An Evolutionary Framework of Acanthaceae Based on Transcriptomes and Genome Skims

Authors: Arias, Joshua D., Manzitto-Tripp, Erin, Kiel, Carrie A., McDade, Lucinda A., and Fisher, Amanda E.

Source: Systematic Botany, 47(3) : 716-728

Published By: The American Society of Plant Taxonomists

URL: https://doi.org/10.1600/036364422X16573019348256

Systematic Botany (2022), 47(3): pp. 716–728
ISSN 0363-6445 (print); ISSN 1548-2324 (online)

DOI 10.1600/036364422X16573019348256
Date of publication September 16, 2022

An Evolutionary Framework of Acanthaceae Based on Transcriptomes and Genome Skims

Joshua D. Arias,[1,2] Erin Manzitto-Tripp,[3] Carrie A. Kiel,[4] Lucinda A. McDade,[4] and Amanda E. Fisher[2,5]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9601 Medical Center Dr., Room 6E610, Rockville, Maryland 20850, USA; josh.arias@nih.gov
[2]Biological Sciences, California State University, Long Beach, 1250 Bellflower Boulevard, Long Beach, California 90807, USA; Amanda.Fisher@csulb.edu
[3]Department of Ecology and Evolutionary Biology and Museum of Natural History, University of Colorado, UCB 350, Boulder, Colorado 80309, USA; erin.tripp@colorado.edu
[4]California Botanic Garden and Claremont Graduate University 1500 North College Avenue, Claremont, California 91711, USA; ckiel@calbg.org; lmcdade@calbg.org
[5]Author for correspondence

Communicating Editor: Martin F. Wojciechowski

Abstract—Acanthaceae is a family of tropical flowering plants with approximately 4900 species. Despite remarkable variation in morphological traits, research on patterns of character evolution has been limited by uncertain relationships among some of the major lineages. We sampled 16 taxa from these major lineages to estimate a phylogenomic framework using a combination of five newly sequenced shotgun genome skims plus seven new and four publicly available transcriptomes. We used OrthoFinder2 to infer a species tree with strong branch support. Except for the placement of *Crabbea*, our results corroborate the most recent chloroplast and nrITS sequence-based topology. Of 587 single copy loci, 10 were recovered for all 16 species; a RAxML tree estimated from these 10 loci resulted in the same topology as other datasets assembled in this study, with the exception of relationships among three sampled species of *Barleria*; however, branch support was lower compared to the tree reconstructed using more data. ABBA-BABA tests were conducted to investigate patterns of introgression involving *Crabbea*; few nucleotides supported alternative topologies. SplitsTree networks of the 587 loci and 6136 orthogroup trees revealed conflict among the branches leading to Andrographideae, Whitfieldieae, and *Neuracanthus*. A principal components analysis in treespace found no distinct clusters of trees. Our results based on combined genome skim and transcriptome sequences strongly corroborate the previously published chloroplast and nr-ITS-based phylogeny of Acanthaceae with increased resolution among Barlerieae, Andrographideae, Whitfieldieae, and *Neuracanthus*. This advance in our knowledge of Acanthaceae relationships will allow us to investigate character evolution and other phenomena within this diverse group of plants in studies with increased taxon sampling.

Keywords—Angiosperms, high throughput sequencing, Lamiales, phylogenetics.

Acanthaceae (Lamiales) is a family of herbs, shrubs, lianas, and trees with an estimated 4900 species (McDade et al. 2008). A revised classification of the family recognizes 191 genera, 19 subtribes, 10 tribes, and four subfamilies and includes *Avicennia* mangroves (Tripp et al. 2022). Most members of Acanthaceae have a woody capsule that is explosively dehiscent (absent in some early-diverging members). Acanthoideae (the "retinaculate clade") have hook-like structures that propel seeds as the fruit dehisces and this is a synapomorphy for the clade (McDade and Moody 1999; McDade et al. 2008; Cooper et al. 2018; Tripp et al. 2022). A great deal remains to be learned about morphology, anatomy, and physiology in the family, in particular with respect to pollination biology and patterns of floral evolution (see McDade and Weeks 2004; Tripp 2007, 2008; Tripp and Manos 2008; Muchhala et al. 2009; Ortegón-Campos et al. 2009; Tripp and Tsai 2017; Zhuang and Tripp 2017b).

As new sequencing and phylogenetic methods have become available, knowledge of Acanthaceae phylogeny and understanding of morphological evolution have increased (McDade et al. 2005; Daniel et al. 2008; Kiel and McDade 2014; Fisher et al. 2015; Kiel et al. 2018; Darbyshire et al. 2019). Yet, there has been no family-wide investigation of phylogenetic relationships since McDade et al. (2008; see also Tripp and McDade 2014). McDade et al. (2008) used four chloroplast regions and nuclear nrITS, and sampled all major lineages across the family. These authors recovered a strongly supported tree, except for relationships among the clades that comprise "BAWN" (i.e. Barlerieae, Andrographideae, Whitfieldieae, *Neuracanthus*; Fig. 1). In particular, the placement of *Neuracanthus* as sister to the remainder of "BAWN" was weakly supported (Bayesian posterior probability [BPP] = 0.72; parsimony bootstrap [PB] < 50%; Decay Index = 1). Tripp and McDade (2014) used nrITS and added two cp regions for a total of six; these authors recovered a tree with all branches > 99% BPP. McDade et al. (2008) found a BAWN clade that is sister to a strongly supported Justicieae + Ruellieae (Fig. 1). In contrast, Tripp and McDade (2014) found that *Neuracanthus* is sister to BAW + (Justicieae + Ruellieae). Although recent studies have advanced our knowledge of evolutionary relationships within a number of lineages (McDade et al. 2005, 2012, 2018; Kiel et al. 2006, 2017, 2018; Tripp 2007; Daniel et al. 2008; Tripp et al. 2013, 2017; Tripp and Darbyshire 2017; Grall and Darbyshire 2021), lack of resolution along the backbone of the Acanthaceae phylogeny is a hindrance to understanding molecular and morphological evolution across the family (McDade et al. 2008; Tripp and McDade 2014). To date, there is one publicly available genome (*Ruellia speciosa*; Zhuang and Tripp 2017a) and 17 transcriptomes of Acanthaceae (Huang et al. 2014; Garg et al. 2015; Yang et al. 2015; Cherukupalli et al. 2016; Zhuang and Tripp 2017b, 2022; Lyu et al. 2018; Morais et al. 2019).

We took a phylogenomic approach to estimate a new phylogeny for Acanthaceae. Our sequencing strategy combines the benefits of genome skims and the benefits of transcriptomes. Major advances in sequencing technology and methods of sub-sampling the genome have enabled the use of thousands of unlinked nuclear loci for phylogenomic studies. While the cost of genome sequencing has decreased over time (Loman et al. 2012), it remains resource-intensive to sequence entire genomes for most plant species, particularly because plants have some of the largest genomes known to

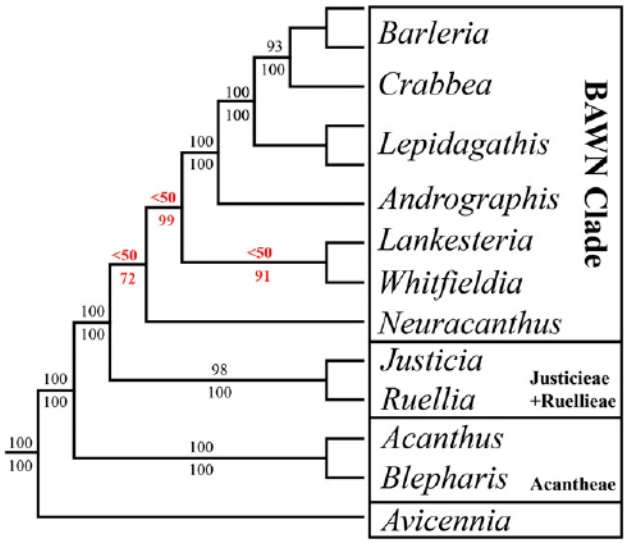


FIG. 1. Phylogenetic estimate of Acanthaceae based on four chloroplast loci and nr-ITS, adapted from McDade et al. (2008), pruned to include only taxa sampled in the present study. Parsimony bootstrap and Bayesian posterior probability values for major clades are shown. The three branches that had low ML bootstrap support in the chloroplast and nr-ITS study are shown with red support values.

date (Dodsworth et al. 2015). An efficient alternative is to sequence and assemble the transcriptome, effectively sampling a subset of the genome. For example, the transcriptome is estimated to compose less than 5% of the total genome in humans (Frith et al. 2005; Eldem et al. 2017; Ungaro et al. 2017). Transcriptome sequences can be mined for orthologous loci (Morais et al. 2019) and used to estimate nuclear phylogenies, whether they were initially sequenced for that purpose or not (Rothfels et al. 2013; Wickett et al. 2014; Garg et al. 2015; Cherukupalli et al. 2016; Hodel et al. 2016; White et al. 2016; Wu et al. 2016). A drawback of transcriptome sequencing is that it must begin with RNA purification and RNA degrades quickly. Homebrew and commercial kits (Salehi and Najafi 2014) are available to preserve RNA from fresh tissue for later extraction, but each species that a researcher wishes to sample must be grown or collected in the field with some forethought to using the tissue for transcriptome sequencing.

When no RNA from living tissue of a plant is available and sequencing transcriptomes is not possible, a convenient alternative is to sequence genome skims from dried plant tissue. In genome skimming, a library of the total genomic DNA is multiplexed with many genomic libraries on a sequencing flow cell, therefore restricting the number of sequence fragments generated for each library. Genome skimming is a relatively inexpensive and efficient method to generate data for phylogenetic studies of non-model organisms (Richter et al. 2015). DNA for genome skimming can be extracted and sequenced from dried leaves, including plant tissue preserved as herbarium specimens (Ripma et al. 2014; Washburn et al. 2015). DNA can be successfully extracted from specimens that are many decades old (Weiß et al. 2016; Yeates et al. 2016; Funk 2018; Jordon-Thaden et al. 2020). Although DNA in leaves has been estimated to decay exponentially (Hart et al. 2016), this loss can be mitigated using modern library preparation and next-generation sequencing (NGS) methods.

Other methods to sub-sample the genome include restriction site-associated sequencing (RAD-seq; Gnirke et al. 2009) and Hyb-seq (Lemmon et al. 2012; Weitemier et al. 2014). These are relatively low-cost skimming methods that have been successfully used for plant phylogenomic studies (Eaton and Ree 2013; McKain et al. 2018; Dodsworth et al. 2019; Hale et al. 2020), including studies of Acanthaceae (Tripp et al. 2017; Daniel and Tripp 2018; Comito et al. 2022; Darbyshire et al. 2020; Tripp and Darbyshire 2020). While RADseq and Hyb-seq are effective at generating sequence data from across the genome for multiple samples, data can only be combined if the same restriction enzymes or baits are used to prepare the samples. This may make it difficult to add new samples later or to combine data across studies, whereas this is relatively easy to do with data from transcriptomes and genome skimming.

Genome skimming, transcriptomes, RADseq, and Hyb-seq have been compared in terms of their costs and efficacy (Wen et al. 2015; Yu et al. 2018; Zimmer and Wen 2015). Transcriptomes are often sequenced to guide the identification of coding sequences for targeted sequencing (Gnirke et al. 2009; Pavey et al. 2015; Washburn et al. 2015; Fisher et al. 2016; Tripp et al. 2017; An et al. 2019; Darbyshire et al. 2020). Sequences generated through genome skimming can be used to compare organellar phylogenetic signals to nuclear signals (Schmickl et al. 2016; Washburn et al. 2017). However, if homologous loci can be identified directly from both genome skims and transcriptomes, then these sequences may be combined in a single matrix to maximize species sampling that may be otherwise constrained by the availability of living tissue for RNA extraction.

To take advantage of data from both transcriptomes and genome skimming, we trialed an approach to combine the two sequencing types to estimate a phylogeny of Acanthaceae. We sequenced seven new transcriptomes and five genome skims, identified orthologous loci across sampled taxa, and estimated a nuclear phylogeny of Acanthaceae. We used this phylogeny to test the topologies of previous studies that used Sanger sequencing (McDade et al. 2008; Tripp and McDade 2014), particularly for those branches of the tree that had low support in earlier studies. We then determined support for the topology using maximum likelihood (ML) bootstrapping, networks, and tree-shape similarity analyses. We compared conflicts between our tree and prior trees based on chloroplast + nrITS loci (McDade et al. 2008; Tripp and McDade 2014) using ML constraint analyses and ABBA-BABA tests for introgression.

