Recent and rapid diversifications are common in the tree of life (Schluter, 2000; Madriñan et al., 2013), and large amounts of data are often required to resolve phylogenetic relationships (Rokas et al., 2003). Multiple independent loci are needed to accurately reconstruct species trees (Small et al., 2004; Leaché and Rannala, 2011) because gene trees from single loci may not reflect species relationships due to incomplete lineage sorting, unrecognized paralogy, and lateral gene transfer or hybridization between species (Maddison, 1997).

Until recently, most phylogenetic studies in plants have relied on a few loci, particularly from the easily amplified and variable plastid and nuclear ribosomal RNA regions. However, each of these regions represents only a single gene history (Small et al., 2004). Additional nuclear loci have been targeted for development in phylogenetic studies because many are relatively fast-evolving and each nuclear locus potentially represents an independent gene history (Yuan et al., 2009). However, their traditional application using PCR and Sanger sequencing is often difficult because of the need to design applications in Plant Sciences 6(3): e1032; http://www.wileyonlinelibrary.com/journal/AppsPlantSci © 2018 Chau et al. Applications in Plant Sciences is published by Wiley Periodicals, Inc. on behalf of the Botanical Society of America. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
primers and test loci for phylogenetic utility in each taxonomic group under study (Hughes et al., 2006; Zimmer and Wen, 2013).

The development of next-generation sequencing technologies allows for the efficient sequencing of large portions of the genome, including in non-model taxa (Egan et al., 2012; Twyford and Ennos, 2012; Soltis et al., 2013). Further increases in efficiency in time and cost can be achieved by combining multiplexing techniques with target enrichment, which reduces the proportion of the genome sequenced to subsets that are more likely to be useful (Grover et al., 2012). Various techniques have been developed for target enrichment, including methods based on restriction enzymes, transcriptomes, PCR, and sequence capture (Cronn et al., 2012; McCormack et al., 2013; Lemmon and Lemmon, 2013).

Targeted sequence capture, or hybrid enrichment, isolates target loci from a pool of fragmented genomic DNA using oligonucleotide probes, which can then be sequenced through next-generation sequencing (Mamanova et al., 2010). Among the strengths of targeted sequence capture are the ability to target known loci, lower stringency in the matching of probes and targets compared to primers for PCR, and the ability to capture sequences even from degraded DNA (Cronn et al., 2012; Lemmon et al., 2012).

Choosing a set of loci to target is an early and crucial part of the sequence capture approach. A broadly used target locus set with universal probes, like ultraconserved elements in amniotes (Faircloth et al., 2012), has not been developed for plants because highly conserved sequences in plants are rare (Reneker et al., 2012; Zheng and Zhang, 2012; but see Buddenhagen et al., 2016). Instead, target locus sets of single-copy nuclear loci with orthologs across a taxonomic group have been developed individually for groups, which requires multiple genomic and/or transcriptomic resources for taxa in the group and bioinformatic expertise (e.g., Mandel et al., 2014; Nicholls et al., 2015; Stephens et al., 2015; Heyduk et al., 2016). We refer to target locus sets identified with these methods as “taxon-specific” because they comprise loci that are shown to be single-copy and have orthologs specifically in the taxa whose genomic information is used. More recently, several pipelines for identifying taxon-specific target loci that require less bioinformatic skills have been developed (Weitemier et al., 2014; Chamala et al., 2015; Schmickl et al., 2016), but multiple genomic resources are still needed.

Because multiple genomic resources are currently available for only a small number of taxa, and new resources can be expensive and time consuming to generate, the development of general target locus sets applicable in many taxa would facilitate the wider use of targeted sequence capture for plant phylogenomics. Several studies have identified loci that are putatively single-copy and have orthologs across large clades of plants by examining genome and transcriptome data from distantly related species. These include the conserved ortholog set (COSII) in euasterids (Wu et al., 2006), shared single-copy nuclear genes (APVO SSC genes) in angiosperms (Duarte et al., 2010), the pentatricopeptide repeat (PPR) gene family in angiosperms (Yuan et al., 2009), other low-copy nuclear genes conserved across angiosperms (Zhang et al., 2012), and universal markers developed for individual families (e.g., Chapman et al., 2007; Curto et al., 2012). The utility of these general locus sets in comparison with taxon-specific locus sets in targeted sequence capture and phylogenomics has not been evaluated (but see Granados Mendoza et al., 2015; Buddenhagen et al., 2016; Lévéillé-Bourret et al., 2018).

Once a target locus set is chosen, probes can be designed to capture the target loci. Probes can also be taxon specific or universal. However, because hybridization efficiency depends on sequence similarity between the probe and target (Cronn et al., 2012; Buddenhagen et al., 2016), and highly conserved sequences are rare in plants (Reneker et al., 2012), taxon-specific probes designed using genomic data from one or more exemplar species in the clade under study usually perform better.

In this study, we identified four locus sets for targeted sequence capture in the genus Buddleja L. (Scrophulariaceae), a clade of 108 species with a crown age of ~20 Ma (Chau, 2017). We used one taxon-specific set that was identified using genomic and transcriptomic data for two species of Buddleja, and three general sets consisting of COSII, APVO SSC, and PPR loci. We compared the four locus sets selected for probe design and evaluated the performance of the locus sets in Buddleja and outgroups in terms of the assembly of target sequences and sequence variation in loci. We also inferred phylogenetic relationships for Buddleja using concatenation and species tree approaches and compared support for relationships from different locus sets and from previous analyses using a non-sequence capture data set.

