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.