MATERIALS AND METHODS

Sampling—Sixteen species of Acanthaceae were sampled in the present study to trial a method of combining genome skims and transcriptomes for phylogenetic analysis. Taxa were chosen to represent the major lineages of Acanthaceae based on McDade et al. (2008) and summarized in Tripp et al. (2022) with an emphasis on sampling representatives from the BAWN clade (10 species) to test for monophyly of *Barleria* and guide our research in this group (Comito et al. 2022). Six sampled taxa represent other major lineages of Acanthaceae whose phylogenies have been recently studied in greater detail, with the exception of *Acanthus* (Tripp 2007; Tripp et al. 2013; Fisher et al. 2015; Kiel et al. 2017, 2018). *Avicennia marina* was used as the out-group (Appendix 1). Transcriptomes for three species (*Acanthus leucostachys, Andrographis paniculata, Avicennia marina*) were downloaded from GenBank (Table 1; note that since the current study was completed, five additional transcriptomes have been released; Morais et al. 2019). The transcriptome for *Ruellia simplex* was sequenced and assembled in E. Tripp's lab at the University of Colorado Boulder (Zhuang and Tripp 2022). New sequences were generated for 12 species

TABLE 1. Taxa sampled and sequence data collection method. Three transcriptomes were downloaded from public databases. E. Tripp provided a *Ruellia simplex* transcriptome assembly and seven additional transcriptomes were sequenced for this study. Five species were sequenced from dried leaf material using shotgun genomic skimming (Gen. skim). NA means the voucher information was not available in the publication or GenBank record. GenBank accession numbers are listed in Appendix 1.

Species	Voucher	Sequencing
Acanthus leucostachys Wall. ex Nees	NA	Transcriptome (Yang et al. 2015)
Andrographis paniculata Nees	NA	Transcriptome (Garg et al. 2015)
Avicennia marina (Forssk.) Vierh.	NA	Transcriptome (Huang et al. 2014)
Barleria albostellata S. Moore	*J.D. Arias 3* (LOB)	Transcriptome; University of California, Irvine
Barleria oenotheroides Dum. Cours.	*J.D. Arias 1* (LOB)	Transcriptome University of California, Irvine
Barleria rotundifolia Oberm.	*A.E. Fisher 412* (LOB)	Transcriptome; University of California, Irvine
Blepharis diversispina (Nees) C. B. Clarke	*M. Stata & R. Sage s.n.*	Transcriptome; University of California, Irvine
Blepharis spinifex Merxm.	*S. Loots et al. s.n.* (WIND) Millennium Seed Bank # 248224	Transcriptome; University of California, Irvine
Crabbea velutina S. Moore	*Manktelow 670* (UPS)	Gen. skim.; University of California, Riverside
Justicia pacifica (Oerst.) A. Gray	*A.E. Fisher 447* (LOB)	Transcriptome; University of California, Irvine
Lankesteria elegans (P. Beauv.) T. Anderson	*Etuge & Thomas 466* (K)	Gen. skim.; Beijing Genomics Institute, Cambridge, MA
Lepidagathis dulcis Nees	*Suddee et al. 999* (BKF)	Gen. skim.; Beijing Genomics Institute, Cambridge, MA
Lepidagathis sessilifolia (Pohl) Kameyama ex Wassh. & J.R.I. Wood	*Daniel et al. 10106* (CAS)	Gen. skim.; Beijing Genomics Institute, Cambridge, MA
Neuracanthus africanus T. Anderson ex S. Moore	*McDade & Balkwill 1258* (ARIZ)	Gen. skim.; University of California, Riverside
Ruellia simplex C. Wright	UC Boulder greenhouse	Transcriptome (Zhuang & Tripp 2022)
Whitfieldia elongata (P. Beauv.) DeWild. & T. Durand.	*A.E. Fisher 443* (LOB)	Transcriptome; University of California, Irvine

(Appendix 1). Of these, mRNA was extracted from fresh tissue of seven species for transcriptome sequencing (*Barleria albostellata, B. oenotheroides, B. rotundifolia, Blepharis diversispina, Blepharis spinifex, Justicia pacifica, Whitfieldia elongata*). DNA was extracted for shotgun genome skim sequencing for five species for which only dried leaf tissue was available (*Crabbea velutina, Lankesteria elegans, Lepidagathis dulcis, Lepidagathis sessilifolia, Neuracanthus africanus*).

Nucleic Acid Extraction—mRNA was extracted using the BrAD-seq (Breath Adapter Directional Sequencing) method from fresh leaves of seven samples grown in the California State University Long Beach greenhouse. Voucher specimens of these plants were deposited in the Long Beach Herbarium (LOB). We followed Townsley et al.'s (2015) protocol except that we used 500 µL of lysate binding buffer and washed streptavidin-bound beads with 200 µL buffer A, buffer B, and low salt buffer. The mRNA was quantified using a Synergy H1 Hybrid Reader (Biotek, Winooski, Vermont).

Genomic DNA was extracted from herbarium samples using a modified CTAB method (Doyle and Doyle 1987). Dried leaf tissue was flash-frozen in liquid nitrogen and added to 3 g of sand in a mortar and ground into a fine powder, then added to 500 µL of extraction buffer in a tube. Tubes were incubated for 15 min in a 37°C dry plate, then in a 65°C water bath for 30–40 min and inverted every 5 min. Next, 5 µL of RNAse A was added to the tube which was then incubated for 30 min in a 37°C dry plate. Next, 500 µL of cold chloroform was added and tubes were centrifuged for 10 min at 12,000 rpm. Next, 400 µL of the supernatant was removed and DNA was precipitated in 1 mL of cold 95% EtOH for 2–3 wk in a −20°C freezer. Finally, samples were spun in a refrigerated (4°C) centrifuge for 15 min at 12,000 rpm to form a pellet, washed in 1 mL cold 95% EtOH, and spun for 5 min at 12,000 rpm. Any remaining supernatant was evaporated using a speed-vacuum centrifuge for 20 min on medium heat. Pellets were resuspended in 100 µL ultra-pure Millipore water. Quality was checked on a 1% agarose gel run for ~45 min at 120 mV and quantified using a Synergy H1 Hybrid Reader (Biotek, Winooski, Vermont).

Library Preparation and Sequencing—Once extracted, the mRNA was fragmented, cDNA was synthesized, and barcode adapters were annealed. The mRNA (7.5 µL) was fragmented in a solution of RT Buffer and 3′ priming adapter, incubated at 25°C for 1 s, 94°C for 1.5 min, and 4°C for 5 min. Of this mRNA solution, 5 µL was mixed with reverse transcriptase. Next, cDNA was synthesized at 25°C for 10 min, 42°C for 50 min, 50°C for 10 min, 70°C for 10 min. The cDNA was isolated using 5 µL of 50-mM EDTA (pH 8.0) and 30 µL Agencourt AMPure XP beads (Beckman Coulter, Brea, California). Clumped beads were washed twice with 300 µL 80% EtOH. Illumina barcode adapters were annealed by adding DNA Polymerase solution, incubating at room temperature for 15 min then adding 10 µL 50-mM EDTA pH 8.0 and 30 µL of new AMPure XP Bead Resuspension Buffer, and washing twice with 300 µL 80% EtOH. cDNA was quantified using a Synergy H1 Hybrid Reader (Biotek, Winooski, Vermont), and by running 2 µL of product on a 1% agarose gel. Only samples that displayed long, bright smears in the gel were submitted for sequencing.

cDNA libraries for seven species (Table 1) were prepared and sent to the University of California–Irvine (UCI) Genomics High-Throughput Facility for sequencing on an Illumina HiSeq 4000 (San Diego, California). Library quality was assessed at UCI with a BioAnalyzer (Agilent, Santa Clara, California) and samples were then multiplexed in one lane for 150 bp, paired-end sequencing.

Total DNA was extracted for *Lankesteria elegans, Lepidagathis dulcis,* and *L. sessilifolia* (Table 1) using the modified CTAB protocol described above and quality was checked with an agarose gel and a Synergy Hybrid Reader. Extractions were sent to the Beijing Genomics Institute Next-Generation Sequencing Lab (Philadelphia, Pennsylvania) where Illumina libraries were sequenced on an Illumina HiSeq 5000 (San Diego, California) for 100 bp, paired-end sequencing. Total DNA was extracted for *Neuracanthus africanus* and *Crabbea velutina* using the modified CTAB protocol and sent to the University of California–Riverside (UCR), Institute for Integrative Genome Biology for Illumina library preparation. Libraries were sequenced on an Illumina NextSeq500 (San Diego, California) with 100 bp, paired-end sequencing.

Sequence Fragment Pre-Processing, De Novo Sequence Assembly, and Quality Assessment—All commands, scripts, and parameters used for the steps below are available at https://github.com/jdarias93/slurmScripts. Output files for 12 newly sequenced species were uploaded to the CyVerse cyberinfrastructure (Merchant et al. 2016). Forward and Reverse reads were trimmed and filtered separately. Sequencing adapters were trimmed using Scythe v. 0.991 (Buffalo 2011). Poor quality reads were filtered using PRINSEQ v. 0.20.4 (Schmieder and Edwards 2011). FastQC v. 0.11.7 (Andrews 2010) was used to ensure that adapters and low quality reads were successfully removed. Sequences were then uploaded to Indiana University's (IU) Galaxy (Afgan et al. 2018) instance for Trinity v. 2.4.0 (Grabherr et al. 2011) de novo transcriptome assembly, which is available at https://galaxy.ncgas-trinity.indiana.edu/. Each sequence file was assembled individually using these parameters: "Trinity -max_memory 240G -CPU 8 –seqType Paired -left <left_file> -right <right_file>." Assembled transcriptomes were downloaded from IU-Galaxy and uploaded to CyVerse for quality assessment.

For five species, de novo assembler SPAdes v. 3.12.0 (Bankevich et al. 2012) was used to assemble contiguous sequences from the raw read genome skim data with k-mer sizes 21, 33, 55, 77, and 99. For each species, contigs assembled with different k-mer sizes were concatenated into a single file.

Combining Transcriptome and Genome Skim Data—Putative coding regions were identified and translated into amino acid sequences to identify orthologous loci across the transcriptome and genome skim assemblies. For the 11 transcriptome assemblies, TransDecoder v. 2.0 (Haas and Papanicolaou 2012) was used to translate the Trinity assemblies into amino acids using the universal genetic code option and identify coding regions using peptides of 100 nucleotides minimum length. CD-HIT v. 4.6 (Fu et al. 2012) removed duplicate amino acid sequences and those with 90% similar global sequence identity. To prepare the five genome skim assemblies, SPAdes contig output files were used for gene prediction in AUGUSTUS v. 3.1 (Stanke and Morgenstern 2005; Keller et al. 2011) trained on *Arabidopsis thaliana* and *Solanum lycopersicum*. Amino acid FASTA files were taken from AUGUSTUS GFF gene predictions using the script "getAnnoFasta.pl" (Stanke 2007). Gene predictions were concatenated and redundant sequences were removed with CD-HIT v. 4.6 (Fu et al. 2012) using 90% global sequence identity. Transcriptome and genome skim completeness was assessed with BUSCO (Benchmarking Universal Single-Copy Orthologs; Simão et al. 2015; Waterhouse et al. 2018).

Identifying Orthologous Loci and Phylogenetic Tree Inference—FASTA metadata were standardized to the format "><species_name>_<gene_id#>" for all species. OrthoFinder v. 2.2.7 (Emms and Kelly 2015, 2019) was used to identify orthologous loci for phylogenetic analysis. This was done by identifying orthogroups that may contain multiple gene copies for each species. Orthogroups are used to infer rooted orthogroup trees and, through an iterative process, a species tree is inferred from the set of orthogroup trees. The orthogroup trees and multiple sequence alignments (MSAs) are then reconciled with the species tree so that they contain a single sequence for each species. Amino acid sequences were analyzed in OrthoFinder v. 2.2.7 (Emms and Kelly 2015, 2019) on the UCR High-Performance Computing Center cluster using sequence similarity determined by Diamond v. 0.9.22 (Buchfink et al. 2014) and two tree estimation approaches: (Run 1) a distance matrix-based approach that uses DendroBLAST (Kelly and Maini 2013) with FastME v. 2.1.5 (Lefort et al. 2015), and (Run 2) an alignment-based approach that uses MAFFT v. 7.271 (Katoh and Standley 2013) with FastTree v. 2.1.10 (Price et al. 2010).

