Syst. Biol. 70(1):162–180, 2021

DOI:10.1093/sysbio/syaa049
Advance Access publication July 3, 2020

Out of Sight, Out of Mind: Widespread Nuclear and Plastid-Nuclear Discordance in the Flowering Plant Genus *Polemonium* (Polemoniaceae) Suggests Widespread Historical Gene Flow Despite Limited Nuclear Signal

JEFFREY P. ROSE[1,2,*], CASSIO A. P. TOLEDO[3], EMILY MORIARTY LEMMON[4], ALAN R. LEMMON[5] AND KENNETH J. SYTSMA[1]

[1]*Department of Botany, University of Wisconsin-Madison, Madison, WI 53706, USA;* [2]*Department of Biology, University of Nebraska at Kearney, Kearney, NE 68849, USA;* [3]*Programa de Pós-Graduação em Biologia Vegetal, Instituto de Biolgia, Universidade Estadual de Campinas-UNICAMP, Rua Monteiro Lobato, 255, Campinas, SP. CEP: 13083-862, Brazil;* [4]*Department of Biological Science, Florida State University, Tallahassee, FL 32306, USA;* [5]*Department of Scientific Computing, Florida State University, Tallahassee, FL 32306, USA*
[*]*Correspondence to be sent to: Department of Botany, University of Wisconsin-Madison, Madison, WI 53706, USA;*
E-mail: jeffrey.rose@wisc.edu

Received 10 February 2019; reviews returned 10 June 2020; accepted 23 June 2020
Associate Editor: Bryan Carstens

Abstract.—Phylogenomic data from a rapidly increasing number of studies provide new evidence for resolving relationships in recently radiated clades, but they also pose new challenges for inferring evolutionary histories. Most existing methods for reconstructing phylogenetic hypotheses rely solely on algorithms that only consider incomplete lineage sorting (ILS) as a cause of intra- or intergenomic discordance. Here, we utilize a variety of methods, including those to infer phylogenetic networks, to account for both ILS and introgression as a cause for nuclear and cytoplasmic-nuclear discordance using phylogenomic data from the recently radiated flowering plant genus *Polemonium* (Polemoniaceae), an ecologically diverse genus in Western North America with known and suspected gene flow between species. We find evidence for widespread discordance among nuclear loci that can be explained by both ILS and reticulate evolution in the evolutionary history of *Polemonium*. Furthermore, the histories of organellar genomes show strong discordance with the inferred species tree from the nuclear genome. Discordance between the nuclear and plastid genome is not completely explained by ILS, and only one case of discordance is explained by detected introgression events. Our results suggest that multiple processes have been involved in the evolutionary history of *Polemonium* and that the plastid genome does not accurately reflect species relationships. We discuss several potential causes for this cytoplasmic-nuclear discordance, which emerging evidence suggests is more widespread across the Tree of Life than previously thought. [Cyto-nuclear discordance, genomic discordance, phylogenetic networks, plastid capture, Polemoniaceae, *Polemonium*, reticulations.]

The increasing use of genomic-level data sets from transcriptomes (Wickett et al. 2014), sequence capture probes such as ultraconserved elements and anchored hybrid enrichment (AHE) (Faircloth et al. 2012; Lemmon et al. 2012), and single nucleotide polymorphisms (SNPs) (Eaton and Ree 2013) has provided a plethora of data for resolving both deep and shallow scale relationships across all branches of the Tree of Life. However, analyzing these data opens the possibility for widespread conflicting phylogenetic signal among loci due to processes such as gene duplication, incomplete lineage sorting (ILS), and gene flow (both hybridization and introgression), necessitating a consideration of the causes of such discordance in analyses. Emerging evidence suggests that gene flow between lineages is important in the evolutionary history of both plants and animals (Cui et al. 2013; Folk et al. 2017, 2018; McVay et al. 2017a; Vargas et al. 2017; Gernandt et al. 2018; Grummer et al. 2018). Discordance among nuclear gene trees may not be the only complicating factor, as there also may be discordance among the nuclear genome and organellar genome(s) (cytoplasmic-nuclear or cyto-nuclear discordance). In plants, this discordance between the nuclear and (usually) the plastid (generally referred to in the literature as chloroplast) genomes has been interpreted as introgressive plastid capture, although uncertainty in phylogenetic inference as an explanation often is not discussed (Smith and Sytsma 1990; Rieseberg and Soltis 1991; Fehrer et al. 2007; Rojas-Andres et al. 2015; Yi et al. 2015). Despite their smaller effective population sizes, organellar genomes are also susceptible to ILS, although this is only beginning to be tested (Folk et al. 2017; Lee-Yaw et al. 2019), and plastid haplotypes that arise in a population via introgression may be maintained in the population as a result of natural selection, genetic drift, or asexual reproduction (Levin 2003; Tsitrone et al. 2003; Bock et al. 2014; Lee-Yaw et al. 2019). Despite the potential importance for gene flow within the evolutionary history of clades, most existing methods for phylogenetic analysis either assume no genomic discordance by concatenating loci or only assume a coalescent model in which all genomic discordance is due to ILS (Heled and Drummond 2010; Liu and Edwards 2010; Chifman and Kubatko 2014; Mirarab and Warnow 2015). However, new computational tools allow the inference of phylogenetic networks that account for both ILS and gene flow (Than et al. 2008; Solís-Lemus and Ané 2016; Kamneva and Rosenberg 2017; Solís-Lemus et al. 2017), although these programs are still currently limited by computational time for even 15 terminals (Solís-Lemus and Ané 2016:Fig. 6; but see Yu and Nakhleh 2015).

