575 (C); *K.A. Shepherd & S.R. Willis* KS1517 (D); *K.A. Shepherd & S.R. Willis* KS1514 (E); *K.A. Shepherd & S.R. Willis* KS1509 (F); *S. Dillon, T. Hammer, K.R. Thiele* SD7000 (G) and *K.A. Shepherd & S.R. Willis* KS1528 (H). Images by: K.A. Shepherd (A–F, H) and S. Dillon (G).

2. Materials and methods

2.1. Sampling, datasets, and molecular methods

We included twenty-four taxa of *Goodenia* s.l.; twenty of these had previously been sequenced by Gardner et al. (2016a), and four were newly sequenced for this study to increase sampling of major groups in clade C: *G. xanthotricha* de Vriese, *G. quadrilocularis* R.Br. and *G. dimorpha* var. *angustifolia* Maiden & Betche from subsect. *Goodenia*, and *G. stobbsiana* F.Muell. from subsect. *Scaevolina* (Table 1). This brought us to a total of 10 exemplar taxa sampled from clade C. Voucher and accession information are included in Appendix S1 (see Supplemental Data with the online version of this article). We constructed five datasets corresponding to organellar or nuclear DNA (mtDNA, cpDNA, NRR, G3PDH, COS), and these datasets included either all 24 taxa (mtDNA, cpDNA), or only the clade C exemplar taxa (NRR, G3PDH, COS).

DNA extractions and genome skimming methods for the four new taxa followed the same protocols as Gardner et al. (2016a). Macrogen, Inc. (Seoul, Korea) performed the library preparation, sequencing, and post-processing including de-multiplexing, adapter removal, and removal of low-quality reads. The four new taxa were sequenced on two lanes of Illumina Highseq 2500 with 150 bp paired-end chemistry.

2.2. Mitochondrial dataset (mtDNA)

Clean, raw reads from genome skimming of each taxon were first filtered against the mitochondrial genome sequence of *Helianthus annuus* L. (KF815390; Bock et al., 2014) using Bowtie 2 (Langmead et al., 2009; Langmead and Salzberg, 2012). Filtered reads were converted to FASTA format and imported into Geneious R9 (http://www.geneious.com, Kearse et al., 2012), where they were assembled against the same *Helianthus* mitochondrion sequence using the built-in Geneious reference-guided assembly algorithm with sensitivity set to medium and iterating up to ten times. For each taxon, a consensus sequence was generated from the assembly that included all positions with minimum 5X coverage. We used 5X as opposed to the 25X cutoff for the cpDNA dataset (see below) because many fewer reads passed the Bowtie filtering step for the mtDNA than the cpDNA (tens of thousands vs. hundreds of thousands to millions), and lowering the coverage cutoff allowed us to include more of the assembly in the consensus and subsequent analyses. Annotations were transferred from the *Helianthus* reference to each consensus sequence. Consensus sequences from all 24 taxa were aligned using the MAFFT plugin in Geneious with default settings. We excluded sites that were mostly or entirely gaps, and positions that aligned ambiguously, using Gblocks (Talavera and Castresana, 2007) following recommended settings. To test for potential intercellular genetic transfer events, the gene sequences were subjected to a BLASTX search against the *nr* database with an e-value cutoff of 10^{-10} to look for non-mitochondrial homologies and none were found.

2.3. Chloroplast dataset (cpDNA)

Clean, raw reads were processed as described above for the mitochondrial dataset, except that filtering in Bowtie2 and subsequent reference-guided assembly in Geneious both used an existing whole plastome sequence from *H. annuus* (NC_007977; Timme et al., 2007). We produced new assemblies for the 20 taxa included in Gardner et al. (2016a) in order to include regions outside the CDS employed in that study (e.g., intergenic spacers). As in Gardner et al. (2016a), we used a 25X coverage cutoff to generate consensus plastome sequences, and annotations were transferred from the *Helianthus* reference. Sequences

Table 1
Taxonomy and clade identity for taxa included in this study (seven taxa included in the unbiased hypothesis tests in bold).

Taxon name	Subgeneric taxonomy (*sensu* Carolin, 1992)	Major clade (*sensu* Gardner et al., 2016a)
Coopernookia polygalacea (de Vriese) Carolin		*Coopernookia*
Coopernookia strophiolata (F.Muell.) Carolin		*Coopernookia*
Goodenia ovata Sm.	subgenus *Goodenia* section *Goodenia* subsection *Goodenia*	*Goodenia* clade A
Goodenia phillipsiae Carolin	subgenus *Goodenia* section *Goodenia* subsection *Ebracteolatae*K.Krause	*Goodenia* clade A
Goodenia tripartita Carolin	subgenus *Goodenia* section *Goodenia* subsection *Goodenia*	*Goodenia* clade A
Goodenia viscida R.Br.	subgenus *Monochila*(G.Don) Carolin	*Goodenia* clade A
Scaevola collaris F.Muell.		*Goodenia* clade A
Selliera radicans Cav.		*Goodenia* clade A
Goodenia filiformis R.Br.	subgenus *Goodenia* section *Goodenia* subsection *Ebracteolatae*K.Krause	*Goodenia* clade B
Goodenia micrantha Hemsl. ex Carolin	subgenus *Goodenia* section *Goodenia* subsection *Ebracteolatae*K.Krause	*Goodenia* clade B
***Goodenia hassallii* F.Muell.**	**subgenus *Goodenia* section *Coeruleae* subsection *Coeruleae*(Benth.) Carolin**	***Goodenia* clade C**
***Goodenia dimorpha* var *angustifolia* Maiden & Betche**	**subgenus *Goodenia* section *Goodenia* subsection *Goodenia***	***Goodenia* clade C**
***Goodenia quadrilocularis* R.Br.**	**subgenus *Goodenia* section *Goodenia* subsection *Goodenia***	***Goodenia* clade C**
***Goodenia xanthotricha* de Vriese**	**subgenus *Goodenia* section *Goodenia* subsection *Goodenia***	***Goodenia* clade C**
***Goodenia decursiva* W.Fitzg.**	**subgenus *Monochila*(G.Don) Carolin**	***Goodenia* clade C**
Goodenia drummondii Carolin	subgenus *Monochila*(G.Don) Carolin	*Goodenia* clade C
Goodenia helmsii(E.Pritzell) Carolin	subgenus *Monochila*(G.Don) Carolin	*Goodenia* clade C
Goodenia pinifolia de Vriese	subgenus *Monochila*(G.Don) Carolin	*Goodenia* clade C
***Goodenia stobbsiana* F.Muell.**	**subgenus *Goodenia* section *Coeruleae* subsection *Scaevolina*Carolin**	***Goodenia* clade C**
***Verreauxia reinwardtii* (de Vriese) Beth.**		***Goodenia* clade C**
Velleia discophora F.Muell.		*Velleia*
Velleia foliosa (Benth.) K.Krause		*Velleia*
Velleia rosea S.Moore		*Velleia*
Scaevola tomentosa Gaudich.		*Scaevola* s.l.

were aligned and ambiguous sites removed as described above for the mtDNA dataset.