**METHODS**

**Whole-genome shotgun sequencing of Buddleja globosa**

One specimen of Buddleja globosa Hope (Washington Park Arboretum accession number: 179-99-A, voucher: R.G. Olmstead 2010-46 [WTU]), a diploid species with a haploid (1C) genome size of 840 Mbp (Hanson et al., 2001), was selected for whole-genome shotgun sequencing. DNA was extracted from fresh young leaves using a modified cetyltrimethylammonium bromide (CTAB) protocol (Doyle and Doyle, 1987) and purified through isopropanol precipitation. DNA was diluted to a concentration of 10 ng/μL and sheared by sonication in a Bioruptor (Diagenode Inc., Denville, New Jersey, USA) with a target size of 300 bp. The sequencing library was prepared with an Illumina TruSeq v2 DNA sample preparation kit (Illumina, San Diego, California, USA), and quality was checked with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA). The library was sequenced with 100-bp paired-end reads on one lane of an Illumina HiSeq 2000 (Illumina) at the QB3 Genomics Sequencing Laboratory at the University of California, Berkeley, USA. Reads were filtered and de novo assembled using CLC Genomics Server 5.0.2 (QIAGEN Bioinformatics, Redwood City, California, USA) with default parameters.

**Selection of loci for targeted sequence capture and probe design**

We took two approaches to selecting loci that are nuclear, single-copy, and have orthologs across Buddleja for sequence capture. For the taxon-specific approach, we used our genomic data for B. globosa and two transcriptomes for B. davidi Franch. (samples GRFT and XRLM) from the 1000 Plants (1KP) initiative (https://sites.google.com/a/ualberta.ca/onekp). Buddleja globosa and B. davidi are in different sections of Buddleja (Chau et al., 2017), so loci found in both species are likely to have orthologs throughout the genus. To select loci, we utilized a modified version of the marker development pipeline Sondovac (Schmickl et al., 2016). Briefly, the pipeline takes genome read data and removes any reads matching a plastome or mitochondrion reference. We used a plastome from Scrophularia takesimensis Nakai (GenBank accession no.: NC_026202), in the same family as Buddleja (Scrophulariaceae), and a mitochondrion from Salvia miltiorrhiza Bunge (GenBank accession no.:
NC_023209), in the same order as Buddleja (Lamiales). Sondovac then removes duplicated transcripts from the transcriptome to identify single-copy loci and finds genome reads matching the remaining unique transcripts, which are de novo assembled. Assembled contigs from genome reads are filtered for length (contig >180 bp, total length of all contigs for a transcript >600 bp) and uniqueness. Remaining contigs are compiled as target sequences. We used the pipeline separately with each of the two transcriptomes of B. davidii and then combined target sequences. Sequences with >90% sequence similarity were identified using cd-hit-est (Li and Godzik, 2006), and the longest sequence in each cluster was retained.

Our other approach for locus selection utilized three general locus sets, which comprise loci that are putatively single-copy and have orthologs across large clades of plants. COSII were identified in the euasterid clade by comparing expressed sequence tag databases for several species of euasterids (in Solanum, Capsicum, and Coffea) and Arabidopsis thaliana (L.) Heynh. (Wu et al., 2006). Sequences for a subset of 369 COSII genes in Solanum lycopersicum L. were downloaded from the Sol Genomics Network (ftp://ftp.sgn.cornell.edu/COSII/Rasmus_s_cleantomatoseq.fasta). Duarte et al. (2010) identified a set of 959 APVO SSC genes shared broadly in angiosperms by comparing genome sequences of three eudicots (Arabidopsis, Populus, and Vitis) and one monocot (Oryza). Yuan et al. (2009) identified 127 loci in the PPR gene family that are single-copy and intronless in both A. thaliana and Oryza sativa L. Coding sequences for APVO SSC and PPR loci in A. thaliana were downloaded from The Arabidopsis Information Resource (www.arabidopsis.org). Sequences in B. globoosa for loci in each of the three locus sets were compiled by conducting BLASTN searches of sequences from Solanum or Arabidopsis against the assembled B. globoosa contigs. Up to five of the top hits with a bit score greater than 70 were retained. For some loci, there were hits that overlapped from different contigs, and these were assembled using de novo assembly in Geneious v9.1.6 (Biomatters, Auckland, New Zealand). If hits in an assembly had pairwise identity <95%, all hits matching that locus were removed because this was inferred to be evidence of the presence of paralogs. Remaining hits and consensus sequences were filtered by length (individual sequence >120 bp, total length of all sequences for a locus >600 bp).

Filtered target sequences from all four locus sets were combined. For any sequences with >90% sequence similarity to another, only the longest sequence was retained as the target sequence using cd-hit-est (Li and Godzik, 2006). Some sequences from different locus sets had significant overlap and were assumed to be from the same locus. These were assembled using de novo assembly in Geneious v9.1.6 if pairwise identity >95%, and consensus sequences were used as the target sequences.

One probe set targeting all four locus sets was designed and manufactured by RAPID Genomics (Gainesville, Florida, USA). Biotinylated RNA probes were 120 bp with 2× tiling density over target sequences. Additional checks were performed to eliminate probes targeting multi-copy loci: probes with more than 10 hits to the assembled B. globoosa genome or with more than 100 matching raw B. globoosa reads were discarded.

### Taxon sampling for targeted sequence capture

Fifty samples were chosen for targeted sequence capture and sequencing (Appendix 1). We sampled 46 species of Buddleja, including 21 of 24 species in section Alternifoliiæ, which had poor resolution in previous phylogenetic analyses with a seven-locus data set (Chau et al., 2017), and at least one representative from each of the other sections of Buddleja. Additionally, we tested the performance of our probes in more distantly related outgroups. We included Teedia lucida Rudolphi (Scrophulariaceae), a member of the sister group to Buddleja; Scrophularia nodosa L. (Scrophulariaceae), in the same family as Buddleja; and Parmentiera aculeata (Kunth) Seem. (Bignoniacaeae) and Lantana leonardiorum Moldenke (Verbenaceae), in the same order (Lamiales) as Buddleja. We also wanted to examine the effectiveness of this method for museum samples, so we included eight samples with DNA extracted from herbarium specimens.

### DNA extraction, sequence capture, and sequencing

DNA was extracted from dried leaf tissue, either preserved in silica gel or from a herbarium specimen, using a modified CTAB protocol and purified by isopropanol precipitation. DNA was run on 1% agarose gels to assess DNA quality. DNA concentration was measured with a Qubit (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Samples were diluted or concentrated to attain a concentration of 50 ng/μL, where possible, although some samples had concentrations as low as 2 ng/μL. Volumes of 35 to 50 μL were submitted for library preparation.

