American Journal of Botany

RESEARCH ARTICLE

Phylogenomic analyses reveal extensive gene flow within the magic flowers (*Achimenes*)

Wade R. Roberts[1,2,3] and Eric H. Roalson[1,2]

Manuscript received 20 September 2017; revision accepted 2 February 2018.

[1] Molecular Plant Sciences Graduate Program, Washington State University, Pullman, Washington 99164-1030, USA

[2] School of Biological Sciences, Washington State University, Pullman, Washington 99164-4236, USA

[3] Author for correspondence (e-mail: wade.roberts@wsu.edu)

Citation: Roberts, W. R. and E. H. Roalson. 2018. Phylogenomic analyses reveal extensive gene flow within the magic flowers (*Achimenes*). *American Journal of Botany* 105(4): 1–15.

doi:10.1002/ajb2.1058

PREMISE OF THE STUDY: The Neotropical Gesneriaceae is a lineage known for its colorful and diverse flowers, as well as an extensive history of intra- and intergeneric hybridization, particularly among *Achimenes* (the magic flowers) and other members of subtribe Gloxiniinae. Despite numerous studies seeking to elucidate the evolutionary relationships of these lineages, relatively few have sought to infer specific patterns of gene flow despite evidence of widespread hybridization.

METHODS: To explore the utility of phylogenomic data for reassessing phylogenetic relationships and inferring patterns of gene flow among species of *Achimenes*, we sequenced 12 transcriptomes. We used a variety of methods to infer the species tree, examine gene tree discordance, and infer patterns of gene flow.

KEY RESULTS: Phylogenomic analyses resolve clade relationships at the crown of the lineage with strong support. In contrast to previous analyses, we recovered strong support for several new relationships despite a significant amount of gene tree discordance. We present evidence for at least two introgression events between two species pairs that share pollinators, and suggest that the species status of *Achimenes admirabilis* be reexamined.

CONCLUSIONS: Our study demonstrates the utility of transcriptome data for phylogenomic analyses, and inferring patterns of gene flow despite gene tree discordance. Moreover, these data provide another example of prevalent interspecific gene flow among Neotropical plants that share pollinators.

KEY WORDS *Achimenes*; gene flow; Gesneriaceae; hybridization; introgression; phylogenetic networks; phylogenomics; transcriptomics.

Botanists have long suspected a significant role for hybridization in diversification processes (Anderson, 1949; Stebbins, 1950; Grant, 1971), where it serves as both a direct source of novel lineages through hybrid speciation and a potential source of new variation through introgression. Molecular studies from across the tree of life have both confirmed the hybrid origin of many organisms (such as wheat and corn) and have revealed a more extensive history of hybridization than previously documented (Mallet et al., 2016; Roda et al., 2017; Vargas et al., 2017). Interspecific gene flow may contribute to lineage convergence because of introgression, occurring more rapidly among closely related species and resulting in an increase in reproductive isolation with genetic divergence (Coyne and Orr, 1997; Moyle et al., 2004; Scopece et al., 2007). Alternatively, hybridization may promote rapid reproductive isolation of sympatric species due to selection against the formation of unfit hybrid offspring ('reinforcement'), thereby accelerating the process of speciation (Servedio and Noor, 2003). Furthermore, favorable new gene combinations may give rise to new species, which may become instantly reproductively isolated from both parental species (Vereecken et al., 2010). There is a general lack of knowledge about the commonality of different pathways to interspecific gene flow, and it is still uncertain whether hybrids contain approximately equal contributions from each parent or whether hybridization events involve asymmetrical contributions. Newly developed approaches to obtain genome-wide estimates of phylogenetic history will considerably improve our understanding of gene flow in plants. We therefore implemented a transcriptome-sequencing based approach to infer phylogenomic patterns of gene flow within *Achimenes*—one of the most colorful and diverse genera of Central American plants.

Achimenes, commonly known as the "magic flowers," is a genus in the large African violet family (Gesneriaceae) and is well known for its floral diversity among closely related species (Fig. 1; Ramírez Roa, 1987; Roalson et al., 2003; Roberts and Roalson, 2017). Recent divergence time estimates and biogeographic reconstructions have

FIGURE 1. Flowers of the *Achimenes* sampled in the current study. (A) *A. admirabilis,* (B) *A. antirrhina,* (C) *A. candida,* (D) *A. cettoana,* (E) *A. erecta,* (F) *A. grandiflora,* (G) *A. longiflora,* (H) *A. misera,* (I) *A. patens* 'Major' and (J) *A. pedunculata*. All photos by Wade R. Roberts.

indicated a Central American origin for *Achimenes* between 7.7 and 14.2 Mya (Roalson and Roberts, 2016). The genus comprises 26 species found throughout Mexico and Central America, with the center of diversity in central and southern Mexico (Fig. 2A, B; Ramírez Roa, 1987). *Achimenes* contains a mixture of widely distributed species, such as *A. grandiflora* (Schltdl.) DC. and *A. longiflora* DC., and many narrow endemics, such as *A. admirabilis* Wiehler and *A. cettoana* H.E. Moore. Many species overlap in distribution and are often found growing in sympatry. Most species of the genus principally inhabit oak and pine forests, and generally are found in seasonally dry areas where the production of scaly rhizomes and propagules allows for seasonal dormancy (Ramírez Roa, 1987). These scaly rhizomes characterize this genus and others in the subtribe Gloxiniinae and offer an important feature useful for the production and cultivation of these plants.

Enormous diversity in floral form among the many closely related *Achimenes* species represents a feature thought to be associated with speciation (Fig. 1; Ramírez Roa, 1987; Roalson et al., 2003). Evidence from phylogenetic, morphological, and ecological studies offers extensive support for pollination syndromes in Neotropical gesneriads (Perret et al., 2007; Martén-Rodríguez et al., 2009; Martén-Rodríguez et al., 2010; Roalson and Roberts, 2016). In *Achimenes,* species are divided into those with hummingbird-, butterfly-, bee-, and female euglossine bee-pollination (Ramírez Roa, 1987; Roalson et al., 2003). The flowers within each of these groups share similar characters. For example, those with hummingbird-pollination tend to have red, tubular flowers with ample nectar provided as a reward, while those with bee-pollination are small, white, funnelform flowers. The genetic basis for transitions in floral form, including traits such as flower color, thought to be closely associated with pollinators, has been studied extensively in many plant systems (Hoballah et al., 2007; Des Marais and Rausher, 2010; Wessinger and Rausher, 2014; Roberts and Roalson, 2017). Along with shared distribution, habitat preferences, and elevational ranges among *Achimenes*, the high level of floral divergence among closely related species in sympatry might argue for speciation being driven by mechanisms such as pollination, hybridization, or time of flowering (Wiehler, 1983).

Achimenes has a rich history of horticultural interest because of the diversity of flowers found among its members. Interest in these plants peaked during the Victorian Era with many new species being identified in Mexico and brought into cultivation, where many

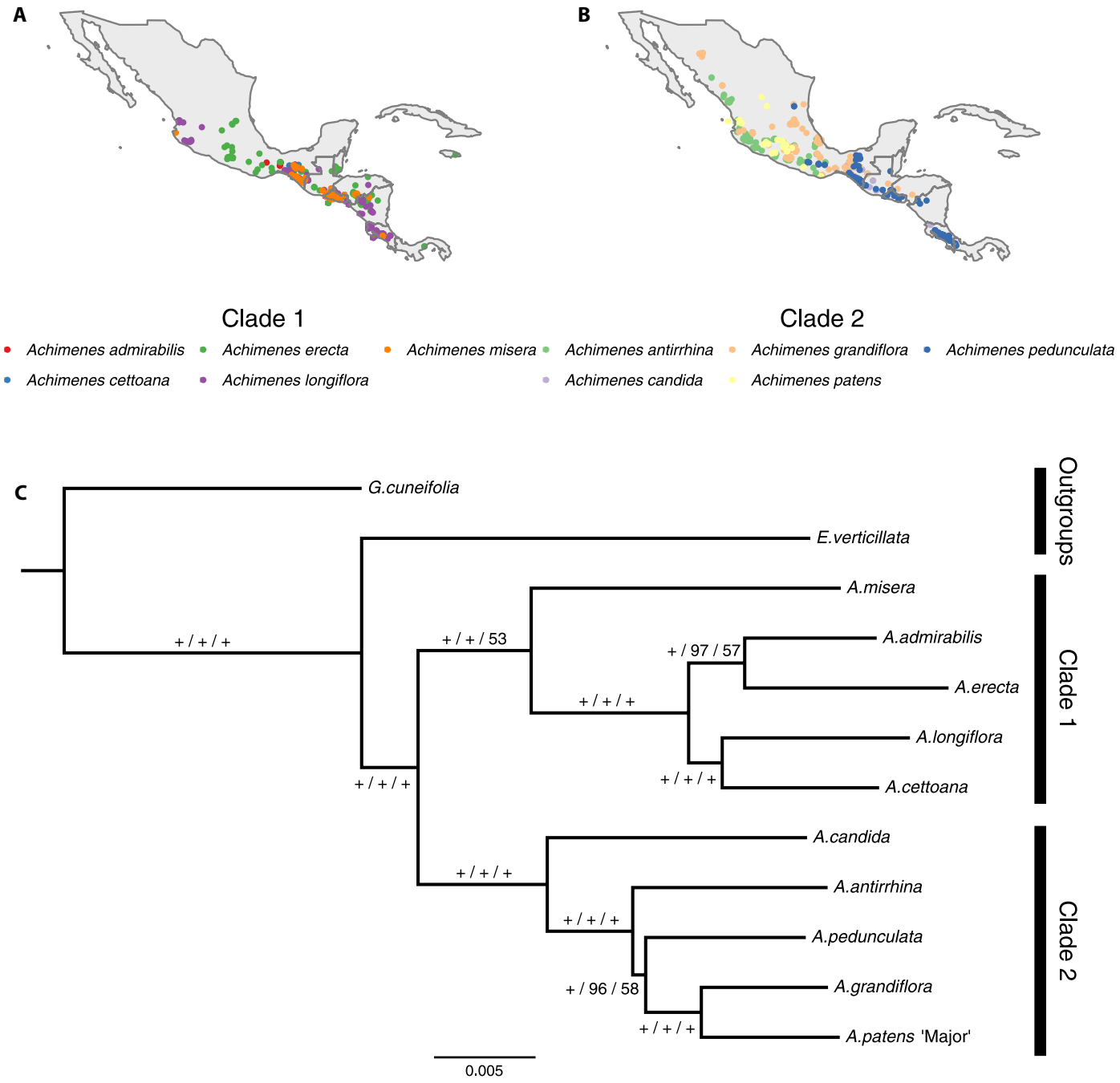

FIGURE 2. Distribution and phylogenetic relationships of *Achimenes*. Range distributions of the species sampled in (A) Clade 1 and (B) Clade 2 of *Achimenes* throughout Mexico and Central America. Location data was downloaded from the Global Biodiversity Information Facility (www.gbif.org). (C) Phylogram summarizing results from phylogenetic analyses of *Achimenes* transcriptomes. *Gesneria cuneifolia* and *Eucodonia verticillata* were used as outgroups. Analyses of concatenated sequences and gene trees converged on the same topology. Branch length units are substitutions per site and are derived from maximum likelihood analysis in RAxML. Numbers at nodes: bootstrap support from RAxML, ASTRAL, and ASTRID. Bootstrap support of 100 is labeled as '+', and support otherwise is listed numerically.

of the earliest hybrid varieties were exhibited at horticultural shows (Gordon, 1846; Moore, 1859). The extraordinary range of desirable colors and shapes in *Achimenes* has provided many growers material to produce well over 200 different hybrids and varieties (Becker, 2008). In addition to the abundance of horticultural hybrids, numerous natural interspecific hybrids among populations of *Achimenes* in Mexico are also known (Wiehler, 1983). Specimens can often display mixed characteristics with other species found in sympatry, particularly with species that share similar pollinators. Given the ease with which hybrids can be produced among *Achimenes*, it is likely that gene flow has occurred more than once during the history of this group.