An unrooted species tree was estimated for all OrthoFinder2 analyses using STAG v. 1.0 (Emms and Kelly 2018). This is done within OrthoFinder2 when DendroBLAST is used (Run 1 above). For Run 2, orthogroup trees were analyzed with the standalone STAG software. STAG filters the orthogroup trees and only uses loci for which all species have at least one sequence. A STAG species tree is estimated from the no-missing-species orthogroup trees using a greedy consensus method. STAG support values are different from bootstrap support values in that they are calculated as the proportion of orthogroup trees that contain a specific bipartition (a clade) out of the total number of orthogroup trees that contain at least one sequence per species. We rooted the STAG species tree with *Avicennia marina* (McDade et al. 2008).

An ML phylogeny was estimated using the OrthoFinder2 MAFFT amino acid alignment of 587 loci that were present for at least 9/16 (56%) species. The loci included in this alignment are single-copy or almost single-copy, using a tree-based assessment of orthology and copy number. OrthoFinder2 (Emms and Kelly 2018) automatically trims positions in the alignment with > 50% missing data. ProtTest v. 3.4.2 (Abascal et al. 2005; Darriba et al. 2011) was used to determine the suitable amino acid substitution models using Akaike Information Criterion (AIC). RAxML v. 8.2.10 (Stamatakis 2014) was accessed in CIPRES (Miller et al. 2010) using the amino acid substitution model JTT+I+G+F with 1000 bootstrap replicates. The analysis was repeated with three sub-optimal models from ProtTest (JTT+G+F, WAG+I+G+F, LG+I+G+F) to assess the robustness of the phylogeny to the model. Maximum likelihood bootstrap support values were considered strong if > 90%, moderate if 80–89%, low if 70–79%, and unsupported if < 70% (Hillis and Bull 1993).

Testing the Robustness of the Species Tree—A neighbor-net network of the 587 locus MSA was constructed in SplitsTree v. 4.14.4 (Huson and Bryant 2006) to graph the character splits in the MSA to identify conflicting characters or alternative branching patterns. The initial networks used all 16 species; to check for conflict in the alignment caused by a specific sample, networks were re-created after sequentially removing *Crabbea velutina* and *Avicennia marina*. *Crabbea velutina* was removed due to its alternative placement in the nuclear tree compared to earlier results using chloroplast and nr-ITS data. Removing *Avicennia marina* was trialed because it was the smallest transcriptome and the initial network showed conflict where it attached to the rest of the network.

Additionally, Newick files of the 6136 orthogroup trees were merged into a single file and imported into SplitsTree4 v. 4.14.4 (Huson and Bryant 2006) in order to construct a tree super-network, assess conflict, and locate alternative signals among the branches in the set of trees. The first iteration of the super network identified *Avicennia marina* and *Crabbea velutina* as introducing conflict; these taxa were sequentially removed and the network analysis was repeated.

The R package treespace (Jombart et al. 2017) was used to further assess conflict among the STAG orthogroup trees, and distances between the STAG orthogroup trees, the STAG species tree, and the cp + nr-ITS tree of McDade et al. (2008). The merged STAG orthogroup tree files were rooted with *Avicennia marina* and analyzed with a principal component analysis (PCA) using the tree distance measure treeVec (Kendall and Colijn 2015). treeVec, like tree distance metrics such as Robinson-Foulds (RF; Robinson and Foulds 1981), quantifies tree topologies using weighted branch lengths; it differs in how it defines distance between a root and the most recent common ancestor, rather than between tips and their most recent common ancestor. This method is potentially an improvement, since RF metrics might lead to high, counter-intuitive distance measures between trees that differ in the placement of even a single tip (Kendall and Colijn 2015). A dendrogram of the PCA groups was used to define 20 clusters of trees and a median consensus tree was constructed to represent each cluster.

To further compare the relationship between *Crabbea velutina* and *Barleria*, the ML score of the 587 loci ML tree was compared with a ML tree that constrained *Crabbea velutina* to be sister to the three sampled *Barleria* species as in McDade et al. (2008). RAxML-HPC Black Box on CIPRES (Miller et al. 2010) was used to estimate an ML tree with the constraint topology. The 587 loci alignment, 587 loci ML tree, and the ML constraint tree were imported into PAUP v. 4.0a165 (Swofford 1998) and likelihood scores calculated for a RELL distribution of 1000 replicates. These scores were used for a one-tailed Shimodaira-Hasegawa test (SH; Shimodaira and Hasegawa 1999) and an Approximately Unbiased test (AU; Shimodaira 2002). The -lnL difference between the two trees and *p* values were calculated to test the null hypothesis that there is no significant difference between the tree likelihoods.

Additionally, ABBA BABA tests in evobiR (Blackmon and Adams 2015) were used to test for potential introgression between lineages. The unexpected placement of *Crabbea velutina* sister to the rest of Barlerieae led us to assess the possibility of introgression between it and *Barleria* and *Lepidagathis*. For 10 combinations of species, alignments were reduced to four taxa and all columns containing gaps were removed. ABBA BABA as implemented in evobiR uses jackknife tests and we chose 1000 replicates, removing a block of 1000 sites each replicate.

Phylogeny Based on Single Copy Nuclear Loci with No Missing Data—We also estimated a phylogeny using 10 nuclear loci that were recovered for all species as a single sequence, leading to the easiest assignment of homology across sequences. The *Ruellia simplex* orthologs for each single-copy locus were the query for a non-redundant BLASTp search to identify the closest matches in GenBank. Amino acid alignments for these loci were trimmed on the 5′ and 3′ ends and concatenated in Geneious v. 11.1.5. ProtTest v. 3.4.2 (Abascal et al. 2005; Darriba et al. 2011) was used to find AIC values and choose the best-fit models for a partitioned alignment. RAxML (HPC2 on XSEDE) on CIPRES (Miller et al. 2010) was used to estimate a ML phylogeny using the JTT+I+G model for each partition with 1000 ML bootstrap replicates.

Results

Sequence/Assembly Statistics and Quality Assessment—For the 12 newly sequenced Acanthaceae species (Table 1), transcriptome and genome skim sequencing resulted in an average of 38,924,165 raw read pairs per newly sequenced species (minimum 20,534,431 raw read pairs in *Lankesteria elegans*; maximum 72,165,270 raw read pairs in *Neuracanthus africanus*; Table 2). Raw sequence reads are available in GenBank (Table 2). For genome skim and transcriptome data combined, there was an average of 317,636 contigs per species after assembly (minimum 61,540 contigs in *Andrographis paniculata*; maximum 1,063,996 contigs in *Crabbea velutina*). Trinity de novo transcriptome assembly and TransDecoder amino acid translation resulted in an average of 110,045 contigs per transcriptome with an average N50 of 1185 (i.e. at least half of the nucleotides in the transcriptome assembly are found in contigs of ≥ 1185). SPAdes genome skim assembly and

TABLE 2. Statistics for transcriptomes and genome skims. Reads were measured using FastQC (Andrews 2010) before (raw read pairs) and after cleaning and filtering with Scythe (Buffalo 2011) and PRINSEQ (Schmieder and Edwards 2011). Genome skimming samples are in gray rows. GC%, number of contigs, and N50 were measured with "Compute Contig Statistics." The upper and lower numbers in the Filtered Reads column for some species are the paired forward and reverse reads, respectively.

Species	Raw read pairs	Filtered reads	GC%	Contigs	N50
Acanthus leucostachys	23,444,377	23,426,349 23,391,691	47	136,292	1365
Andrographis paniculata	NA NA	NA NA	NA NA	61,540	2620
Avicennia marina	NA NA	NA NA	NA NA	89,238	532
Barleria oenotheroides	36,877,487	36,719,869 36,510,792	46	138,937	1726
Barleria rotundifolia	41,744,166	41,140,846 39,063,165	48	104,131	435
Barleria albostellata	43,640,293	43,457,038 43,208,248	47 46	159,291	1675
Blepharis diversispina	51,012,971	50,231,649 48,083,943	47	99,649	442
Blepharis spinifex	58,704,594	57,686,086 55,820,618	48	110,113	459
Crabbea velutina	48,751,507	47,839,355 47,581,991	36	115,749	175
Justicia pacifica	33,484,158	33,349,030 33,160,332	47	78,137	1858
Lankesteria elegans	20,534,431	20,483,792 14,368,248	37	145,280	214
Lepidagathis dulcis	20,512,122	20,493,988 20,473,610	38	351,484	184
Lepidagathis sessilifolia	20,540,453	20,504,642 20,479,807	37	330,240	174
Neuracanthus africanus	72,165,270	70,411,714 69,931,939	36	95,509	1136
Ruellia simplex	NA NA	NA NA	NA NA	76,276	989
Whitfieldia elongata	34,573,993	34,418,469 34,216,538	47 46	80,058	1703

AUGUSTUS amino acid prediction resulted in an average of 789,703 contigs per genome skim with an average N50 of 377. The average BUSCO score was 74.88% complete for transcriptomes and 73.42% complete for genome skims (Table 3).

Ortholog Inference, Multiple Sequence Alignment, and Gene/Species-Tree Estimation—Across all species, 76.8% of loci were placed in an orthogroup (Fig. 2; Table 4) and OrthoFinder2 yielded 1,257,148 orthogroups. Of these, 6136 orthogroups contained at least one sequence for all 16 species. Only 10 loci were present as only one sequence for all 16 species; these are putative single copy orthogroups (SCOGs). There were 587 single-copy orthogroups with one sequence present for at least nine (56.2%) species (i.e. allowing for ≤ 7 species to be missing or to have multiple sequences for that locus; this data set is referred to as Single-Copy locus with Missing Data [SCMD]). The 587 SCMD loci MSAs were concatenated, and the resulting alignment had an average un-gapped length of 105,926 amino acids (minimum length = 51,181; maximum length = 133,413), 877 (0.5%) identical sites), a mean pairwise identity of 39.5%, and 36% missing data (952,548 missing amino acids). The alignments are available in Dryad (Arias et al. 2022).

A RAxML tree based on the 587 loci SCMD MSA and rooted with *Avicennia marina* recovered Acantheae as sister to the cystolith clade (Fig. 3). *Justicia* + *Ruellia* are sister and together are sister to the BAWN clade. Barlerieae are monophyletic and *Crabbea* is sister to *Lepidagathis* + *Barleria*. All branches in the RAxML tree had strong ML bootstrap support (> 98%). The 6136 orthogroup trees were used to estimate a STAG species tree in OrthoFinder2 (Fig. S1, Arias et al. 2022); this tree had the same topology as the SCMD RAxML tree. A STAG species tree (Fig. S2, Arias et al. 2022) from 28,316 FastME trees using the DendroBLAST distance matrix of 6136 orthogroups returned a topology with *Crabbea velutina* sister to *Lepidagathis*, and these together sister to *Barleria*, within which *B. oenotheroides* and *B. albostellata* were sister (Fig. S2).

Single Copy Loci with No Missing Data—OrthoFinder2 found 10 single copy loci with no missing data (SCOGs); the top three BLAST hits for each SCOG locus are presented in Table 5. The RAxML tree of the 10 SCOGs (Fig. 4) corroborated the topology of the 6136 loci STAG (Fig. S1) and of the 587 loci SCMD RAxML trees (Fig. 3), except that *Barleria rotundifolia* is sister to *B. oenotheroides* + *B. albostellata* in the 10 SCOG tree (Fig. 4) and *B. oenotheroides* is sister to *B. rotundifolia* + *B. albostellata* in the STAG (Fig. S1) and SCMD trees (Fig. 3).

Phylogenetic Robustness Tests—Three sub-optimal models of amino acid evolution were used for RAxML estimation with the 587 loci SCMD MSA: JTT+G+F, WAG+I+G+F, and LG+I+G+F, which were the second, ninth, and 24th best models rated by ProtTest, respectively. Each model returned the same topology as the best model (Fig. S3, Arias et al. 2022).