Library preparation, sequence capture, and sequencing (Capture-Seq) were done by RAPID Genomics. For each sample, 250 to 1000 ng of genomic DNA, where available, was fragmented to a target size of 400 bp. DNA from herbarium specimens was not additionally fragmented if gel images showed that the DNA was already degraded. DNA libraries were constructed by end-repairing the sheared DNA, A-tailing and adapter ligation, barcoding, and PCR amplification. Libraries were pooled in equimolar ratios by ploidy of species, and probes were hybridized to pools to enrich for targets. Enriched pools were combined in equimolar ratios for sequencing, and 100-bp paired-end reads were sequenced on ~16% of one lane of an Illumina HiSeq 3000 (Illumina).

### Read processing and assembly

De-multiplexed reads were obtained from RAPID Genomics. Sequence quality was checked using FastQC v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc). Using modified scripts from the pipeline SqCL (https://github.com/singhal/SqCL), Trimmomatic 0.36 (Bolger et al., 2014) was used to remove adapters, barcodes, and poor quality bases using the setting LEADING:20 TRAILING:20 SLIDINGWINDOW:5:20 MINLEN:36.

Remaining paired and unpaired reads were assembled using the pipeline HybPiper (Johnson et al., 2016). Briefly, the reads_first.py script sorts reads by target sequence for each sample using the Burrows–Wheeler Alignment tool (BWA) (Li and Durbin, 2009), assemblies mapped reads for each target sequence using the assembler SPAdes (Bankevich et al., 2012), and finally extracts the coding sequence from the assembled contig using Exonerate (Slater and Birney, 2005). For the target reference sequences used by BWA, we used our target sequences for probe design.

Assembled coding sequences from each sample were compiled and sorted by target sequence using the HybPiper script retrieve_sequences.py. If multiple long-length contigs were assembled by SPAdes for a target sequence for a sample, one contig was chosen based on higher sequencing coverage depth or higher percent identity to the reference sequence. Data on lengths of assembled coding sequences for each target sequence and statistics on assembly efficiency were calculated using the scripts get_seq_lengths.py and hybpiper_stats.py. Multiple
long-length contigs for a target sequence for a sample, which may represent paralogs, were identified using the script paralog_investigator.py.

After removing two samples with failed sequencing, differences among locus sets for Buddleja samples in the average percentage of target sequences with an assembled sequence and in the average percentage of the total target length assembled were evaluated with one-way analysis of variance tests blocked by sample and Tukey multiple comparison tests conducted in R (R Core Team, 2015).

**Phylogenetic analyses**

Any samples with assembled coding sequences for less than 50% of target sequences were excluded from further analyses. A custom script (Appendix S1) was used to filter out loci with missing data or with paralogs in any remaining sample.

For each sequence set, sequences were aligned with MAFFT (Katoh and Standley, 2013) using default parameters. Sites with more than 50% missing data were removed from alignments using the “clean” function in Phyutility (Smith and Dunn, 2008). Concatenated alignments were generated for each of the four locus sets using the “concat” function in Phyutility.

Because a target locus might comprise multiple target sequences, we created concatenated alignments for each locus using the “concat” function in Phyutility. The percentage of identical sites was calculated for each locus in Geneious v9.1.6 (Biomatters). Differences among locus sets in the average percentage of identical sites were evaluated with one-way analysis of variance tests and Tukey multiple comparison tests in R.

Maximum likelihood (ML) trees were inferred for the concatenated alignments for each locus set using RAxML v8.0.7 (Stamatakis, 2014). We searched for the best-scoring ML tree and conducted 100 rapid bootstraps. Data sets were unpartitioned, and the multispecies coalescent tree model was used. Ambiguities collected using Quartet FM (QFM) quartet assembly (Raez et al., 2014), (Swofford, 2003). All possible quartets were evaluated. Trees were selected using singular value decomposition scores for species quartets (SVDquartets) (Chifman and Kubatko, 2014) in PAUP* v4.0a157 (Swofford, 2003). The GTR + gamma model of rate heterogeneity was used for each analysis, 100 bootstraps were performed.

**RESULTS**

**Whole-genome shotgun sequencing of Buddleja globosa**

One lane of sequencing produced 292,788,924 100-bp paired reads. After filtering, 257,538,046 reads (88%) remained and were used in the de novo assembly. Of these, 246,968,396 reads (84.4%) were assembled into 311,304 contigs that had a total length of 343,339,138 bp. Contigs ranged in length from 118 to 166,512 bp and had an N50 of 2390 bp. Raw reads are available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (BioProject ID PRJNA419550, SRA SRP125846). Assembled contigs are available in the Dryad Digital Repository (available at https://doi.org/10.5061/dryad.v6q0p; Chau et al., 2018).

**Locus sets for targeted sequence capture**

In total, 2906 target sequences from 1049 loci with a total length of 1,010,028 bp were submitted for probe design (Table 1). Of these, 1880 target sequences from 708 loci with a total length of 580,437 bp were identified in **Buddleja** using the taxon-specific approach. The remaining 1026 target sequences from 341 loci with a total length of 429,591 bp were in the three general locus sets. There were 67 COSII loci, 162 APVO SSC loci, and 112 PPR loci. The average target locus length was higher in the three general sets (COSII: 1119 bp, APVO SSC: 1079 bp, PPR: 1605 bp) than in the taxon-specific set (820 bp). Target locus sequences are available in the Dryad Digital Repository (https://doi.org/10.5061/dryad.v6q0p; Chau et al., 2018).

**DNA quality, sequence capture, and sequencing**

All but one of our samples (**B. rinchonensis** (Mayfield)) J. H. Chau) had at least 250 ng of DNA for library preparation (Appendix 2). DNA from silica gel–preserved tissues were of high molecular weight. DNA from herbarium specimen tissue varied in quality, but most were degraded (Appendix S2).