2.4. Assembly of single-locus nuclear datasets

For hypothesis testing within *Goodenia* clade C, we assembled sequences of NRR and G3PDH for the four new taxa and combined these with existing sequences for clade C members previously generated by Gardner et al. (2016a). The NRR dataset included eleven taxa and the G3PDH dataset included ten (assembly of *G. helmsii* for G3PDH was unsuccessful). *Velleia discophora* F.Muell. was included as an outgroup for both datasets. We used *Verreauxia*, a member of clade C sequenced by Gardner et al. (2016a), as a reference in Bowtie2 to extract reads for both markers from the genome skimming data for the four new species. These reads were then processed as above using reference-guided assembly to the appropriate marker from *Verreauxia*.

2.5. Assembly of COS dataset

For the ten taxa sampled from clade C, we used the HybPiper pipeline with default settings (Johnson et al., 2016) to assemble reads corresponding to COS loci based on the probe set published by Mandel et al. (2014). Of the 1061 orthologous loci represented in the COS probe set, we retained only those loci that were assembled for five or more of the ten clade C exemplar taxa. Sequences for each individual locus were aligned with MAFFT with default settings in Geneious. Individual taxa with less than 33% of the total alignment length for a given locus were removed from that alignment, and any gene that had fewer than five of the ten clade C taxa present after this step was also removed from further analysis.

2.6. Phylogenetic analyses

Scaevola tomentosa Gaudich. was included as an outgroup in analyses of the 24-taxon datasets (mtDNA and cpDNA), and *Velleia discophora* in the 10-taxon datasets (NRR, G3PDH, COS), based on the results of Jabaily et al. (2012) and Gardner et al. (2016a). For the mtDNA and cpDNA datasets, the bounds of coding and noncoding regions in the alignments were determined based on annotations transferred from *Helianthus* as described above. We used PartitionFinder v. 2.0 (Lanfear et al., 2012) to determine the optimal partitioning scheme and the best set of nucleotide substitution models for the mtDNA and cpDNA datasets using the AICc criterion and greedy algorithm. We used the AICc criterion in jModeltest v. 2.0 (Guindon and Gascuel, 2003; Darriba et al., 2012) to determine the optimal model of molecular evolution for the individual COS loci, NRR, and G3PDH. For all datasets (mtDNA, cpDNA, individual COS loci, NRR, and G3PDH), we performed maximum likelihood analyses in Garli v. 2.01 (Zwickl, 2006) to estimate the best tree and analyze bootstrap replicate datasets, using the optimal model(s) of molecular evolution and optimal partitioning schemes for each dataset. We completed 100 bootstrap replicates for each dataset. We considered ML bootstrap values higher than 90% to constitute strong support for clades and relationships.

For the COS dataset, we used the ML results from the individual loci to estimate a species tree with local posterior probabilities (Sayyari and Mirarab, 2016) with ASTRAL-II and its default settings for multi-locus bootstrapping (Mirarab and Warnow, 2015). The input tree for each locus was that locus's ML tree annotated with the results of the 100 bootstrap replicates completed in Garli as described above. Nodes supported by less than 33% bootstrap support in these analyses were collapsed in the ML trees for the individual loci.

2.7. Clade C hypothesis testing

To investigate support for topological incongruences between the mtDNA, cpDNA, NRR, G3PDH, and COS datasets in relationships within *Goodenia* clade C, we performed a series of approximately unbiased (AU) tests (Shimodaira, 2002). Because these analyses require a sequence matrix as input rather than a tree, we could not use the ASTRAL-II species tree for the COS dataset, and instead used a single concatenated matrix of all the COS loci that had been used individually to produce the species tree. For each of the five datasets, we first constructed subalignments that contained only seven taxa that represent the major lineages in clade C (Table 1). We generated all unrooted, fully bifurcating topologies for these seven taxa (10,395 total) using the phangorn package (Schliep, 2011) in R (R Development Core Team, 2008). We then used RAxML (Stamatakis, 2014) to compute per-site log likelihood values ("-f G" setting) for all trees under each alignment,

Table 2
Results of approximate unbiased (AU) tests for five datasets of seven taxa that represent the major lineages in *Goodenia* clade C. The number and percentage of topologies with $P \geq 0.05$, and the clades present (if any) in a strict consensus of those topologies, are given. Three subclades (C1-C3) were present in the individual highest likelihood topologies for the various datasets. For each of these subclades, the number and proportion of trees from each dataset in which that subclade appears are given, as is the *P*-value of the topology in each dataset where all trees with that value and higher yield the clade.

Dataset	Number and % of total topologies with $P \geq 0.05$	Clades in strict consensus of topologies with $P \geq 0.05$	Clade C1 (*G. hassallii*, *G. xanthotricha*)	Clade C2 (*G. stobbsiana*, *G. decursiva*)	Clade C3 (*G. quadrilocularis*, *G. dimorpha* var. *angustifolia*)
mtDNA	88 (0.8%)	Clade C1 (*G. hassallii*, *G. xanthotricha*)	88/88 ($P \geq 0.05$)	21/88 ($P \geq 0.323$)	not present $P \geq 0.05$
cpDNA	60 (0.6%)	seven taxon polytomy	9/60 ($P \geq 0.559$)	not present $P \geq 0.05$	55/60 ($P \geq 0.091$)
COS	67 (0.6%)	Clade C3 (*G. quadrilocularis*, *G. dimorpha* var. *angustifolia*)	not present $P \geq 0.05$	not present $P \geq 0.05$	67/67 ($P \geq 0.05$)
NRR	531 (5.1%)	seven taxon polytomy	33/531 ($P \geq 0.499$)	21/531 ($P \geq 0.857$)	250/531 ($P \geq 0.117$)
G3PDH	2739 (26.3%)	seven taxon polytomy	1/2739 ($P \geq 0.865$)	2/2739 ($P \geq 0.865$)	1794/2739 ($P \geq 0.077$)

which also assigns a likelihood score to each topology for each dataset. The per-site log likelihood values for each dataset were then analyzed using Consel (Shimodaira and Hasegawa, 2001), which performs a set of topology tests to evaluate whether or not all trees explain the data equally well. We report results of the approximately unbiased test, calculated from a multiscale bootstrap. For each dataset, we ranked all trees by AU *P*-value. Contrary to typical analyses where a result at $P \leq 0.05$ is desirable, in this case we were interested in trees with the highest *P*-values, which are most consistent with the data. Trees with *P*-values ≤ 0.05 are in fact those statistically rejected by the underlying alignments. We sorted all topologies by *P*-value, and identified the *P*-value at which a strict consensus of all topologies with *P* greater than or equal to that value would include a particular clade of interest (see Table 2). The topology with the next lowest *P*-value would have lacked that clade.

3. Results

3.1. Phylogenetic analyses of 24-taxon mtDNA & cpDNA datasets

We generated partial mtDNA genome sequences for all 24 taxa, with a total alignment length of 41,552 bp. Less than half of the total sequence length was annotated as coding based on the *Helianthus* reference. We also generated partial cpDNA genomes for all 24 taxa, and this alignment was 98,088 bp long (87% longer than the 52,388 plastome CDS alignment from Gardner et al. (2016a)). The partial cpDNA and mtDNA assemblies are available from Dryad (https://doi.org/10.5061/dryad.j0nf1sn). The percent parsimony informative characters (PIC) in the total mtDNA dataset was only 19% of the total cpDNA dataset, and the *Goodenia* clade C mtDNA dataset PIC was 23.8% of the *Goodenia* clade C cpDNA dataset PIC (Table 3). Maximum likelihood analyses in Garli yielded an mtDNA phylogeny with lnL = −77,764.53 and cpDNA phylogeny with lnL = −340,173.02 (Fig. 2). The optimal partitioning schemes identified by PartitionFinder for both datasets are provided in Appendix S2.