The 587 loci data set rejects the alternative topology of *Crabbea velutina* sister to *Barleria* ($p < 0.001$) via both the SH and AU tests. An ABBA BABA test found relatively few positions supporting ABBA or BABA splits (Table S1, Arias et al. 2022). The results of the introgression analysis varied depending on which species of *Barleria* was used for the test. Using *Barleria oenotheroides* or *B. albostellata* (as P1) led to a finding of significant introgression between *Crabbea* and *Lepidagathis*, but the test was not significant when *Barleria rotundifolia* was used.

SplitsTree networks from the 587 loci SCMD MSA (Fig. S4, Arias et al. 2022) and 6136 trees (Fig. S5, Arias et al. 2022) found long, thin branches leading to Acantheae and Barlerieae (arrows in Fig. S5), suggesting little character conflict. The networks found a "net" of character conflict among *Ruellia* + *Justicia* and 'AWN' (Figs. S4, S5). Independent taxon removal analyses of *Crabbea velutina* and *Avicennia marina* found no substantial reduction in character conflict among the remaining taxa for the species MSA (Figs. S6, S7, Arias et al. 2022) or the STAG trees (Figs. S8, S9, Arias et al. 2022).

A PCA based on treeVec distances created from 6,136 STAG trees reconciled with the species tree did not show distinct clusters of trees (Fig. 5). Instead, there is one large cluster of trees with little variation in principal components across the possible tree topologies. Although some tree topologies are distant from each other in the PCA, the pattern in the PCA suggests there are no distinct, competing evolutionary signals among the trees. There were 23 trees with the same topology as the 6136 loci STAG tree. There were no trees with the same topology as the McDade et al. (2008) tree. Median consensus trees were found for 20 different clusters in the PCA (Fig. 5). Trees that were most similar to the STAG tree and the McDade et al. (2008) tree were from clusters 10 and 16 (n = 499 and 797 trees, respectively; Fig. 5), and both of these consensus trees had a topology that matched the STAG topology, which places *Crabbea* sister to *Barleria* + *Lepidagathis*.

TABLE 3. BUSCO completeness scores. BUSCO (Simão et al. 2015) assesses the completeness of genome and transcriptome assemblies using hidden Markov model profiles from a curated database of universal and highly-conserved single-copy orthologs. Genome skimming samples are in gray rows. We assessed each assembly with the BUSCO v. 3 land plant (Embryophyta) database (*odb9*) of 1440 orthologs. High-identity protein matches between the assembly and the 1440 orthologs in the database are considered complete, partial matches are fragmented, and low-identity matches (or lack thereof) against database proteins are missing from the BUSCO v. 3 database.

Species	% Complete + fragmented	Complete (% complete)	# Fragmented	# Missing
Acanthus leucostachys	85.6	1159 (80.5%)	74	207
Andrographis paniculata	87.7	1197 (83.1%)	67	176
Avicennia marina	51.8	351 (24.4%)	395	694
Barleria oenotheroides	90.7	1230 (85.4%)	76	134
Barleria rotundifolia	49.7	388 (26.9%)	328	724
Barleria albostellata	90.1	1221 (84.8%)	77	142
Blepharis diversispina	51.7	462 (32.1%)	283	695
Blepharis spinifex	48.3	415 (28.8%)	281	744
Crabbea velutina	89.4	1176 (81.7%)	111	153
Justicia pacifica	89.9	1199 (83.3%)	95	146
Lankesteria elegans	93.6	1309 (90.9%)	39	92
Lepidagathis dulcis	77.3	878 (61.0%)	235	327
Lepidagathis sessilifolia	65.8	684 (47.5%)	264	492
Neuracanthus africanus	90.7	1239 (86.0%)	67	134
Ruellia simplex	67.5	785 (54.5%)	187	468
Whitfieldia elongata	89.9	1217 (84.5%)	78	145

DISCUSSION

We here present a well-supported phylogeny among major lineages of Acanthaceae subfamily Acanthoideae by combining predicted coding sequences derived from transcriptomes with sequences from shotgun genome skimming. We used de novo assembly and translation to compare amino acid sequences and find orthogroups. This study shows that samples from these two sequencing methods can be combined, increasing our ability to include taxa for which tissues are available only from herbarium specimens and subsampling the genome for conserved regions (Morais et al. 2019). Morais et al. (2019) used MarkerMiner (Chamala et al. 2015) on eight Acanthaceae transcriptomes to find 1619 low copy nuclear genes with an average of four species per gene. In comparison, we found 6136 gene trees and 587 single copy loci present for at least nine of 16 species.

The tree estimated from 587 loci (Fig. 3) largely corroborates the chloroplast + nrITS based phylogeny (McDade et al. 2008) with greatly increased branch support for the placement of Whitfieldieae and *Neuracanthus*. All of the trees estimated using a multiple sequence alignment approach, including the RAxML trees reconstructed using 587 loci (Fig. 3), corroborate McDade et al.'s (2008) cp + nr-ITS topology, except for the placement of *Crabbea* (discussed below); branching order within *Barleria* also differed in the SCOGs tree (Fig. 4). Recovery of the same topology using primarily chloroplast (McDade et al. 2008) and primarily nuclear data (this study), and different estimation methods suggests that there is strong phylogenetic signal for these relationships that is genome independent and neither random nor based on sampling error (Penny and Hendy 1986). Corroboration of the chloroplast + nr-ITS topology with

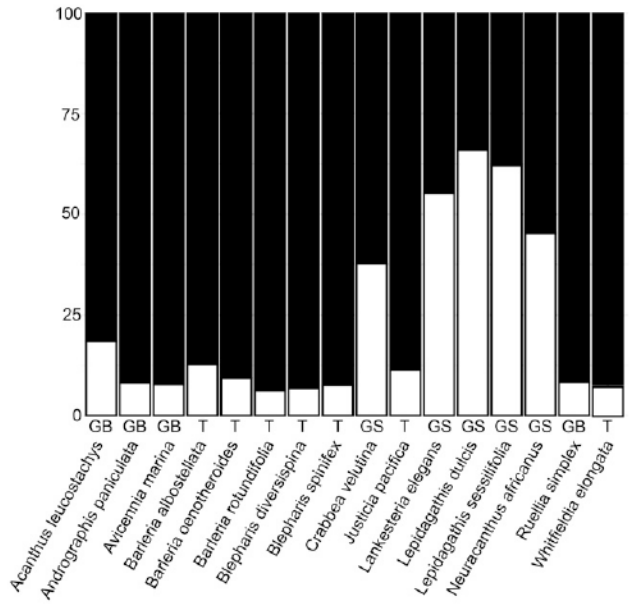


FIG. 2. Percentage of transcriptome and genome skim loci sorted into orthogroups by OrthoFinder2 for each species. Black bars = percent of loci sorted into orthogroups, white bars = percent unassigned loci. The letters below each bar indicate whether the species was a transcriptome sequenced for the present study (T), a genome skim sequenced for the present study (GS), or a publicly-available transcriptome from GenBank (GB). Samples with genome skim sequences had a higher percentage of unassigned loci (white) than transcriptomes, but still had 30–50% of loci sorted to orthogroups.

TABLE 4. Distribution of orthologous loci within orthogroups and loci not assigned to orthogroups for each species. Overall, genome-skimming (gray rows) resulted in more loci in orthogroups and more unassigned loci.

Species	Number of loci in orthogroups	Number of unassigned loci
Acanthus leucostachys	33,934	7733
Andrographis paniculata	24,176	2153
Avicennia marina	27,358	2301
Barleria albostellata	42,399	6229
Barleria oenetheroides	39,847	4076
Barleria rotundifolia	33,480	2224
Blepharis diversispina	30,697	2209
Blepharis spinifex	32,338	2631
Crabbea velutina	61,783	37,426
Justicia pacifica	27,052	3452
Lankesteria elegans	56,342	69,318
Lepidagathis dulcis	96,985	187,873
Lepidagathis sessilifolia	106,815	174,733
Neuracanthus africanus	45,702	37,777
Ruellia simplex	24,745	2234
Whitfieldia elongata	28,736	2390

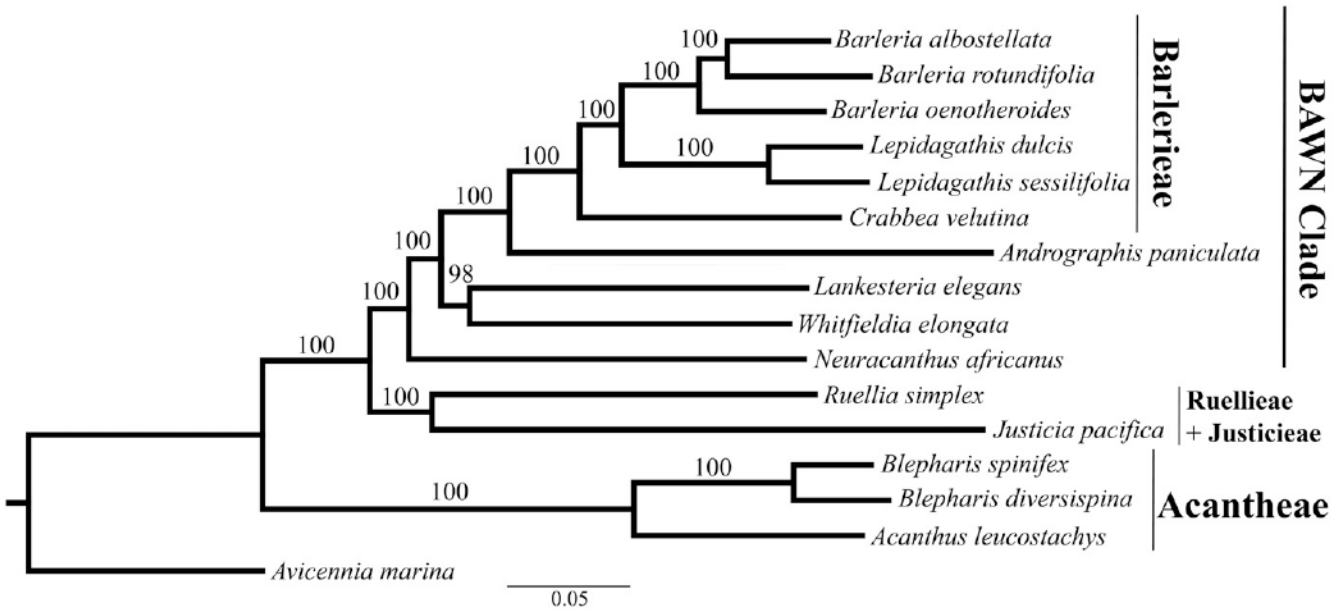


FIG. 3. Most likely tree for Acanthaceae derived from RAxML analysis and based on 587 nuclear loci for which sequences were present for at least nine species (model of evolution = JTT+I+G+F); support values are from 1000 ML bootstrap replicates. The same topology was recovered with FastTree (Fig. S1), 6136 gene trees (Fig. S2), and three alternative models of protein evolution (Fig. S3).

phylogenomic data suggests that low support values in the cp + nr-ITS tree (McDade et al. 2008) were the result of too few loci, although the sequences sampled were among the fastest evolving markers available at the time. The cp + nr-ITS five-loci dataset was 5518 nucleotides long, with 0.1–31.7% pairwise identity (McDade et al. 2008). In contrast, the 587 loci phylogenomic dataset used in our study was 30x longer (166,335 amino acids), with a mean of 39.5% pairwise identity. It is not clear whether low MP bootstrap support (< 50%) for a few branches in McDade et al. (2008) was due to low numbers of potentially informative characters or conflicting phylogenetic signal possibly caused by hybridization or incomplete lineage

TABLE 5. Top BLASTp hits for the 10 single copy orthogroups (SCOGs) that were present in all species as identified by OrthoFinder2. The *Ruellia* locus from each SCOG was searched against the nr database and the top three hits are shown. Note that "chloroplastic" and "mitochondrial" means nuclear encoded and expressed in the chloroplast or mitochondria. "PREDICTED" sequences contain an open reading frame that would produce a protein if they are translated, but they have not yet been experimentally verified to produce the described protein.