Here, we evaluate the importance of gene flow in the evolutionary history of the flowering plant genus *Polemonium* L. (Polemoniaceae), a lineage with documented cases of interspecific hybridization under greenhouse conditions (Ostenfeld 1929; Clausen 1931, 1967) and with historical gene flow hypothesized based on morphological intergradation (Davidson 1950; Anway 1968; Sledge and Anway 1970; Grant 1989).

We reconstruct the evolutionary history of *Polemonium* using both coalescent and phylogenetic network approaches with data gathered from AHE of up to 499 nuclear genes. In addition, we examine the genealogical history of the plastid genome (ptDNA, often referred to in the literature as cpDNA) and nuclear ribosomal DNA (nrDNA) independent of the low or single copy nuclear genes, both obtained from the AHE raw reads. Importantly, this study includes multiple samples for most of the morphological species in *Polemonium*, an approach undertaken in only a few studies using data of this type (cf. Folk et al. 2017; Morales-Briones et al. 2018; Stubbs et al. 2018). In studies containing multiple individuals per morphological species, these proposed taxa are not always reciprocally monophyletic in both the nuclear and plastid genomes (Folk et al. 2017; Pham et al. 2017; Lee-Yaw et al. 2019).

The Genus Polemonium

Polemonium is a genus of temperate, largely diploid (Grant 1989), primarily perennial herbs. It is the most widespread genus in Polemoniaceae, mostly occurring in Eurasia and North America, with one species (*P. micranthum*) disjunct between western North America and southern South America (Davidson 1950; Grant 1959; Johnson and Porter 2017). Infraspecific variability in floral traits and pollination mode has spurred research on trait heritability and natural selection by pollinators in several species (Galen and Kevan 1980; Galen and Newport 1987; Galen et al. 1987, 1991; Kulbaba et al. 2013), necessitating a robust understanding of the evolutionary history of the genus. While *Polemonium* consistently has been recovered as monophyletic (Johnson et al. 1996, 2008; Prather et al. 2000), the placement of the genus in Polemoniaceae is ambiguous (Steele and Vilgalys 1994; Johnson et al. 1996, 2008; Porter 1996; Porter and Johnson 1998; Prather et al. 2000). The currently accepted hypothesis is a sister relationship of *Polemonium* to tribe Phlocideae (Johnson et al. 2008), but support for this relationship varies considerably between optimality criteria, with posterior probabilities (PP) from Bayesian inference consistently finding strong support but maximum parsimony finding contrastingly low support. Based on our current understanding of the taxonomy of *Polemonium*, the genus radiated ca. 7.3–10.8 Ma (Landis et al. 2018; Rose et al. 2018) and consists of at least 27 species and as many as 38, although new species continue to be recognized (Murray and Elven 2011; Irwin et al. 2012; Stubbs and Patterson 2013). Uncertainty in the number of species has largely depended on the circumscription of closely related Eurasian populations, which continue to be poorly understood (Davidson 1950; Hultén 1971; Vassiljev 1974; Feulner et al. 2001). Davidson (1950) was the first to tackle interspecific relationships in *Polemonium* in a systematic way. Based on his taxonomic conclusions, subsequent workers have focused on four major species complexes: 1) *Polemonium caeruleum*: erect plants of northern wetlands with paniculate inflorescences, coplanar leaflets, and yellow anthers; 2) *P. pulcherrimum*: caespitose mid- to high-elevation plants with corymbose inflorescences, coplanar leaflets, and white anthers (Wherry 1967; Grant 1989); 3) *P. viscosum*: caespitose plants of alpine environments with paniculate to spicate inflorescences, densely glandular-pubescent verticillate leaflets, and white to orange anthers (Grant 1989); and 4) *P. foliosissimum*: erect plants of midelevation conifer forests with corymbose inflorescences, coplanar leaflets, and yellow anthers (Anway 1968).