Of our 50 samples, 48 were successfully sequenced (Appendix S2). One sample (**B. rinchonensis**) had a very low amount of DNA available and produced no mapped reads. Another sample (**B. macrostachya** Benth.) had sufficient starting DNA, but sequencing failed for unknown reasons. For the 48 remaining samples, between 372,898 and 4,963,618 paired reads were produced (Appendix 2). There were no issues with read quality when checked in FastQC. On average, 96% of reads were retained after trimming for low-quality bases and adapter and barcode sequences. Raw reads are available in the NCBI Sequence Read Archive (BioProject ID PRJNA419999, SRA SRP125765).

**Read assembly**

For each sample, between 44% and 49% of total reads were mapped to the target sequences (Appendix 2). For species of **Buddleja**, the HybPiper pipeline produced assembled coding sequences for 91% to 99% of all target sequences. For outgroups, the number of target sequences with assemblies decreased with increasing phylogenetic resolving power.

**TABLE 1.** Characteristics of target locus sets for probe design.

| Locus sets | Total target sequencesa | Total target locib | Total target length (bp) | Average target sequence length (bp) | Average target locus length (bp) |
|------------|-------------------------|-------------------|--------------------------|------------------------------------|---------------------------------|
| Taxon-specific | 1880                   | 708               | 580,437                  | 309                                | 820                             |
| General     | 1034                   | 344               | 431,226                  | 419                                | 1260                            |
| COSII       | 280                    | 67                | 74,988                   | 268                                | 1119                            |
| APVO SSC    | 572                    | 162               | 174,848                  | 306                                | 1079                            |
| PPR         | 174                    | 112               | 179,755                  | 1033                               | 1605                            |
| Total       | 2906                   | 1049              | 1,010,028                | 348                                | 963                             |

aA target sequence is a single consensus sequence from reads mapped to a target locus. There may be multiple target sequences for a target locus if sequences do not overlap.
bA target locus is from a single transcript (taxon-specific approach) or a single gene (general approach).
distance from *Buddleja*. For other Scrophulariaceae, 76% to 90% of target sequences had assemblies. For other families, 24% to 47% of target sequences had assemblies.

All locus sets had high assembly success in *Buddleja* samples, with an average of 94% to 98% of target sequences with assemblies (Table 2). Significant differences between locus sets were found in the average percentage of target sequences with assemblies ($P < 0.01$) and the average percentage of total target length assembled ($P < 0.01$). The PPR and taxon-specific locus sets had the highest success, with an average of 98% of target sequences having an assembly for both data sets. The COSII locus set had the lowest success, with an average of 94% of sequences having an assembly.

For the four outgroup samples, there was a larger difference between assembly of taxon-specific and general locus sets, with greater assembly success in general locus sets (Table 2). On average, 50% of taxon-specific target sequences had assemblies, whereas 65% to 74% of target sequences in general locus sets had assemblies.

Samples from herbarium specimens performed similarly to samples from silica-preserved tissue (Appendices 1 and 2). The average percentage of targets with assembled sequences was 96.1% in our seven samples from herbarium specimens with adequate DNA for library preparation versus 97.6% in samples from silica-preserved tissue.

**Sequence filtering for phylogenetic analyses**

We used 44 samples in *Buddleja* and two outgroups in Scrophulariaceae with high assembly success in phylogenetic analyses. Of the 2906 target sequences, 1524 did not have an assembled coding sequence for at least one sample and 538 had paralogous coding sequences for at least one sample. These were removed from further analyses under our filtering criteria. Thus, 800 taxon-specific target sequences and 400 general target sequences remained for phylogenetic analyses (Table 3).

The PPR locus set had the largest percentage of target sequences (58%) remaining after filtering, whereas the COSII locus set had the smallest percentage (29%). The taxon-specific locus set had an intermediate percentage of target sequences (43%) retained, although the absolute number of taxon-specific target loci (511) was higher than in any of the three general sets (50–128). The PPR locus set had the longest average length of filtered assembled loci (1453 bp), whereas the taxon-specific locus set had the shortest average length (526 bp).

The PPR and taxon-specific loci showed the greatest sequence variability (i.e., lowest percentage of identical sites). Variable sites comprised 35.17% and 36.07%, respectively, of the locus sequences on average, which was significantly higher than the average percentages for the other two general locus sets ($P < 0.01$). The COSII loci had the lowest percentage of variable sites (27.96%).

The trimmed concatenated alignments had a total length of 268,710 bp for the taxon-specific locus set and 28,332 to 120,635 bp for the general locus sets. Concatenated alignments for each locus set are available in the Dryad Digital Repository (https://doi.org/10.5061/dryad.v6q0p; Chau et al., 2018) and TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S21931).

**Phylogenetic analyses**

ML analyses with different locus sets produced trees with different levels of node support (Appendix 3). The tree from the taxon-specific data set was substantially more well-supported, with 93% of nodes with bootstrap support (BP) ≥ 90% (Fig. 1). Trees from general data sets had 61% to 75% of nodes with BP ≥ 90%, with the tree from COSII sequences being the least well-supported. In section *Alternifoliae*, 100% of nodes had BP ≥ 90% with the taxon-specific data set, and 78% to 89% of nodes had such high support with general data sets. Among ML trees from different locus sets, there were several well-supported topological differences, including

**TABLE 2.** Average performance of locus sets in assembly for 44 *Buddleja* samples (excludes two *Buddleja* samples with failed sequencing) and four outgroup samples.*

| Locus sets          | No. of sequences | Total length | No. of sequences | Total length |
|---------------------|------------------|--------------|------------------|--------------|
| Taxon-specific      | 1845 (98%)       | 567,161 (98%)| 992 (53%)       | 287,344 (50%)|
| General             | 1984 (95%)       | 421,307 (98%)| 733 (71%)       | 312,062 (72%)|
| COSII               | 264 (94%)        | 72,207 (96%) | 184 (66%)       | 48,390 (65%) |
| APVO SSC            | 548 (96%)        | 170,224 (74%)| 425 (74%)       | 130,143 (74%)|
| PPR                 | 171 (98%)        | 178,876 (100%)| 124 (71%)     | 133,524 (74%)|
| Total               | 2829 (97%)       | 988,468 (98%)| 1724 (59%)     | 599,405 (59%)|

*Shown are the average number of target sequences with assembled coding sequence and the average total length of assembled coding sequences. In parentheses are the percentages of total target sequences used for probe design. Superscript letters show significant differences in averages at the 0.05 level among locus sets for Tukey multiple comparison tests with blocking by sample.