Table 3
Alignment length and parsimony informative character (PIC) information for cpDNA, mtDNA, and nuclear datasets COS, NRR, and G3PDH. Average values across all 85 COS loci are given.

Dataset	Alignment length (basepairs)	Total number parsimony informative characters (PIC)	Percent PIC
cpDNA (all 24 taxa)	98,088	10,365	10.57
cpDNA (clade C)		1843	1.88
mtDNA (all 24 taxa)	41,552	836	2.01
mtDNA (clade C)		186	0.45
COS (clade C)	average 483	average 7.85	0.83
NRR (clade C)	6494	93	1.43
G3PDH (clade C)	861	20	2.32

The topologies of the mtDNA and cpDNA phylogenies are identical in the composition and arrangement of major clades across *Goodenia* s.l. (Fig. 2A, B). With maximum support in both datasets, *Coopernookia* is sister to the remainder of *Goodenia* s.l., followed stepwise by *Goodenia* clades A, B, *Velleia*, and *Goodenia* clade C. Relationships that resolve with lower ML bootstrap support, and which differ in topology between the mtDNA and cpDNA trees, are found within *Goodenia* clades A and C. In clade A, *Selliera radicans* Cav. is sister to *Scaevola collaris* in the cpDNA topology with 100% support; in the mtDNA topology, *Selliera* is strongly supported as sister to *Goodenia ovata* plus *G. viscida* R.Br. Within clade C, four and six nodes, respectively, are resolved with greater than 90% bootstrap support in the cpDNA and mtDNA topologies (Fig. 2). Both topologies recover a monophyletic subg. *Monochila*, with *G. drummondii* Carolin, *G. helmsii* (E.Pritz.) Carolin, *G. pinifolia* de Vriese, and *G. decursiva* W.Fitzg. forming a strongly supported subclade. Both phylogenies further support a sister relationship between *G. drummondi* and *G. helmsii*. The cpDNA topology also groups *G. dimorpha* var. *angustifolia* and *G. quadrilocularis* (both subsect. *Goodenia*) with strong support, while the mtDNA resolves *G. hassallii* (subsect. Coeruleae) and *G. xanthotricha* (subsect. *Goodenia*) as sister species, with *Verreauxia* sister to the pair, and places *G. stobbsiana* (subsect. *Scaevolina*) sister to the subg. *Monochila* subclade. Throughout the two phylogenies, branch lengths are substantially shorter in the mtDNA than the cpDNA tree (Fig. 2A, B).

3.2. Phylogenetic analyses of nuclear datasets

The aligned matrices for G3PDH and NRR were 861 and 6494 base pairs long, respectively (Table 3), and are available on Dryad (https://doi.org/10.5061/dryad.j0nf1sn). Within clade C, the G3PDH alignment had the highest percentage of parsimony informative characters (and mtDNA the lowest; Table 3). JModelTest identifed HKY + G and TIM2 + I + G as the best models of evolution for the G3PDH and NRR datasets, respectively. ML analysis of the NRR dataset produced a tree with lnL = −11,458.89; the G3PDH tree had lnL = −1,866.66. Both analyses, which included only clade C taxa, recovered a monophyletic subg. *Monochila* with 100% bootstrap support (*G. helmsii* was not included in the G3PDH analysis because of assembly issues), and a sister-species pairing of *G. quadrilocularis* and *G. dimorpha* var. *angustifolia* (subsect. *Goodenia*), but with less than maximal support (Fig. 2C, D). The NRR dataset also placed *G. stobbsiana* (subsect. *Scaevolina*) sister to subg. *Monochila*, and found *G. xanthotricha* (subsect. *Goodenia*) and *G. hassallii* (subsect. Coeruleae) to be sister to one another; both relationships were also recovered by the mtDNA dataset, but with stronger support there than in the NRR.

Eighty-five COS loci, with an average length of 483 bases, met our 5-taxon inclusion benchmark for analysis (Table 3). Average PIC per locus was 7.85, and ranged from one to seventy. The ASTRAL-II analysis