Although not explicitly stated, both Wherry (1942) and Grant (1989) implied an hypothesis that the *P. caeruleum* species complex exhibits the most plesiomorphic character states in the genus. In addition, Grant (1989) explicitly hypothesized that alpine species of *Polemonium* have evolved multiple times from low to midelevation species. Molecular phylogenetic investigations into *Polemonium* have provided limited but important insights into the evolution and taxonomy of the group. These studies used sequence data (de Geofroy 1998; Timme 2001; Irwin et al. 2012) or amplified fragment length polymorphisms (AFLPs) (Worley et al. 2009). When multiple individuals per species have been sampled, these studies generally support the monophyly of species recognized by Davidson (1950). However, inferring interspecific relationships in *Polemonium* has been exceedingly difficult with weak to nonexistent support along the backbone of the genus (de Geofroy 1998; Timme 2001; Worley et al. 2009; Irwin et al. 2012). Furthermore, phylogenetic reconstructions based on sequence and AFLP data have been strongly conflicting. Taken as a whole, previous results suggest that a rapid radiation, recent divergence, or widespread gene flow (either present day or historical), separately or in concert, may contribute to a complex evolutionary history of *Polemonium*.

Materials and Methods

Data Availability

Scripts for nuclear locus assembly and processing are available from the Dryad Digital Repository (https://doi.org/10.5061/dryad.5hqbzkh2r) as well as GitHub (https://github.com/jrosaceae/polemonium_AHE). Raw reads have been deposited in the NCBI Sequence Read Archive (BioProject PRJNA638804). Tree files and pre- and postfiltered alignments for all analyses are available from the Dryad Digital Repository.

Taxonomic Sampling

Taxonomic sampling was designed so as to both elucidate the phylogenetic placement of *Polemonium*, as well as examine the evolutionary history of the genus itself. In addition to *Polemonium*, we sampled one species each from the two tribes possibly sister to the clade: Gilieae (*Gilia capitata*) and Phlocideae (*Leptosiphon montanus*), with tribe Loeselieae (*Ipomopsis aggregata*) treated as the outgroup based on the results of Johnson et al. (2008). Infrageneric sampling in *Polemonium* was tailored to meet our primary goal of resolving deeper relationships

in the genus, although thorough taxonomic sampling was also desirable. We therefore focused our taxonomic sampling on major lineages indicated by previous phylogenetic studies. As previous studies have lacked sufficient sampling of the *P. foliosissimum* and *P. viscosum* species complexes, we also targeted these putative clades in which additional phylogenetic diversity may exist. For New World *Polemonium*, leaf tissue was obtained for all but one North American species and several Mexican species, totaling 45 accessions of *Polemonium* (Table 1). In many cases, multiple geographically distant individuals were sampled for each morphological species. Genomic DNA was extracted from silica-dried leaf tissue and two herbarium samples using the Qiagen DNeasy Plant Mini Kit (Valencia, CA, USA) following the recommended protocol.

Library Preparation and Sequencing

We utilized an anchored phylogenomics approach, AHE, in collaboration with the Center for Anchored Phylogenomics at Florida State University (www.anchoredphylogeny.com). This method targets conserved "anchor" regions in the nuclear genome and generates data for ~500 loci including exons and introns (Lemmon et al. 2012; Prum et al. 2015). The pipeline employed here uses probes designed from the genomes of 25 broadly sampled angiosperms and targets 499 moderately conserved exons with some intronic regions (Buddenhagen et al. 2016). This pipeline has been used successfully to infer infrageneric relationships across distantly related groups (Cardillo et al. 2017; Léveillé-Bourret et al. 2017; Mitchell et al. 2017; Kriebel et al. 2019). Library preparation was carried out on the extracted genomic DNA. DNA was sonicated to a fragment size of between 200 and 600 bp before indexed library preparation and indexing following a modified protocol of Meyer and Kircher (2010). Library preparation was performed on a Beckman Coulter FXp liquid handling robot. Indexed samples were pooled and enriched using the Angiosperm v.1 enrichment kit (Buddenhagen et al. 2016). Sequencing was performed on one PE150 Illumina HiSeq 2500 lane (~50 Gb total yield) at the Translational Science Laboratory, College of Medicine, Florida State University.

