Enabling evolutionary studies at multiple scales in Apocynaceae through Hyb-Seq

Shannon C. K. Straub1,4, Julien Boutte1, Mark Fishbein2, and Tatyana Livshultz3

PREMISE: Apocynaceae is the 10th largest flowering plant family and a focus for study of plant–insect interactions, especially as mediated by secondary metabolites. However, it has few genomic resources relative to its size. Target capture sequencing is a powerful approach for genome reduction that facilitates studies requiring data from the nuclear genome in non-model taxa, such as Apocynaceae.

METHODS: Transcriptomes were used to design probes for targeted sequencing of putatively single-copy nuclear genes across Apocynaceae. The sequences obtained were used to assess the success of the probe design, the intrageneric and intraspecific variation in the targeted genes, and the utility of the genes for inferring phylogeny.

RESULTS: From 853 candidate nuclear genes, 835 were consistently recovered in single copy and were variable enough for phylogenomics. The inferred gene trees were useful for coalescent-based species tree analysis, which showed all subfamilies of Apocynaceae as monophyletic, while also resolving relationships among species within the genus Apocynum. Intraspecific comparison of Elytropus chilensis individuals revealed numerous single-nucleotide polymorphisms with potential for use in population-level studies.

DISCUSSION: Community use of this Hyb-Seq probe set will facilitate and promote progress in the study of Apocynaceae across scales from population genomics to phylogenomics.

KEY WORDS: Apocynaceae; dogbane; genome reduction; Hyb-Seq; low-copy nuclear genes; milkweed; phylogenomics; targeted sequencing.
other downstream applications (Chamala et al., 2015). This approach has been successful in aiding development of targeted sequencing probes in multiple plant lineages (Villaverde et al., 2018; Morais et al., 2019; Jantzen et al., 2020).

The milkweed and dogbane family, Apocynaceae, has publicly available transcriptomes from several species. With ca. 5300 species, Apocynaceae is the 10th largest family of flowering plants and has a worldwide distribution (Endress et al., 2018). Its members are widely known for their production of secondary metabolites that function in defense against herbivores, some of which have evolved mechanisms to detoxify and sequester these compounds (Malcolm and Brower, 1989). Some metabolites, such as vinblastine and vincristine, are used as human medicines (Cragg and Newman, 2005). Pollination biology and floral evolution have been another focal area of study due to the presence of complex derived floral structures, such as pollinia, gynostegia, and coronas (Endress, 1994; Endress and Bruyns, 2000; Fishbein, 2001; Fishbein et al., 2018). Complete resolution of the backbone of the phylogeny of Apocynaceae has proven to be difficult using traditional molecular systematics methods, a supermatrix approach, and even plastome data sets (Livshultz, 2010; Fishbein et al., 2018). However, understanding the evolution of secondary metabolite biosynthetic pathways and other traits, such as floral structure and pollen aggregation, in these plants requires a well-resolved phylogeny, which could be inferred based on information from hundreds of nuclear genes using a target capture sequencing approach.

Previously, Weitemier et al. (2014) used the Hyb-Seq pipeline to identify loci for probe design targeted for sequencing in one genus of Apocynaceae, Asclepias L. They identified 768 single-copy genes and demonstrated their utility in two subtribes of Asclepiadaceae (Asclepiadoideae), and this set has subsequently been utilized to successfully resolve relationships among Asclepias species (Bouette et al., 2019). However, Weitemier et al. (2014) also showed that there would be greatly decreased enrichment success due to DNA sequence divergence from Asclepias in members of the rauvolfioid grade (formerly Rauvolfioideae; Simões et al., 2016) of Apocynaceae. In this study, we produced a set of hybridization probes for targeted sequencing of hundreds of nuclear genes that work across the whole of Apocynaceae, and can be used for future studies at multiple scales ranging from higher-level phylogenomics to intraspecific studies.

**METHODS**

**Probe design**

Five transcriptomes from across Apocynaceae were utilized for probe design for targeted enrichment of single- or low-copy genes. The sampled species represent two of the five subfamilies and four different tribes of Apocynaceae: Asclepias syriaca L. (Weitemier et al., 2019) and Calotropis procera (Aiton) W. T. Aiton (Kwon et al., 2015) of Asclepiadaceae (Asclepiadoideae), and Rauvolfia serpentina (L.) Benth. ex Kurz (Gōngora-Castillo et al., 2012), Rhazya stricta Decne. (Yates et al., 2014), and Tabernaemontana elegans Stapf (Xiao et al., 2013) of the rauvolfioid-grade tribes Vinciae, Amsoniae, and Tabernaemontanaceae, respectively. These exemplars should capture the diversity of the family given that their most recent common ancestor is close to the crown node of the family (Fishbein et al., 2018).