**TABLE 3.** Characteristics of assembled sequence data sets used for phylogenetic analyses.*

| Locus sets          | Total sequences | Total loci | Average sequence length (bp) | Average locus length (bp) | Average total length: unaligned (bp) | Total length: aligned, trimmed (bp) | Average % variable sites |
|---------------------|-----------------|------------|-----------------------------|---------------------------|-------------------------------------|-----------------------------------|------------------------|
| Taxon-specific      | 800 (43%)       | 511 (72%)  | 336                         | 526                       | 266,710 (46%)                      | 268,603 (46%)                   | 36.07%                 |
| General             | 400 (39%)       | 261 (76%)  | 605                         | 928                       | 242,161 (56%)                      | 242,359 (56%)                   | 30.55%                 |
| COSII               | 82 (29%)        | 50 (75%)   | 346                         | 567                       | 28,332 (38%)                       | 28,380 (38%)                    | 27.96%                 |
| APVO SSC            | 217 (38%)       | 128 (79%)  | 429                         | 728                       | 93,194 (53%)                       | 93,253 (53%)                    | 28.56%                 |
| PPR                 | 101 (58%)       | 83 (74%)   | 1194                        | 1453                      | 120,635 (67%)                      | 120,726 (67%)                   | 35.17%                 |
| Total               | 1200 (41%)      | 772 (74%)  | 425                         | 661                       | 510,579 (51%)                      | 510,962 (51%)                   | 34.20%                 |

*Superscript letters show significant differences in averages at the 0.05 level among the 0.05 level among loci sets for Tukey multiple comparison tests.

A sequence is assembled to a single target sequence. There may be multiple target sequences for a target locus if target sequences do not overlap.
in the positions of species *B. asiatica, B. alternifolia, B. crispa*, and *B. myriantha* in section *Alternifoliae*.

The SVDQuartets analyses produced trees with similar topologies to the ML analyses but with less support at the nodes (Appendix 4). Support varied from 65% of nodes with BP ≥ 90% in the tree from PPR sequences to 47% of nodes with BP ≥ 90% in the tree from COSII sequences. Topological incongruencies between trees from ML and SVDQuartets analyses generally occurred at nodes weakly supported in the SVDQuartets trees. All tree files are available in TreeBASE (http://purl.org/phylo/treebase/phylostudy/TV2:S21931).

**DISCUSSION**

**Comparison of locus sets**

We were able to develop a substantially greater number of taxon-specific target loci than general target loci. This was not unexpected because closely related species, which were used to develop the taxon-specific locus set, are expected to share more loci than distantly related species, which were used to develop the general locus sets.

Recovery of assembled coding sequences for our target sequences was high overall. In *Buddleja*, no sample had less than 80% of target sequences with an assembled sequence for any locus set. Although statistically significant differences were found among average assembly success in different locus sets, the difference was not large, ranging from 94% to 98%. Outgroups showed a different pattern, with general locus sets consistently outperforming the taxon-specific locus set. This pattern is consistent with the fact that the taxon-specific locus set was designed using genomic resources in *Buddleja*, thus it is unknown whether these loci are single-copy or taxon-specific. Values at nodes indicate bootstrap support. (A) indicates node subtending clade of section *Alternifoliae*.

![Maximum likelihood phylogram from RAxML analysis with concatenated sequences from taxon-specific locus set](http://www.wileyonlinelibrary.com/journal/AppsPlantSci)

**FIGURE 1.** Maximum likelihood phylogram from RAxML analysis with concatenated sequences from taxon-specific locus set. Values at nodes indicate bootstrap support. (A) indicates node subtending clade of section *Alternifoliae*.

**Recommendations for use of targeted sequence capture in plant phylogenomics**

In groups where multiple genomic resources do not exist to design a taxon-specific locus set for sequence capture, using general locus sets is a suitable alternative. In our ingroup, both general and taxon-specific locus sets had high assembly success. In our outgroups, recovery of sequences was higher in general locus sets than in the taxon-specific locus set. Information content of general locus sets is a suitable alternative. In our ingroup, both general and taxon-specific locus sets had high assembly success. In our outgroups, recovery of sequences was higher in general locus sets than in the taxon-specific locus set. Information content of general locus sets is a suitable alternative.
of loci can potentially still be targeted, which may be sufficient to resolve relationships (Leaché and Rannala, 2011).

Even in groups where taxon-specific locus sets can be designed, researchers may consider adding general locus sets to their sequence capture targets. In addition to having greater or comparable assembly efficiency and information content per locus, utilizing general locus sets can facilitate the combination of data from different studies because known loci can be used in different taxa. In particular, the PPR loci have other characteristics that make them desirable for phylogenetic analysis, including a lack of introns that facilitates unambiguous alignment (Yuan et al., 2009). The PPR loci, with their greater average length and high proportion of variable sites, are also suited to the inference of well-supported gene trees, which are necessary for a number of species tree methods (e.g., Accurate Species TRee AlgoritHm [ASTRAL]; Mirarab and Warnow, 2015).

Designing a general target locus set for a group does not require multiple genomic resources, but some source of genomic sequence data is still necessary to design probes whose sequences will adequately complement targets in the group of interest. The most closely related taxon whose genome is available should be used, but in our study, probes designed for a different genus or different family from a sample were still able to recover 56% to 95% of sequences in general locus sets given adequate starting DNA. Many genomic resources for plants are now publicly available for probe design, including genomes (e.g., Phytozome; https://phytozome.jgi.doe.gov/pz/portal.html) and transcriptomes (e.g., 1KP initiative; https://sites.google.com/a/ualberta.ca/onekp/).

Targeted sequence capture is a suitable method even for samples from herbarium specimens or those that otherwise have degraded DNA. In our study, sequence recovery was nearly the same in our herbarium samples as in our silica-preserved samples. For several of these herbarium samples, PCR amplification of low-copy nuclear loci had not been successful (unpublished data), but the targeted sequence capture method generated large amounts of sequence data suitable for phylogenetic analyses.