Nuclear DNA Assembly and Alignment

Paired reads were merged before assembly, following Rokyta et al. (2012). Reads were mapped to the probe regions using *Arabidopsis thaliana, Billbergia nutans*, and *Carex lurida* as references and the assembly was extended into the flanks using a quasi-de novo assembly approach described by Hamilton et al. (2016). Consensus sequences were calculated for each assembly cluster with contigs based on fewer than 20 reads removed. The orthology of consensus sequences at each locus was assessed using a pairwise distance matrix among homologs and used to cluster sequences with a neighbor-joining algorithm to assess if gene duplication occurred prior to or following the crown of the clade. To minimize cases of miscalled orthology, we applied a conservative filter by removing clusters containing evidence of gene duplication if they contained <90% of samples (see Hamilton et al. 2016 for details). Sequences in each orthologous cluster were aligned in MAFFT v. 7.023b (Katoh and Standley 2013). To remove poorly aligned regions, raw alignments were trimmed and masked using the following procedure from Prum et al. (2015). Sites with the same character in >40% of sequences were considered "conserved." A 20 bp sliding window was then moved across the alignment, and regions with <14 characters matching the common base at the corresponding conserved site were masked. Sites with <25 unmasked bases were removed. Finally, the masked alignments were inspected by eye in Geneious v. 10.2.3 (Kearse et al. 2012). Regions considered obviously misaligned or likely paralogous were removed and poorly aligned sections in a given alignment were deleted.

Plastid and Nuclear Ribosomal DNA Assembly and Alignment

To obtain the ptDNA and nrDNA, we used Geneious to map all recovered forward and reverse reads to reference sequences. For the ptDNA, we used the whole plastid genome of *Saltugilia latimeri* (Polemoniaceae, GenBank accession KT921175 from Landis et al. 2016) as a reference. For the nrDNA, we used a concatenated ETS/18S/ITS/26S reference compiled from published and unpublished Sanger-derived sequences from multiple *Polemonium* species. Raw reads were trimmed and assembled using iterative refinement of up to five times with the default Geneious mapper and medium sensitivity. As the reference plastid genome is fairly distantly related, we assembled plastomes using a multistep process. First, reads were mapped to the *Saltugilia* reference. Next, we used the sample with the most reads mapped to it (*P. pectinatum*) as a second reference to map published and unpublished Sanger-derived reads to improve our reference. We then remapped all reads to this *Polemonium* reference using the same parameters mentioned above. For the ptDNA, consensus sequences were generated using the strict consensus approach. If coverage for a particular site was <7, the consensus nucleotide was scored as a gap. Unmapped regions were treated as missing data and reads mapped to multiple positions were excluded from consensus calculations. To account for potential paralogous copies of the nrDNA (Buckler et al. 1997; Poczai and Hyvönen 2010), we applied a more stringent consensus criterion of base calls matching 65% of sequences. Coverage and unmapped regions were treated as above. Sequences were aligned using MAFFT with default parameters. After alignment, ambiguously aligned or called regions were removed by hand. The mitochondrial genome was not extracted as the closest

TABLE 1. Voucher and sequence information for taxa of *Polemonium* and outgroups used in this study