MarkerMiner version 1.2 (Chamala et al., 2015) was employed to identify putatively single-copy genes using the *Vitis vinifera* L. proteome reference and a minimum similarity of 85%. The default minimum transcript length of 900 bp was used because it approximates the setting used by Weitemier et al. (2014) for target capture sequencing probe design in *Asclepias* using the Hyb-Seq pipeline. Intron/exon boundaries were determined by comparison with the *V. vinifera* genome. This work was completed utilizing the High Performance Computing Center facilities of Oklahoma State University at Stillwater.

MarkerMiner output alignments were submitted to MYcroarray (now Daicel Arbor Biosciences, Ann Arbor, Michigan, USA) for the design and synthesis of myBaits biotinylated RNA probes of 120 nucleotides with 2× tiling. Probes with multiple BLAST hits to the *A. syriaca* nuclear genome assembly (Weitemier et al., 2019) or those containing repetitive sequence were excluded from the final probe set of 48,972 probes. An additional 2707 probes for targeting two nuclear genes, paralogs *dlhs* and *hss*, were present in the probe set used for enrichment that generated data for a separate project investigating pyrrolizidine alkaloid biosynthesis (Livshultz et al., 2018). These two genes are not considered further in the reported results of this paper.

The genes targeted across Apocynaceae were compared with the set of genes targeted in *Asclepias* by Weitemier et al. (2014) and the universal angiosperm probe set (Johnson et al., 2019) using reciprocal BLASTN (E-value threshold of 10⁻⁶; minimum percentage of identity of 80%; Altschul et al., 1990) and Python custom scripts (Python version 2.7.12; Python Software Foundation, 2016).

**Taxon sampling, library preparation, and targeted sequencing**

Probes were tested for target enrichment success in 15 species of Apocynaceae spanning the diversity of the family and including representatives of the rauvolfioid grade, the apocynoid grade, Secamonoideae, and Asclepiadoideae (Table 1). Four species of *Apopnynum* L. and two individuals of *Elytropus chilensis* (A. DC.) Müll. Arg. were sampled to assess the utility of genes at the infrageneric and intraspecific levels, respectively. One outgroup species, *Gelsemium sempervirens* (L.) J. St.-Hil. (Gelsemiaceae), from a closely related family in Gentianales was also included.

Total genomic DNA was extracted using previously described methods (Livshultz et al., 2007; Fishbein et al., 2018) from either silica-dried leaf tissue or herbarium specimens (Table 1). DNA was quantified using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and visualized with GelRed nucleic acid stain (Phenix Research Products, Candler, North Carolina, USA) after agarose gel electrophoresis. Genomic DNA was fragmented to approximately 400 bp using sonication in a Bioruptor Pico (Diagenode, Denville, New Jersey, USA). Illumina-compatible libraries were constructed using NEBNext Ultra II kits (New England Biolabs, Ipswich, Massachusetts, USA) in one-half volume reactions with the NEB adapters replaced by NEXTflex-HT barcoded adapters (Bio Scientific Corporation, Austin, Texas, USA). Libraries were pooled in equimolar ratios and sent to Arbor Biosciences for solution hybridization with the biotinylated RNA probes described above. Enriched pools were combined and 150-bp paired-end reads were sequenced on a NextSeq 500 (Illumina, San Diego, California, USA) at the Genomics Core Facility of the Drexel University College of Medicine.
TABLE 1. Sampling, sequencing, and assembly success for 853 nuclear genes in Apocynaceae.a