Only a few attempts have been made to reconstruct a molecular phylogeny of *Achimenes* (Roalson et al., 2003) and no studies have tested specific hypotheses about the role of hybridization in

affecting the evolutionary history of the lineage. Gene flow between closely related, or more distantly related, species often results in a strong conflict between gene trees with the traditional hierarchical (bifurcating) representation of species and is best represented by a reticulation network (McBreen and Lockhart, 2006; Huson et al., 2011; Solís-Lemus and Ané, 2016). Reconstruction of phylogenetic relationships using genome-wide data can be problematic because hybridization is a major cause of topological incongruence between gene trees (McBreen and Lockhart, 2006; Folk et al., 2017). Studying these incongruences offers an opportunity to detect hybrid speciation. However, confounding population genetic processes such as lineage sorting might mislead inference of the real contribution of hybridization to the observed patterns of gene tree incongruence (Linder and Rieseberg, 2004; Kubatko, 2009; Goulet et al., 2017). This is especially common among closely related species where lineage sorting is not complete, leading to nonmonophyletic species assemblages (Schmidt-Lebuhn et al., 2012).

Sampling exemplars of 10 species, including several of the most broadly distributed species and several narrow endemics, our initial goal was to reevaluate the phylogenetic hypothesis of Roalson et al. (2003) using thousands of loci derived from transcriptome sequencing. Recent advancements in methods used to detect gene flow using genomic data now offer new opportunities to understand the process of speciation (Joly et al., 2009; Martin et al., 2015; Rosenzweig et al., 2016). Transcriptome sequencing provides excellent resources for evolutionary studies, but has so far been little used for studies of gene flow and speciation (Roda et al., 2017). Here, we employ transcriptome sequencing from across 10 species of *Achimenes* in three complementary strategies to uncover phylogenetic patterns and detect the presence and direction of gene flow. We hypothesized that some level of gene flow occurred within this group in the past, particularly between *A. grandiflora* and *A. patens* Benth., given their close morphological similarities and sympatric range. First, we used gene trees to estimate phylogenetic networks, which allow reticulation events such as hybridization (Solís-Lemus and Ané, 2016). Second, we investigated patterns of divergence at multiple informative sites using the *D*-statistic (also known as the 'ABBA-BABA' test [Green et al., 2010]). Finally, we investigated variation in gene flow by comparing the relative divergence of genes with discordant gene trees across multiple samples. Our results suggest instances of gene flow and potential introgression among several species of *Achimenes*, while also calling into question the species status of *A. admirabilis*.

MATERIALS AND METHODS

Taxon sampling and tissue collection

We sampled 10 of the 26 currently recognized species of *Achimenes,* including the entirety of Clade 1 and a subset of Clade 2 (sensu Roalson et al., 2003; Fig. 1; Appendix 1). We sampled 5 of 13 species in Clade 2, which included the entire Clade 2a of Roalson et al. (2003) as well as *A. patens*. Those species in Clade 2 not included in the current study are ones that are highly endemic in southern Mexico and not in cultivation. Each of these clades contain some of the most widespread species (e.g., *A. longiflora*) and some narrow endemic species (e.g., *A. admirabilis*; Figs. 2A, B). These 10 species were chosen based on their hypothesized close relationships and their diversity in floral form and pollination syndrome. Additionally, we used the variety *A. patens* 'Major' as our exemplar of *A. patens.* On the basis of previous molecular work with Sanger-sequenced loci (Roalson et al., 2008; Roalson and Roberts, 2016), two species were chosen as outgroups to represent related lineages in the Gesnerieae: *Eucodonia verticillata* (M. Martens & Galeotti) Wiehler and *Gesneria cuneifolia* (DC.) Fritsch (Appendix 1). Plants for all sampled species were grown in standard greenhouse conditions, under 16 h days, 24–27°C, and 80–85% humidity. Flower buds from two developmental stages were sampled: an immature bud stage (Bud) and an intermediate stage (D) pre-anthesis. Three biological replicates (accessions) each were sampled for both Bud and D stages in all species, contributing to a total of six samples for each species. This sampling scheme will allow additional studies to examine gene expression during development as it relates to the evolution of floral form and pollination syndrome. Tissue was collected from plants and immediately flash frozen in liquid nitrogen. Total RNA was extracted from each frozen sample using an RNEasy Plant Kit (Qiagen, Valencia, California, USA) according to the manufacturer's directions. The quality and quantity of the RNA samples were assessed with 1.0% agarose gels and the Fragment Analyzer (Advanced Analytical Technologies, Ankeny, Iowa, USA) at the Washington State University Biotechnology Core Lab (Pullman, Washington, USA). The RNA samples having 28S/18S rRNA ratios approximately 2:1 and RNA Quality Numbers (RQN) ≥8 were found to be high quality and were used for library preparation.

Library preparation and sequencing

Ribosomal depleted RNA samples were prepared using the RiboMinus Plant Kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA), using ≤10 μg total RNA as input, followed by an ethanol precipitation to concentrate the RNA and ensure recovery of smaller (<200 nt, nucleotides) RNA. The ethanol precipitation was performed by adding the following to the ribosomal depleted RNA: 1 μl glycogen (20 μg/μl; Thermo Fisher Scientific, Waltham, Massachusetts, USA), 1/10 sample volume of 3M sodium acetate, and 2.5× sample volume of 100% ethanol. Samples were then incubated at –80°C for 1 hour, followed by centrifuging for 15 min at ≥12,000 × *g*, and washing twice with 70% ethanol and centrifuging for 5 min at ≥12,000 × *g*. The supernatant was removed after each centrifugation step. The resulting ribosomal depleted RNA was eluted into 30 μl nuclease-free water and was quantified using a Qubit 2.0 Fluorometer (RNA HS assays; Thermo Fisher Scientific, Waltham, Massachusetts, USA).

We prepared stranded RNA libraries for the 64 samples using the NEBNext Ultra Directional RNA Library Kit (New England Biolabs, Ipswich, Massachusetts, USA). For each sample, 10 ng of eluted RNA was fragmented to 400 nt and primed using 1 μl of random hexamers and 4 μl of NEBNext First Strand Synthesis Reaction Buffer (New England Biolabs, Ipswich, Massachusetts, USA), incubating at 94° C for 15 min. Then, we performed first-strand cDNA synthesis by combining the fragmented RNA with 0.5 μl Murine RNase Inhibitor (New England Biolabs, Ipswich, Massachusetts, USA), 1 μl of ProtoScript II Reverse Transcriptase (New England Biolabs, Ipswich, Massachusetts, USA), and 5 μl of Actinomycin D (0.1 μg/μl; Sigma-Aldrich, St. Louis, Missouri, USA). The solutions were incubated for 10 min at 25°C, 15 min at 42°C, and 15 min at 70°C. For second-strand synthesis, we added 8 μl of NEBNext Second Strand Synthesis Reaction Buffer (New England Biolabs, Ipswich, Massachusetts, USA) and 4 μl of NEBNext Second Strand Synthesis Enzyme Mix (New England Biolabs, Ipswich, Massachusetts, USA). The reaction was incubated at 16°C for 1 hour. After second-strand

synthesis, the reaction was cleaned twice up using 1.8X Agencourt AMPure XP beads (Beckman Coulter, Indianapolis, Indiana, USA) and eluted into 55.5 µl of nuclease-free H_2O. The double-stranded cDNA was input for end-repair, dA-tailing, and adapter ligation with Illumina TruSeq barcoded adapters. The ligation reaction was purified using 1.0X AMPure XP beads and eluted into 17 µl of nuclease-free H_2O. For the library amplification reaction, we ran the initial denaturation at 98°C for 30 s, followed by 12 cycles of denaturation at 98° C for 10 s, annealing and extension at 65°C for 75 s, and final extension at 65°C for 5 min.

Prepared libraries were checked for quality and quantity using three methods: the Qubit (dsDNA HS assays; Thermo Fisher Scientific Waltham, Massachusetts, USA), the Agilent BioAnalyzer 2100 (High Sensitivity DNA Kit; Agilent Technologies, Santa Clara, California, USA), and qPCR (KAPA Library Quantification Kit; Kapa Biosystems, Wilmington, Massachusetts, USA) performed on a Bio-Rad CFX96 Touch Real-Time PCR machine (Bio-Rad Laboratories, Hercules, California, USA). These assays measured library concentration (ng/µl) and library molarity (nM/l; BioAnalyzer and qPCR only) to create library pools. Libraries were pooled based on nanomolar concentrations and purified once using 0.6X AMPure XP beads. The library pool was sequenced on one lane of an Illumina HiSeq2500 at the Washington State University Genomics Core Lab (Spokane, Washington, USA) for paired-end 101 bp reads. All sequence data were deposited in the Sequence Read Archive on GenBank (BioProject: PRJNA401042).

Transcriptome assembly and homology inference

The per base quality of the raw reads, adapter content, and per base sequence content was first assessed for each sample using the tools implemented in FastQC (Andrews, 2010) to determine if any samples returned poor quality reads. Trimmomatic v0.36 (Bolger et al. 2014) was then used to filter low-quality paired-end sequence reads, performing the following steps in this order: removing Illumina adapters (TruSeq3-PE-2.fa file provided by Trimmomatic), removing the first 13 bases of each read, removing leading and trailing low quality or N bases (below quality 3), scanning each read with a 4-base wide sliding window and cutting when average quality per base drops below 15, and removing reads that are less than 50 bases long after these steps (ILLUMINACLIP:TruSeq3-PE2.fa:2:30:10 HEADCROP:13 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50). The first 13 bases at the 5′-end from each read were trimmed based on the per base sequence content data from FastQC. The use of random hexamers during Illumina library preparation can create biases in nucleotide composition, affecting the uniformity of the read location along expressed transcripts (Hanson et al., 2010). Filtered reads generated in the current study were combined with the sequence data generated from Roberts and Roalson (2017). Reads were de novo assembled into contigs using Trinity v2.3.0 (Grabherr et al., 2011; Haas et al., 2013). We generated reference transcriptomes for each species by using all samples derived from an individual species. For assembly, we used default settings in Trinity and specified reverse-forward read alignment (–SS_lib_type RF). These assemblies were not filtered using expression levels (FPKM or TPM) because our downstream analyses are not considering gene expression and we did not wish to discard any biologically relevant transcripts from our data. We predicted open reading frames (ORF) and identified coding sequences (CDS) and peptide sequences for each transcriptome assembly with TransDecoder v3.0.1 (Haas et al., 2013), using a BLASTp search against the SwissProt database (www.uniprot.org) to increase the sensitivity of functional coding region identification. While TransDecoder may include 5′- and 3′-untranslated regions (UTRs) in its output, we used both coding sequence (CDS) and peptide sequences for all further analyses. Transcriptome sequence redundancy was additionally reduced with CDHIT-EST v4.6 (Li and Godzik 2006) using a clustering threshold of 0.99 and a word size of 5.