Taxon	Voucher	Provenance	Reads (10^6)	No. of loci	Mean coverage
Gilia capitata Sims	Rose s.n. (WIS)	Cultivated	12.57	448	274
Ipomopsis aggregata (Pursh) V.E. Grant	Rose 14-221 (WIS)	ID, USA	7.34	359	76
Leptosiphon montanus (Greene) J.M. Porter and L.A. Johnson	Rose s.n. (WIS)	Cultivated	4.14	350	80
P. acutiflorum Willd.	Bennett 14-0056 (WIS)	BC, CAN	7.49	402	176
P. boreale Adams	Bennett 14-0081 (WIS)	YT, CAN	4.73	413	140
P. brandegeei (A. Gray) Greene	Rose 14-245 (WIS)	CO, USA	7.73	400	170
P. californicum Eastw.	Stubbs 18 (SFSU)	CA, USA	6.96	411	164
P. californicum	JR009 (WIS)	WA, USA	2.49	371	89
P. carneum A. Gray	Stubbs 10 (SFSU)	CA, USA	5.09	367	116
P. carneum	Stubbs 8 (SFSU)	OR, USA	3.52	327	63
P. chartaceum Mason	Stubbs 23 (SFSU)	CA, USA	7.93	398	179
P. chartaceum	Stubbs 24 (SFSU)	CA, USA	6.26	378	115
P. confertum A. Gray	Rose 14-240 (WIS)	CO, USA	5.76	416	221
P. delicatum Rydb.	Rose 14-239 (WIS)	CO, USA	6.22	401	159
P. delicatum	Rose 12-233b (WIS)	UT, USA	7.06	396	142
P. eddyense Stubbs	Stubbs 15 (SFSU)	CA, USA	8.81	405	159
P. elegans Greene	Rose 13-53 (WIS)	WA, USA	5.01	398	124
P. elusum J.J. Irwin and R.L. Hartm.	Irwin 5148 (RM)	ID, USA	6.22	333	78
P. elusum	Irwin 5496 (RM)	ID, USA	4.89	355	80
P. eximium Greene	Stubbs 22 (SFSU)	CA, USA	3.93	247	51
P. eximium	Stubbs 14 (SFSU)	CA, USA	5.84	378	129
P. foliosissimum A. Gray var. *alpinum* Brand	Rose 14-230 (WIS)	UT, USA	5.14	395	123
P. foliosissimum var. *flavum* (Greene) Anway	Rose 14-260 (WIS)	AZ, USA	7.05	366	120
P. foliosissimum A. Gray var. *f*.	Rose 14-256 (WIS)	NM, USA	6.08	372	145
P. foliosissimum var. *molle* (Greene) Anway	Rose 14-250 (WIS)	CO, USA	5.77	378	143
P. foliosissimum var. *molle*	Rose 14-235a (WIS)	UT, USA	4.78	333	75
P. foliosissimum var. *molle*	Rose 14-254a (WIS)	NM, USA	4.86	369	120
P. foliosissimum var. nov. (aff. *P. filicinum* Greene)	Rose 14-258 (WIS)	AZ, USA	4.40	389	97
P. grandiflorum Benth.	MEXU630068	Mexico	5.37	210	27
P. micranthum Benth.	Stubbs 4 (SFSU)	CA, USA	5.33	405	151

TABLE 1. (Continued)

Taxon	Voucher	Provenance	Reads (10^6)	No. of loci	Mean coverage
P. nevadense Wherry	Rose 14-219 (WIS)	NV, USA	4.26	349	78
P. occidentale Greene subsp. *o.*	Rose 14-226 (WIS)	ID, USA	4.00	351	89
P. occidentale subsp. *o.*	Stubbs 17 (SFSU)	CA, USA	8.00	397	185
P. occidentale subsp. *lacustre* Wherry	MIN816873	MN, USA	3.22	372	117
P. pauciflorum S. Wats.	Fishbein 2438 (ARIZ)	AZ, USA	8.18	357	110
P. pectinatum Greene	JR95 (WIS)	WA, USA	9.10	369	143
P. pulcherrimum subsp. *lindleyi* (Wherry) V.E. Grant	Bennett 14-0039 (WIS)	YT, CAN	6.15	388	148
P. pulcherrimum Hook. subsp. *p.*	DiNicola 2013-53 (WIS)	ID, USA	7.97	396	173
P. pulcherrimum subsp. *p.*	Stubbs 20 (SFSU)	CA, USA	4.70	348	80
P. pulcherrimum var. *shastense* (Eastw.) Stubbs	Stubbs 16 (SFSU)	CA, USA	8.24	386	149
P. reptans L. var. *r.*	Rose 13-44 (WIS)	OH, USA	7.21	331	93
P. reptans var. *villosum* E.L. Braun	JR11 (WIS)	OH, USA	4.76	386	113
P. vanbruntiae Britton	JRVB2 (WIS)	VT, USA	6.16	372	129
P. vanbruntiae	JR002 (WIS)	NY, USA	7.41	382	198
P. vanbruntiae	JRVB3 (WIS)	QB, CAN	5.30	376	111
P. vanbruntiae	Rose 13-46 (WIS)	WV, USA	4.25	368	116
P. viscosum Nutt.	Stubbs 25 (SFSU)	NV, USA	7.16	396	147
P. viscosum	Rose 14-232 (WIS)	UT, USA	7.63	411	205
P. viscosum	Rose 12-243 (WIS)	NM, USA	5.44	376	104