| Subfamily                  | Tribe            | Sample name                  | Voucher specimen (Herbarium)b | Trimmed reads | Reads on target | Genes with reads mapped | Exons assembled | Exon sequence assembled (bp) | Splash zone sequence assembled (bp) | Paralog warnings | Final gene occupancy | Plastome sequencing depth |
|----------------------------|------------------|------------------------------|--------------------------------|---------------|-----------------|-------------------------|-----------------|----------------------------|--------------------------------------|-----------------|------------------------|------------------------|
| Rauvolfioideae             | Vinceae          | Ochrosia coccinea (Teijm. & Binn.) Miq. | Takeuchi 14422 (A) | 2,717,939 | 73% | 846 | 65,59 | 1,386,894 | 2,675,319 | 3 | 834 | 93x | |
| Tabernaemontaneae          | Tabernaemontana  | Tabernaemontana              | Middleton et al. 1749 (A)    | 1,350,506 | 71% | 843 | 7715 | 1,375,668 | 2,825,977 | 3 | 832 | 40x | |
| Melodineae                 | Caspitoxyspermum | verticatum Bqer ex ADC      | Schoningenber et al. D36 (UPS) | 383,099 | 79% | 825 | 49,16 | 1,070,728 | 1,168,586 | 2 | 816 | 18x | |
| Hurteriæae                 | Hurteria zeylanica | Gardner ex Thwaites         | Middleton et al. 3816 (E)    | 1,295,655 | 77% | 845 | 6186 | 1,314,036 | 2,212,866 | 2 | 834 | 36x | |
| Amsoniæae                  | Amsonia orientalis Deene | Prestaria portellensis (Beurl.) Woodson | Endless sn. (Z) | 3,663,743 | 84% | 848 | 68,39 | 1,394,925 | 2,419,790 | 1 | 835 | 30x | |
| Apocynoideae               | Echiteae         | Prestaria portellensis      | Ventura 21262 (NY)           | 3,581,141 | 78% | 844 | 48,55 | 1,191,183 | 2,053,075 | 1 | 830 | 156x | |
| Odontadeniæae              | Elytropus chensis Müll. | Elytropus chensis | Sobel & Strudwick 2740 (NY)  | 909,288 | 82% | 844 | 55,11 | 1,166,106 | 1,447,041 | 1 | n/a | 49x | |
| Apocyneae                  | Apocynum androsaemifolium L. | Apocynum androsaemifolium | T. Livshultz 03-32 (BH)       | 1,931,260 | 78% | 847 | 54,08 | 1,234,158 | 1,823,085 | 1 | 834 | 155x | |
| Asclepiadoideae            | Fockeae          | Fockea edulis                | T. Livshultz 03-28 (BH)       | 1,385,483 | 79% | 847 | 56,86 | 1,272,198 | 2,590,109 | 1 | 832 | 198x | |
| Marsdenieae                | Marsdenia glabra | costantin                   | D. J. Middleton et al. 1123 (A) | 3,138,002 | 78% | 847 | 42,82 | 989,700 | 1,216,606 | 2 | 812 | 19x | |
| Asclepiadeae               | Tassadia berteroana | G. W.D. Stevens | Fuentes 3904 (OKLA)           | 6,652,102 | 86% | 844 | 57,27 | 2,043,461 | 1,829 | 1 | 829 | 102x | |
| Outgroup                   | Gelsemium sempervirens (L.) J. St.-Hil. | F. Konta & N. Abjusalik 35959 (NY) | 1,751,248 | 69% | 842 | 57,27 | 1,205,457 | 2,430,178 | 1 | 829 | 102x | |

Mean (SD) 2,134,189 (1,653,721) 76% (6%) 840 (10) 507 (120) 1,183,039 (165,106) 2,007,757 (533,739) 14 823 (16) 80x (61x)

Median 1,751,248 78% 844 5686 1,191,183 2,053,075 1 829.5 54x

aDNA extractions were performed from herbarium specimens for vouchers indicated in boldface.
bHerbarium acronyms are according to Index Herbariorum (http://sweetgum.nybg.org/science/ih/).
**Sequence clean-up, assembly, and alignment**

Illumina reads were quality trimmed using Trimmomatic version 0.36 (Bolger et al., 2014). Adapter sequences were deleted and reads with a quality score ≤ Q20 were removed. Then, a sliding window approach was employed to trim the 3′-end of the read with a quality score ≤ Q20 for a window of 5 bp. Only reads of at least 36 bp were retained for downstream analyses. Cleaned reads were assembled using HybPiper version 1.2 (Johnson et al., 2016) with the Burrows-Wheeler Aligner (BWA) option. Supercontigs were produced and intron positions were detected using the HybPiper intronrate module. Genes for which paralogous sequences were detected using the HybPiper paralog Investigator module and genes with <75% terminal occupancy were removed from downstream analyses. Read depth and HybPiper statistics were obtained using the depth_calculator and hybpiper_stat modules, respectively. Work reported here was run on hardware supported by Drexel’s University Research Computing Facility.

Putatively orthologous sequences were aligned using MAFFT version 1.3.6 (Katoh et al., 2002) using an iterative refinement method (FFT-NS-i, two cycles). The BAM output files created by HybPiper were converted to PILEUP files using SAMtools (Li et al., 2009) to detect single-nucleotide polymorphisms (SNPs). Intraspecific SNPs were identified in *E. chilensis* and intrageneric SNPs were identified in *Apocynum* using the SNP detection method of Boutte et al. (2016) with a minimum read depth of 25 and a SNP detection threshold of 30 as well as additional custom Python scripts. Read pools for each sample were assessed for the potential to assemble plastomes using off-target reads. Trimmed reads were mapped to a reference plastome from *Oncinotis tenuiloba* Stapf (NC_025657), a member of the apocynoid grade, in Geneious. Oncinotis tenuiloba mapped to a reference plastome from *Stapf* assemble plastomes using off-target reads. Trimmed reads were 30 as well as additional custom Python scripts.

**RESULTS**

**Targeted sequencing**

A total of 853 putatively single-copy genes with a cumulative exon length of 1,545,593 bp were identified in Apocynaceae. Among these 853 genes, probes were designed for 359 genes considering at least one transcript from an Asclepiadoideae species and a Rauvolfioideae species. Probes were designed for 341 genes without a transcript from Asclepiadoideae present, and for 153 genes without a Rauvolfioideae transcript present. On average 76% (±6%) of reads were on target, a mean of 98% (±1%) of genes were at least partially assembled, very few potential paralogs (0–3) were detected, and samples sourced from herbarium specimens produced results similar to silica-dried samples (Table 1, Fig. 1). For the Apocynaceae ingroup, assembly of targeted exons ranged from 4288 to 7715, with an average of 6.8 and a median of six exons per gene, and yielded between 989,706 bp and 1,394,925 bp of sequence. Assembly of untargeted splash zones (i.e., intron and intergenic spacers) produced from 1,168,586 bp to 2,825,097 bp of additional sequence per sample (Table 1). The outgroup, *G. sempervirens*, did not have the fewest partially assembled genes, but did produce the fewest assembled exons and the least amount of total assembled sequence data, and had the lowest percentage of reads on target (Table 1, Fig. 1). Mapping results indicated that from the off-target reads, the plastome had been sequenced to an average depth of 80× (Table 1).