To identify homolog and ortholog sequences, we used the tree-based identification pipeline described in Yang and Smith (2014) across all 12 species. All peptide sequences were clustered using an all-by-all BLASTp search performed using DIAMOND v0.8.29.91 (–evalue 1e-6 –outfmt 6; Buchfink et al., 2015). The BLASTp results were then clustered with MCL v14-137 (Enright et al., 2002) using an inflation value of 1.4. Each of these clusters was then aligned using MAFFT v7.271 (–genafpair –maxiterate 1000; Katoh and Standley, 2013) and the resulting alignments were trimmed using Phyutility v2.7.1 (Smith and Dunn, 2008) with minimal column occupancy of 0.1. Codon alignments were produced for each peptide cluster using PAL2NAL v14 (Suyama et al., 2006) and the nonredundant CDS sequences derived above. Tree inference was then performed for each trimmed alignment using FastTree v2.1.8 (Price et al., 2010) using the general time-reversible (GTR) model. Because these trees may contain long branches resulting from misassembly, paralogy, or recombination, we trimmed branches that were more than 10 times longer than its sister or longer than 0.2 substitutions per site. Final homolog group alignments were created using MAFFT and used for tree inference using RAxML v8.2.9 (Stamatakis, 2014) and the GTRCAT approximation of the GTR model. Ortholog groups were inferred using the monophyletic outgroups (MO) method of Yang and Smith (2014) with *G. cuneifolia* as outgroup. This method looks for clusters with monophyletic outgroups, roots the tree, and infers gene duplication events to identify the subtree with the highest taxa membership (Yang and Smith, 2014). We filtered orthologous groups to contain at least 12 taxa and an alignment length of at least 300 bp. Boostrap branch support in the ortholog trees was assessed using 200 rapid bootstrap replicates in RAxML. The resulting ortholog alignments, trees, and bootstrap replicates were used in downstream analyses.

Genomewide differences between the species

Read mapping and SNP detection—We used BWA v0.7.12 (Li and Durbin, 2009) to align trimmed Illumina reads from each of the 12 species to the ortholog sequences we created (as described above) with the default parameters. All samples from both the Bud and D stages derived from a species were separately aligned to the ortholog sequences with high sensitivity. We then sorted and combined alignments from the different samples using SAMTOOLS v1.3.1 (Li et al., 2009) to call variable positions in each species. Only paired aligned reads were used for SNP calling, using the combined alignments from all samples in each species. The genotype likelihoods for each individual site were calculated, and allele frequencies were estimated. The 'MPILEUP' command in SAMTOOLS was used to identify SNPs with the parameters '-q 30 –C 50 –t SD –t DP –Q 30 –m 2 –F 0.002 –uf'. Genotypes with Phred-scaled genotype likelihoods below 30 were treated as missing, corresponding to a genotyping accuracy of at least 99.9%. To reduce the false positive discovery rate, sites with depth (DP) <30 were also considered to be missing. VCFtools was used to filter false positive SNPs or paralogs in each species by excluding sites with depth that was 3× greater than the mean depth of that species (Danecek et al., 2011).

Diversity and divergence—We used the consensus sequence alignments from orthology inference (without heterozygous positions) to calculate sequence diversity, d_{XY}, for all pairwise comparisons of *Achimenes, Eucodonia*, and *Gesneria*. d_{XY} is defined as the number of diverged sites between two species divided by the alignment length. We then calculated the relative node depth (RND) of taxa pairs compared to an outgroup (Feder et al., 2005; Rosenzweig et al., 2016). The RND is calculated as the divergence between two species divided by the average divergence between each species and an outgroup (*G. cuneifolia*).

Species tree estimations

An initial estimate of phylogenetic relationships was performed using all 1306 ortholog groups aligned individually and concatenated for a maximum likelihood analysis in RAxML under the GTRGAMMA model. We used 200 rapid bootstrap replicates combined with a tree search from every fifth bootstrap tree (option –f a) to assess clade support. Separate runs were performed using a no partitioning scheme or a scheme partitioning the alignment with loci treated as separate partitions (n = 1306). No difference in topology or bootstrap support was found under different partitioning schemes.

Under scenarios of high incomplete lineage sorting (ILS), concatenation has been demonstrated to have lower power to reconstruct phylogenetic relationships (Mirarab et al., 2014; Chou et al., 2015). We addressed the possibility of lineage sorting using two recently developed methods of coalescent phylogenetic inference: ASTRAL v4.10.12 (Mirarab and Warnow, 2015) and ASTRID v1.4 (Vachaspati and Warnow, 2015). Both coalescent methods take advantage of unrooted gene trees to infer species trees under the coalescent and in the presence of ILS. ASTRAL seeks to find the tree that maximizes the number of induced quartets in a set of gene trees that are shared by the species tree (Mirarab and Warnow, 2015), while ASTRID is an ILS-aware distance-based method that uses the BIONJ algorithm (Gascuel, 1997; Vachaspati and Warnow, 2015). We also performed 200 multilocus bootstrapping replicates in both ASTRAL and ASTRID to assess branch support values. Additionally, the quartet scores were calculated for the ASTRAL species tree to measure the amount of gene tree conflict around each branch. The quartet scores provide the percentage of quartets in the gene trees that agree with each branch.

Analyses of introgression

Phylogenetic inference of introgression—Gene flow among *Achimenes* can be inferred using a phylogenetic network that allows for reticulation events (Huson et al., 2011). We created a phylogenetic network using PhyloNetworks v0.5.1 (Solís-Lemus and Ané, 2016; Solís-Lemus et al., 2017) and all 1306 gene trees that contained all 12 samples with *G. cuneifiolia* as the outgroup. We used SNaQ (Species Networks applying Quartets; Solís-Lemus and Ané, 2016) to evaluate the most likely network (given the species tree and gene trees) and to calculate γ, the vector of inheritance probabilities describing the proportion of genes inherited by a hybrid node from one of its parents. We performed nested analyses that allowed for 0, 1, 2, 3, 4, or 5 hybridization (h) events and compared the negative log pseudolikelihood score. Optimization in each nested analysis was performed for 10 independent runs. The run with the lowest negative log pseudolikelihood score was kept as the best estimate. A sharp improvement in score is expected until h reaches the best value and a slower, linear improvement thereafter.

D-statistic—Incongruence between *Achimenes* species relationships was tested using the *D*-statistic, also known as the ABBA-BABA test, which compares counts of discordant site patterns (Durand et al., 2011). For these analyses, we used ortholog alignments that contained all 12 samples that were at least 300 bp long and removed all gaps. We had a total of 1306 orthologs consisting of 407,343 aligned base pairs and 48,321 variable sites.

The R package HYBRIDCHECK (Ward and Oosterhout, 2015) was used to count the number of ABBA and BABA site patterns in four-population phylogenies. The *D*-statistic was calculated across 12 four-population phylogenies where quartet scores from ASTRAL or SNaQ indicated admixture or hybridization events. We expect equal counts of the two site patterns (ABBA and BABA) when incomplete lineage sorting (ILS) causes discordance. On the other hand, if discordance is caused by gene flow, we expect the ABBA site patterns to be more prevalent than the other (i.e., *D* values will be positive). Differences in discordant site pattern counts were tested using the *D*-statistic as implemented in HYBRIDCHECK. Although a full genome alignment (or other linkage information) is currently not available for *Achimenes*, our sampled loci likely represent a random sample of mostly unlinked markers from across the genome. Under these circumstances, a jackknife approach was used to test for genomewide variation in incongruence (Meyer et al., 2012; Eaton and Ree, 2013).

Testing for recent gene flow—As discussed below in the Results section, we found evidence of gene flow between two pairs of species and conflicting signal of gene flow between *A. admirabilis* and *A. erecta* (Lam.) H.P. Fuchs. We hypothesized this inconsistency is due to recent gene flow in sympatry, such as between *A. candida* Lindl. and *A. misera* Lindl., and between *A. grandiflora* and *A. patens* 'Major', or patterns of incomplete lineage sorting between *A. admirabilis* and *A. erecta*. Recent work has estimated each of these species originated less than 7 Mya (Roalson and Roberts, 2016). To test these hypotheses, we investigated patterns of relative divergence in genes for which *A. admirabilis* was sister to *A. erecta*, *A. candida* sister to *A. misera*, and *A. grandiflora* sister to *A. patens* 'Major'. This analysis is based on two assumptions. First, more recent gene flow likely results in geographic variation in introgression because, compared to ancient polymorphism and ILS, there is less time for novel alleles to spread across populations. Second, loci that were recently exchanged between populations will show higher sequence similarity than loci undergoing ILS because introgressed loci have less time to diverge in each lineage. Therefore, we predict that sequences from sympatric samples will have more discordant gene tree topologies as sister taxa, and those sequences showing discordance only in sympatry will have lower between-species divergence than genes showing the same pattern of discordance in both allopatric and sympatric samples.

Within each of the three pairs of species with evidence of gene flow, there were many gene trees showing each species as sister to the other. Both ILS and introgression can result in the same gene tree topology when *D* values are positive (excess of ABBA site patterns), but relative divergence between sequences in different taxa showing introgression is predicted to be much less than divergence of sequences that underwent ILS. We predict that if the excess of genes showing discordant gene trees in each sister pair

compared to their relationships in the species tree is due to gene flow, these genes will have lower relative divergence. If the gene tree discordance in each sister pair is due to ILS, we predict these genes will not have lower relative divergence compared to the species tree. To test this prediction, we used the set of orthologs containing all 12 samples (a total of 1306 genes) to create six sets of gene trees, two for each sister species comparison. The first set of gene trees contained all trees where species A and species B were sister. The second set of gene trees contained all trees from the first set, along with all gene trees that showed the alternate topology found in the species tree or the SNaQ tree (in the case of *A. grandiflora* and *A. patens* 'Major'). The RND was calculated for both sets of gene trees for each species pair and significance was assessed using a Student's t-test. Because our gene tree categories contained different numbers of genes, in addition to testing for RND differences between categories, we calculated 95% bootstrap confidence intervals.

RESULTS

Transcriptome assembly and orthology inference

Over 255 million reads and more than 51 Gb were sequenced from the libraries constructed for this study (Appendix S1; see Supplemental Data with this article). Libraries that were sequenced previously (Roberts and Roalson 2017) for the Bud and D stages in four species (*A. cettoana, A. erecta, A. misera,* and *A. patens* 'Major') were also used in the current study. We carried out de novo transcriptome assembly for the 12 species and identified orthologs in these assemblies to study the evolutionary history of 10 species of *Achimenes*. Our assembled transcriptomes contained between 58,000 and 111,000 putative transcripts with a mean N50 of 1698 (Appendix S1). Open reading frames were detected in 61–74% of the transcripts, resulting in between 21,000 and 24,000 putative genes with 2.5 ± 0.50 putative isoforms (Table 1). For orthology inference and phylogenetic reconstruction, we used CDS and excluded noncoding transcripts to minimize the amount of missing sequence data resulting from RNA degradation or sequencing errors. Using the monophyletic outgroups (MO) orthology inference (Yang and Smith, 2014), we identified 1306 ortholog clusters containing all 12 samples. Together, these 1306 ortholog clusters contain nearly 2 million aligned sites and an overall matrix alignment occupancy greater than 87% (Appendix S2).

Diversity and divergence

Pairwise estimates of sequence similarity (d_{XY}) for the ingroup ranged from 0.0301 ± 0.0016 (*A. longiflora* vs. *A. pedunculata* Benth.; Appendix S3) to 0.0090 ± 0.0001 (*A. grandiflora* vs. *A. patens* 'Major'; Appendix S3). The average pairwise sequence similarity across all ingroup species was 0.0212 ± 0.0013 (Appendix S3). Estimates of pairwise divergence using the RND statistic ranged from 0.9406 ± 0.0381 (*A. longiflora* vs. *A. pedunculata*; Appendix S3) to 0.3268 ± 0.0223 (*A. grandiflora* vs. *A. patens* 'Major'; Appendix S3). The average pairwise divergence across all ingroup species was 0.7263 ± 0.0324 (Appendix S3).