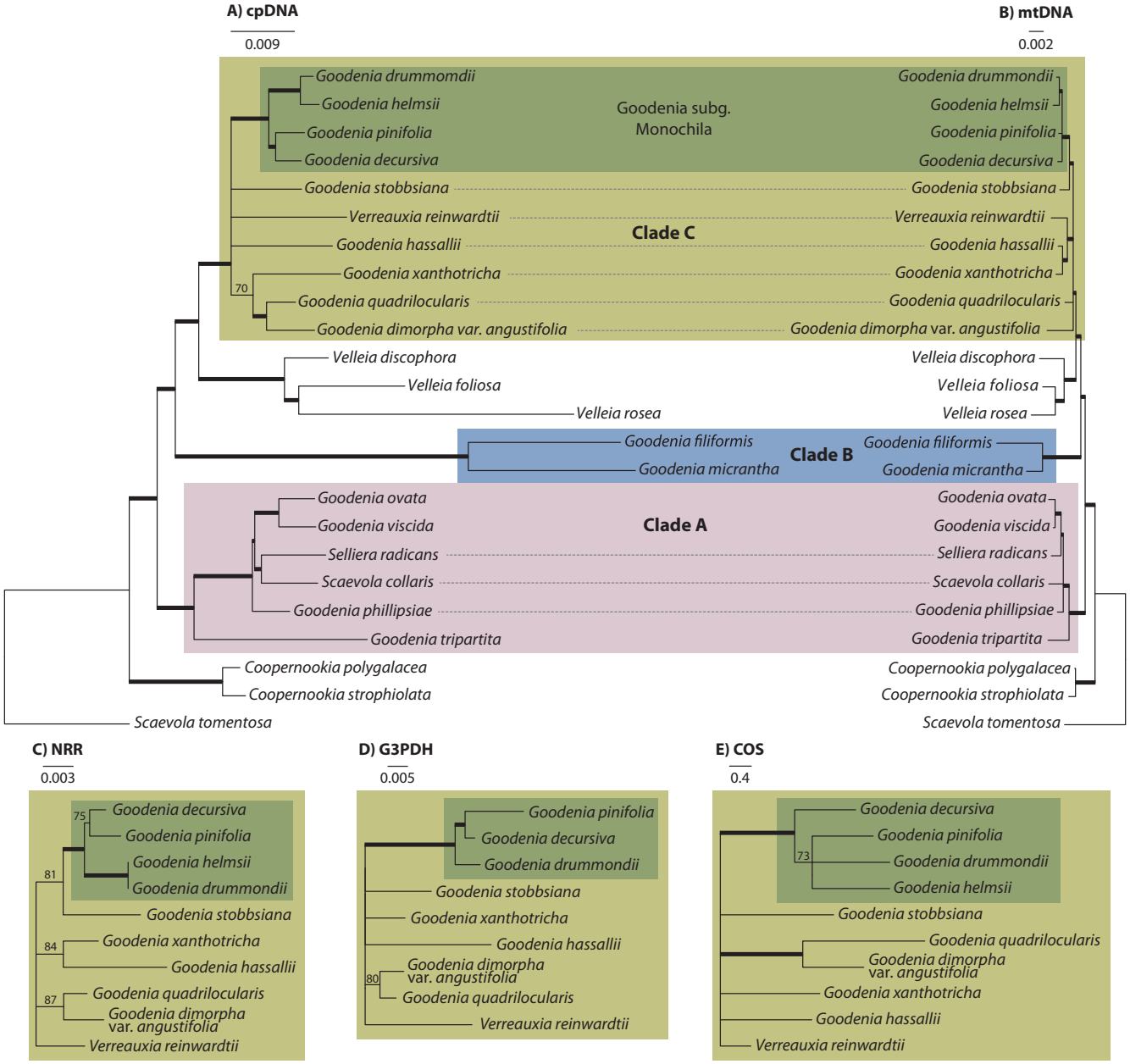

Fig. 2. Topologies produced by ML analyses with bootstrapping for four datasets: (A) cpDNA, (B) mtDNA, (C) NRR, (D) G3PDH, and the ASTRAL-II species tree for (E) COS loci. Scale bars indicate number of substitutions per site, and the cpDNA and mtDNA trees are shown at the same scale. Thick branches have ≥90% bootstrap support (BS). Branches with less than 70% BS are collapsed and shown as polytomies. The NRR, G3PDH, and COS trees were rooted with *Velleia discophora*, which is removed in the figure. Fully resolved topologies for all five trees are provided in Appendix S3.

of the individual COS loci yielded a species tree in which *G. quadrilocularis* and *G. dimorpha* var. *angustifolia* form a subclade with 100% local posterior probability, and with a well-supported (96% local posterior probability) monophyletic subg. *Monochila* (Fig. 2). All other relationships received less than 80% local posterior probability. Aligned matrices for all of the COS loci are available on Dryad (https://doi.org/10.5061/dryad.j0nf1sn).

3.3. Hypothesis testing of alternate topologies

The highest likelihood unrooted topologies for the five datasets are congruent with the broader sampled phylogenies from the same datasets (Fig. 3) and are found within the subset of $P \geq 0.05$ topologies for their respective dataset. Approximate unbiased tests of all possible topologies (10,395) for the seven representative lineages of clade C found that the mtDNA, cpDNA, and COS datasets rejected (at $P \geq 0.05$) more of the total possible topologies than the NRR and G3PDH datasets (Table 2). A strict consensus of all topologies with $P \geq 0.05$ produced a seven-way polytomy for all datasets except mtDNA and COS (Table 2). When we sequentially removed individual topologies starting with the lowest *P*-values above 0.05, this removal gradually resulted in strict consensus trees with substructure of specific sister relationships. These subclades are recovered at different *P*-values (Table 2). Maximum likelihood and AU analyses determined that all datasets except for COS and G3PDH recover a sister relationship of *Goodenia xanthotricha* (subsect. *Goodenia*) and *G. hassallii* (subsect. *Coeruleae*) (subclade C1), with the mtDNA dataset recovering this subclade in all topologies above $P = 0.05$. The mtDNA and NRR datasets place *G. decursiva* (subg. *Monochila*) and *G. stobbsiana* (subsect. *Scaevolina*) as sister (subclade C2), with the greatest percentage of trees above $P = 0.05$ having this

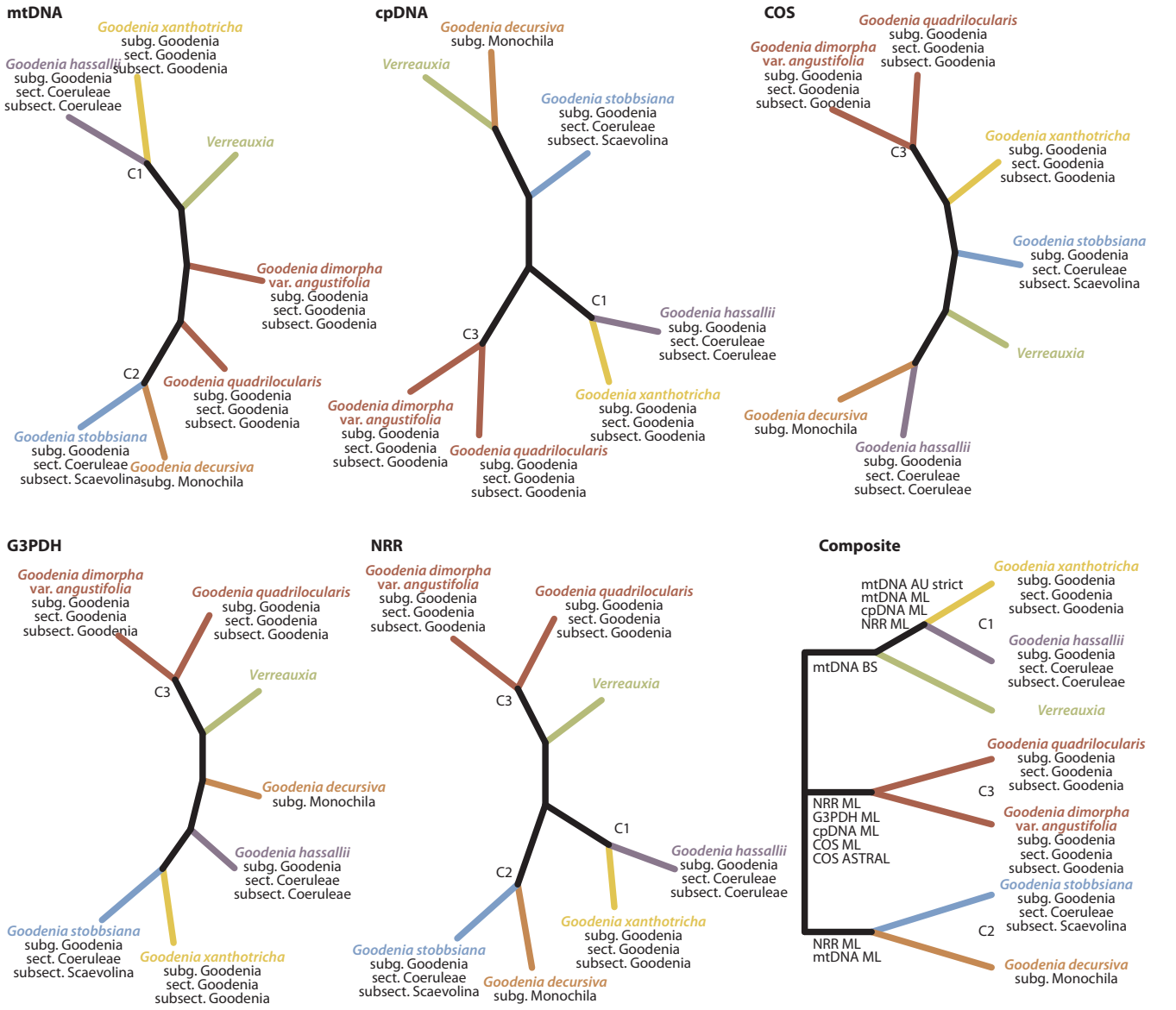

Fig. 3. For all five datasets, the single highest likelihood topology from the set of 10,395 possible unrooted topologies for seven taxa. The seven taxa present represent all major clades in *Goodenia* clade C. A composite at lower right shows the subclades and placement of *Verrauxia* that receive support from one or more analyses, which are indicated above the branches.

relationship in the mtDNA dataset (Table 2). This subclade is not found in the highest ML tree for G3PDH, cpDNA, or COS, but is found in the sets of highest trees *P*-value trees for each dataset. All but the mtDNA dataset place *G. dimorpha* var. *angustifolia* (subsect. *Goodenia*) and *G. quadrilocularis* (subsect. *Goodenia*) as sister (subclade C3), with the COS dataset recovering this subclade in all topologies above $P = 0.05$. Two additional sister relationships, partially in conflict with hypothesized clades C1-C3 and both involving placement of *Verreauxia*, are supported by single datasets each. The placement of *Verreauxia* changes between all datasets, except NRR and G3PDH, in both of which it is sister to subclade C3; however, the only position of *Verreauxia* that receives support from multiple datasets is its placement in the mtDNA ML bootstrap tree (Fig. 2), where it is sister to subclade C1 with strong support. A composite tree showing the subclades that are supported by our various analyses is shown in Fig. 3.