Eighteen genes were removed from downstream analyses because four of the targeted genes were not recovered in any species,

**Phylogenetic analyses**

For phylogenetic analyses, only the *E. chilensis* individual with the highest number of assembled genes was retained. Gene trees were estimated in a maximum likelihood framework using the GTR+GAMMA model of sequence evolution in IQ-TREE version 1.5.5 (Nguyen et al., 2015). One thousand ultrafast bootstrap replicates (Hoang et al., 2018) were conducted for each sequence matrix. Species trees were inferred using ASTRAL II version 4.10.12 with default parameters (Mirarab and Warnow, 2015). For species tree analyses, three sampling strategies were used: (1) all gene trees based on exons only, (2) exon-only gene trees with very poorly resolved gene trees (total of ≥75% of nodes with <75% bootstrap support) removed, and (3) all genes with exons and splash zones included (supercontigs). Phyparts with the bootstrap filter set to 50% (Smith et al., 2015) was used to assess gene tree conflict, which was represented using pie charts using the PhyParts_PieCharts script (https://github.com/mossmatters/MJPythonNotebooks/blob/master/phypartspiecharts.py) for all trees that contained the outgroup, *Gelsemium sempervirens*.

**FIGURE 1.** Heatmap of target enrichment, sequencing, and assembly success for 853 nuclear genes in Apocynaceae.
an additional nine genes had <75% terminal occupancy, and five
genes had putative paralogs present in one or more sampled indi-
viduals. Of the 853 targeted genes, 81 (9.5%) overlapped with the
353 genes in the universal angiosperm set and 69 (8.1%) overlapped
with the 768 genes in the Asclepias-specific set.

Alignment and phylogeny
Cumulatively, the exon-only sequence alignments had 1,922,030
sites with an average of 2302 (±1397) sites per gene and a median
length of 1900 sites. For the ingroup, there were 30% variable and
12% parsimony informative characters. The aligned exon-plus-
splash-zone matrices had 7,748,184 characters with an average of
9387 (±5949) characters per matrix and a median length of 7650
characters. There were 34% variable and 14% parsimony informa-
tive characters for the ingroup.

Most gene trees were inferred with strong bootstrap support for
the relationships among taxa. Only 18 of the exon-only trees had less
than 75% of the nodes with greater than 75% bootstrap support. The
species trees inferred using all exon-only (n = 835), exon-only with
very poorly resolved trees removed (n = 817; Fig. 2), and the all ex-
on-plus-splash-zone trees (n = 835) had identical topologies, with all
subfamilies (including Rauvolfioideae) monophyletic. They differed
slightly in the local posterior probabilities within Rauvolfioideae
(data not shown). Notably, the local posterior probability for a mono-
phyletic Rauvolfioideae was much lower than all other local poste-
rior probabilities (0.69 vs. ≥0.98). The normalized quartet score for
the species tree was 0.84, indicating a moderate level of discordance.

![FIGURE 2. ASTRAL species tree of Apocynaceae inferred from 817 nuclear gene trees. Numbers next to the nodes are local posterior probabilities. The scale bar is in coalescent units. For each node, pie charts show the proportions of the gene tree bipartitions that are concordant with the species tree (blue), the most frequently observed alternative bipartition (green), all other bipartitions (red), and uninformative or missing (gray). The proportions were calculated using all gene trees containing the outgroup Gelsemium sempervirens (n = 761). Numbers of concordant/conflicting gene trees are shown next to or below each pie chart.](http://www.wileyonlinelibrary.com/journal/AppsPlantSci)
among gene trees. For most of the relationships within and among Apocynoideae, Secamonoideae, and Asclepiadoideae, the majority of gene trees were informative and concordant with the species tree. However, among Rauvolfioideae, there were many uninformative gene trees, and more gene trees were in conflict with the species tree than were concordant, despite there being relatively high local posterior probabilities for most splits.

**Interspecific and intraspecific variation**

For the four species of *Apocynum*, 171,323 intrageneric SNPs were detected, and were much more frequent in splash zones than in exons (*n* = 834 genes; Table 2, Fig. 3A). The phylogenetic information present in these nucleotide substitutions led to complete resolution of the evolutionary relationships among these species (Fig. 2). In the comparison of the sequences from the two *E. chilensis* individuals, 38,555 intraspecific SNPs were observed, and they were more common in splash zones than in exons (*n* = 831 genes; Table 2, Fig. 3B).