Phylogenetic relationships among *Achimenes* species

We inferred the phylogenetic relationships among the sampled *Achimenes* species using maximum likelihood and two coalescent-based methods. With 10 of 26 *Achimenes* species sampled, all our analyses strongly supported a monophyletic *Achimenes* with *Eucodonia* as sister (BS = 100/100/100; Fig. 2C). *Achimenes misera* is strongly supported as sister to the species in Clade 1 in both ML and ASTRAL analyses, which includes *A. erecta* and *A. cettoana*, with much lower support in the ASTRID analysis (BS = 100/100/53; Fig. 2C). In Clade 1, *A. cettoana* and *A. longiflora* are strongly supported as sister in all analyses (BS = 100/100/100; Fig. 2C), while *A. admirabilis* and *A. erecta* had strong support in ML and ASTRAL and much lower support in ASTRID (BS = 100/97/57; Fig. 2C). In Clade 2, *A. candida* was separated from the other members of the clade (Fig. 2C). The branch leading to *A. antirrhina* (DC.) C.V. Morton, *A. pedunculata, A. grandiflora,* and *A. patens* 'Major' was strongly supported in all analyses (BS = 100/100/100; Fig. 2C), while the branch separating *A. antirrhina* from the other three was less supported in both coalescent analyses (BS = 100/96/58; Fig. 2C).

Examining the quartet scores for each branch produced from ASTRAL found nearly all with high scores (>60) for the species tree topology, and two branches that showed much lower scores (<50) (Appendix S4). The quartet score is proportional to the percentage of induced quartet trees found in the species tree. Higher quartet scores indicate a larger proportion of the gene trees that share the same topology as the inferred species tree. The first branch with a low quartet score unites *A. admirabilis* and *A. erecta* and quartet scores for the species tree topology, and the first alternate shows that nearly equal proportion of induced quartets support either topology (node 5; Appendix S5). The second branch with low quartet score separates *A. antirrhina* from *A. pedunculata, A. grandiflora*, and *A. patens* 'Major' (node 2; Appendix S5). The quartet scores

TABLE 1. Summary of *Achimenes* CDS transcriptome assemblies.

Species	Number genes	Number transcripts	Mean length	N50 length	Assembled bases	Number of SNPs
Achimenes admirabilis	24,510	73,023	1077	1401	78,615,048	13,252
Achimenes antirrhina	21,610	54,684	993	759	54,299,127	75,207
Achimenes candida	22,051	60,605	960	1209	58,181,853	93,710
Achimenes cettoana	23,278	41,426	1018	1320	42,154,371	11,121
Achimenes erecta	21,583	43,221	926	1167	40,041,570	86,406
Achimenes grandiflora	21,895	48,833	1058	1368	51,677,433	25,097
Achimenes longiflora	21,127	66,491	1052	1356	69,919,131	45,167
Achimenes misera	21,619	47,464	943	1194	44,749,797	84,758
Achimenes patens 'Major'	22,110	47,729	918	1149	43,798,476	90,377
Achimenes pedunculata	21,755	69,325	1024	1311	71,019,474	44,551
Eucodonia verticillata	21,436	48,695	1066	1380	51,907,482	36,917
Gesneria cuneifolia	22,169	68,351	957	1209	65,395,389	97,633

for the primary topology, first alternate, and second alternate show proportions at this branch varied between 25 and 40 (Appendix S5). Additionally, the node separating Clade 1 and Clade 2 (node 9; Appendix S5) had quartet scores very close to 50.

Evidence for gene flow among several species pairs

Phylogenetic inference using networks—We used a recently developed method (Solís-Lemus and Ané, 2016) to infer a phylogenetic network of the *Achimenes* samples from individual gene trees (Fig. 3). Unlike in the ABBA-BABA tests (described below), we used all 12 samples for this analysis. The inferred phylogenetic network that best fits our data included three hybrid branches with vectors of inheritance probabilities (γ) estimated for each (Solís-Lemus and Ané, 2016; Fig. 3; Appendix S6). One hybrid branch led from *A. misera* to *A. candida* (γ = 0.24), another led from *A. admirabilis* to *A. erecta* (γ = 0.47), and a third led from *A. grandiflora* to *A. patens* 'Major' (γ = 0.40; Fig. 3). The hybrid branches leading from *A. misera* and *A. grandiflora* additionally had high bootstrap support of 100 (Fig. 3), while the hybrid branch connecting *A. admirabilis* and *A. erecta* had very low boostrap support of 46 (Fig. 3). Additionally, the placement of *A. patens* 'Major' differs in this analysis (Fig. 3) than in the phylogenetic analyses above (Fig. 2C). Here, *A. patens* 'Major' is placed sister to *A. antirrhina* (Fig. 3) with moderately strong bootstrap support (Fig. 3).

D-statistic—We explicitly tested for asymmetry in discordance patterns using the *D*-statistic (Green et al., 2010; Durand et al., 2011). Because this test required sets of four populations, we conducted several tests using different combinations of taxa and clades (Table 2). We observed several highly significant positive *D* values for analyses that compared populations both within and between Clades 1 and 2. Within Clade 1, we observed a significant positive *D* value between *A. admirabilis* and *A. erecta* ($D = 0.82$, $P < 0.001$) and nonsignificant *D* values between *A. cettoana* and either *A. admirabilis* or *A. erecta*. Within Clade 2, significant positive *D* values were observed between *A. grandiflora* and *A. patens* 'Major' ($D = 0.72$, $P < 0.001$) and between *A. antirrhina* and *A. patens* 'Major' ($D = 0.55$, $P < 0.001$), while nonsignificant *D* values were calculated between *A. antirrhina* and either *A. pedunculata* or *A. grandiflora*. Comparing populations between Clade 1 and Clade 2, we also observed a significant positive *D* value between *A. candida* and *A. misera* ($D = 0.85$, $P < 0.001$). It is important to note that the *D*-statistic is useful to suggest the presence of population admixture but cannot be used to determine absolute rates of gene flow. The *D*-statistic often cannot distinguish site pattern discordance that is due to ancient polymorphism/ILS or introgression (Feder et al., 2005; Goulet et al., 2017). Therefore, we applied additional analyses to determine the most likely gene flow events between *Achimenes* populations and to distinguish ILS from gene flow.

TABLE 2. Summary of ABBA-BABA tests for population admixture in *Achimenes*.

P1	P2	P3	O	ABBA	BABA	*D*	*P*-value
AA	*AE*	*AC, AL*	*AM*	89	895	−0.82	< 0.001
AA	*AE*	*AC*	*AL*	214	136	0.22	1.00
AC, AL	*AA*	*AE*	*AM*	895	89	0.82	< 0.001
AD	Clade 2	*AM*	Clade 1	89	998	−0.85	< 0.001
AG	*AN*	*AT*	*AD, AP*	345	178	0.32	0.98
AL	*AC*	*AA*	*AE*	136	214	−0.22	1.00
AN	*AG*	*AT*	*AD, AP*	178	345	−0.32	0.98
AP	*AG*	Clade 2	Clade 1	284	215	0.14	0.02
AP	*AT*	Clade 2	Clade 1	240	586	−0.42	< 0.001
Clade 2	*AD*	*AM*	Clade 1	998	82	0.85	< 0.001
Clade 2	*AP*	*AG*	Clade 1	1334	284	0.65	< 0.001
Clade 2	*AP*	*AT*	Clade 1	823	240	0.55	< 0.001

Note: AA, A. admirabilis; AC, A. cettoana; AD, A. candida; AE, A. erecta; AG, A. grandiflora; AL, A. longiflora; AM, A. misera; AN, A. pedunculata; AP, A. patens 'Major'; *AT, A. antirrhina*; Clade 1, *AA, AE, AC, AL*; Clade 2, *AD, AG, AN, AP, AT*; P1, population 1; P2, population 2; P3, population 3; O, outgroup population.

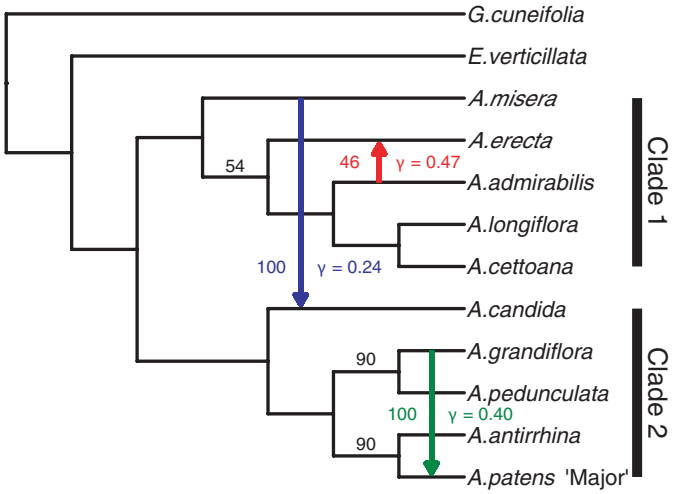

FIGURE 3. Introgression model of *Achimenes* estimated with SNaQ. From 1306 genes, *h* = 3, and rooted with *Gesneria cuneifolia*. Black branches: major tree (including hybrid branches with γ > 0.5). Colored arrows: minor hybrid branches annotated by γ, the vector of heritance probabilities estimated with SNaQ. Black numbers: bootstrap support for branches in the major tree, if different from 100. Color numbers: bootstrap support for the placement of minor hybrid branches.

Using levels of divergence to test for gene flow—The results from both *D*-statistics and the phylogenetic network analyses indicate admixture between three species pairs: *A. admirabilis* and *A. erecta*, *A. candida* and *A. misera*, and *A. grandiflora* and *A. patens* 'Major'. First, we investigated the divergence in genes with topologies showing each species pair as sister with the topologies of each species pair in the species tree. This method compares the genetic divergence between genes with discordant topologies to distinguish between ILS and introgression (Fig. 4). We predict that recently introgressed loci will have discordant tree topologies and display low interspecific divergence. We found that genes that have *A. candida* sister to *A. misera* had a significantly lower RND than genes where each is sister to their respective clades ($t = 8.682$, df = 801, $P < 0.001$; Fig. 4A). We also found that genes that have *A. grandiflora* sister to *A. patens* have significantly lower RND than genes where *A. antirrhina* is sister to *A. patens* ($t = 4.727$, df = 1404, $P < 0.001$; Fig. 4B). Genes showing *A. admirabilis* sister to *A. erecta* have RND values nearly indistinguishable from genes showing the alternate topology where *A. admirabilis* is sister to *A. cettoana* and *A. longiflora* ($t = 2.306$, df = 527, $P = 0.022$; Fig. 4C). These findings suggest *A. candida* and *A. patens* experienced recent introgression from *A. misera* and *A. grandiflora*, respectively. These findings additionally suggest some level of ILS between *A. admirabilis* and *A. erecta* causing gene tree discordance between these populations.

DISCUSSION

Hybridization and gene flow are frequent evolutionary forces that influence the process of speciation. New genomic tools provide an exciting opportunity to test hypotheses on the effect of gene flow during lineage diversification (Gompert and Buerkle, 2016; Payseur and Rieseberg, 2016; Vallejo-Marín and Hiscock, 2016). Here we take advantage of transcriptome sequencing and demonstrate its utility to reassess phylogenetic relationships in *Achimenes* and to ask whether gene flow occurred during the evolution of this lineage. Using multiple analyses of phylogenetic discordance, we show that gene flow occurred between two pairs of sympatric sister species, and provide evidence that questions the species status of *A. admirabilis*.