Assembly of sequencing reads can be accomplished with a number of different programs and pipelines. Although HybPiper successfully generated assembled coding sequences for the vast majority of target sequences, it did not assemble separate contigs for paralogs of target sequences where they were expected to occur in polyploid species (data not shown). For groups where polyploidization or hybridization are important parts of the evolutionary history, testing of other assembly methods is suggested.

CONCLUSIONS

We show in this study that targeted sequence capture using general or taxon-specific locus sets is an effective method for gathering sequence data for phylogenetic studies that can significantly increase resolution and support of relationships versus non-sequence capture data sets consisting of only several loci. General target loci are simpler and require fewer resources to develop, but can be as effective as taxon-specific loci in terms of assembly success and phylogenetic informativeness. Applying general locus sets widens the opportunity to use targeted sequence capture in more plant groups with few genomic resources.

ACKNOWLEDGMENTS

The authors thank A. Leaché, T. Gill, and J. Grummer for technical assistance in preparing genomic libraries, the use of laboratory and computing equipment, and advice on genomic laboratory techniques and phylogenomic analyses; P. Lu-Irving for spearheading the initial foray into genomic work; M. Deyholos, D. Soltis, and the 1KP initiative for providing access to transcriptome data; and funding from the National Science Foundation (no. DEB-1311111 [Doctoral Dissertation Improvement Grant] and no. DEB-1020369).

DATA ACCESSIBILITY

Raw reads from whole genome shotgun sequencing of Buddleja globosa are available in the NCBI Sequence Read Archive, BioProject ID PRJNA419550, SRA SRP125846 (https://www.ncbi.nlm.nih.gov/sra/?term=SRP125846). Assembled contigs from whole genome shotgun sequencing of Buddleja globosa are available on Dryad Digital Repository (https://doi.org/10.5061/dryad.v6q0p). Target locus sequences in Buddleja globosa used for probe design for targeted sequence capture are available on Dryad Digital Repository (https://doi.org/10.5061/dryad.v6q0p). Raw reads from sequencing after targeted sequence capture for 50 samples are available in the NCBI Sequence Read Archive, BioProject ID PRJNA419999, SRA SRP125765 (https://www.ncbi.nlm.nih.gov/sra/?term=SRP125765). Concatenated alignments for each target locus set are available on Dryad Digital Repository (https://doi.org/10.5061/dryad.v6q0p) and TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S21931). Tree files from maximum likelihood and SVDquartets analyses are available on TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S21931).

SUPPORTING INFORMATION

Additional Supporting Information (Appendices S1 and S2) may be found online in the supporting information tab for this article.

LITERATURE CITED

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, et al. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19: 455–477.

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

Buddenhagen, C., A. R. Lemmon, E. M. Lemmon, J. Bruhl, J. Cappa, W. L. Clement, M. J. Donoghue, et al. 2016. Anchored phylogenomics of angiosperms 1: Assessing the robustness of phylogenetic estimates. bioRxiv https://doi.org/10.1101/086298.

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

Chapman, M. A., J. C. Chang, D. Weisman, R. V. Kesseli, and J. M. Burke. 2007. Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae). Theoretical and Applied Genetics 115: 747–755.

Chau, J. H. 2017. Systematics of Buddleja (Scrophulariaceae): Phylogenetic relationships, historical biogeography, and phylogenomics. Ph.D. dissertation, University of Washington, Seattle, Washington, USA.

Chau, J. H., N. O’Leary, W.-B. Sun, and R. G. Olmstead. 2017. Phylogenetic relationships in tribe Buddleeae (Scrophulariaceae) based on multiple nuclear and plastid markers. Botanical Journal of the Linnean Society 184: 137–166.

Chau, J. H., W. A. Rahnfeldt, and R. G. Olmstead. 2018. Data from: Comparison of taxon-specific versus general locus sets for targeted sequence capture in Buddleja globosa. Applications in Plant Sciences 2018 6(3): e1032 Chau et al.—Sequence capture locus sets • 7 of 14

http://www.wileyonlinelibrary.com/journal/AppsPlantSci © 2018 Chau et al.
plant phylogenomics. Dryad Digital Repository https://doi.org/10.5061/dryad.v6q0p.

Chifman, J., and L. Kubatko. 2014. Quartet inference from SNP data under the coalescent. Bioinformatics 30: 3317–3324.

Cronn, R., B. J. Knaus, A. Liston, P. J. Maughan, M. Parks, J. V. Syring, and J. Udall. 2012. Targeted enrichment strategies for next-generation plant biology. American Journal of Botany 99: 291–311.

Curto, M. A., P. Puppo, D. Ferreria, M. Nogueira, and H. Meimberg. 2012. Development of phylogenetic markers from single-copy nuclear genes for multi locus, species level analyses in the mint family (Lamiaceae). Molecular Phylogenetics and Evolution 63: 758–767.

Degnan, J. H., and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology and Evolution 24: 332–340.

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11–15.

Duarte, J. M., P. K. Wall, P. P. Edger, L. L. Landherr, H. Ma, J. C. Pires, J. Leebens-Mack, et al. 2010. Identification of shared single copy nuclear genes in Arabidopsis, Populus, Vitis and Oryza and their phylogenetic utility across various taxonomic levels. BMC Evolutionary Biology 10: 61.

Egan, A. N., J. Slueter, and D. M. Spooner. 2012. Applications of next-generation sequencing in plant biology. American Journal of Botany 99: 175–185.

Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield, A. N. Egan, J. Schlueter, and D. M. Spooner. 2012. Applications of next-generation enrichment from transcriptomic and genome skim data: The pipeline and its application in southern African Oxalis (Oxalidaceae). Molecular Ecology Resources 16: 1124–1135.

Hedges, E. E., R. J. Eastwood, and C. D. Bailey. 2006. From famine to feast: Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. Philosophical Transactions of the Royal Society B 361: 211–225.

Johnson, M. G., E. M. Gardner, Y. Liu, R. Medina, B. Goffinet, A. J. Shaw, N. J. C. Zerega, et al. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Applications in Plant Sciences 4: 1600016.

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.