**DISCUSSION**

The probes designed from transcriptome sequences of Apocynaceae were effective for targeted sequencing of 835 putatively single-copy nuclear genes from an initial set of 853 genes. The assembled sequences contained information that was useful for resolving the phylogeny of Apocynaceae from the higher-level subfamily relationships down to species relationships within a single genus, *Apocynum*. The probe set should therefore be useful for phylogenetic studies across the family at multiple scales, and potentially in other families of Gentianales given the high success rate for assembling nuclear genes in the outgroup, a representative of Gelsemiaceae. Most samples had sufficient plastome sequencing depth from off-target reads to suggest that reference-guided or de novo assembly of the plastome would likely be successful (Straub et al., 2012) and add another potential source of information for inferring phylogeny (see Fishbein et al., 2018). Furthermore, given the numerous SNPs detected in both exonic and intronic gene regions for *Elytropus* Müll. Arg. individuals, the probes will also be useful for generating data appropriate for population genetics and evolutionary genomics studies.

The new bait set for targeted sequencing in Apocynaceae presented here provides an example of a successful lineage-specific probe design. MarkerMiner yielded loci that were conserved enough to be targeted across the whole family, and very few of the loci turned out not to be single copy, as has been observed in other studies that compared MarkerMiner-based designs to those generated by other pipelines (Kadlec et al., 2017; Vatanparast et al., 2018). Additionally, several studies that have compared the success of taxon-specific and universal probe sets have found that taxon-specific designs yield the largest number of most useful loci (Kadlec et al., 2017; Chau et al., 2018; Jantzen et al., 2020), although this has not been universally observed (Larridon et al., 2020). When feasible, a combination of lineage-specific and universal probe sets, such as the Angiosperms353 probe set (Johnson et al., 2019), can yield the largest pool of nuclear loci appropriate for phylogenomic studies (Jantzen et al., 2020).

In agreement with previous molecular phylogenetic studies that utilized chloroplast regions, whole plastomes, and/or few nuclear

---

**TABLE 2.** Number of single-nucleotide polymorphisms per 100 bp of exonic and splash zone sequence.

| Taxon             | Sequence type | Mean  | Median | 95% Confidence interval       |
|-------------------|---------------|-------|--------|-------------------------------|
| *Apocynum*        | Exons         | 0.450 | 0.357  | [0.426, 0.474]                |
|                   | Splash zones  | 3.242 | 2.392  | [3.127, 3.357]                |
| *Elytropus chilensis* | Exons        | 0.110 | 0.068  | [0.101, 0.119]                |
|                   | Splash zones  | 0.746 | 0.327  | [0.676, 0.825]                |

**FIGURE 3.** Frequency and distribution of intrageneric and intraspecific single-nucleotide polymorphisms (SNPs) in targeted nuclear genes. (A) Intrageneric SNPs among four species of *Apocynum*. Sites where one or more individuals were heterozygous were excluded from the graph. (B) Intraspecific SNPs in two individuals of *Elytropus chilensis*. 
genes for inferring the phylogeny of Apocynaceae (Verhoeven et al., 2003; Livshultz et al., 2007; Livshultz, 2010; Fishbein et al., 2018). Asclepiadoideae was monophyletic and sister to Secamonoideae with local posterior probabilities of 1. Based on previously published analyses, Rauvolfioideae and Apocynoideae have been understood to be paraphyletic grades within the family (Livshultz et al., 2007; Simões et al., 2007, 2016; Livshultz, 2010; Fishbein et al., 2018), but were unexpectedly recovered as monophyletic in this analysis. For Apocynoideae, one might have expected *Apocynum* to be sister to Secamonoideae plus Asclepiadoideae with *Elytropus* (Odontadenieae) plus *Prestonia* (Echiteae) sister to that group, and for sampled Rauvolfioideae to be paraphyletic with the *Amsonia-Hunteria* clade sister to the Apocynoideae-Secamonoideae-Asclepiadoideae clade (Fishbein et al., 2018). Here, the monophyly of Apocynoideae in the species tree had high local posterior probability and was supported by a majority of gene trees (Fig. 2). In contrast, the monophyly of Rauvolfioideae was only weakly supported in the species tree, and the proportion of gene trees with bipartitions that conflicted with the species tree was much higher than the proportion that supported the species tree topology.

The phylogenomic results observed here are based on data from many nuclear genes and may give better insight into the evolution of Apocynaceae than previous studies due to greatly increased character sampling; however, even in phylogenomic studies taxon sampling has been shown to affect species tree topology (Walker et al., 2017). Sparse taxon sampling has been demonstrated to have the potential to negatively affect phylogenetic reconstruction (Hillis et al., 2003; Nabhan and Sarkar, 2012), even though coalescent approaches may be less affected by this issue than are other analytical methods for species tree inference (Bravo et al., 2019). Given that previous studies of both plastid and nuclear loci have provided strong evidence that Rauvolfioideae and Apocynoideae are paraphyletic grades, their surprising monophyly observed in this study is likely to be shown to be anomalous with more thorough taxon sampling. Future phylogenomic studies using these probes will employ more extensive sampling, and engage in deeper exploration of the data to assess the impacts of various strategies on species tree inference. Potential strategies include: trimming of multiple sequence alignments and detection of errors in homology assessment, partitioning sequences, evaluating models of sequence evolution prior to phylogenetic analysis to infer gene trees, data filtering such as detecting outlier genes, and employing multiple approaches to species tree inference (Bravo et al., 2019).