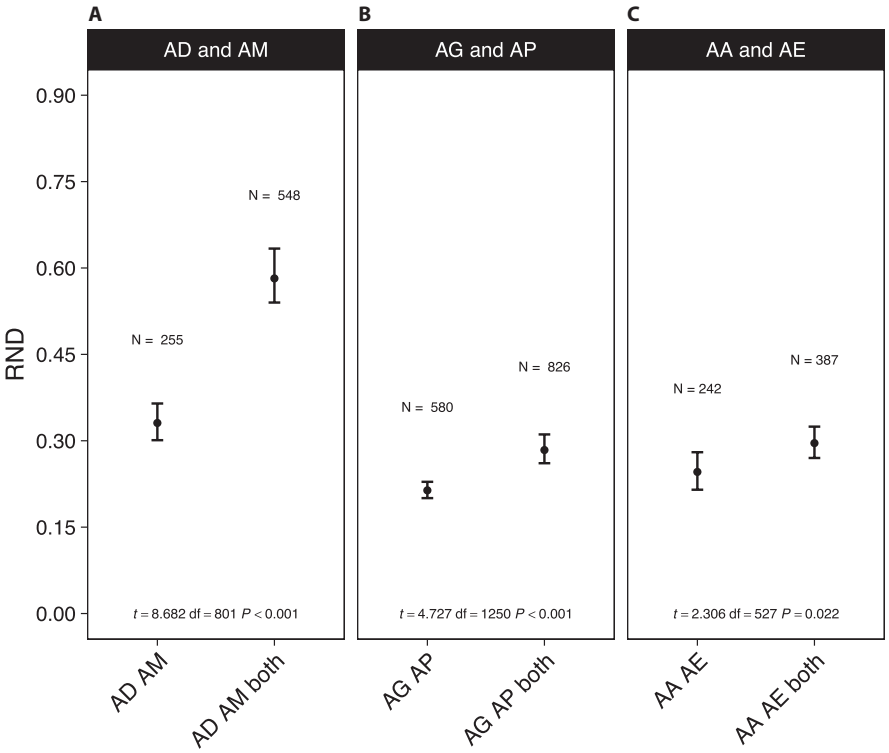

FIGURE 4. Tests of recent introgression. (A) Interspecific divergence for genes that show *A. candida* as sister to *A. misera* ('AD, AM') and as sister to both *A. misera* and Clade 2 ('AD, AM both'). (B) Interspecific divergence for genes that show *A. patens* 'Major' as sister to *A. grandiflora* ('AG, AP') and as sister to both *A. grandiflora* and *A. antirrhina* ('AG, AP both'). (C) Interspecific divergence for genes that show *A. admirabilis* as sister to *A. erecta* ('AA, AE') and as sister to both *A. erecta* and *A. cettoana/A. longiflora* ('AA, AE both'). Results from *t*-tests comparing the means are shown. Bars indicate 95% confidence intervals as calculated using bootstrap resampling.

Phylogenetic relationships among *Achimenes* species

Previous phylogenetic hypotheses (Roalson et al., 2003; Roalson and Roberts, 2016) do not agree with the phylogenetic hypothesis presented here. These studies have indicated moderate to strong support for species placement in *Achimenes* within three distinct clades (Clade 1, Clade 2, and Clade 3; sensu Roalson et al., 2003). Resolution was lacking, however, at the crown of the genus to indicate how these three clades were related to one another (Roalson et al., 2003; Roalson et al., 2005; Roalson and Roberts, 2016). The previous phylogenetic hypothesis of Roalson et al. (2003) was generated using two loci (nuclear ITS and plastid *trnL-F* spacer), while the hypothesis presented here uses >1300 loci and demonstrates the utility of transcriptome-based phylogenomic approaches to reassess previous Sanger-based hypotheses. The level of discordance between the phylogenetic hypothesis of this study and Roalson et al. (2003) could be due to gene sampling effects. The two loci used by Roalson et al. (2003) likely belong to gene families with alternate histories from the species tree presented here. Our results suggest two distinct clades in *Achimenes*, Clade 1 and Clade 2 (Fig. 2C), with strong bootstrap support from all four methods employed (BS = 100/100/100/100; Fig. 2C). While these estimates provide strong bootstrap support, the quartet score for the primary topology at the node separating Clade 1 and Clade 2 was close to 50 (node 9; Appendix S4). While not indicative of strong ILS at this node, the decreased quartet score could reflect lower support for the placement of *A. misera* in Clade 1 (Fig. 2C). An increased sampling of species, particularly some that were placed in the original Clade 3 of Roalson et al. (2003), might shed more light on this apparent discordance. Lastly, the addition of samples from *Smithiantha,* a small herbaceous genus closely related to *Eucodonia* and *Achimenes,* and *Solenophora,* a genus of woody shrubs thought to be closely allied to *Achimenes,* would also provide additional data to test the apparent monophyly of *Achimenes* presented here.

Results also indicate that *A. patens* 'Major' is strongly supported as sister to *A. grandiflora* in all phylogenetic analyses (BS = 100/100/100; Fig. 2C). This result adds support to our hypothesis that *A. patens* 'Major' experienced some level of gene flow with *A. grandiflora*, particularly given the numerous morphological similarities between them. Given the abundance of discordant gene trees, it is not surprising that previous analyses based on limited gene sampling suggested different relationships. Additional discordant patterns of topology and branch support were found, particularly when comparing species tree methods. All relationships were strongly supported with BS = 100 when using ML (Fig. 2C), while two branches were slightly less well supported using ASTRAL (BS ≥ 96; Fig. 2C), and three branches had weak support when using ASTRID (BS ≤ 58; Fig. 2C). The branch showing a relationship between *A. admirabilis* and *A. erecta* and the branch separating *A. antirrhina* from *A. grandiflora, A. patens* 'Major', and *A. pedunculata,* were both the same branches where quartet support was low (Appendix S4), indicating ILS or admixture at these branches (Fig. 2C). Both ASTRAL and ASTRID, using different approaches to estimate a species tree, have been shown in some data sets to reduce branch support when there is high variance in gene tree topologies (Esselstyn et al., 2017).

Additionally, SNaQ analysis showed *A. erecta* sister to the rest of Clade 1, minus *A. misera* (BS = 57; Fig. 3), rather than sister to *A. admirabilis* (BS = 100/97/57; Fig. 2C). Furthermore, the relationship of *A. misera* to Clade 1 was strongly supported in the SNaQ

analysis (BS = 100; Fig. 3), but weakly supported in the ASTRID analysis (BS = 53; Fig. 2C). Many coalescent-based methods (such as ASTRAL and ASTRID) work under the assumption that ILS is the only source of gene tree discordance (Mirarab and Warnow, 2015; Vachaspati and Warnow, 2015), while ignoring the presence of gene flow. These species tree methods are not robust to such violations even with large numbers of well-constructed trees and may be inconsistent under gene flow (Solís-Lemus et al., 2016). SNaQ allows for both ILS and gene flow (Solís-Lemus and Ané, 2016). The branches showing very weak support in ASTRID were also branches on the tree where other analyses indicated ILS or gene flow. The distance-based algorithm of ASTRID may be less robust under high gene flow than the quartet-based algorithm of ASTRAL (Davidson et al., 2015). Our results demonstrate that estimating evolutionary relationships in lineages where both ILS and gene flow occur (such as *Achimenes*) remains a challenging endeavor and advocates for more extensive use of network-based approaches that can account simultaneously for both processes (Solís-Lemus and Ané, 2016).

Effects of gene flow during speciation

Phylogenetic network analysis indicated the evolutionary history of these species includes some level of gene flow (Fig. 3). Additionally, four-taxon analyses of population admixture provided additional evidence for gene flow between two *Achimenes* species pairs, from *A. misera* to *A. candida*, and from *A. grandiflora* to *A. patens* 'Major' (Table 2). While the *D*-statistic can detect, but not quantify, introgression or admixture, many studies have shown it to be robust when used on a genome-wide scale (Green et al., 2010; Eaton and Ree, 2013). However, it can also be stochastic when applied over small windows and is sensitive to within-species diversity (Martin et al., 2015), sometimes producing conflicting results. Therefore, we caution the over interpretation of its meaning outside of providing an indication of potential admixture that should be more extensively evaluated using alternative measures, such as the $\hat{f}_d$ statistic (Martin et al., 2015). Using exemplars for this study provides initial evidence that gene flow occurred within these species pairs sometime in the past and provides a starting point for additional sampling of individuals to explore the prevalence of gene flow and hybridization within sympatric populations.

Combining the results of each analysis, we found two consistent sister pairs that exhibited signs of gene flow: *A. candida* and *A. misera*, and *A. grandiflora* and *A. patens* 'Major'. Each pair consists of species that display remarkable similarity in floral form (Fig. 1). Gene flow among species that share similar pollinators has been observed in other plant groups, including many orchid lineages (Cortis et al., 2009; Gögler et al., 2015). Our estimates of gene flow from *A. candida* to *A. misera* and from *A. grandiflora* to *A. patens* 'Major' are too low for an early hybrid, and are comparable to levels found in other species with moderate levels of hybridization (Cahill et al., 2016; Solís-Lemus and Ané, 2016). Together with the similarities in floral form, these results indicate gene flow between the two pairs is not unexpected and might suggest that gene flow occurred through visitations by a common pollinator.

Hybrids between *A. grandiflora* and *A. patens* are observed in the field (Wiehler, 1983; Ramírez Roa, 1987) and have high fertility when crossed in the greenhouse (Cooke and Lee, 1966). Experimental crosses between the two species indicate that these hybridization events produce stainable pollen at rates of 88% (Cooke and Lee, 1966). *Achimenes grandiflora* is among the most widespread species in the genus and is sympatric with numerous species (Fig. 2B; Wiehler, 1983; Ramírez Roa, 1987). Before *A. patens* was formally described as a species in 1840, it was considered synonymous with *A. grandiflora* because of the striking similarity of their flowers and their overlapping geographic distributions (Gordon, 1846). While these similarities in floral form are certainly an example of floral convergence, there are some differences. Particularly, the corolla spur in *A. patens* can be pointed and elongated, whereas, the corolla spur in *A. grandiflora* is blunt and short. Our sample of *A. patens* 'Major' showed intermediate characters, particularly *A. patens* 'Major' has a similar short, blunt spur to *A. grandiflora,* while retaining other floral and vegetative characteristics of *A. patens.* We initially hypothesized that *A. patens* 'Major' may have experienced some introgression in the past, likely from *A. grandiflora*. Both *A. grandiflora* and *A. patens* 'Major' share the closest genetic similarity among all pairwise comparisons (Appendix S3) and were found sister in all species tree reconstructions (Fig. 2C), contrary to previous analyses (Roalson et al., 2003). Our results from all analyses of gene flow indicate that introgression has likely occurred between these species at some point in the past.