SCOG	Accession	Description	E-value
259	PIN18591.1	Hypothetical protein CDL12_08745 [*Handroanthus impetiginosus*]	3.00e-92
	XP_011092637.1	Uncharacterized protein LOC105172756 [*Sesamum indicum*]	4.00e-89
	XP_022870794.1	Uncharacterized protein LOC111390040 [*Olea europaea* var. *sylvestris*]	9.00e-06
267	XP_011101381.1	Protein MITOFERRINLIKE 1, chloroplastic [*Sesamum indicum*]	0
	PIN10553.1	Mitochondrial carrier protein PET8 [*Handroanthus impetiginosus*]	0
	XP_012829567.1	PREDICTED: protein MITOFERRINLIKE 1, chloroplastic [*Erythranthe guttata*]	0
323	XP_011098347.1	UDP-sulfoquinovose synthase, chloroplastic isoform X1 [*Sesamum indicum*]	0
	XP_011098411.1	UDP-sulfoquinovose synthase, chloroplastic isoform X2 [*Sesamum indicum*]	0
	PIN13226.1	UDP-glucose 4-epimerase/UDP-sulfoquinovose synthase [*Handroanthus impetiginosus*]	0
338	XP_022897574.1	Probable 3-hydroxyisobutyrate dehydrogenase-like 1, mitochondrial [*Olea europaea* var. *sylvestris*]	2.00e-73
	XP_011093637.1	Probable 3-hydroxyisobutyrate dehydrogenase-like 1, mitochondrial [*Sesamum indicum*]	3.00e-73
	PIN04844.1	Putative dehydrogenase [*Handroanthus impetiginosus*]	4.00e-71
369	KZV29525.1	Hypothetical protein F511_00803 [*Dorcoceras hygrometricum*]	7.00e-59
	XP_011097904.1	Uncharacterized protein LOC105176708 [*Sesamum indicum*]	9.00e-51
	XP_009781954.1	PREDICTED: uncharacterized protein LOC104230777 [*Nicotiana sylvestris*]	7.00e-50
408	XP_022871722.1	Serine hydroxymethyltransferase 4 [*Olea europaea* var. *sylvestris*]	0
	TEY91911.1	Glycine hydroxymethyltransferase [*Salvia splendens*]	0
	XP_011084218.1	Serine hydroxymethyltransferase 4 [*Sesamum indicum*]	0
410	PIN14942.1	Hypothetical protein CDL12_12417 [*Handroanthus impetiginosus*]	2.00e-175
	PIN09737.1	Hypothetical protein CDL12_17684 [*Handroanthus impetiginosus*]	2.00e-174
	XP_020552957.1	Probable transcription factor At3g04930 [*Sesamum indicum*]	3.00e-169
425	XP_011096504.1	Uncharacterized protein LOC105175673 [*Sesamum indicum*]	0
	PIN13816.1	Hypothetical protein CDL12_13544 [*Handroanthus impetiginosus*]	0
	XP_012849311.1	PREDICTED: probable methyltransferase PMT4 [*Erythranthe guttata*]	0
447	XP_011079822.1	Protein ROOT PRIMORDIUM DEFECTIVE 1 [*Sesamum indicum*]	5.00e-96
	XP_022890569.1	Protein WHAT'S THIS FACTOR 1 homolog [*Olea europaea* var. *sylvestris*]	9.00e-94
	KZV18513.1	Protein ROOT PRIMORDIUM DEFECTIVE 1 [*Dorcoceras hygrometricum*]	1.00e-92
483	PIN04627.1	Hypothetical protein CDL12_22836 [*Handroanthus impetiginosus*]	7.00e-156
	PIN20300.1	Hypothetical protein CDL12_07016 [*Handroanthus impetiginosus*]	9.00e-142
	EYU34960.1	Hypothetical protein MIMGU_mgv1a026890mg [*Erythranthe guttata*]	4.00e-135

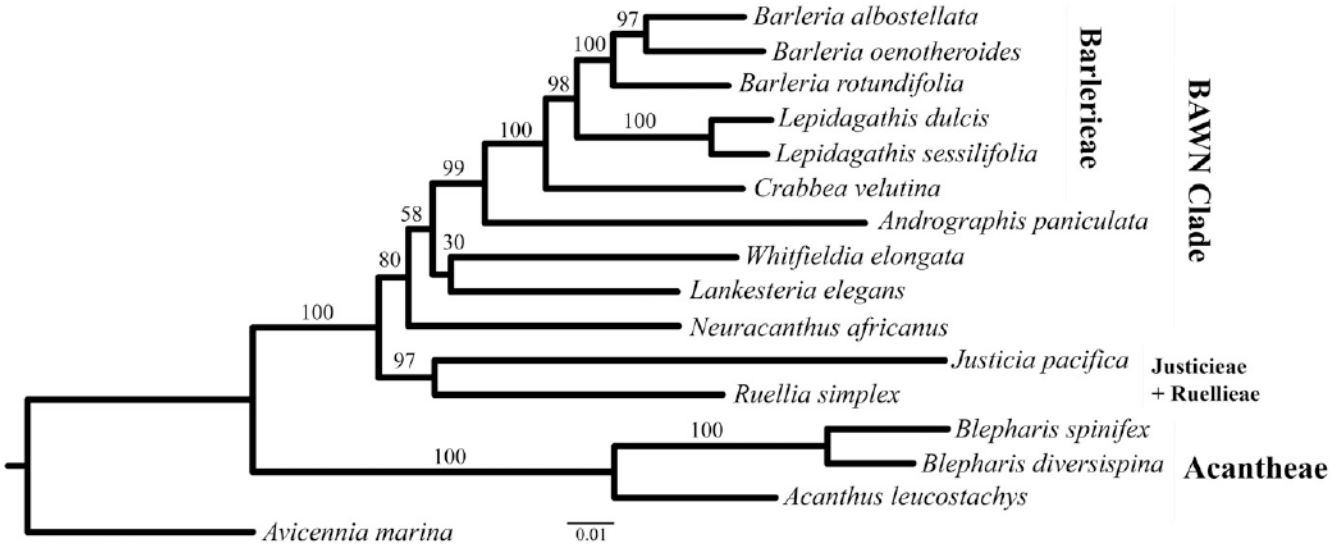


FIG. 4. RAxML phylogeny of Acanthaceae based on 10 single-copy orthologs (SCOGs) as identified by OrthoFinder2 and present for all species sampled. ML bootstrap support is above branches. The topology is similar to trees using more loci (Fig. 3; Fig. S2); however, the branching order within *Barleria* is different and the branches leading to *Neuracanthus* and Whitfieldieae have low bootstrap support.

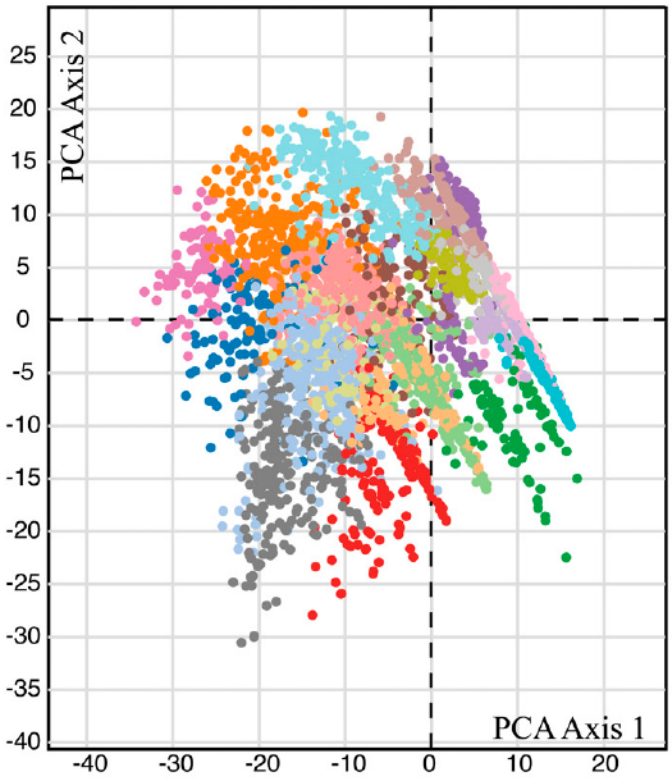
sorting (Page and Holmes 1998; Susko 2015), but in light of our phylogenomic data, the McDade et al. (2008) phylogenetic hypothesis is here confirmed as a good estimate of the species phylogeny. We did not find any support for *Neuracanthus* as sister to BAW + (Justicieae + Ruellieae) as in Tripp and McDade (2014), which was based on a BEAST analysis of Sanger sequences that did not sample as densely from BAWN as did McDade et al. (2008). It may be that taxon sampling differences or the additional parameters of the BEAST analysis led to the alternative placement of *Neuracanthus*. The tree presented here (Fig. 3) is a framework that can be used to estimate better-sampled phylogenies of the family.

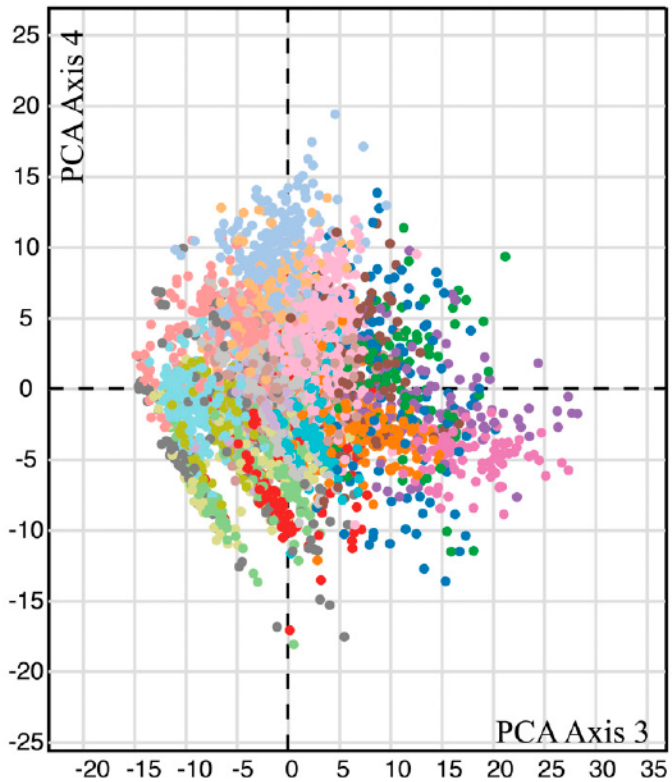
Combined transcriptome and GSS data increased support for some branches along the backbone of McDade et al.'s (2008) phylogeny that were weakly supported. Notably, the branches leading to the BAWN clade, the BAW clade and the Whitfieldeae are here more strongly supported. As in the cp + nr-ITS tree, the 10 single-copy loci RAxML tree returned moderate to no support for these branches, and the branching order of *Barleria* (Fig. 4) is inconsistent with topologies from the 587 loci data set (Fig. 3), as well as with other studies of *Barleria* (Darbyshire et al. 2019; Comito et al. 2022). Ten single-copy loci do not adequately resolve this part of the tree.

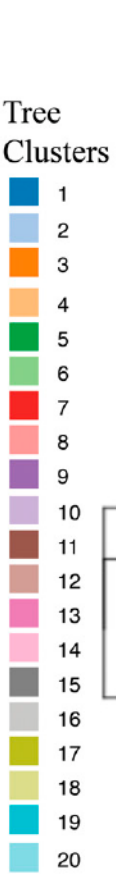
In phylogenetic studies using large numbers of loci, bootstrap values can become artificially inflated when data supporting alternative topologies are present at low levels (Seo 2008; Narechania et al. 2012). We tested for the possibility that our ML bootstrap values (Fig. 3) were inflated by changing the model of amino acid evolution used to estimate a tree in RAxML (Fig. S3) and by using networks and trees to explore conflict among branches (Figs. S4–S9). We also explored tree space with a PCA based on tree-distances (treeVec; Fig. 5). Changing the model of evolution had no effect on the topology and the PCA suggested that there were not distinct evolutionary signals in the data that would support different and conflicting tree topologies. If there were distinct, competing tree topologies, one should see more pockets of disparate trees ("tree islands") in the PCA; instead there is one large, homogenous cluster of trees, as would be expected from estimating the phylogeny based on different nuclear loci.