4. Discussion

4.1. Summary of relationships in Goodenia s.l.

We aimed to produce a fully resolved, fully supported estimate of phylogeny for *Goodenia* s.l., using data from multiple genomic compartments and building on the resolution achieved in previous studies (Jabaily et al., 2012; Gardner et al., 2016a). We were partially able to do this, with data from both cpDNA and mtDNA strongly supporting *Coopernookia* as sister to the remaining members of *Goodenia* s.l., as also found by Gardner et al. (2016a). This is not surprising, as *Coopernookia* is supported by a number of synapomorphies including distinctive strophiolate seeds and a base chromosome number of $x = 7$ (vs $x = 8$ for the remainder of *Goodenia* s.l.) (Peacock, 1963). The relationships between clades A, B, and C of *Goodenia* s.l. are also congruent between both the cpDNA and mtDNA datasets and receive similar, high support values (Fig. 2). *Velleia* is supported as monophyletic and consistently placed sister to *Goodenia* clade C, while *Selliera* is embedded in clade A and *Verreauxia* also in clade C (Fig. 2).

Despite the addition of more base pairs in the cpDNA dataset compared to previous analyses (Gardner et al., 2016a), additional data from a new genomic compartment (mtDNA), and extensive sampling of nuclear loci, relationships within clade C still prove difficult to resolve, and no single dataset provides full resolution with strong support. Only *Goodenia* subg. *Monochila* was consistently recovered as monophyletic in all analyses (Fig. 2). The results of the exhaustive tree searches and extensive hypothesis tests we performed allow us to draw some further conclusions. In these analyses, which contained only seven taxa representing the major lineages in clade C, based on previous studies (Jabaily et al., 2012; Gardner et al., 2016a), we considered all possible unrooted topologies for seven taxa and how each topology performed under each dataset. All datasets except mtDNA and COS produced an uninformative, seven-way polytomy in a strict consensus of all topologies with $P \geq 0.05$ (Table 2). However, the combined results from the best maximum likelihood topologies (Fig. 3), standard bootstrapping analyses (Fig. 2), and sister relationships found in strict consensus trees supported by higher P-values (Table 2) provide support for further structure among these relationships, and can be synthesized into a hypothesis of relationships in clade C that draws on the best supported subclades from all analyses, and which has reliable implications for taxonomy.

The results of all of our analyses are congruent with three subclades within clade C (Fig. 3, Table 2): "**C1**": *G. hassallii* (subsect. *Coeruleae*) + *G. xanthotricha* (subsect. *Goodenia*), "**C2**": *G. stobbsiana* (subsect. *Scaevolina*) + *G. decursiva* (subg. *Monochila*), and "**C3**": *G. quadrilocularis* (subsect. *Goodenia*) + *G. dimorpha* var. *angustifolia* (subsect. *Goodenia*), *Verreauxia*. These subclades are compatible with each other, though in the maximum likelihood topologies for the five datasets they are only found together in the NRR tree (Fig. 3), and in the sets of trees with $P \geq 0.05$, all three subclades occur among those trees only for G3PDH and NRR (Table 2). Among the trees with $P \geq 0.05$, subclades C1 and C3 are present in more trees and from more datasets than C2 (Table 2). Subclade C2 is only recovered in one or more of the $P \geq 0.05$ trees in the mtDNA, NRR, and G3PDH analyses, but it is present in the single maximum likelihood topologies for mtDNA and NRR (Fig. 3). Several of these subclades, and the relationships between them, highlight hitherto unrecognised relationships within clade C. A composite of the subclades supported by one or more of our analyses is shown in Fig. 3.

Goodenia section *Coeruleae* as a whole includes the blue-flowered species of *Goodenia* (Fig. 1D, G) in which the septum of the ovary is at least 2/3 as long as the locule. Despite these shared characters, it is clear from our analyses that this section is not monophyletic, as the representatives of subsects. *Coeruleae* and *Scaevolina* (in subclades C1 and C2, respectively) are never recovered as sister to one another (Fig. 3). Carolin (1992) recognised subsect. *Scaevolina* as including the predominantly northern Australian perennial species, while subsect. *Coeruleae* encompasses the remaining southwest Western Australian species of the section. The subsections also differ in several aspects of morphology and habit, and these discrepancies further support their recognition as separate taxonomic entities.

Subclade C1 in our analyses also includes *G. xanthotricha*, which Carolin (1992) had placed in the yellow-flowered sect. *Goodenia*, despite the fact that it has blue flowers like members of sect. *Coeruleae* (Fig. 1C). Carolin believed that *G. xanthotricha* was not allied to that section due to its seeds having an aculeate testa and a shrub-like habit. However, our results show that *G. xanthotricha* is in fact closely allied with subsect. *Coeruleae*. Subclade C2 includes the representatives of subsect. *Scaevolina* and subg. *Monochila*. The latter was consistently recovered as monophyletic when multiple representatives of subg. *Monochila* were included in all analyses (Fig. 2). Superficially, the groups that comprise subclade C2 do not appear to be allied; however, some members of subsect. *Scaevolina* have flowers tending towards a fan-shaped form (Gardner et al., 2016a) (Fig. 1G) and a similar narrow indusium as observed in members of subg. *Monochila* (K.A. Shepherd, Western Australian Herbarium, unpublished data). Subclade C3 includes a subset of yellow-flowered species previously included by Carolin (1992) in the typical subsection *Goodenia*, the remainder of which we now recognize as belonging to *Goodenia* clade A (Jabaily et al., 2012; Gardner et al., 2016a). The subset of yellow-flowered species in clade C includes a number of diploid and polyploid taxa from eastern Australia (we sampled only diploids here) (Peacock, 1963; K. Shepherd, Western Australian Herbarium, unpublished data), as well as the Western Australian *G. quadrilocularis* R.Br. (Fig. 1A, B). The fact that subclade C3 occurs in four of the five highest-likelihood topologies (Fig. 3) supports the monophyly of this group. However, the taxonomic conflict with C1, which also includes a member of this section (*G. xanthotricha*; Fig. 3), and the well-supported polyphyly of its members between clade C and clade A, clearly indicates that revision of the typical section and subsection are in order.