reference genome available was too distantly related to extract a large and phylogenetically informative amount of the genome (*Vaccinium macrocarpon*, Ericaceae).

Checking for Tree-Like Structure

We began by analyzing the 45 accessions of *Polemonium* alone using 142 loci without any missing ingroup data. We first estimated a phylogenetic tree from a concatenated version of this data set using maximum likelihood in RAxML (Stamatakis 2014) under the GTR + G model of sequence evolution. We conducted 10 separate runs to find the most likely tree as well as 500 rapid bootstrap (BS) replicates.

Next, to examine if an assumption of underlying tree-like structure was appropriate for our data set, we employed the Tree Incongruence Checking in R (TICR) pipeline (Stenz et al. 2015) to obtain concordance factors (CFs) for each possible quartet and to use these data to infer an optimal population tree for *Polemonium*. Briefly, for the TICR pipeline, we ran MrBayes v. 3.2.6 (Ronquist et al. 2012) on all genes using a batch script. The best gene tree for each locus was inferred under the GTR + I + G model of sequence evolution using 3 runs of 3 chains each for 6 million generations with sampling every 6000 generations with a chain temperature of 0.4 and a 30% burnin. Standard deviations of split frequencies from MrBayes were usually <0.010. Following the MrBayes analysis, Bayesian concordance analysis on the posterior sample of gene trees was conducted in BUCKy v. 1.4.4 (Ané et al. 2007; Larget et al. 2010) with 100 000 postburnin generations. This analysis calculates all possible quartets and prunes the MrBayes gene trees to all but the four terminals of interest. Then, BUCKy is run on each pruned gene tree to generate a table of all quartet CFs and their standard errors. The parameter α in BUCKy is the a priori amount of discordance between gene trees. Although empirical studies have shown that CFs and branch lengths in coalescent units are robust to this prior (Cranston et al. 2009; McVay et al. 2017a; Everson et al. 2016; Crowl et al. 2017; Folk et al. 2017), we used three values of α (0.5, 1, and 2) to ensure the robustness of our results as in Folk et al. (2017).

Next, Quartet MaxCut (Snir and Rao 2012) was used to generate a starting population tree based on these CFs. The TICR test (Stenz et al. 2015) was employed using the R package phylolm (Ho and Ané 2014) to examine if any or all edges in the optimal population tree could be modeled under panmixia using the test.one.species.tree and stepwise.test.tree functions. The stepwise.test.tree function starts from either a fully resolved tree or a fully collapsed tree and uses a stepwise search of edges to determine the best resolution of the tree. We conducted two searches with a maximum number of iterations of 10 000 each with forward and backwards searches at each step and considering 1000 population trees at each step. One search started from panmixia and another started from a fully resolved tree.