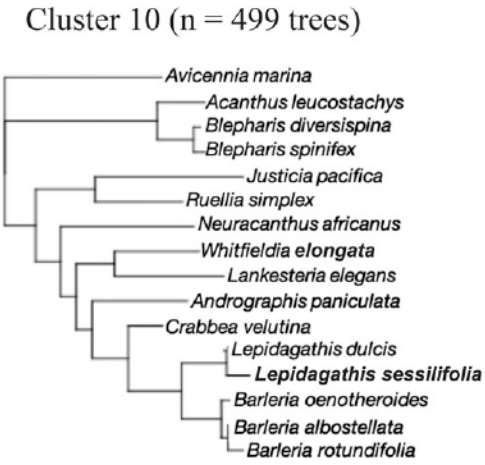
Networks of alignments (Figs. S4–S9) and the OrthoFinder2 trees with STAG support (Figs. S1, S2) suggest some degree of conflict in the branching order of *Andrographis*, *Neuracanthus*, and Whitfieldieae. Removing potentially problematic taxa in the networks, namely *Avicennia marina* (lowest-quality transcriptome used in this study) and *Crabbea velutina* (because it changes position in the chloroplast and nuclear topologies), did not visibly reduce conflict in the networks (Figs. S4–S9). The presence of conflict in these branches suggests that denser species sampling in Andrographideae, *Neuracanthus*, and Whitfieldieae may stabilize the topology. We also found that *Ruellia simplex* and *Justicia pacifica* did not have as many characters supporting their sister relationship as did the branches leading to Barlerieae and Acantheae (Figs. S4–S9). This may be because we sampled only a single species of each of these large lineages.

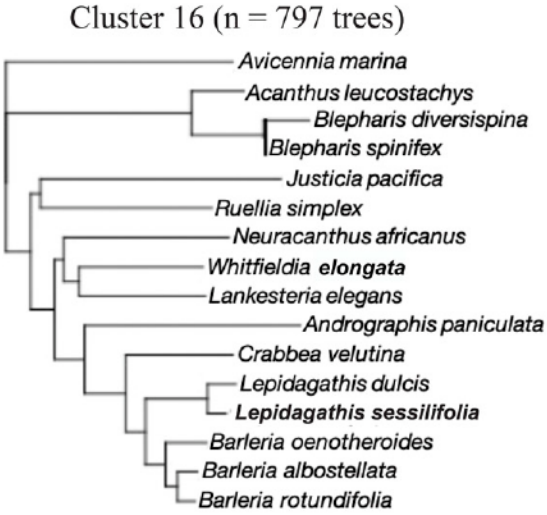
The Acanthaceae phylogenomic tree (Fig. 3) disagrees with McDade et al.'s (2008) chloroplast + nr-ITS topology in the placement of *Crabbea*. This genus of 16 species of perennial herbs and shrublets is distributed across Africa (Bidgood and Brummitt 1985; Thulin 2004; Darbyshire et al. 2010, 2015). De Gouveia et al. (2017) sampled seven *Crabbea* species from South Africa and assembled data from two chloroplast loci and morphology. Using *Barleria repens*, *Andrographis paniculata*, and *Thunbergia erecta* as outgroups and rooting the tree on the branch leading to the outgroups, these authors recovered a monophyletic *Crabbea*. *Crabbea velutina* was sister to the rest (De Gouveia et al. 2017). In McDade et al.'s (2008) cp + nr-ITS topology, *C. acaulis* is strongly supported as sister to *Barleria*. We sampled *C. velutina* and found it to be sister to the rest of Barlerieae (*Barleria* + *Lepidagathis*) when we used an alignment-based approach (MAFFT, Fig. S1), but not when we used a distance-matrix approach (DendroBLAST, Fig. S2). Neither the DendroBLAST nor MAFFT trees resolve *C. velutina* sister to *Barleria*, as was *C. acaulis* in McDade et al. (2008). Additionally, a RADseq study of *Barleria* (Comito et al. 2022) sampled *C. hirsuta* and a second, unidentified *Crabbea* species. These were together sister to the rest of sampled Barlerieae (*Barleria* + *Lepidagathis*), as in our trees. It is possible that nuclear and chloroplast data disagree about the











	STAG species tree	McDade cp + nrITS tree
Best tree distance	0 (23 trees)	6
Median tree distance	17	16
Farthest tree distance	57	53

FIG. 5. Principal component analysis of 6136 gene trees in treespace based on treeVec-distances. Top left: PC1 vs. PC2; bottom left: PC3 vs. PC4. Colors correspond to 20 tree clusters. There are no groups of tree topologies that are distinct from the main group, suggesting subtle differences in the evolutionary histories of these loci, but no strong alternative to the majority signal. The median trees for clusters 10 and 16, right, matched the nuclear topology (Fig. 3). Best, Median, and Farthest Tree Distances were calculated for all 6136 trees vs. the STAG species tree and the McDade et al. (2008) cp + nr-ITS tree. Twenty-three of the 6136 gene trees had the same topology as the STAG species tree.

placement of *Crabbea* owing to introgression between the ancestors of this genus and other Barlerieae lineages (represented here by *Barleria*, *Lepidagathis*) and our ABBA-BABA test found few alignment positions supporting *C. velutina* as sister to *Barleria* (Table S1). It is also possible that *Crabbea* is not monophyletic: one or more *Crabbea* species may be sister to *Barleria*, whereas others are sister to all Barlerieae. The results of De Gouveia et al. (2017) suggest otherwise, but their sampling outside of *Crabbea* was limited and monophyly of the genus was not tested. Phylogenetic resolution of *Crabbea* will require increased sampling of *Crabbea* and *Lepidagathis* s.l., and of Malagasy genera of Barlerieae (*Boutonia*, *Lasiocladus*, *Pericalypta*, *Podorungia*, *Pseudodicplitera*) to determine the limits of *Crabbea* and consider its relationships to other Barlerieae.

In our study, low taxon sampling may have affected resultant topologies. Increased taxon sampling has not necessarily provided high bootstrap support in previous studies of this group using chloroplast + nr-ITS Sanger sequences, but increased sampling may provide a stable phylogeny with larger nuclear datasets. We limited taxon sampling in this study to focus on generating more loci and to trial the approach of combining transcriptome and genome skim data. Now that these genomic sequences are available, they can be used for future studies involving more species in each of these lineages to further refine the phylogeny of Acanthaceae.

A highly resolved and strongly supported phylogeny of Acanthaceae will help guide reclassification of the family (Tripp et al. 2022), identify potential synapomorphies for clades, and better understand morphological evolution (e.g. explosive seed capsules, hygroscopic trichomes on seeds, corolla morphology). Our results suggest that the placement of *Crabbea* warrants further consideration and future studies should increase taxon sampling across the BAWN lineages of Barlerieae, Andrographideae, Whitfieldeae, and *Neuracanthus*, as these groups have not yet been well sampled in any molecular phylogenetic study.

Acknowledgments

This work was done in partial fulfillment of Joshua Arias' M.S. degree at California State University, Long Beach (CSULB); his committee members, Judy Brusslan and Renaud Berlemont, gave insightful feedback on his thesis. Funding for this work came from the CSULB RISE program (NIH-RISE Award 2R25GM071638), the State of California, and A. Fisher's NSF Award #1754792. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Science Foundation nor the National Institutes of Health. EMT acknowledges support from NSF (Award #s 1354963 & 1754493). This research was not preregistered in any institutional registry. Thank you to CSULB Botany technician Brian Thorson and CSULB students Nneka Arouma, Laymon Ball, Robert Comito, Emily Do, Key Keymanesh, Amy Nguyen, and Jean-Paul Urenda for assistance in the lab and greenhouse. Judy Brusslan suggested the BRaD-seq method of RNA extraction. Mark Porter of California Botanic Garden suggested the ABBA-BABA analysis. Special thanks go to Carrie Ganote of the Indiana University Galaxy Trinity instance, Roger Barthelson and Upendra Devissety of CyVerse, and Charles Forsyth of the UC Riverside HPCC for their guidance with analyses.

Author Contributions

JA did the molecular lab work and analyses, wrote the initial draft of the manuscript as a M.S. thesis, and designed the figures. EMT provided the *Ruellia* transcriptome before public release and revised the manuscript. LM contributed preserved specimens. EMT, CK, and LM contributed to writing and editing the manuscript. AF designed the project, supervised molecular lab work and analyses, provided funding, and revised and edited the manuscript.

All new specimens gathered for this study were collected legally and with permission.

Literature Cited

Abascal, F., R. Zardoya, and D. Posada. 2005. ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* 21: 2104–2105.

Afgan, E., D. Baker, B. Batut, M. Van Den Beek, D. Bouvier, M. Ech, J. Chilton, D. Clements, N. Coraor, B. A. Grüning, A. Guerler, J. Hillman-Jackson, S. Hiltemann, V. Jalili, H. Rasche, N. Soranzo, J. Goecks, J. Taylor, A. Nekrutenko, and D. Blankenberg. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research* 46: W537–W544.

An, H., X. Qi, M. L. Gaynor, Y. Hao, S. C. Gebken, M. E. Mabry, A. C. McAlvay, G. R. Teakle, G. C. Conant, M. S. Barker, T. Fu, B. Yi, and J. C. Pires. 2019. Transcriptome and organellar sequencing highlights the complex origin and diversification of allotetraploid *Brassica napus*. *Nature Communications* 10: 1–12.

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. (0.11.5). Babraham Institute. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Arias, J. D., E. Manzitto-Tripp, C. A. Kiel, L. A. McDade, and A. E. Fisher. 2022. Data from: An evolutionary framework of Acanthaceae based on transcriptomes and genome skims. Dryad Digital Repository. doi: 10.5061/dryad.59zw3r27x.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, and P. A. Pevzner. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455–477.

Bidgood, S. and R. K. Brummitt. 1985. *Acanthostelma*: A new genus of Acanthaceae from Somalia. *Kew Bulletin* 40: 855–858.

Blackmon, H. and R. A. Adams. 2015. EvobiR: Tools for comparative analyses and teaching evolutionary biology. https://github.com/coleoguy/evobir.

Buchfink, B., C. Xie, and D. H. Huson. 2014. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12: 59–60.

Buffalo, V. 2011. Scythe: A Bayesian adapter trimmer (0.994). https://github.com/vsbuffalo/scythe.

Chamala, S., N. García, G. T. Godden, V. Krishnakumar, I. E. Jordon-Thaden, R. De Smet, W. B. Barbazuk, D. E. Soltis, and P. S. Soltis. 2015. MarkerMiner 1.0: A new application for phylogenetic marker development using Angiosperm transcriptomes. *Applications in Plant Sciences* 3: 1400115.

Cherukupalli, N., M. Divate, S. R. Mittapelli, V. R. Khareedu, and D. R. Vudem. 2016. De novo assembly of leaf transcriptome in the medicinal plant *Andrographis paniculata*. *Frontiers in Plant Science* 7: 1203.

Comito, R., I. Darbyshire, C. A. Kiel, L. A. McDade, and A. E. Fisher. 2022. A RADseq Phylogeny of *Barleria* (Acanthaceae) resolves fine-scale relationships. *Molecular Phylogenetics and Evolution* 169: 107428.

Cooper, E. S., M. A. Mosher, C. M. Cross, and D. L. Whitaker. 2018. Gyroscopic stabilization minimizes drag on *Ruellia ciliatiflora* seeds. *Journal of the Royal Society, Interface* 15: 20170901.

Daniel, T. F. and E. A. Tripp. 2018. *Louteridium* (Acanthaceae: Acanthoideae: Ruellieae: Trichantherinae): Taxonomy, phylogeny, reproductive biology, and conservation. *Proceedings of the California Academy of Sciences* 65: 41–106.

Daniel, T. F., L. A. McDade, M. Manktelow, and C. A. Kiel. 2008. The "Tetramerium Lineage" (Acanthaceae: Acanthoideae: Justicieae): Delimitation and Intra-lineage Relationships Based on cp and nrITS Sequence Data. *Systematic Botany* 33: 416–436.

Darbyshire, I., K. Vollesen, and E. Kelbessa. 2010. Acanthaceae (Part 2). Pp. 297–300 in *Flora of Tropical East Africa*, ed. H. J. Beentje. Kew: Royal Botanic Gardens.

Darbyshire, I., K. Vollensen, and E. Kelbessa. 2015. Acanthaceae. Pp. 12–19 in *Flora Zambesiaca* vol. 8 (6), eds. J. R. Timberlake and E. S. Martins. Kew: Royal Botanic Gardens.