The phylogenetic position of *Verreauxia* changes between almost all datasets in the highest-likelihood topologies (Fig. 3) and in the $P \geq 0.05$ topologies for each dataset. While there is consistency between its position in the G3PDH and NRR ML topologies, the only dataset that strongly supported a placement for *Verreauxia* was the mtDNA bootstrap analysis, which placed it sister to subclade C1 (Fig. 3). None of the $P \geq 0.05$ topologies in mtDNA placed it as sister to any other lineage. Sanger sequencing of more taxon-rich datasets (Jabaily et al., 2012; K. Shepherd, Western Australian Herbarium, unpublished data) has consistently found *Verreauxia*, which is a small genus (3 spp.) from southwestern Western Australia, to be monophyletic and sister to monotypic *Pentaptilon* (Carolin, 1992). These data suggest that *Verreauxia* (and *Pentaptilon*) form a distinct lineage within clade C, and our best hypothesis at present is that this group is closely allied to subclade C1.

4.2. Examining incongruence

Despite high overall congruence between the cpDNA and mtDNA topologies across *Goodenia* s.l. (Fig. 2), there are several points of conflict within *Goodenia* clade C, both between the cpDNA and mtDNA, and between one or more of the organellar and nuclear datasets (Fig. 3). The lack of support for subclade C1 in the COS and G3PDH analyses and for C3 in the mtDNA analyses is particularly notable, given the strong support for these relationships in the other datasets. The *P*-values at which these relationships were retained in strict consensus topologies from the other datasets were low, indicating strong relative support (Table 2).

Strong topological conflict between phylogenies derived from cytoplasmic organelles (mitochondria and chloroplast) and/or nuclear phylogenies may be evidence for incomplete lineage sorting due to rapid radiation, or may reflect that these genomes have tracked different evolutionary histories as a result of phenomena such as hybridization or polyploidy. The mitochondrial genome in angiosperms is almost always maternally inherited, as is the chloroplast genome, historically the workhorse of plant phylogenetics, and both are single, non-recombining molecules (Barr et al., 2005). For these reasons, the chloroplast and mitochondrial genomes may be assumed to track similar evolutionary histories, and can potentially be combined to enhance phylogenetic resolving power. While this has proved useful in several previous studies (e.g., Malé et al., 2014; Zhang et al., 2015), concatenation of the two alignments in the current study was not attempted given the topological incongruence between the datasets. We feel further justified retaining the independence of the datasets given that other studies have found conflicting signals between these compartments. For example, Straub et al. (2012) identified a novel phylogenetic reconstruction using partial mitochondrial genomes that was not supported by other molecular or morphological data. Henriquez et al. (2014) similarly found significant conflict between a genome-skimming-derived mitochondrial dataset compared to cpDNA, and the former did not produce strong support for phylogenetic relationships. In

contrast, Rydin et al. (2017) determined that mtDNA sequences from reference-guided assembly of genome skimming data produced a highly resolved phylogenetic reconstruction for Rubiaceae that was congruent with cpDNA, however there was conflict at several key nodes.

In diploid taxa, incongruence between the plastid and mitochondrial genomes can be evidence for paternal, or even biparental, inheritance of mitochondria, with or without subsequent recombination, which can further complicate phylogenetic inference (Barr et al., 2005; Apitz et al., 2012; Govindarajulu et al., 2015). In cases where the mitochondria are known to be differentially inherited compared to the chloroplast (e.g. in *Pinus,* which has paternal chloroplast inheritance; Wang and Wang, 2014), incongruence can reflect past introgression, or genome capture without recombination. Ancient or recent polyploidy events could also contribute to incongruence (Govindarajulu et al., 2015), and polyploidy has been documented in subsect. *Goodenia* taxa belonging to clade C (Peacock, 1963). However, we found no evidence for multiple alleles in any of the nuclear loci, and the lack of resolution in the NRR, G3PDH, and COS trees (Fig. 2) does not appear to be driven by polyploidy. There are no hard incongruences between them that could reflect hybridization or polyploidy differentially affecting the histories of these unlinked markers.

In *Goodenia* clade C, it seems most likely that a history of rapid radiation and subsequent incomplete lineage sorting, potentially combined with hybridization, has led to the intractability of backbone relationships in this group. The lack of resolution with strong support coincides with nodes having short branch lengths in all topologies (Fig. 2). Jabaily et al. (2014) provided a historical biogeographic reconstruction of the family, and found that clade C most likely originated towards the beginning of the Miocene, with the major lineages studied in the current paper originating throughout the Miocene (approximately 23–5 mya). Although the relative branch lengths reconstructed within clade C were similar to or longer than those in other groups across the family, the pattern of poor support from multiple genomic compartments suggests that widespread incomplete lineage sorting may be at work.

4.3. Relative contributions of organellar vs. nuclear data recovered by genome skimming

Given the inability of either the chloroplast or mitochondrial partial genomes to fully support backbone relationships within *Goodenia* clade C, it is important to question whether complete genome sequences of either organelle would add resolving power or support at these crucial positions. Full chloroplast genomes for land plants are between 107 and 218 kb in size (Wicke and Schneeweiss, 2015; 150 kb in *Helianthus* (Timme et al., 2007)), with 25% of that sequence occurring in non-coding regions in *H. annuus*. Our annotated cpDNA datasets include an average of 151 genetic elements (genes, tRNAs, rRNAs), which is within the range of expected full counts. The overall length of our average cpDNA sequence is 65% the total length of *H. annuus*, and is 25% non-coding. We have attempted *de novo* reconstructions of the plastomes from our genome skimming data, but the resulting assemblies are similar in length and coverage to the ones built from Bowtie2-mapped reads that are presented here. The missing sections, when compared to the *Helianthus* reference, are primarily non-coding sequences. Targeted sequencing of these faster-evolving regions could provide additional, more useful and phylogenetically informative characters. Phylogenetic support values along the backbone of *Goodenia* clade C greatly increased when the entire chloroplast spacer region *trnL-F* was analyzed across the clade, compared to a portion of the spacer plus the gene *matK* (Jabaily et al., 2012). *De novo* construction of full plastomes is possible from genome skimming libraries (Straub et al., 2011), though computational requirements are considerably higher than for reference-guided assembly. Our library preparations seem not to have captured a substantial portion of the plastome, frustrating our attempts at assembling full chloroplast genomes. Alternative methods of target enrichment or capture probes focused on spacer-rich regions may increase the likelihood of producing full chloroplast genomes (Wicke and Schneeweiss, 2015).

It is also unclear if assembling a greater percentage of the complete mitochondrial genomes of *Goodenia* s.l. would have brought us closer to a fully resolved clade C phylogeny. The percentage of parsimony-informative characters in the mtDNA dataset was substantially lower than in the cpDNA dataset (Table 3), and we did not find evidence of high rates of molecular evolution in the mtDNA regions that had previously been documented in unigene analyses of mitochondrial sequences of Goodeniaceae (Cho et al., 2004; Qiu et al., 2010). It is uncertain what proportion of the complete mitochondrial genome we assembled, as the overall size of these genomes can vary greatly, even between close relatives. Our assemblies seem to have suffered a similar fate as the cpDNA, likely missing huge tracts of sequence presumably present in the mitochondrial genomes of these species. The reference of *H. annuus* is 300,945 (Bock et al., 2014) and 19.5% was annotated as coding, similar to estimates from other completed mitochondrial genomes (Wicke and Schneeweiss, 2015). Our alignment, which mapped to less than 14% of the total reference length, had about 55% of the genetic elements annotated in the reference, though *H. annuus* also has many fewer genetic regions annotated than other species (e.g. *Daucus carota*, Iorizzo et al., 2012).