Going forward, the large number of putatively single-copy genes identified across the Apocynaceae and the probe set designed in this study will both serve as resources for the community of researchers interested in answering questions about Apocynaceae phylogenetics, phylogeography, and population genetics. This new gene set does not greatly overlap with the universal 353 gene set for angiosperms (Johnson et al., 2019) or the *Asclepias*-specific 768 gene set (Weitemier et al., 2014) that works well across one Apocynaceae subfamily (Asclepiadoideae), and therefore increases the range of options for researchers to target subsets of single-copy genes appropriate to the level of divergence and questions asked in a particular study.

ACKNOWLEDGMENTS

The authors thank A. Simões (Universidade Estadual de Campinas) and M. Endress (University of Zurich) for tissue of *A. orientalis*; J. Teisher and D. Chin (Drexel University), and J. Schafer (Oklahoma State University) for computational support; C. Smith (Drexel University), and A. Foote, M. Cullinan, M. Steinfeldt, K. Kostović, and C. Chung (Hobart and William Smith Colleges) for laboratory help; and A. Devault and J. Enk (Daicel Arbor BioSciences) for probe design assistance. Funding was provided by the National Science Foundation (grant DEB 1655223/1655553 to S.C.K.S. and T.L. and 1457473/1457510 to S.C.K.S. and M.F.), and a Drexel University Clinical and Translational Research Institute (CTRI) seed grant to T.L.

DATA AVAILABILITY

Raw Illumina sequence data are available from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (BioProject PRJNA660787). Nuclear gene probe sequences ([https://doi.org/10.6084/m9.figshare.12830579.v1](https://doi.org/10.6084/m9.figshare.12830579.v1) and [https://doi.org/10.6084/m9.figshare.12830648.v1](https://doi.org/10.6084/m9.figshare.12830648.v1)) are available from Figshare. Custom Python scripts are available upon request.

LITERATURE CITED

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

Bagley, J. C., S. Uribe-Convers, M. M. Carlsen, and N. Muchhala. 2020. Utility of targeted sequence capture for phylogenomics in rapid, recent angiosperm radiations: Neotropical *Barnesieria* bellflowers as a case study. *Molecular Phylogenetics and Evolution* 152: 106769.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

Boutte, J., I. Ferreira de Carvalho, M. Rousseau-Gueutin, J. Poulain, C. Da Silva, P. Wincker, M. Ainouche, and A. Salmon. 2016. Reference transcriptomes and detection of duplicated copies in hexaploid and allododecaploid *Spartina* species (Poaceae). *Genome Biology and Evolution* 8: 3030–3044.

Boutte, J., M. Fishbein, A. Liston, and S. C. K. Straub. 2019. NGS-Indel Coder: A pipeline to code indel characters in phylogenomic data with an example of its application in milkweeds (*Asclepias*). *Molecular Phylogenetics and Evolution* 139: 106534.

Bravo, G. A., A. Antonelli, C. D. Bacon, K. Bartoszek, M. P. K. Blom, S. Huyghn, G. Jones, et al. 2019. Embracing heterogeneity: Coalescing the Tree of Life and the future of phylogenomics. *Peer J*: e6399.

Chamala, S., N. Garcia, G. T. Godden, V. Krishnakumar, I. E. Jordon-Thaden, R. DeSmet, W. B. Barbazuk, and et al. 2015. MarkerMiner 1.0: A new application for phylogenetic marker development using angiosperm transcriptomes. *Applications in Plant Sciences* 3: 1400115.

Chau, J. H., W. A. Rahfeldt, and R. G. Olimstead. 2018. Comparison of taxon-specific versus general locus sets for targeted sequence capture in plant phylogenomics. *Applications in Plant Sciences* 6: e1032.

Couvreur, T. L. P., A. J. Helmstetter, E. J. M. Koenen, K. Bethune, R. D. Brandão, S. A. Little, H. Sauquet, and R. H. J. Erkins. 2019. Phylogenomics of the major tropical plant family Annonaceae using targeted enrichment of nuclear genes. *Frontiers in Plant Science* 9: 1941.

Cragg, G. M., and D. J. Newman. 2005. Plants as a source of anti-cancer agents. *Journal of Ethnopharmacology* 100: 72–79.