Darbyshire, I., A. E. Fisher, C. A. Kiel, and L. A. McDade. 2019. Phylogenetic relationships among species of *Barleria* (Acanthaceae, Lamiales): Molecular data reveal complex patterns of morphological evolution and support a revised classification. *Taxon* 68: 92–111.

Darbyshire, I., C. A. Kiel, C. M. Astroth, K. G. Dexter, F. M. Chase, and E. A. Tripp. 2020. Phylogenomic study of *Monechma* reveals two

divergent plant lineages of ecological importance in the African savanna and succulent biomes. *Diversity (Basel)* 12: 237–262.

Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2011. ProtTest-HPC: Fast selection of best-fit models of protein evolution. *Bioinformatics* 27: 1164–1165.

De Gouveia, A., M. Jackson, and L. Joubert. 2017. Systematics of *Crabbea* (Acanthaceae) in southern Africa. *South African Journal of Botany* 109: 331–356.

Dodsworth, S., A. R. Leitch, and I. J. Leitch. 2015. Genome size diversity in angiosperms and its influence on gene space. *Current Opinion in Genetics & Development* 35: 73–78.

Dodsworth, S., L. Pokorny, M. G. Johnson, J. T. Kim, O. Maurin, N. J. Wickett, F. Forest, et al. 2019. Hyb-Seq for flowering plant systematics. *Trends in Plant Science* 24: 887–891.

Doyle, J. J. and J. L. Doyle. 1987. Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochemical Bulletin* 19: 11–15.

Eaton, D. A. and R. H. Ree. 2013. Inferring phylogeny and introgression using RADseq data: An example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology* 62: 689–706.

Eldem, V., G. Zararsiz, T. Taşçi, I. P. Duru, Y. Bakir, and M. Erkan. 2017. Transcriptome analysis for non-model organisms: Current status and best-practices. Pp. 55–78 in *Applications of RNA-Seq and Omics Strategies - From Microorganisms to Human Health*, eds. F. A. Marchi, P. D. R. Cirillo, and E. C. Mateo. Intech Open. DOI:10.5772/intechopen.68983.

Emms, D. M. and S. Kelly. 2015. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* 16: 157.

Emms, D. M. and S. Kelly. 2018. STAG: Species tree inference from all genes. *BioRxiv* 267914. https://doi.org/10.1101/267914.

Emms, D. M. and S. Kelly. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology* 20: 238.

Fisher, A. E., L. A. McDade, C. A. Kiel, R. Khoshravesh, M. A. Johnson, M. Stata, T. L. Sage, and R. F. Sage. 2015. Evolutionary history of *Blepharis* (Acanthaceae) and the origin of C_4 photosynthesis in section *Acanthodium*. *International Journal of Plant Sciences* 176: 770–790.

Fisher, A. E., K. M. Hasenstab, H. L. Bell, E. Blaine, A. L. Ingram, and J. T. Columbus. 2016. Evolutionary history of chloridoid grasses estimated from 122 nuclear loci. *Molecular Phylogenetics and Evolution* 105: 1–14.

Frith, M. C., M. Pheasant, and J. S. Mattick. 2005. Genomics: The amazing complexity of the human transcriptome. *European Journal of Human Genetics* 13: 894–897.

Fu, L., B. Niu, Z. Zhu, S. Wu, and W. Li. 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28: 3150–3152.

Funk, V. A. 2018. Collections-based science in the 21st century. *Journal of Systematics and Evolution* 56: 175–193.

Garg, A., L. Agrawal, R. C. Misra, S. Sharma, and S. Ghosh. 2015. *Andrographis paniculata* transcriptome provides molecular insights into tissue-specific accumulation of medicinal diterpenes. *BMC Genomics* 16: 659.

Gnirke, A., A. Melnikov, J. Maguire, P. Rogov, E. M. LeProust, W. Brockman, T. Fennell, G. Giannoukos, S. Fisher, C. Russ, S. Gabriel, D. B. Jaffe, E. S. Lander, and C. Nusbaum. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* 27: 182–189.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.

Grall, A. and I. Darbyshire. 2021. A synopsis of the African genus *Whitfieldia* (Acanthaceae: Whitfieldieae) and a key to the species. *Kew Bulletin* 76: 191–221.

Haas, B. J. and A. Papanicolaou. 2012. TransDecoder (find coding regions within transcripts). https://github.com/TransDecoder/TransDecoder/wiki.

Hale, H., E. M. Gardner, J. Viruel, L. Pokorny, and M. G. Johnson. 2020. Strategies for reducing per-sample costs in target capture sequencing for phylogenomics and population genomics in plants. *Applications in Plant Sciences* 8: e11337.

Hart, M. L., L. L. Forrest, J. A. Nicholls, and C. A. Kidner. 2016. Retrieval of hundreds of nuclear loci from herbarium specimens. *Taxon* 65: 1081–1092.

Hillis, D. M. and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* 42: 182–192.

Hodel, R. G. J., M. A. Gitzendanner, C. C. Germain-Aubrey, X. Liu, A. A. Crowl, M. Sun, J. B. Landis, M. C. Segovia-Salcedo, N. A. Douglas, S. Chen, D. E. Soltis, and P. S. Soltis. 2016. A new resource for the development of SSR markers: Millions of loci from a thousand plant transcriptomes. *Applications in Plant Sciences* 4: 1600024.

Huang, J., X. Lu, W. Zhang, R. Huang, S. Chen, and Y. Zheng. 2014. Transcriptome sequencing and analysis of leaf tissue of *Avicennia marina* using the Illumina platform. *PLoS One* 9: e108785.

Huson, D. H. and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies, molecular biology and evolution. *Molecular Biology and Evolution* 23: 254–267.

Jombart, T., M. Kendall, J. Almagro-Garcia, and C. Colijn. 2017. treespace: Statistical exploration of landscapes of phylogenetic trees. *Molecular Ecology Resources* 17: 1385–1392.

Jordon-Thaden, I. E., J. B. Beck, C. A. Rushworth, M. D. Windham, N. Diaz, J. T. Cantley, C. T. Martine, and C. J. Rothfels. 2020. A basic ddRADseq two-enzyme protocol performs well with herbarium and silica-dried tissues across four genera. *Applications in Plant Sciences* 8: e11344.

Katoh, K. and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.

Keller, O., M. Kollmar, M. Stanke, and S. Waack. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* 27: 757–763.

Kelly, S. and P. K. Maini. 2013. DendroBLAST: Approximate phylogenetic trees in the absence of multiple sequence alignments. *PLoS One* 8: e58537.

Kendall, M. and C. Colijn. 2015. A tree metric using structure and length to capture distinct phylogenetic signals. arXiv:1507.05211.

Kiel, C. A. and L. A. McDade. 2014. The *Mirandea* Clade (Acanthaceae, Justicieae, Tetramerium Lineage): Phylogenetic signal from molecular data and micromorphology makes sense of taxonomic confusion caused by remarkable diversity of floral form. *Systematic Botany* 39: 950–964.

Kiel, C. A., L. A. McDade, T. F. Daniel, and D. Champluvier. 2006. Phylogenetic delimitation of Isoglossinae (Acanthaceae: Justicieae) and relationships among constituent genera. *Taxon* 55: 683–694.

Kiel, C. A., T. F. Daniel, I. Darbyshire, and L. A. McDade. 2017. Unraveling relationships in the morphologically diverse and taxonomically challenging "justicioid" lineage (Acanthaceae: Justicieae). *Taxon* 66: 645–674.

Kiel, C. A., T. F. Daniel, and L. A. McDade. 2018. Phylogenetics of New World 'Justicioids' (Justicieae: Acanthaceae): Major lineages, morphological patterns, and widespread incongruence with classification. *Systematic Botany* 43: 459–484.

Lefort, V., R. Desper, and O. Gascuel. 2015. FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Molecular Biology and Evolution* 32: 2798–2800.

Lemmon, A. R., S. A. Emme, and E. M. Lemmon. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology* 61: 727–744.

Loman, N. J., C. Constantinidou, J. Z. M. Chan, M. Halachev, M. Sergeant, C. W. Penn, E. R. Robinson, and M. J. Pallen. 2012. High-throughput bacterial genome sequencing: An embarrassment of choice, a world of opportunity. *Nature Reviews. Microbiology* 10: 599–606.

Lyu, H., X. Li, Z. Guo, Z. He, and S. Shi. 2018. De novo assembly and annotation of the *Avicennia officinalis* L. transcriptome. *Marine Genomics* 39: 3–6.

McDade, L. A. and M. L. Moody. 1999. Phylogenetic relationships among Acanthaceae: Evidence from noncoding trnL-trnF chloroplast DNA sequences. *American Journal of Botany* 86: 70–80.

McDade, L. A. and J. A. Weeks. 2004. Nectar in hummingbird-pollinated neotropical plants II: Interactions with flower visitors. *Biotropica* 36: 216.

McDade, L. A., T. F. Daniel, C. A. Kiel, and K. Vollesen. 2005. Phylogenetic Relationships among Acantheae (Acanthaceae): Major lineages present contrasting patterns of molecular evolution and morphological differentiation. *Systematic Botany* 30: 834–862.

McDade, L. A., T. F. Daniel, and C. A. Kiel. 2008. Toward a comprehensive understanding of phylogenetic relationships among lineages of Acanthaceae s.l. (Lamiales). *American Journal of Botany* 95: 1136–1152.

McDade, L. A., T. F. Daniel, C. A. Kiel, and A. J. Borg. 2012. Phylogenetic placement, delimitation, and relationships among genera of the enigmatic Nelsonioideae (Lamiales: Acanthaceae). *Taxon* 61: 637–651.

McDade, L. A., T. F. Daniel, and C. A. Kiel. 2018. The Tetramerium lineage (Acanthaceae, Justicieae) revisited: Phylogenetic relationships