While the high sequence-level similarity between plant mitochondrial genes can make reference-guided assembly and alignment straightforward, and produce at least partial assemblies of flanking or intron regions (Malé et al., 2014), the positional homology of these assembled sequence is often difficult to discern. The overall structure of mitochondrial genomes in seed plants is highly variable and evolutionarily dynamic, and thus gene order and intergenic spacer length and content can vary dramatically, even among close relatives (Knoop, 2004; Grewe et al., 2014; Gualberto et al., 2014; Liu et al., 2014; Guo et al., 2016). Intergenic regions in particular are responsible for most of the enormous variety in size seen between sequenced mitochondrial genomes (Kubo and Mikami, 2007). This is in sharp contrast to the small gene content and highly conserved gene order of animal mitochondrial genomes, and the generally conserved gene order and intergenic spacer size in plant chloroplasts (Kubo and Mikami, 2007). Given the difficulties this structural complexity poses for designing primers to sequence intergenic spacers, additional, targeted sequencing of variable mtDNA regions may prove challenging. However, given that our mitochondrial dataset was generally concordant with the cpDNA results (at least, outside of clade C), we predict that more complete mtDNA genome sequences could similarly increase phylogenetic support (though not necessarily congruence with cpDNA). Assembling more complete mitochondrial genomes from genome skimming data presents substantial bioinformatics challenges (Wicke and Schneeweiss, 2015), but will likely be worth the effort.

Turning to the nucleus, we were surprised that our analysis of over eighty COS loci yielded a species tree with low overall statistical support and which had little in common with the trees for NRR and G3PDH (Fig. 3). Even though the COS dataset as a whole provided far more sequence data and informative characters than the smaller nuclear datasets (Table 3), the individual COS loci were relatively uninformative. Léveillé-Bourret et al. (2017) found that gene tree incongruence, which can result from analysis of short or low-information loci, can decrease support in ASTRAL species tree analyses, and our individual COS loci were generally below the ~25 PIC per locus recommended as a threshold for ASTRAL analysis using ultra-conserved elements (Meiklejohn et al., 2016). We also recovered only a handful of the complete dataset of 1061 loci that were targeted in the cross-Asteraceae probe development study (Mandel et al., 2014). This may have been caused by a lack of representation of these low-copy loci in our unenriched genome skimming data, or by the probes themselves failing bioinformatically due to the relative divergence of Goodeniaceae from Asteraceae, though the latter seems unlikely given the expected,

conserved nature of these sequences (Mandel et al., 2014). An alternate data generation approach, such as anchored phylogenomics (Léveillé-Bourret et al., 2017) or PCR-enrichment ahead of sequencing (Uribe-Convers et al., 2016) would likely yield a higher number of more informative loci that could lead to a better-resolved and supported species tree.

4.4. Conclusions

This study has increased our understanding of the evolutionary history of Core Goodeniaceae, and particularly *Goodenia* clade C, by expanding on our previous analyses primarily through the addition of mitochondrial and nuclear sequence data. These new data, combined with extensive hypothesis testing of datasets from all three genomic compartments, allow us to recover a reasonably consistent set of relationships between the major clades of clade C, and further support our hypothesis that this group represents an ancient, rapid radiation. The partial mitochondrial genomes are comparable to chloroplast genomes at resolving relationships outside of clade C, despite the fact that relatively few studies have used mitochondrial sequences in plant systematics. Several clades and relationships within *Goodenia* clade C are supported by multiple analyses, which provides support for taxonomic recognition of the major subclades in clade C, however, due to the lack of overall backbone support it is unlikely that this will be at the generic level. Additional sequencing of hundreds or thousands of nuclear loci, which is becoming an increasingly cost-effective strategy, will most likely be the only way to fully untangle the evolutionary history of this perplexing and enigmatic group of plants.

Acknowledgements

The authors thank José Miguel Ponciano (Univ. Florida); Matt Johnson (Texas Tech. University), Liam Revell (Univ. Massachusetts-Boston), and Brian O'Meara (Univ. Tennessee-Knoxville) for assistance with topological analyses, and Juliet Wege (Western Australian Herbarium) and Spencer Willis for providing field support. We gratefully acknowledge the invaluable contributions of Austin T. Bellini. Funding: This work was supported by the National Science Foundation of the United States of America [grant DEB RUI #1256946 to RSJ].

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ympev.2018.05.005.

References

Apitz, J., Weihe, A., Pohlheim, F., Börner, T., 2012. Biparental inheritance of organelles in *Pelargonium*: evidence for intergenomic recombination of mitochondrial DNA. Planta 237, 509–515.

Barr, C.M., Neiman, M., Taylor, D.R., 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. New Phytol. 168, 39–50.

Bock, D.G., Kane, N.C., Ebert, D.P., Rieseberg, L.H., 2014. Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop species: neither from Jerusalem nor an artichoke. New Phytol. 201, 1021–1030.

Carolin, R.C., 1992. Goodenia. In: In: George, A.S. (Ed.), Flora of Australia, vol. 35. Australian Government Publishing Services, Canberra, Australia, pp. 147–281.

Cho, Y., Mower, J.P., Qiu, Y.-L., Palmer, J.D., 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. Proc. Natl. Acad. Sci. U.S.A. 101, 17741–17746.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9 (8), 772.

Drouin, G., Daoud, H., Xia, J., 2008. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol. Phylogenet. Evol. 49, 827–831.

Gardner, A.G., Sessa, E.B., Michener, P., Johnson, E., Shepherd, K.A., Howarth, D.G., Jabaily, R.S., 2016a. Utilizing next-generation sequencing to resolve the backbone of the Core Goodeniaceae and inform future taxonomic and floral form studies. Mol. Phylogenet. Evol. 94 (Part B), 605–617. http://dx.doi.org/10.1016/j.ympev.2015.10.003.

Gardner, A.G., Gerald, J.N.F., Menz, J., Shepherd, K.A., Howarth, D.G., Jabaily, R.S., 2016b. Characterizing floral symmetry in the core goodeniaceae with geometric morphometrics. PloS one 11, e0154736. http://dx.doi.org/10.1371/journal.pone.0154736.

Govindarajulu, R., Parks, M., Tennessen, J.A., Liston, A., Ashman, T.-L., 2015. Comparison of nuclear, plastid, and mitochondrial phylogenies and the origin of wild octoploid strawberry species. Am. J. Bot. 102, 544–554.

Grewe, F., Edger, P.P., Keren, I., Sultan, L., Pires, J.C., Ostersetzer-Biran, O., Mower, J.P., 2014. Comparative analysis of 11 Brassicales mitochondrial genomes and the mitochondrial transcriptome of Brassica oleracea. Mitochondrion 19 (Part B), 135–143.

Gualberto, J.M., Mileshina, D., Wallet, C., Niazi, A.K., Weber-Lotfi, F., Dietrich, A., 2014. The plant mitochondrial genome: dynamics and maintenance. Biochimie 100, 107–120.

Guindon, S., Gascuel, O., 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Sys. Biol. 52, 696–704.

Guo, W., Grewe, F., Fan, W., Young, G.J., Knoop, V., Palmer, J.D., Mower, J.P., 2016. Ginkgo and Welwitschia Mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. Mol. Biol. Evol. 33 (6), 1448–1460. http://dx.doi.org/10.1093/molbev/msw024.