Achimenes misera has been considered a reproductively isolated species (Cooke and Lee, 1966; Ramírez Roa, 1987). While nearly all of *Achimenes* are diploids (n = 11; Cooke and Lee, 1966; Ramírez Roa, 1987), *A. misera* (and *A. erecta*) is a polyploid (n = 22; Cooke and Lee, 1966; Ramírez Roa, 1987). It was thought that this difference in chromosome number might be the limiting factor in the formation of hybrids (Ramírez Roa, 1987). The suggestion by our analyses that gene flow occurred from *A. misera* to *A. candida* was therefore surprising given these reports. *Achimenes candida* was previously shown to have the ability to produce stainable hybrid pollen with many other species, although only with species belonging to Clade 2 (Fig. 2B; Cooke and Lee, 1966; Wiehler, 1983), which does not include *A. misera*. When we considered the putative paralog contributions of each species to the others' genome, we found 323 putative paralogs of *A. candida* and *A. misera* (25% of 1306 orthologs) in our data set. Within that group of paralogs, 62 (19% of the paralogs or 5% of the total) were contributed from *A. candida* to *A. misera* and 184 (57% of the paralogs or 14% of the total) were contributed from *A. misera* to *A. candida*. These numbers would suggest that introgression occurred between these species sometime in the recent past. In order for ~5% of the *A. candida* genome to have introgressed from *A. misera*, a possible route may have been through an initial hybridization event with *A. misera* (both diploids at this point) followed by several backcrossing events to *A. candida* (Fig. 5A). Similarly for ~14% of the *A. misera* genome to be introgressed from *A. candida*, one possible route would be through an initial hybridization event with *A. candida* (both diploids) followed by several backcrossing events to *A. misera*, eventually ending with an autopolyploidy event that creates the polyploid *A. misera* (Fig. 5B). These two scenarios fit extremely well with the data and offer testable hypotheses for future studies. More extensive population sampling of these two species in Mexico, particularly in areas where they are sympatric, would provide further insight into how extensive introgression and gene flow contributed to patterns of diversity.

Given that *A. candida* and *A. misera* have rather small, unassuming flowers, the kind typically of little interest to horticultural hybridizers of *Achimenes*, we did not initially hypothesize that gene flow or hybridization would have occurred during the history of this pair. Phylogenetic network analyses found a strongly supported hybrid branch between these species (BS = 100; Fig. 3), while the

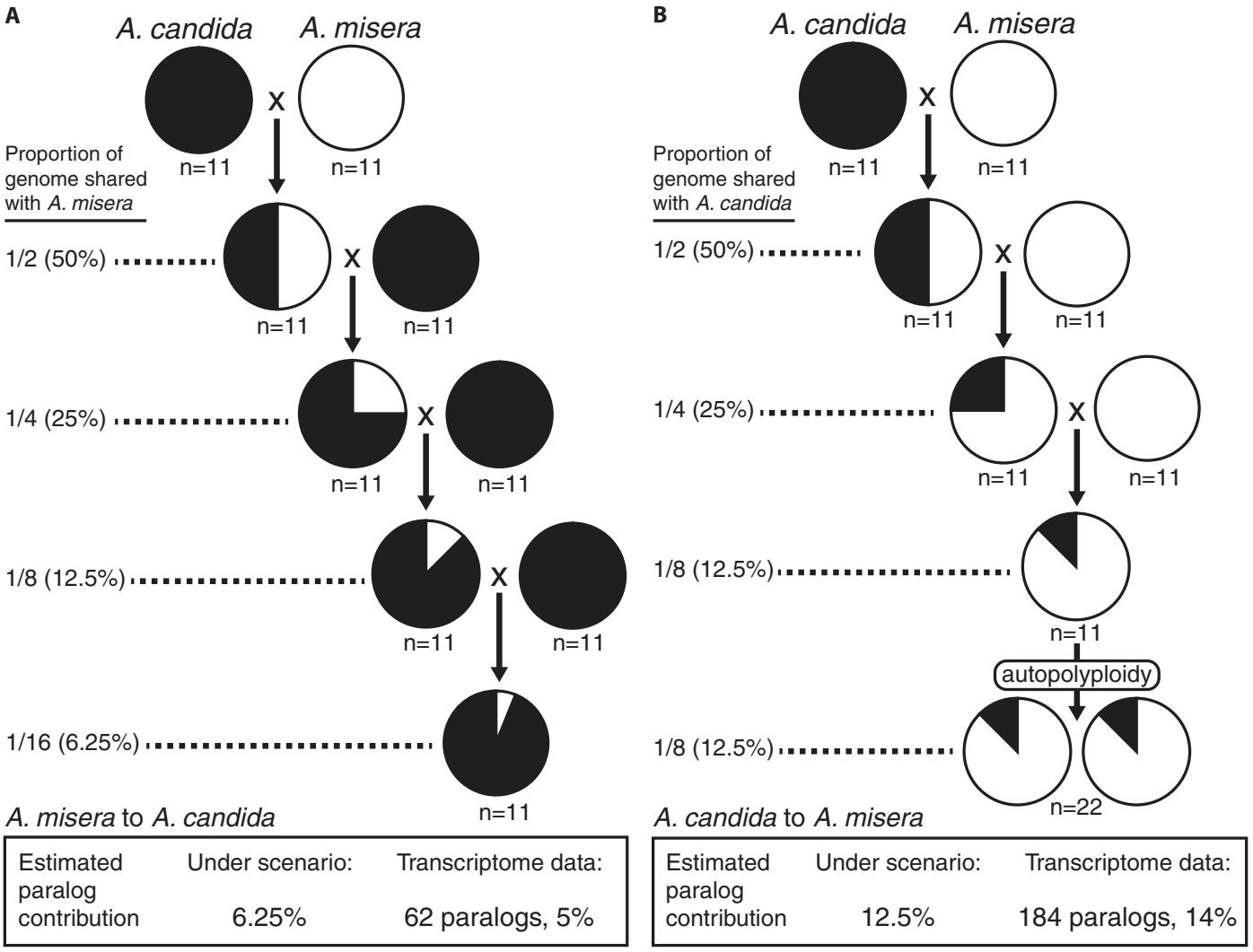

FIGURE 5. Scenarios of introgression between *A. candida* and *A. misera*. (A) An introgression model for *A. candida* begins with a hybridization event with *A. misera* and is followed by several backcrossing events to *A. candida*. This scenario leads to an estimated 6.25% of the *A. candida* genome being shared with *A. misera*, closely matching the 5% estimated from the transcriptome data. (B) An introgression model for *A. misera* begins with a hybridization event with *A. candida,* followed by several backcrossing events to *A. misera*, and ending with a recent autopolyploidy event that creates the polyploid *A. misera*. This scenario leads to an estimated 12.5% of the *A. misera* genome being shared with *A. candida*, closely matching the 14% estimated from the transcriptome data. Pie charts indicate the estimated proportion of the genome shared with *A. candida* or *A. misera* (colored in black and white, respectively). The ploidy of each individual is these scenarios is indicated below each circle.

genetic similarity between the two species was also moderate compared to other pairwise comparisons (Appendix S3). *Achimenes misera* has also received interest for its close affinity to another species, *A. warszewicziana* (Regel) H.E. Moore. This species also displays similar floral form to both *A. candida* and *A. misera*, is pollinated by bees, and is found in sympatric locations with both *A. candida* and *A. misera*. Until very recently, *Achimenes warszewicziana* was considered to be synonymous with *A. misera.* It is now regarded as a distinct species. Sampling of these three species may reveal more extensive patterns of gene flow and introgression between sympatric populations that have similar flowers and similar pollinators.

Gene flow among sympatric species that share pollinators has been studied for decades in many different plant systems (Beattie, 1976; Campbell, 1985; Soliva and Widmer, 2003; Gögler et al., 2015). In gesneriads, the contributions of pre- and post-mating barriers to gene flow in the maintenance of species barriers has not been extensively studied, outside of sympatric Hawaiian *Cyrtandra* (Johnson et al., 2015). Our results offer preliminary genomic evidence of extensive gene flow within *Achimenes,* and provide a starting point to address other patterns of sympatric gene flow within other Neotropical gesneriads such as *Sinningia* (Perret et al., 2007). Premating reproductive isolation between sympatric species is often associated with differences in flowering time (Soliva and Widmer, 1999; Savolainen et al., 2006), but changes in flower morphology also contribute to reproductive isolation through specialization to different pollinators (Grant, 1971; Ramsey et al., 2003; Fenster et al., 2004), or by limiting pollen transfer between species with similar pollinators (Armbruster et al., 1994; Wolf et al., 2001). Hybridization experiments in Neotropical gesneriads reveal high pollen stainability and fertility among interspecific hybrids and indicate that the

effective isolating mechanisms between species may not be genetic, but external physiological, spatial, or ecological barriers (Wiehler, 1983). No studies have looked at whether pre- or post-mating barriers contribute to reproductive isolation among closely related species in *Achimenes* or other Neotropical gesneriads.

Asymmetric gene flow

Our phylogenetic analyses suggest that the evolutionary history of these lineages involved asymmetric gene flow from *A. misera* into *A. candida*, and from *A. grandiflora* into *A. patens* 'Major' (Table 2; Fig. 3). This would presumably occur because hybrids backcrossed into *A. candida* and *A. patens* 'Major' more than with *A. misera* and *A. grandiflora*, respectively. The direction of gene flow between these two species pairs agrees with our predictions of gene flow based on morphological similarities and range overlap in central Mexico.

Directional bias in gene flow from one species into another can be important for determining the direction of evolutionary change and species succession (Petit et al., 2004). There has been longstanding interest in understanding the factors that drive asymmetrical gene flow in plants. Some of the underlying factors can include mating system variation (Lewis and Crowe, 1958), the relative proportions of parent species (Burgess et al., 2005), and differences in the fitness of reciprocal crosses (Tiffin et al., 2001). Alternatively, asymmetric gene flow could reflect demographic processes related to species range expansion, which has been implicated as a major determinant of gene flow in a wide array of plants and animals (Currat et al., 2008). Our results indicate asymmetries in gene flow from two of the most widely distributed species.

Achimenes grandiflora has one of the largest ranges throughout Mexico and Central America, is sympatric with a high number of species, and can produce hybrids with the largest number of other species (Cooke and Lee, 1966; Ramírez Roa, 1987). *Achimenes misera* is also widespread throughout Central America, but had previously been considered reproductively isolated (Ramírez Roa, 1987). In many natural plant populations, selection remains the primary mechanism implicated in determining patterns of hybridization and introgression (Lexer et al., 2005; Whitney et al., 2006). Further exploration and sampling of populations sympatric with both *A. grandiflora* and *A. misera* will provide data allowing us to determine how species boundaries are maintained and what role selection and demographic processes play in the patterns found in *Achimenes*. Additional genomic sequencing of more neutral loci than those used in the current study would further allow more sophisticated estimates of migration and gene flow (Gronau et al., 2011).

Geographic and ecological patterns of gene flow and divergence

We performed genomic analyses of gene flow using 10 transcriptomes that represent 10 species of *Achimenes*. These plants are often found in sympatry and allopatry throughout Mexico and Central America. Species can be found from sea level upward to 3000 m growing in *Quercus* and *Pinus* forests, but also found in transitional zones between forests and arid subtropical shrubs. Geographic patterns can additionally be found among the pollination syndromes. Those species that are bee-pollinated (e.g., *A. candida* and *A. misera*) tend to be more narrowly endemic in western and southern Mexico (Figs. 2A, B), while species pollinated by butterflies (e.g., *A. grandiflora* and *A. longiflora*) and hummingbirds (e.g., *A. antirrhina* and *A. erecta*) are more widespread throughout Mexico and Central America (Figs. 2A, B). These emerging patterns contribute to our understanding of diversification processes in this lineage.

Achimenes candida and *A. misera* can be found in Chiapas, Mexico and inhabit similar forest habitats and elevations throughout their ranges (Figs. 2A, B; Ramírez Roa, 1987). These overlaps also extend to flowering times. *Achimenes misera* flowers from April through October and *A. candida* flowers from July through August (Ramírez Roa, 1987). These ecological and phenological factors may contribute to gene flow in contact areas where species may be pollinated by similar bees during similar flowering times. More extensive work will need to be done in Chiapas to test these hypotheses within sympatric populations.