Coalescent Species Tree Estimation

Because methods that explicitly account for reticulate evolution can only realistically work on many fewer terminals than those for which we collected data, we first explored relationships under ILS alone to investigate the phylogenetic position of *Polemonium* within Polemoniaceae as well as phylogenetic relationships within all of our samples of *Polemonium*. This data set consisted of all 45 accessions of *Polemonium* plus three outgroups (*Gilia capitata, Ipomopsis aggregata*, and *Leptosiphon montanus*). For this data set, we allowed up to 24 missing sequences per locus (50% missing sequences), which resulted in a data set of 48 terminals and 316 loci (123 loci without missing data for all taxa, 142 loci without missing data for *Polemonium*). We employed two methods for inferring a coalescent species tree. First, we used SVDquartets (Chifman and Kubatko 2014) as implemented in PAUP* version 4.0a163 (Swofford 2002) on the entire data set, including those with missing data. SVDquartets is primarily designed for use with SNP data and the theoretical background for this assumes that each SNP is unlinked. However, Chifman and Kubatko (2014) demonstrated that the statistical power of this method is not greatly affected by using linked SNPs. We evaluated all possible quartets with ambiguities treated as missing data. Support for the SVDquartets population tree was assessed with 500 nonparametric BS replicates. As an additional method for reconstructing phylogenies under ILS alone, we also used ASTRAL-III (Mirarab et al. 2014; Zhang et al. 2017). Using a batch script, we first conducted maximum likelihood as implemented in RAxML on each aligned locus under the GTR + G model of sequence evolution with 100 rapid BS replicates. We then ran these gene trees in ASTRAL-III and accounted for uncertainty in the estimated population tree using 100 BS replicates from the RAxML BS trees, as well as calculating ASTRAL quadripartition support (QS): that is, the fraction of quartet trees in the set of gene trees for a particular quadripartition (Sayyari and Mirarab 2016). Lastly, we conducted a polytomy test to see if the null hypothesis that a branch is a polytomy could be rejected using the −t 10 option in ASTRAL (Sayyari and Mirarab 2018). This test uses the fact that under panmixia, the three topologies of an unrooted quartet of species each should be present in equal frequency and should follow a chi-squared distribution. We set our α value for this test to a less conservative cutoff of 0.10 based on Figure 7 of Sayyari and Mirarab (2018), which suggests that in phylogenies reconstructed using less than several thousand genes and which contain many short branches (in coalescent units) there may not be enough resolution for this test.

Phylogenetic Network Estimation

To explore the possibility for gene flow as a cause of discordance in *Polemonium*, we utilized 17 exemplars of *Polemonium*, a sampling which represents a good compromise between taxonomic coverage and computational cost. For this data set, we allowed up to six missing sequences per locus (up to 33% missing sequences/locus). Missing sequences were scored as a single "N" for each individual missing the locus. Individuals for this data set were selected based on taxonomic representativeness and having high numbers of loci captured. We first used the TICR pipeline as described above in the subsection "Checking for Tree-Like Structure," but with the MrBayes parameters changed to 2 million generations with sampling every 1000 generations. Standard deviations of split frequencies from MrBayes were usually less than 0.010 and always less than 0.017. We ran BUCKy on each quartet for 1 million postburnin generations, again testing values of $\alpha = 0.5$, 1, and 2, summarizing the results as an optimal population tree in Quartet MaxCut. As for the 45-taxon data set, we used phylolm to test for an unresolved or partially resolved population tree (see "Checking for Tree-Like Structure" section). Following these steps, we analyzed the BUCKy CFs and the Quartet MaxCut population tree as a starting tree using the SNaQ function in the Julia package PhyloNetworks (Solís-Lemus and Ané 2016; Solís-Lemus et al. 2017) to examine the contribution of ILS and reticulation to the phylogenetic history of *Polemonium*. This package uses maximum pseudolikelihood to fit a network while also accounting for ILS. PhyloNetworks considers quartet topologies only and does not take into account information from branch lengths in individual gene trees. Furthermore, PhyloNetworks assumes a one-level network: a network where each hybrid node only has one lineage transferring genetic material horizontally. We first tested the fit of models allowing from 0 to 5 reticulation events (h) and compared models using their pseudolikelihood score. The best network model was selected by examining at what value of h the pseudolikelihood score plateaus, as per the recommendation of Solís-Lemus et al. (2017). For each value of h, the best network over 20 search replicates was selected. We examined branch support on the best phylogenetic network using the bootsnaq function with 100 runs of 10 replicates each.