reveal polyphyly of many New World genera accompanied by rampant evolution of floral morphology. *Systematic Botany* 43: 97–116.
McKain, M. R., M. G. Johnson, S. Uribe-Convers, D. Eaton, and Y. Yang. 2018. Practical considerations for plant phylogenomics. *Applications in Plant Sciences* 6: e1038.
Merchant, N., E. Lyons, S. Goff, M. Vaughn, D. Ware, D. Micklos, and P. Antin. 2016. The iPlant Collaborative: Cyberinfrastructure for enabling data to discovery for the life sciences. *PLoS Biology* 14: e1002342.
Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Pp. 1–8 in *Proceedings of the Gateway Computing Environments Workshop (GCE)*. New Orleans: Gateway Computing.
Morais, E. B., J. Schönenberger, E. Conti, A. Antonelli, and P. Szövényi. 2019. Orthologous nuclear markers and new transcriptomes that broadly cover the phylogenetic diversity of Acanthaceae. *Applications in Plant Sciences* 7: e11290.
Muchhala, N., A. Caiza, J. C. Vizuete, and J. D. Thomson. 2009. A generalized pollination system in the tropics: Bats, birds and *Aphelandra acanthus*. *Annals of Botany* 103: 1481–1487.
Narechania, A., R. H. Baker, R. Sit, S. O. Kolokotronis, R. DeSalle, and P. J. Planet. 2012. Random addition concatenation analysis: A novel approach to the exploration of phylogenomic signal reveals strong agreement between core and shell genomic partitions in the cyanobacteria. *Genome Biology and Evolution* 4: 30–43.
Ortegón-Campos, I., V. Parra-Tabla, L. Abdala-Roberts, and C. M. Herrera. 2009. Local adaptation of *Ruellia nudiflora* (Acanthaceae) to biotic counterparts: Complex scenarios revealed when two herbivore guilds are considered. *Journal of Evolutionary Biology* 22: 2288–2297.
Page, R. D. M. and E. C. Holmes. 1998. *Molecular Evolution: A Phylogenetic Approach*. Oxford: Blackwell Publishing, Ltd.
Pavey, S. A., J. Gaudin, E. Normandeau, M. Dionne, M. Castonguay, C. Audet, and L. Bernatchez. 2015. RAD sequencing highlights polygenic discrimination of habitat ecotypes in the panmictic American Eel. *Current Biology* 25: 1666–1671.
Penny, D. and M. D. Hendy. 1986. Estimating the reliability of evolutionary trees. *Molecular Biology and Evolution* 3: 403–417.
Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. FastTree 2: Approximately maximum-likelihood trees for large alignments. *PLoS One* 5: e9490.
Richter, S., F. Schwarz, L. Hering, M. Böggemann, and C. Bleidorn. 2015. The utility of genome skimming for phylogenomic analyses as demonstrated for glycerid relationships (Annelida, Glyceridae). *Genome Biology and Evolution* 7: 3443–3462.
Ripma, L. A., M. G. Simpson, and K. Hasenstab-Lehman. 2014. Geneious! Simplified genome skimming methods for phylogenetic systematic studies: A case study in *Oreocarya* (Boraginaceae). *Applications in Plant Sciences* 2: 1400062.
Robinson, D. F. and L. R. Foulds. 1981. Comparison of phylogenetic trees. *Mathematical Biosciences* 53: 131–147.
Rothfels, C. J., A. Larsson, F. W. Li, E. M. Sigel, L. Huiet, D. O. Burge, M. Ruhsam, S. W. Graham, D. W. Stevenson, G. K. S. Wong, P. Korall, and K. M. Pryer. 2013. Transcriptome-mining for single-copy nuclear markers in ferns. *PLoS One* 8: e76957.
Salehi, Z. and M. Najafi. 2014. RNA preservation and stabilization. *Biochemistry & Physiology* 3: 126.
Schmickl, R., A. Liston, V. Zeisek, K. Oberlander, K. Weitemier, S. C. K. Straub, R. C. Cronn, L. L. Dreyer, and J. Suda. 2016. Phylogenetic marker development for target enrichment from transcriptome and genome skim data: The pipeline and its application in southern African *Oxalis* (Oxalidaceae). *Molecular Ecology Resources* 16: 1124–1135.
Schmieder, R. and R. Edwards. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27: 863–864.
Seo, T. K. 2008. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. *Molecular Biology and Evolution* 25: 960–971.
Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* 51: 492–508.
Shimodaira, H. and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16: 1114–1116.
Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–3212.
Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.
Stanke, M. 2007. getAnnoFasta.pl: Creates FASTA sequence files from AUGUSTUS output (3.2.1). https://github.com/nextgenusfs/augustus/blob/master/scripts/getAnnoFasta.pl.
Stanke, M. and B. Morgenstern. 2005. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research* 33: W465–W467.
Susko, E. 2015. Bayesian long branch attraction bias and corrections. *Systematic Biology* 64: 243–255.
Swofford, D. L. 1998. PAUP* Phylogenetic analysis using parsimony (* and other methods), v. 4. Sunderland, Massachusetts: Sinauer Associates.
Thulin, M. 2004. Expansion of *Crabbea* (Acanthaceae) and the description of two new species from Somalia. *Nordic Journal of Botany* 24: 501–506.
Townsley, B. T., M. F. Covington, Y. Ichihashi, K. Zumstein, and N. R. Sinha. 2015. Breath Adapter Directional sequencing (BRaD-seq): A streamlined, ultra-simple and fast library preparation protocol for strand specific mRNA library construction. *Frontiers in Plant Science* 6: 366.
Tripp, E. A. 2007. Evolutionary relationships within the species-rich genus *Ruellia* (Acanthaceae). *Systematic Botany* 32: 628–649.
Tripp, E. A. 2008. *Systematics and Pollination System Evolution in* Ruellia *(Acanthaceae)*. Ph.D. dissertation. Durham, North Carolina: Duke University.
Tripp, E. A. and I. Darbyshire. 2017. Phylogenetic relationships among Old World *Ruellia* L.: A new classification and reinstatement of the genus *Dinteracanthus* Schinz. *Systematic Botany* 42: 470–483.
Tripp, E. A. and I. Darbyshire. 2020. Mcdadea: A new genus of Acanthaceae endemic to the Namib Desert of southwestern Angola. *Systematic Botany* 45: 200–211.
Tripp, E. A. and P. S. Manos. 2008. Is floral specialization an evolutionary dead-end? Pollination system transitions in *Ruellia* (Acanthaceae). *Evolution* 62: 1712–1737.
Tripp, E. A. and L. A. McDade. 2014. A rich fossil record yields calibrated phylogeny for Acanthaceae (Lamiales) and evidence for marked biases in timing and directionality of intercontinental disjunctions. *Systematic Biology* 63: 660–684.
Tripp, E. A. and Y. H. E. Tsai. 2017. Disentangling geographical, biotic, and abiotic drivers of plant diversity in neotropical *Ruellia* (Acanthaceae). *PLoS One* 12: e0176021.
Tripp, E. A., T. F. Daniel, S. Fatimah, and L. A. McDade. 2013. Phylogenetic relationships within Ruellieae (Acanthaceae) and a revised classification. *International Journal of Plant Sciences* 174: 97–137.
Tripp, E. A., I. Darbyshire, T. F. Daniel, C. A. Kiel, and L. A. McDade. 2022. Revised Classification of Acanthaceae and worldwide dichotomous keys. *Taxon* 71: 103–153.
Tripp, E. A., Y. H. E. Tsai, Y. Zhuang, and K. G. Dexter. 2017. RADseq dataset with 90% missing data fully resolves recent radiation of *Petalidium* (Acanthaceae) in the ultra-arid deserts of Namibia. *Ecology and Evolution* 7: 7920–7936.
Ungaro, A., N. Pech, J. F. Martin, R. J. S. McCairns, J. P. Mévy, R. Chappaz, and A. Gilles. 2017. Challenges and advances for transcriptome assembly in non-model species. *PLoS One* 12: e0185020.
Washburn, J. D., J. C. Schnable, G. C. Conant, T. P. Brutnell, Y. Shao, Y. Zhang, M. Ludwig, G. Davidse, and J. C. Pires. 2017. Genome-guided phylo-transcriptomic methods and the nuclear phylogentic tree of the Paniceae grasses. *Scientific Reports* 7: 13528.
Washburn, J. D., J. C. Schnable, G. Davidse, and J. C. Pires. 2015. Phylogeny and photosynthesis of the grass tribe Paniceae. *American Journal of Botany* 102: 1493–1505.
Waterhouse, R. M., M. Seppey, F. A. Simão, M. Manni, P. Ioannidis, G. Klioutchnikov, E. V. Kriventseva, and E. M. Zdobnov. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution* 35: 543–548.
Weiß, C. L., V. J. Schuenemann, J. Devos, G. Shirsekar, E. Reiter, B. A. Gould, J. R. Stinchcombe, J. Krause, and H. A. Burbano. 2016. Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens. *Royal Society Open Science* 3: 160239.
Weitemier, K., S. C. Straub, R. C. Cronn, M. Fishbein, R. Schmickl, A. McDonnell, and A. Liston. 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 2: 1400042.
Wen, J., A. N. Egan, R. B. Dikow, and E. A. Zimmer. 2015. Utility of transcriptome sequencing for phylogenetic inference and character evolution. Pp. 1–42 in *Next-Generation Sequencing in Plant Systematics, Regnum Vegetabile* vol. 158, eds. E. Hörandl and M. S. Appelhans. Königstein, Germany: Koeltz Botanical Books.

White, O. W., B. Doo, M. A. Carine, and M. A. Chapman. 2016. Transcriptome sequencing and simple sequence repeat marker development for three Macaronesian endemic plant species. *Applications in Plant Sciences* 4: 1600050.

Wickett, N. J., S. Mirarab, N. Nguyen, T. Warnow, E. J. Carpenter, N. Matasci, S. Ayyampalayam, M. S. Barker, J. G. Burleigh, M. A. Gitzendanner, B. R. Ruhfel, E. Wafula, J. P. Der, S. W. Graham, S. Mathews, M. Melkonian, D. E. Soltis, P. S. Soltis, N. W. Miles, and J. H. Leebens-Mack. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences USA* 111: E4859–E4868.

Wu, C., R. N. Crowhurst, A. B. Dennis, V. G. Twort, S. Liu, R. D. Newcomb, H. A. Ross, and T. R. Buckley. 2016. De novo transcriptome analysis of the common New Zealand stick insect *Clitarchus hookeri* (Phasmatodea) reveals genes involved in olfaction, digestion and sexual reproduction. *PLoS One* 11: e0157783.

Yang, Y., S. Yang, J. Li, Y. Deng, Z. Zhang, S. Xu, W. Guo, C. Zhong, R. Zhou, and S. Shi. 2015. Transcriptome analysis of the holly mangrove, *Acanthus ilicifolius*, and its terrestrial relative, *Acanthus leucostachyus*, provides insights into adaptation to intertidal zones. *BMC Genomics* 16: 1–12.

Yeates, D. K., A. Zwick, and A. S. Mikheyev. 2016. Museums are BioBanks: Unlocking the genetic potential of the three billion specimens in the world's biological collections. *Current Opinion in Insect Science* 18: 83–88.

Yu, X., D. Yang, C. Guo, and L. Gao. 2018. Plant phylogenomics based on genome-partitioning strategies: Progress and prospects. *Plant Diversity* 40: 158–164.

Zhuang, Y. and E. A. Tripp. 2017a. The draft genome of *Ruellia speciosa* (Beautiful Wild Petunia: Acanthaceae). *DNA Research* 24: 179–192.

Zhuang, Y. and E. A. Tripp. 2017b. Genome-scale transcriptional study of hybrid effects and regulatory divergence in an F1 hybrid *Ruellia* (Wild Petunias: Acanthaceae) and its parents. *BMC Plant Biology* 17: 1–13.

Zhuang, Y. and E. A. Tripp. 2022. Co-expression network analyses of anthocyanin biosynthesis genes in *Ruellia* (Wild Petunias; Acanthaceae). *BMC Ecology and Evolution* 22: 1–17.

Zimmer, E. A. and J. Wen. 2015. Using nuclear gene data for plant phylogenetics: Progress and prospects II. Next-gen approaches. *Journal of Systematics and Evolution* 53: 371–379.

Appendix 1. Taxon information. Voucher information for species sampled in this study. ***Taxon*** Authority; GenBank SRA Accession; source, specimen voucher collector and collection number (herbarium); Acanthaceae lineage.

Newly sequenced genome skims:

Crabbea velutina S. Moore; SRR7806557; Tanzania, *Manketelow 670* (UPS); Barlerieae. ***Lankesteria elegans*** (P. Beauv.) T. Anderson; SRR7806555; Cameroon, *Etuge & Thomas 466* (K); Whitfieldieae. ***Lepidagathis dulcis*** Nees; SRR7798775; Kew Botanic Garden, *Suddee et. al. 999* (K); Barlerieae. ***Lepidagathis sessilifolia*** (Pohl) Kameyama ex Wassh. & J.R.I. Wood (= *Lophostachys pubiflora* Lindau); SRR7806564; Bolivia, *Daniel et. al. 10106* (CAS); Barlerieae. ***Neuracanthus africanus*** T. Anderson ex S. Moore; SRR7806563; South Africa, *McDade & Balkwill 1258* (ARIZ); *Neuracanthus*.

Newly sequenced transcriptomes:

Barleria albostellata S. Moore; SRR7806560; cultivated at Long Beach State University, *J.D. Arias 3* (LOB); Barlerieae. ***Barleria oenotheroides*** Dum. Cours.; SRR7806562; cultivated at Long Beach State University, *J.D. Arias 1* (LOB); Barlerieae. ***Barleria rotundifolia*** Oberm.; SRR7806561; cultivated at Long Beach State University, *A.E. Fisher 412* (LOB); Barlerieae. ***Blepharis diversispina*** (Nees) C. B. Clarke; SRR7806559; *M. Stata & R. Sage s.n.*; Acantheae. ***Blepharis spinifex*** Merxm.; SRR7806558; Millennium Seed Bank #248224, *S. Loots et. al. s.n.* (WIND); Acantheae. ***Justicia pacifica*** (Oerst.) A. Gray; SRR7806556; cultivated at Long Beach State University, *A.E. Fisher 447* (LOB); Justicieae. ***Whitfieldia elongata*** (P. Beauv.) DeWild. & T. Durand; SRR7806565; cultivated at Long Beach State University, *A.E. Fisher 443* (LOB); Whitfieldieae.

Publicly available transcriptomes:

Acanthus leucostachys Wall. Ex Nees; SRR1793319; Wenchang, Hainan, China; Acantheae. ***Andrographis paniculata*** Nees; SRR1519324; Andrographideae. ***Avicennia marina*** Wall. Ex Nees; SRR6533719; Wenchang, Hainan, China; Avicennieae. ***Ruellia simplex*** C. Wright; SRR7124544; Ruellieae.