Crow, A. A., C. Myers, and N. Cellinese. 2017. Embracing discordance: Phylogenomic analyses provide evidence for allopolyploidy leading to cryptic diversity in a Mediterranean *Campanula* (Campanulaceae) clade. *Evolution* 71: 913–922.

de la Harpe, M., J. Hess, O. Loiseau, N. Salamin, C. Lerox, and M. Paris. 2019. A dedicated target capture approach reveals variable genetic markers across
micro- and macro-evolutionary time scales in palms. *Molecular Ecology Resources* 19: 221–234.

DeSmet, R., K. L. Adams, K. VandePoule, M. C. E. V. Montagu, S. Maere, and Y. Y. de Peer. 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proceedings of the National Academy of Sciences*, USA 110: 2988–2993.

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

Edwards, S. V., Z. Xi, A. Janke, B. C. Faircloth, J. E. McCormack, T. C. Glenn, B. Zhong, et al. 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. *Molecular Phylegenetics and Evolution* 94: 447–462.

Endress, M. E., and P. V. Bruyns. 2000. A revised classification of the Apocynaceae s.l. *Botanical Review* 66: 1–56.

Endress, M. E., U. Meve, D. J. Middleton, and S. Liede-Schumann. 2018. Apocynaceae. In K. Kubitzki [ed.], The families and genera of vascular plants: Flowering plants, Eudicots, Apiales, Gentianales (except Rubiaceae). 207–411. Springer International Publishing, Cham, Switzerland.

Endress, P. K. 1994. Diversity and evolutionary biology of tropical flowers. Cambridge University Press, Cambridge, United Kingdom.

Fisher, M. 2001. Evolutionary innovation and diversification in the flowers of Asclepiadaceae. *Annals of the Missouri Botanical Garden* 88: 603–623.

Fisher, M., T. Livshultz, S. C. K. Straub, A. O. Simões, J. Boute, A. McDonnell, and A. Foote. 2018. Evolution on the backbone: Apocynaceae phylogenomics and new perspectives on growth forms, flowers, and fruits. *American Journal of Botany* 105: 495–513.

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.

Gongora-Castillo, E., K. L. Childs, G. Fedewa, J. P. Hamilton, D. K. Liscombe, M. Magallanes-Lundback, K. K. Mandali, et al. 2012. Development of transcriptomic resources for investigating the biosynthesis of monoterpenoid indole alkaloids in medicinal plant species. *PLoS ONE* 7: e32506.

Herrando-Moraira, S., J. A. Calleja, M. Galbany-Casals, N. Garcia-Jacas, J.-Q. Liu, J. López-Alvarado, J. López-Pujol, et al. 2019. Nuclear and plastid DNA phylogeny of tribe Cardueae (Compositae) with Hyb-Seq data: A new subtribal classification and a temporal diversification framework. *Molecular Phylogenetics and Evolution* 137: 313–332.

Heyduk, K., D. W. Trapnell, C. F. Barrett, and J. Leebens-Mack. 2016. Phylogenomic analyses of species relationships in the genus *Salal* (Araucaceae) using targeted sequence capture. *Biological Journal of the Linnean Society* 117: 106–120.

Hillis, D. M., D. D. Pollock, J. A. McGuire, and D. J. Zwickl. 2003. Is sparse taxon sampling a problem for phylogenetic inference? *Systematic Biology* 52: 124–126.

Hoang, D. T., O. Chernomor, A. von Haeseler, B. Q. Minh, and L. S. Vinh. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution 35*: 518–522.

Jantzen, J. R., P. Amarasinghe, R. A. Folk, M. Reginato, F. A. Michelangeli, D. W. Trapnell, C. F. Barrett, and J. Leebens-Mack. 2016. Phylogenomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. *PLoS ONE* 7: e32506.

Katoh, K., K. Misawa, K. Kuma, and T. Miyata. 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30: 3059–3066.

Kwon, C. W., K.-M. Park, B.-C. Kang, D.-H. Kweon, M.-D. Kim, S. W. Shin, Y. H. Je, and P.-S. Chang. 2015. Cysteine protease profiles of the medicinal plant *Calotropis procera* R. Br. revealed by de novo transcriptome analysis. *PLoS ONE* 10: e0119328.

Larridon, I., T. Villaverde, A. R. Zuntini, L. Pokorny, G. E. Brewer, N. Epitawalage, I. Fairlie, et al. 2020. Tackling rapid radiations with targeted sequencing. *Frontiers in Plant Science* 10: 1655.

Lemmon, E. M., and A. R. Lemmon. 2013. High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 44: 99–121.

Léveillé-Bourret, E., J. R. Starr, B. A. Ford, E. Moriarty Lemmon, and A. R. Lemmon. 2018. Resolving rapid radiations within angiosperm families using anchored phylogenomics. *Systematic Biology* 67: 94–112.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079.

Livshultz, T. 2010. The phylogenetic position of milkweeds (Apocynaceae subfamilies Secamonoideae and Asclepiadoideae): Evidence from the nucleus and chloroplast. *Taxon* 59: 1016–1030.

Livshultz, T., D. J. Middleton, M. E. Endress, and J. K. Williams. 2007. Phylogeny of Apocynaceae and the APSA clade (Apocynaceae s.l.). *Annals of the Missouri Botanical Garden* 94: 324–359.