Leaché, A. D., and B. Rannala. 2011. The accuracy of species tree estimation under simulation: A comparison of methods. Systematic Biology 60: 126–137.

Maddison, W. P. 1997. Gene trees in species trees. Systematic Biology 46: 523–536.

Maddron, A. A., A. J. Cortés, and J. E. Richardson. 2013. Páramo is the world’s fastest evolving and coolest biodiversity hotspot. Frontiers in Genetics 4: 192.

Magallon, S., S. Gómez-Acevedo, L. L. Sánchez-Reyes, and T. Hernández-Hernández. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytologist 207: 437–453.

Mamanova, L., A. I. Coffey, C. E. Scott, I. Kozarewa, E. H. Turner, A. Kumar, E. Howard, et al. 2010. Target-enrichment strategies for next-generation sequencing, Nature Methods 7: 111–118.

Mandel, J. R., R. B. Dikow, A. F. Funk, R. R. Masala, S. E. Staton, A. Kozik, R. W. Michelmore, et al. 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. Applications in Plant Sciences 2: 1300085.

McMordie, J. E., S. M. Hird, A. J. Zellmer, B. C. Carstens, and R. T. Brunfield. 2013. Applications of next-generation sequencing to phycogeography and phylogenetics. Molecular Phylogenetics and Evolution 66: 526–538.

Miranar, S., and T. Warnow. 2015. ASTRA-LI: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31:144–152.

Nicholls, J. A., R. T. Pennington, E. J. M. Koenen, C. E. Hughes, J. Hearrn, L. Bunnefeld, K. G. Dexter, et al. 2015. Using targeted enrichment of nuclear genes to infer phylogenetic resolution in the neotropical rain forest genus Inga (Leguminosae: Mimosoideae). Frontiers in Plant Science 6: 710.

R Core Team. 2015. R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org/ [accessed 26 May 2017].

Raz, R., M. S. Bayzid, and M. S. Rahman. 2014. Accurate phylogenetic tree reconstruction from quartets: A heuristic approach. PLoS One 9: e104008.

Reneker, J., E. Lyons, G. C. Conant, J. C. Pires, M. Freeing, C.-R. Shyu, and D. Korkin. 2012. Long identical multispecies elements in plant and animal genomes. Proceedings of the National Academy of Sciences USA 109: E1183–E1191.

Rokas, A., B. L. Williams, N. King, and S. B. Carroll. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenetics. Nature 425: 790–804.

Schlueter, D. 2000. The ecology of adaptive radiation. Oxford University Press, New York, New York, USA.

Schmickl, R., A. Liston, V. Zeisek, K. Oberlander, K. Weitemier, S. C. K. Straub, R. C. Cronn, et al. 2016. Phylogenetic marker development for target enrichment from transcriptomic and genome skin data: The pipeline and its application in southern African Oxalis (Oxalidaceae). Molecular Ecology Resources 16: 1124–1135.

Slater, G., and E. Birney. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6: 31.

Small, R. L., R. C. Cronn, and J. F. Wendel. 2004. Use of nuclear genes for phylogeny reconstruction in plants. Australian Systematic Botany 17: 145–170.

Smith, S. A., and C. W. Dunn. 2008. Phyutility: A phyloinformatics tool for trees, alignments and molecular data. Bioinformatics 24: 715–716.

Soltis, D. E., M. A. Gitzendanner, G. Stull, M. Chester, A. Chandlerbali, S. Chamala, I. Jordon-Thaden, et al. 2013. The potential of genomes in plant systematics. Taxon 62: 886–898.

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312–1313.

Stephens, J. D., W. L. Rogers, C. M. Mason, L. A. Donovan, and R. L. Malmberg. 2015. Species tree estimation of diploid Helianthus (Asteraceae) using target enrichment. American Journal of Botany 102: 910–920.

Swoford, D. L. 2003.PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts, USA.

Twyford, A. D., and R. A. Ennos. 2012. Next-generation sequencing as a tool for plant ecology and evolution. Plant Ecology and Diversity 5: 411–413.

Weitemier, K., S. C. K. Straub, R. C. Cronn, M. Fishheim, R. Schmickl, A. McDonnell, and A. Liston. 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenetics. Applications in Plant Sciences 2: 1400042.

Wu, F. L., A. Mueller, D. Crouzillat, V. Périard, and S. D. Tanksley. 2006. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative evolutionary and systematic studies: A test case in the eusterid plant clade. Genetics 174: 1407–1420.
**APPENDIX 1.** Specimens used for targeted sequence capture, with voucher information, infrageneric or familial classification, and sample source (herbarium specimen or silica gel preserved).