The specimen of *A. patens* sampled in this study was *A. patens* 'Major', originally described as a "first-class variety" (Moore, 1859). The origins of this variety are unclear, but it displays similarity to both *A. grandiflora* and *A. patens,* with the most obvious difference being that *A. patens* 'Major' contains a short, blunt corolla spur similar to *A. grandiflora*. Given the clear similarities between *A. patens* 'Major' and *A. grandiflora*, we hypothesized that *A. patens* 'Major' may represent a case of introgression from *A. grandiflora*. We envision two potential scenarios for the origins of this variety. First, this variety could represent a natural hybrid that was brought into cultivation from Mexico. Rhizomes of *A. patens* were first brought to England by a Mr. Hartweg in 1846 from Zitacuaro, Mexico, a location where populations of *A. grandiflora* and *A. patens* are found in sympatry (Gordon, 1846). Both species inhabit similar forest habitats in higher elevations, upwards of 1800 m (Ramírez Roa, 1987). Second, this variety could represent a horticultural hybrid whose origins might be found among hybridizations that took place during the peak popularity of magic flowers in Victorian Era England. Our results from all analyses provide support for introgression (Table 2; Figs. 2B and 3) between *A. patens* 'Major' and *A. grandiflora*, but without a better record for *A. patens* 'Major' we can only speculate about its origins.

Achimenes erecta is one of the most widespread species throughout Central America and some populations can be found in the Caribbean (Fig. 2A; Ramírez Roa, 1987). The original type specimen of this species was sent to England from Jamaica in 1778 (Fuchs, 1963). Considerable morphological variation is known from across the range of this species, including vegetative and reproductive characters (Wiehler, 1983; Ramírez Roa, 1987). Many varieties collected in different locations in Mexico, Central America, and the Caribbean have been brought into cultivation. Before *A. admirabilis* was described as a species (Wiehler, 1992), it was considered a variety of *A. erecta*, but having characteristics similar to *A. cettoana* (Ramírez Roa, 1987). Leaves in both *A. cettoana* and *A. admirabilis* are elliptic-linear and flowers are curved and glabrous, while *A. erecta* has lanceolate leaves and slightly curved, puberulent flowers. *Achimenes admirabilis* is only found from a few populations within Oaxaca and our analyses indicate that the patterns of gene tree discordance between *A. admirabilis* and *A. erecta* we found may partly be due to ILS, possibly with a low level of gene flow (Figs. 2C and 3). These results suggest the status of *A. admirabilis* should be reevaluated with additional sampling of *A. erecta* and *A. admirabilis* from throughout their ranges. The high variation in *A. erecta* may suggest that other varieties exist. Potential scenarios for the status of *A. admirabilis* exist, and given more extensive sampling of *A. erecta* may include two possibilities. (1) *A. erecta* may include multiple lineages that are as distinct as *A. admirabilis*, possibly leading to the

recognition of more species lineages. (2) *Achimenes erecta* may be a monophyletic lineage and *A. admirabilis* is in the process of divergence from *A. erecta* sensu lato. The evidence for gene flow between the two samples included here suggests that sampling *A. erecta* from across its range and morphological variability, in combination with *A. admirabilis* samples, will be necessary to more fully address the relationship between these two putative lineages.

We took advantage of the sister relationship between *A. admirabilis* and *A. erecta* to elucidate whether the observed gene tree discordances were due to ILS or introgression. *D*-statistics alone often cannot distinguish site patterns resulting from ILS and ancient polymorphism or recent hybridization (Feder et al., 2005; Rosenzweig et al., 2016; Goulet et al., 2017). Therefore, we compared divergence in genes with our discordant phylogenetic signals (Fig. 4; Roda et al., 2017). We found nearly equal numbers of genes supporting *A. admirabilis* as sister to *A. erecta* or supporting *A. admirabilis* sister to *A. cettoana* and *A. longiflora* (Appendices S4 and S5). Furthermore, we found that those genes showing *A. admirabilis* sister to *A. erecta* were not significantly less diverged than those gene trees showing both the discordant topology and the species tree topology (Fig. 4C). These results might suggest that the status of *A. admirabilis* being a distinct species from *A. erecta* be reconsidered.

As with other genome-wide analyses inferring gene flow, our study depends on analyzing patterns across many loci in a limited number of individuals. Sequencing many more individuals and species from across the range of *Achimenes* throughout Mexico and Central America will provide stronger estimates of the timing and amount of gene flow across the landscape, particularly for populations found in allopatry and sympatry with other species.

CONCLUSIONS

Our transcriptome analyses provide evidence of gene flow and introgression during the evolution of *Achimenes*. Multiple phylogenomic analyses of gene flow indicate introgression has occurred between at least two species pairs that share pollinators and are found in sympatry. These analyses also call into question the species status of *A. admirabilis*. Although the analyses applied here were originally designed for a small number of samples (Green et al., 2010; Payseur and Rieseberg, 2016), we acknowledge that some of our results could benefit by increasing the sampling. Particularly, sampling populations from across the range of *Achimenes* in Mexico, including both allopatric and sympatric individuals, will allow for a better quantification of the timing, direction, and magnitude of gene flow. Transcriptome sequencing approaches provide extensive genomic resources useful for studies of biodiversity that allows us to investigate both the patterns and processes involved in the evolution of tropical lineages. Lastly, the current study highlights interesting patterns of gene flow among species of *Achimenes* and provides the basis for further phylogenomic and phylogeographic studies into the evolution and diversification of this colorful and diverse lineage of gesneriads.

ACKNOWLEDGEMENTS

We thank Associate Editor John Freudenstein, Lucy Allison, Kimberly Hansen, Nan Jiang, Joseph Kleinkopf, and two anonymous reviewers for thoughtful comments on the manuscript; Joanna Kelley for access to laboratory facilities and helpful discussion on the manuscript; Corey Quackenbush for valuable insight on library preparation; Michael Neff and Karen Sanquinet for access to their Real-Time PCR machine; and Chuck Cody for maintaining the growth and happiness of our gesneriad collection. The molecular work of this study was conducted in the Kelley and Roalson Labs in the School of Biological Sciences, Washington State University. The sequencing work of this study was conducted in the Genomics Core Lab at Washington State University, Spokane. The Elvin McDonald Research Endowment Fund from The Gesneriad Society [to W.R.R], the Global Plant Sciences Initiative Fellowship [to W.R.R], and a NSF Doctoral Dissertation Improvement Grant DEB-1601003 [to W.R.R. and E.H.R.] supported this research.

DATA ACCESSIBILITY

Raw reads for the 64 sequenced libraries generated in this study are deposited in the NCBI Sequence Read Archive (BioProject: PRJNA401042). Raw reads for the 12 sequenced libraries of Roberts and Roalson (2017) are deposited in the NCBI Sequence Read Archive (BioProject: PRJNA340450). Assembled sequences, data files, alignments, and trees are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.9202s. Scripts and data for analyses are available from http://www.github.com/wrroberts/Achimenes-Phylogenomics-2017.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

LITERATURE CITED

Anderson, E. 1949. Introgressive hybridization. Wiley, New York.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc [accessed 12 December 2016].

Armbruster, W. S., M. E. Edwards, and E. M. Debevec. 1994. Floral character displacement generates assemblage structure of western Australian triggerplants (*Stylidium*). *Ecology* 75: 315–329.

Beattie, A. J. 1976. Plant dispersion, pollination, and gene flow in *Viola*. *Oecologia* 25: 291–300.

Becker, J. 2008. The Gesneriad Register 2008: *Achimenes*. The Gesneriad Society, Inc.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2: 8–13.

Buchfink, B., C. Xie, and D. H. Huson. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12: 59–60.

Burgess, K. S., M. Morgan, L. Deverno, and B. C. Husband. 2005. Asymmetrical introgression between two *Morus* species (*M. alba, M, rubra*) that differ in abundance. *Molecular Ecology* 14: 3471–3483.

Cahill, J. A., Z. Fan, I. Gronau, J. Robinson, J. P. Bollinger, B. Shapiro, J. Well, and R. K. Wayne. 2016. Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf. *Science Advances* 2: e1501714.

Campbell, D. R. 1985. Pollen and gene dispersal: the influences of competition for pollination. *Evolution* 39: 418–431.

Chou, J., A. Gupta, S. Yaduvanshi, R. Davidson, M. Nute, S. Mirarab, and T. Warnow. 2015. A comparative study of SVDquartets and other

coalescent-based species tree estimation methods. *BMC Genomics* 16(Suppl 10): S2.

Cooke, J. F., and R. E. Lee. 1966. Hybridization within and between *Achimenes* P. Br. and *Smithiantha* Kuntze (Gesneriaceae). *Baileya* 14: 92–101.

Cortis, P., N. J. Vereecken, F. P. Schiestl, M. R. Barone Lumaga, A. Scrugli, and S. Cozzolino. 2009. Pollinator convergence and the nature of species' boundaries in sympatric Sardinian *Ophrys* (Orchidaceae). *Annals of Botany* 104: 497–506.

Coyne, J. A., and H. A. Orr. 1997. "Patterns of speciation in *Drosophila*" revisited. *Evolution* 51: 295–303.

Currat, M., M. Reudi, R. J. Petit, and L. Excoffier. 2008. The hidden side of invasions: massive introgression by local genes. *Evolution* 62: 1908–1920.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 15: 2156–22158.

Davidson, R., P. Vachaspati, S. Mirarab, and T. Warnow. 2015. Phylogenomic species tree estimation in the presence of incomplete lineage sorting and horizontal gene transfer. 2015. *BMC Genomics* 16 (Suppl 10): S1.

Des Marais, D. L., and M. D. Rausher. 2010. Parallel evolution at multiple levels in the origin of hummingbird pollinated flowers in *Ipomoea*. *Evolution* 64: 2044–2054.

Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution* 28: 2239–2252.

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

Enright, A. J., S. Van Dongen, and C. A. Ouzounis. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research* 30: 1575–1584.

Esselstyn, J. A., C. H. Oliveros, M. T. Swanson, and B. C. Faircloth. 2017. Investigating difficult nodes in the placental mammal tree with expanded taxon sampling and thousands of ultraconserved elements. *Genome Biology and Evolution* 9: 2308–2321.

Feder, J. L., X. Xie, J. Rull, S. Velez, A. Forbes, B. Leung, H. Dambroski, et al. 2005. Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in *Rhagoletis*. *Proceedings of the National Academy of Sciences, USA* 102: 6573–6580.

Fenster, C. B., W. S. Armbruster, P. Wilson, M. R. Dudash, and J. D. Thomson. 2004. Pollination syndromes and floral specialization. *Annual Review of Ecology, Evolution, and Systematics* 35: 375–403.

Folk, R. A., J. R. Mandel, and J. V. Freudenstein. 2017. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. *Systematic Biology* 66: 320–337.

Fuchs, H. P. 1963. *Achimenes erecta*. *Acta Botanica Neerlandica* 12: 15.

Gascuel, O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution* 14: 685–695.

Gögler, J., J. Stökl, P. Cortis P, H. Beyrle, M.R. Barone Lumaga, S. Cozzolino, and M. Ayasse. 2015. Increased divergence in floral morphology strongly reduced gene flow in sympatric sexually deceptive orchids with the same pollinator. *Evolutionary Ecology* 29: 703–717.

Gompert, Z., and C. A. Buerkle. 2016. What, if anything, are hybrids: enduring truths and challenges associated with population structure and gene flow. *Evolutionary Applications* 9: 909–923.