Henriquez, C.L., Arias, T., Pires, J.C., Croat, T.B., Schaal, B.A., 2014. Phylogenomics of the plant family Araceae. Mol. Phylogenet. Evol. 75, 91–102.

Iorizzo, M., Senalik, D., Szklarczyk, M., Grzebelus, D., Spooner, D., Simon, P., 2012. De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. BMC Plant Biol. 12, 61. http://dx.doi.org/10.1186/1471-2229-12-61.

Jabaily, R.S., Shepherd, K.A., Gustafsson, M.H.G., Sage, L.W., Krauss, S.L., Howarth, D.G., Motley, T.J., 2012. Systematics of the Austral-Pacific family Goodeniaceae: establishing a taxonomic and evolutionary framework. Taxon 61, 419–436.

Jabaily, R.S., Shepherd, K.A., Gardner, A.G., Gustafsson, M.H.G., Howarth, D.G., Motley, T.J., 2014. Historical biogeography of the predominantly Australian plant family Goodeniaceae. J. Biogeogr. 41, 2057–2067.

Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J.C., Wickett, N.J., 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment1. Appl. Plant Sci. 4 (7). http://dx.doi.org/10.3732/apps.1600016. apps.1600016.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28 (12), 1647–1649. http://dx.doi.org/10.1093/bioinformatics/bts199.

Knoop, V., 2004. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Curr. Genet. 46, 123–139.

Kubo, T., Mikami, T., 2007. Organization and variation of angiosperm mitochondrial genome. Physiol. Plant. 129, 6–13.

Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S., 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25.

Langmead, B., Salzberg, S., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.

Léveillé-Bourret, É., Starr, J.R., Ford, B.A., Moriarty Lemmon, E., Lemmon, A.R., 2017. Resolving rapid radiations within angiosperm families using anchored phylogenomics. Sys. Biol. syx050. http://dx.doi.org/10.1093/sysbio/syx050.

Liu, Y., Medina, R., Goffinet, B., 2014. 350 my of mitochondrial genome stasis in mosses, an early land plant lineage. Mol. Biol. Evol. 31 (10), 2586–2591. http://dx.doi.org/10.1093/molbev/msu199.

Malé, P.J., Bardon, L., Besnard, G., Coissac, E., Delsuc, F., Engel, J., Lhuillier, E., Scotti-Saintagne, C., Tinaut, A., Chave, J., 2014. Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. Mol. Ecol. Resour. 14, 966–975.

Mandel, J.R., Dikow, R.B., Funk, V.A., Masalia, R.R., Staton, S.E., Kozik, A., Michelmore, R.W., Rieseberg, L.H., Burke, J.M., 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. Appl. Plant Sci. 2 (2). http://dx.doi.org/10.3732/apps.1300085. apps.1300085.

Meiklejohn, K.A., Faircloth, B.C., Glenn, T.C., Kimball, R.T., Braun, E.L., 2016. Analysis of a rapid evolutionary radiation using ultraconserved elements: evidence for a bias in some multispecies coalescent methods. Sys. Biol. 65, 612–627.

Mirarab, S., Warnow, T., 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31, i44–i52.

Mower, J.P., Touzet, P., Gummow, J.S., Delph, L.F., Palmer, J.D., 2007. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. BMC Evol. Biol. 7, 135.

Palmer, J.D., Herbon, L.A., 1988. Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. J. Mol. Evol. 28, 87–97.

Parkinson, C.L., Mower, J.P., Qiu, Y.-L., Shirk, A.J., Song, K., Young, N.D., Depamphilis, C.W., Palmer, J.D., 2005. Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. BMC Evol. Biol. 5, 73.

Peacock, W.J., 1963. Chromosome numbers and cytoevolution in the Goodeniaceae. Proc. Linn. Soc. New South Wales 88, 8–27.

Qiu, Y.-L., Li, L., Wang, B., Xue, J.-Y., Hendry, T.A., Li, R.-Q., Brown, J.W., Liu, Y., Hudon, G.T., Chen, Z.-D., 2010. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. J. Syst. Evol. 48, 391–425.

Rydin, C., Wikström, N., Bremer, B., 2017. Conflicting results from mitochondrial genomic data challenge current views of Rubiaceae phylogeny. Am. J. Bot. Available at: http://www.amjbot.org/content/early/2017/10/18/ajb.1700255 (accessed 27.

10.17).
Sayyari, E., Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33, 1654–1668.
Schliep, K.P., 2011. Phangorn: phylogenetic analysis in R. Bioinformatics 15, 592–593.
Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Sys. Biol. 51, 492–508.
Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247.
Stamatakis, A., 2014. RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Straub, S.C., Fishbein, M., Livshultz, T., Foster, Z., Parks, M., Weitemier, K., Cronn, R.C., Liston, A., 2011. Building a model: developing genomic resources for common milkweed (Asclepias syriaca) with low coverage genome sequencing. BMC Genom. 12, 211. 10.1186/1471-2164-12-211.
Straub, S.C., Moore, M.J., Soltis, P.S., Soltis, D.E., Liston, A., Livshultz, T., 2014. Phylogenetic signal detection from an ancient rapid radiation: Effects of noise reduction, long-branch attraction, and model selection in crown clade Apocynaceae. Mol. Phylogenet. Evol. 80, 169–185.
Straub, S.C.K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C., Liston, A., 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. Am. J. Bot. 99, 349–364.
Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Sys. Biol. 56, 564–577.
Tank, D.C., Donoghue, M.J., 2010. Phylogeny and phylogenetic nomenclature of the Campanulidae based on an expanded sample of genes and taxa. Syst. Bot. 35, 425–441.
Timme, R.E., Kuehl, J.V., Boore, J.L., Jansen, R.K., 2007. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. Am. J. Bot. 94, 302–312.
Wicke, S., Schneeweiss, G.M., 2015. Next-generation organellar genomics: Potentials and pitfalls of high-throughput technologies for molecular evolutionary studies and plant systematics. In: Hörandl, E., Appelhans, M.S. (Eds.), Next-Generation Sequencing in Plant Systematics. International Association for Plant Taxonomy (IAPT), pp. 1–42.
Uribe-Convers, S., Settles, M.L., Tank, D.C., 2016. A phylogenomic approach based on PCR target enrichment and high throughput sequencing: resolving the diversity within the South American Species of Bartsia L. (Orobanchaceae). PLoS One 11(2), e0148203. https://doi.org/10.1371/journal.pone.0148203.
Wang, B., Wang, X.-R., 2014. Mitochondrial DNA capture and divergence in Pinus provide new insights into the evolution of the genus. Mol. Phylogenet. Evol. 80, pp. 20–30.
Wolfe, K.H., Li, W.H., Sharp, P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. PNAS 84, 9054–9058.
Zhang, N., Wen, J., Zimmer, E.A., 2015. Congruent Deep Relationships in the Grape Family (Vitaceae) Based on Sequences of Chloroplast Genomes and Mitochondrial Genes via Genome Skimming. PLoS One 10 (12), e0144701. http://dx.doi.org/10.1371/journal.pone.0144701.
Zhu, A., Guo, W., Jain, K., Mower, J.P., 2014. Unprecedented Heterogeneity in the Synonymous Substitution Rate within a Plant Genome. Mol. Biol. Evol. 31(5), pp. 1228–1236. 10.1093/molbev/msu079.
Zwickl, D.J., 2006. Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets under the Maximum Likelihood Criterion. Ph.D. dissertation, The University of Texas at Austin.