| Species | Voucher (Herbarium)* | Section in *Buddleja*, or family if outside *Buddleja* | Herbarium sample? |
|---------|-----------------------|-------------------------------------------------------|------------------|
| *Buddleja alboflora* Hemsl. | J. Chau 260 (WTU, A) | Alternifoliae | No |
| *Buddleja alternifolia* Maxim. | J. Chau 262 (WTU, A) | Alternifoliae | No |
| *Buddleja americana* L. | L. Frost 148 (WTU) | *Buddleja* | No |
| *Buddleja arnoldii* Kunze | J. Chau 224 (WTU, LPB) | *Buddleja* | No |
| *Buddleja aromatica* J. Rémy | J. Chau 206 (WTU, LPB) | *Buddleja* | No |
| *Buddleja asiatica* Loc. | J. Chau 157 (WTU) | Alternifoliae | No |
| *Buddleja auriculata* Bentham. | J. Chau 246 (WTU) | Chilianthus | No |
| *Buddleja blattaria* J. F. Macbr. | J. Chau 101 (WTU) | *Buddleja* | No |
| *Buddleja brachystachya* Diels | (KUN 22547)* | Alternifoliae | Yes |
| *Buddleja caryophyloides* W. W. Sm. | J. Chau 171 (WTU) | Alternifoliae | No |
| *Buddleja colvillii* Hook. f. | J. Chau 42 (WTU) | Alternifoliae | No |
| *Buddleja coriacea* J. Rémy | J. Chau 194 (WTU, LPB) | *Buddleja* | No |
| *Buddleja crispa* Bentham. | J. Chau 170 (WTU) | Alternifoliae | No |
| *Buddleja curviflora* Hook. & Arn. | R. Olmstead 2010-49 (WTU) | Alternifoliae | No |
| *Buddleja davidi* Franch. | J. Chau 177 (WTU) | Alternifoliae | No |
| *Buddleja delavayi* L. F. Gagnep. | J. Chau 165 (WTU) | Alternifoliae | No |
| *Buddleja dyssphylla* (Benth.) Radlk. | J. Chau 233 (WTU) | Chilianthus | No |
| *Buddleja elegans* Champ. & Schult. subsp. elegans | R. Olmstead 2010-214 (ICN) | *Buddleja* | No |
| *Buddleja fallowiana* Bal. f. & W. W. Sm. | J. Chau 166 (WTU) | Alternifoliae | No |
| *Buddleja forestii* Diels | J. Chau 161 (WTU) | Alternifoliae | No |
| *Buddleja gloriamorata* H. Wendl. | J. Chau 254 (WTU) | incertae sedis | No |
| *Buddleja interrupta* Kunth | J. Chau 123 (WTU) | *Buddleja* | No |
| *Buddleja japonica* Hemsl. | J. Wood 124-2014 (A) | Alternifoliae | No |
| *Buddleja lindleyana* Fortune | J. Wood & K. Richardson 125-2014 (A) | Alternifoliae | No |
| *Buddleja lucioceata* Leeuwenberg | J. Chau 253 (WTU) | Chilianthus | No |
| *Buddleja macrostachya* Bentham. | J. Chau 159 (WTU) | Alternifoliae | No |
| *Buddleja madagascariensis* Lam. | J. Chau 256 (WTU) | Nicodemia | No |
| *Buddleja manubribifolia* Bentham. | M. Moore 1567 (WTU, MEXU) | *Buddleja* | No |
| *Buddleja microstachya* E. D. Liu & H. Peng | E. Liu 925 (KUN) | Alternifoliae | Yes |
| *Buddleja myriantha* Diels | J. Chau 158 (WTU) | Alternifoliae | No |
| *Buddleja nitida* Bentham. | J. Chau 150 (WTU) | *Buddleja* | No |
| *Buddleja nivea* Duthie | R. Olmstead 2010-47 (WTU) | Alternifoliae | No |
| *Buddleja normaniae* J. H. Chau | D. Riskind 23860 (TEX) | *Buddleja* | Yes |
| *Buddleja officinalis* Maxim. | J. Chau 179 (WTU) | Alternifoliae | No |
| *Buddleja polystachya* Fresen. | G. Simon 308 (MO) | Nicodemia | Yes |
| *Buddleja pulchella* N. E. Br. | I. Nanni 319 (NBG) | Pulchellae | Yes |
| *Buddleja racemosa* Torr. | J. Chau 324 (WTU) | *Buddleja* | No |
| *Buddleja rinconensis* (Mayfield) J. H. Chau | S. Aguilar Ruiz 164 (TEX) | *Buddleja* | Yes |
| *Buddleja saligna* Willd. | J. Chau 231 (WTU) | Chilianthus | No |
| *Buddleja salviifolia* (L.) Lam. | J. Chau 240 (WTU) | Salviifoliae | No |
| *Buddleja sessiliflora* Kunth | G. Webster 31455 (DAV) | *Buddleja* | Yes |
| *Buddleja subcapitata* E. D. Liu & H. Peng | H. Peng 5153 (KUN) | Alternifoliae | Yes |
| *Buddleja tucumanensis* Griseb. | J. Chau 212 (WTU, LPB) | *Buddleja* | No |
| *Buddleja utahensis* Coville | J. Chau 322 (WTU) | *Buddleja* | No |
| *Buddleja virginata* L. f. | J. Chau 180 (WTU) | Gomphostigma | No |
| *Buddleja yunnanensis* L. F. Gagnep. | J. Chau 178 (WTU) | Alternifoliae | No |
| *Teedia lucida* Rudolphi | J. Chau 318 (WTU) | Scorpiulariaceae | No |
| *Scrophularia nodosa* L. | J. Chau 228 (WTU) | Scorpiulariaceae | No |
| *Parmentiera aculeata* (Kunth) Seem. | S. Grose 93 (WTU) | Bignoniaceae | No |
| *Lantana leonardianum* Moldenke | P. Lu-Irving 2012-105 (WTU) | Verbenaceae | No |

*Herbarium acronyms are per Index Herbariorum (http://sweetgum.nybg.org/science/ih/).

*Collector unknown. Herbarium accession number provided.

http://www.wileyonlinelibrary.com/journal/AppsPlantSci © 2018 Chau et al.
## APPENDIX 2. Amount of DNA available for library preparation, sequencing output, and assembly efficiency for each sample used in targeted sequence capture.

| Sample                  | Total DNA (ng) | Total raw reads (% of total) | Total assembled reads (% of total) | Total collapsed reads (% of total) | Total collapsed reads (% of total) | Total assembled coding sequences (% of total) | Total target sequences (% of total) | Total assembled coding sequences (% of total) | Total target sequences (% of total) | Total assembled coding sequences (% of total) | Total target sequences (% of total) |
|-------------------------|----------------|-------------------------------|------------------------------------|------------------------------------|------------------------------------|---------------------------------------------|--------------------------------|---------------------------------------------|--------------------------------|---------------------------------------------|--------------------------------|
APPENDIX 3. Maximum likelihood phylograms from RAxML analyses with concatenated sequences from individual locus sets: (A) taxon-specific, (B) conserved ortholog set (COSII), (C) shared single-copy nuclear (APVO SSC) genes, and (D) pentatricopeptide repeat (PPR) genes. Values at nodes indicate bootstrap support.
**APPENDIX 4.** Best trees from SVDquartets analyses with concatenated sequences from individual locus sets: (A) taxon-specific, (B) conserved ortholog set (COSII), (C) shared single-copy nuclear (APVO SSC) genes, and (D) pentatricopeptide repeat (PPR) genes. Values at nodes indicate bootstrap support.