Gordon, G. 1846. Some accounts of *Achimenes patens*, with its cultivation, and that of the species allied to it. *The Journal of the Horticultural Society of London* 1: 257–260.

Goulet, E. B., F. Roda, and R. Hopkins. 2017. Hybridization in plants: old ideas, new techniques. *Plant Physiology* 173: 65–78.

Grabherr, M.G., B.J. Haas, M. Yassour M, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, et al. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology* 29: 644.

Grant, V. 1971. Plant speciation. Columbia University Press, New York, NY.

Green, R. E., J. Krause, A. W. Biggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328: 710–722.

Gronau, I., M. J. Hubisz, B. Gulko, C. G. Danko, and A. Siepel. 2011. Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics* 43: 1031–1034.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8: 1494–1512.

Hanson, K. D., S. E. Brenner, and S. Dudoit. 2010. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Research* 36: e105.

Hoballah, M. E., T. Gübitz, J. Stuurman, L. Broger, M. Barone, T. Mandel, A. Dell'Olivo, et al. 2007. Single gene-mediated shift in pollination attraction in *Petunia*. *The Plant Cell* 19: 779–790.

Huson, D. H., R. Rupp, and C. Scornavacca. 2011. Phylogenetic networks: concepts, algorithms, and applications. Cambridge University Press, Cambridge, United Kingdom.

Johnson, M. A., D. K. Price, J. P. Price, and E. A. Stacy. 2015. Postzygotic barriers isolate sympatric species of *Cyrtandra* (Gesneriaceae) in Hawaiian montane forest understories. *American Journal of Botany* 102: 1870–1882.

Joly, S., P. A. McLenachan, and P. J. Lockhart. 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting. *The American Naturalist* 174: E54–E70.

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.

Kubatko, L. S. 2009. Identifying hybridization events in the presence of coalescence via model selection. *Systematic Biology* 58: 478–488.

Lewis, D., and L. K. Crowe. 1958. Unilateral interspecific incompatibility in flowering plants. *Heredity* 12: 233–256.

Lexer, C., M. F. Fay, J. A. Joseph, M. S. Nica, and B. Heinze. 2005. Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. *Molecular Ecology* 14: 1045–1057.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.

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.

Li, W., and A. Godzik. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658–1659.

Linder, C. R., and L. H. Rieseberg. 2004. Reconstructing patterns of reticulate evolution in plants. *American Journal of Botany* 91: 1700–1708.

Mallet, J., N. Besansky, and M. W. Hahn. 2016. How reticulated are species? *BioEssays* 38: 140–149.

Martén-Rodríguez, S., A. Almarales-Castro, and C. B. Fenster. 2009. Evaluation of pollination syndromes in Antillean Gesneriaceae: evidence for bat, hummingbird and generalized flowers. *Journal of Ecology* 97: 348–359.

Martén-Rodríguez, S., C. B. Fenster, I. Agnarsson, L. E. Skog, and E. A. Zimmer. 2010. Evolutionary breakdown of pollination specialization in a Caribbean plant radiation. *New Phytologist* 188: 403–417.

Martin, S. H., J. W. Davey, and C. D. Jiggins. 2015. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular Biology and Evolution* 32: 244–257.

McBreen, K., and P. J. Lockhart. 2006. Reconstructing reticulate evolutionary histories of plants. *Trends in Plant Science* 11: 398–404.

Meyer, M., M. Kircher, M.-T. Gansauge, H. Li, F. Racimo, S. Mallick, J. G. Schraiber, et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338: 222–226.

Mirarab, S., M. Shamsuzzoha Bayzid, and T. Warnow. 2014. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. *Systematic Biology* 65: 366–380.

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

Moore, T. 1859. Report on the varieties of Achimenes. *Proceedings of the Royal Horticultural Society London* 1: 454–464.

Moyle, L. C., M. S. Olson, and P. Tiffin. 2004. Patterns of reproductive isolation in three angiosperm genera. *Evolution* 58: 1195–1208.

Payseur, B. A., and L. H. Rieseberg. 2016. A genomic perspective on hybridization and speciation. *Molecular Ecology* 25: 2337–2360.

Perret, M., A. Chautems, R. Spichiger, T. G. Barraclough, and V. Savolainen. 2007. The geographical pattern of speciation and floral diversification in the Neotropics: the tribe Sinningieae (Gesneriaceae) as a case study. *Evolution* 62: 1641–1660.

Petit, R. J., C. Bodénès, A. Ducousso, G. Roussel, and A. Kremer. 2004. Hybridization as a mechanism of invasion in oaks. *New Phytologist* 161: 151–164.

Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5: e9490.

Ramírez Roa, M.A. 1987. Revision de Achimenes (Gesneriaceae). Ph.D. dissertation, Universidad Nacional Autonoma de Mexico, Mexico City, Mexico.

Ramsey, J., H. D. Bradshaw, and D. W. Schemske. 2003. Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* 57: 1520–1534.

Roalson, E. H., J. K. Boggan, L. E. Skog, and E. A. Zimmer. 2005. Untangling Gloxinieae (Gesneriaceae). I. Phylogenetic patterns and generic boundaries inferred from nuclear, chloroplast, and morphological cladistic datasets. *Taxon* 54: 389–410.

Roalson, E. H., and W. R. Roberts. 2016. Distinct processes drive diversification in different clades of Gesneriaceae. *Systematic Biology* 65: 662–684.

Roalson, E. H., L. E. Skog, and E. A. Zimmer. 2003. Phylogenetic relationships and the diversification of floral form in *Achimenes* (Gesneriaceae). *Systematic Botany* 28: 593–608.

Roalson, E. H., L. E. Skog, and E. A. Zimmer. 2008. Untangling Gloxinieae (Gesneriaceae). II. Reconstructing biogeographic patterns and estimating divergence times among New World continental and island lineages. *Systematic Botany* 33: 159–175.

Roberts, W. R., and E. H. Roalson. 2017. Comparative transcriptome analyses of flower development in four species of *Achimenes* (Gesneriaceae). *BMC Genomics* 18: 240.

Roda, F., F. K. Mendes, M. W. Hahn, and R. Hopkins. 2017. Genomic evidence for gene flow during reinforcement in Texas *Phlox*. *Molecular Ecology* 26: 2317–2330.

Rosenzweig, B. K., J. B. Pease, N. J. Besansky, and M. W. Hahn. 2016. Powerful methods for detecting introgressed regions from population genomic data. *Molecular Ecology* 25: 2387–2397.

Savolainen, V., M.-C. Anstett, C. Lexer, I. Hutton, J. J. Clarkson, M. V. Norup, M. P. Powell, et al. 2006. Sympatric speciation in palms on an oceanic island. *Nature* 441: 210.

Schmidt-Lebuhn, A. N., B. Keller, J. M. De Vos, and E. Conti. 2012. Phylogenetic analysis of *Primula* section *Primula* reveals rampant non-monophyly among morphologically distinct species. *Molecular Phylogenetics and Evolution* 65: 23–34.

Scopece, G., A. Muscacchio, A. Widmer, and S. Cozzolino. 2007. Patterns of reproductive isolation in Mediterranean deceptive orchids. *Evolution* 61: 2623–2642.

Servedio, M. R., and M. A. F. Noor. 2003. The role of reinforcement in speciation: theory and data. *Annual Review of Ecology, Evolution, and Systematics* 34: 339–364.

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

Solís-Lemus, C., and C. Ané. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genetics* 12: e1005896.

Solís-Lemus, C., P. Bastide, and C. Ané. 2017. PhyloNetworks: a package for phylogenetic networks. *Molecular Biology and Evolution* 34: 3292–3298.

Solís-Lemus, C., M. Yang, and C. Ané. 2016. Inconsistency of species tree methods under gene flow. *Systematic Biology* 65: 843–851.

Soliva, M., and A. Widmer. 1999. Genetic and floral divergence among sympatric populations of *Gymnadenia conopsea* sl (Orchidaceae) with different flowering phenology. *International Journal of Plant Sciences* 160: 897–905.

Soliva, M., and A. Widmer. 2003. Gene flow across species boundaries in sympatric, sexually deceptive *Ophrys* (Orchidaceae) species. *Evolution* 57: 2252–2261.

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

Stebbins, G. L. 1950. Variation and evolution in plants. Columbia University Press, New York.

Suyama, M., D. Torrents, and P. Bork. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* 34 (Suppl 2): W609–612.

Tiffin, P., M. S. Olson, and M. C. Moyle. 2001. Asymmetrical crossing barriers in angiosperms. *Proceedings of the Royal Society B* 268: 861–867.

Vachaspati, P., and T. Warnow. 2015. ASTRID: accurate species trees from internode distances. *BMC Genomics* 16: S3.

Vallejo-Marín, M., and S. J. Hiscock. 2016. Hybridization and hybrid speciation under global change. *New Phytologist* 211: 1170–1187.

Vargas, O. M., E. M. Ortiz, and B. B. Simpson. 2017. Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). *New Phytologist* 214: 1736–1750.

Vereecken, N. J., S. Cozzolino, and F. P. Schiestl. 2010. Hybrid floral scent novelty drives pollinator shifts in sexually deceptive orchids. *BMC Evolutionary Biology* 10: 103.

Ward, B. J., and C. Oosterhout. 2015. HYBRIDCHECK: software for the rapid detection, visualization and dating of recombinant regions in genome sequence data. *Molecular Ecology Resources* 16: 534–539.

Wessinger, C. A., and M. D. Rausher. 2014. Predictability and irreversibility of genetic changes associated with flower color evolution in *Penstemon barbatus*. *Evolution* 68: 1058–1070.

Whitney, K. D., R. A. Randell, and L. H. Rieseberg. 2006. Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. *The American Naturalist* 167: 794–807.

Wiehler, H. 1983. A synopsis of the neotropical Gesneriaceae. *Selbyana*. 6: 1–219.

Wiehler, H. 1992. New species of Gesneriaceae from the Neotropics. *Phytologia* 73: 220–241.

Wolf, P. G., D. R. Campbell, N. M. Waser, S. D. Sipes, T. R. Toler, and J. K. Archibald. 2001. Tests of pre- and postpollination barriers to hybridization between sympatric species of *Ipomopsis* (Polemoniaceae). *American Journal of Botany* 88: 213–219.

Yang, Y., and S. A. Smith. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Molecular Biology and Evolution* 31: 3081–3092.

APPENDIX 1. **Taxa used for the analyses**; Collector; *Voucher number*; WS = Washington State University.

Achimenes admirabilis Wiehler; W.R. Roberts; *WR0569*; WS. ***Achimenes antirrhina*** (DC.) C.V. Morton; W.R. Roberts; *WR0570*; WS. ***Achimenes candida*** Lindl.; W.R. Roberts; *WR0571*; WS. ***Achimenes cettoana*** H.E. Moore; W.R. Roberts; *WR0155*; WS. ***Achimenes erecta*** (Lam.) H.P. Fuchs; W.R. Roberts; *WR0156*; WS. ***Achimenes grandiflora*** (Schltdl.) DC.; W.R. Roberts; *WR0572*; WS. ***Achimenes longiflora*** DC.; W.R. Roberts; *WR0573*; WS. ***Achimenes misera*** Lindl.; W.R. Roberts; *WR0157;* WS. ***Achimenes patens*** Benth.; W.R. Roberts; *WR0158*; WS. ***Achimenes pedunculata*** Benth.; W.R. Roberts; *WR0574*; WS. ***Eucodonia verticillata*** (M. Martens & Galeotti) Wiehler; W.R. Roberts; *WR0575*; WS. ***Gesneria cuneifolia*** (DC.) Fritsch; W.R. Roberts; *WR0576*; WS.