Livshultz, T., E. Kaltenegger, S. C. K. Straub, K. Weitemier, E. Hirsch, K. Koval, L. Mema, and A. Liston. 2018. Evolution of pyrrolizidine alkaloid biosynthesis in Apocynaceae: Revisiting the defence de-escalation hypothesis. *New Phytologist* 218: 762–773.

Malcolm, S. B., and L. P. Brower. 1989. Evolutionary and ecological implications of cardenolide sequestration in the monarch butterfly. *Experimental* 45: 284–295.

Mandel, J. R., R. B. Dikow, and V. A. Funk. 2015. Using phylogenomics to resolve mega-families: An example from Compositae. *Journal of Systematics and Evolution* 53: 391–402.

Mirarab, S., and T. Warnow. 2015. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31: 144–152.

Moraes, E. B., J. Schönengerber, E. Conti, A. Antonelli, and P. Szövényi. 2019. Orthogonal nuclear markers and new transcriptomes that broadly cover the phylogenetic diversity of Acanthaceae. *Applications in Plant Sciences* 7: e11290.

Nabhan, A. R., and I. N. Sarkar. 2012. The impact of taxon sampling on phylogenetic inference: A review of two decades of controversy. *Briefings in Bioinformatics* 13: 122–134.

Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274.

Peng, Z., W. Fan, L. Wang, D. Paudel, D. Leventini, B. L. Tillman, and J. Wang. 2017. Target enrichment sequencing in cultivated peanut (*Arachis hypogaea* L.) using probes designed from transcript sequences. *Molecular Genetics and Genomics* 292: 955–965.

Python Software Foundation. 2016. Python language reference. version 2.7.12. Website http://www.python.org [accessed 20 October 2020].

Simões, A. O., T. Livshultz, E. Conti, and M. E. Endress. 2007. Phylogeny and systematics of the Rauvolfioideae (Apocynaceae) based on molecular and morphological evidence. *Annals of the Missouri Botanical Garden* 94: 268–297.

Simões, A. O., L. S. Kinoshita, I. Koch, M. J. Silva, and M. E. Endress. 2016. Systematics and character evolution of Vinceae (Apocynaceae). *Taxon* 65: 99–122.

Smith, S. A., M. J. Moore, J. W. Brown, and Y. Yang. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. BMC *Evolutionary Biology* 15: 150.

Straub, S. C. K., M. Parks, K. Weitemier, M. Fishbein, R. C. Cronn, and A. Liston. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* 99: 349–364.

http://www.wileyonlinelibrary.com/journal/AppsPlantSci © 2020 Straub et al.
Straub, S. C. K., M. J. Moore, P. S. Soltis, D. E. Soltis, A. Liston, and T. Livshultz. 2014. Phylogenetic signal detection from an ancient rapid radiation: Effects of noise reduction, long-branch attraction, and model selection in crown clade Apocynaceae. *Molecular Phylogenetics and Evolution* 80: 169–185.

Vatanparast, M., A. Powell, J. J. Doyle, and A. N. Egan. 2018. Targeting legume loci: A comparison of three methods for target enrichment bait design in Leguminosae phylogenomics. *Applications in Plant Sciences* 6: e1036.

Verhoeven, R. L., S. Liede, and M. E. Endress. 2003. The tribal position of *Fockea* and *Cibirhiza* (Apocynaceae: Asclepiadoideae): Evidence from pollinium structure and cpDNA sequence data. *Grana* 42: 70–81.

Villaverde, T., L. Pokorny, S. Olsson, M. Rincón-Barrado, M. G. Johnson, E. M. Gardner, N. J. Wickett, et al. 2018. Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytologist* 220: 636–650.

Walker, J. F., Y. Yang, M. J. Moore, J. Mikenas, A. Timoneda, S. F. Brockington, and S. A. Smith. 2017. Widespread paleopolyploidy, gene tree conflict, and recalcitrant relationships among the carnivorous Caryophyllales. *American Journal of Botany* 104: 858–867.

Weitemier, K., S. C. K. 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.

Weitemier, K., S. C. K. Straub, M. Fishbein, C. D. Bailey, R. C. Cronn, and A. Liston. 2019. A draft genome and transcriptome of common milkweed (*Asclepias syriaca*) as resources for evolutionary, ecological, and molecular studies in milkweeds and Apocynaceae. *PeerJ* 7: e7649.

Xiao, M., Y. Zhang, X. Chen, E.-I. Lee, C. J. S. Barber, R. Chakrabarty, I. Desgagné-Penix, et al. 2013. Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *Journal of Biotechnology* 166: 122–134.

Yates, S. A., I. Chernukhin, R. Alvarez-Fernandez, U. Bechtold, M. Baeshen, N. Baeshen, M. Z. Mutwakil, et al. 2014. The temporal foliar transcriptome of the perennial C₃ desert plant *Rhazya stricta* in its natural environment. *BMC Plant Biology* 14: 2.