Cyto-nuclear Discordance

NrDNA and ptDNA for the entire 48-taxon set were analyzed with MrBayes under the GTR + I + G model of sequence evolution using 4 runs of 3 chains each for 25 million generations with sampling every 100 000 generations and a 30% burnin. To examine if any observed topological discordance between the estimated population tree based on nuclear data and individual genome trees can be explained by ILS alone, we implemented two nested approaches. First, we modified the approach of Olave et al. (2017) that takes a given species tree with branch lengths in $4N_e$ units and simulates gene trees using the program MS (Hudson 2002). The number of extra lineages between the species tree and gene trees is then calculated using the function deep_coal_count in Phylonet v. 2.4 (Than and Nakhleh 2009). The approach of Olave et al. (2017) generates gene trees under different values of the migration (m) parameter and then compares the likelihood of simulated gene trees under various values of m to actual gene trees to infer gene flow between lineages. However, this method assumes that branch lengths are accurately estimated. We used the best fitting major network topology from our 17-taxon phylogenetic network analysis and the optimal population tree from our 45-taxon analysis to simulate gene trees in MS for both the nuclear and plastid genomes. To obtain correct branch lengths for nuclear gene trees, we multiplied branch lengths in coalescent units by a factor of two to get branch lengths in $4N_e$ units. To simulate gene trees for the plastid genome, branch lengths in this starting nuclear species tree were divided by a factor of four before simulating these gene trees to account for the faster coalescence times of uniparentally inherited genes. Second, we generated 5000 gene trees simulated under a coalescent process with no gene flow (meaning all discordance is due to ILS) for both the nuclear and organellar 17 and 45 taxon species trees. Extra lineages were counted on the 5000 simulated gene trees and pruned nrDNA and ptDNA trees with edges supported by PP < 0.95 collapsed. If there was evidence for hybridization in our gene trees, we would expect our observed gene trees to have more extra lineages than the simulated trees. To get a sense of which observed clades may be supported under an ILS-only scenario, we assumed our nrDNA and ptDNA trees were "true" species trees and assessed bipartition support by treating our 5000 simulated trees for each genome as BS replicates.

Results

Overall Metrics of AHE Nuclear Loci

Metrics of number of reads and loci generated from AHE sequencing as well as locus coverage are provided in Table 1. After filtering, our pipeline recovered 360 putatively orthologous regions. After cleaning to remove loci with miscalled orthology, few potentially parsimony informative sites, and/or a large amount of ambiguous base calls, we recovered 316 loci with less than 50% missing sequences, 123 of which are not missing any sequences. These loci vary greatly in taxonomic coverage, length, and number of potentially parsimony informative sites. Locus coverage ranges from 22 to 48 taxa (mean of 44 taxa). Loci range from 254 to 1704 aligned nucleotides in length, with a median length of 593.5 nucleotides. Within each of these loci, the number of potentially parsimony informative sites varies from 3 to 151, with a median value of 33.

Evidence for Underlying Tree-Like Structure

Inference of the optimal population tree from the set of major quartets in Quartet MaxCut for the 45-terminal data set shows no topological differences between values of α and only slight differences in branch lengths and CFs. We therefore focus on the results of the analysis in which $\alpha = 1$ when generating CFs in BUCKy. CFs on the optimal population tree suggest a large amount of gene tree discordance within *Polemonium*, with CFs ranging from 0.95 to 0.34 with a mean CF of 0.53 (Fig. 1). The TICR test suggests that ILS does not fully explain the observed quartet CFs ($X^2 = 2359.6, P = 0.0$). Furthermore, panmixia is strongly rejected ($X^2 = 33\ 150.13$, $P = 0.0$). Stepwise addition of edges starting from panmixia or a fully resolved tree converge on the same answer and suggests that all edges should be retained except the edge supporting a closer relationship of *P. californicum* from California to *P. delicatum* than to *P. californicum* from Washington (Fig. 1) ($X^2 = 2368.3$, $P = 0.0$).

Estimating Bifurcating Coalescent Trees

The optimal population tree from our ASTRAL analysis of 48 terminals and all 316 loci suggests an overall well-supported population tree (BS and QS > 80) (Fig. 2a). The normalized quartet score of this tree is 0.553, suggesting strong discordance between loci. Although there is clearly tree-like structure to the data set, the polytomy test on our data suggests several edges could be collapsed ($P > 0.10$) (Supplementary Fig. 1 available on Dryad at https://doi.org/10.5061/dryad.5hqbzkh2r). Our ASTRAL analysis suggests the closest relative of *Polemonium* is tribe Gilieae rather than Phlocideae (BS $= 1.0$, QS $= 1.0$). Within *Polemonium*, *P. micranthum* is recovered as sister to the remainder of the genus (BS $= 1.0$, QS $= 1.0$). We recover two major clades within the remainder of *Polemonium*. The first clade consists of members of the *P. caeruleum* (caeruleum clade) and *P. pulcherrimum* species complexes (pulcherrimum clade) (BS $= 1.0$, QS $= 0.9$). The second clade is less supported (BS $= 0.94$, QS $= 0.78$, edge best collapsed into a polytomy). *Polemonium elegans* is sister to the remainder of this clade (BS $= 1.0$, QS $= 0.98$), and *P. carneum* is sister

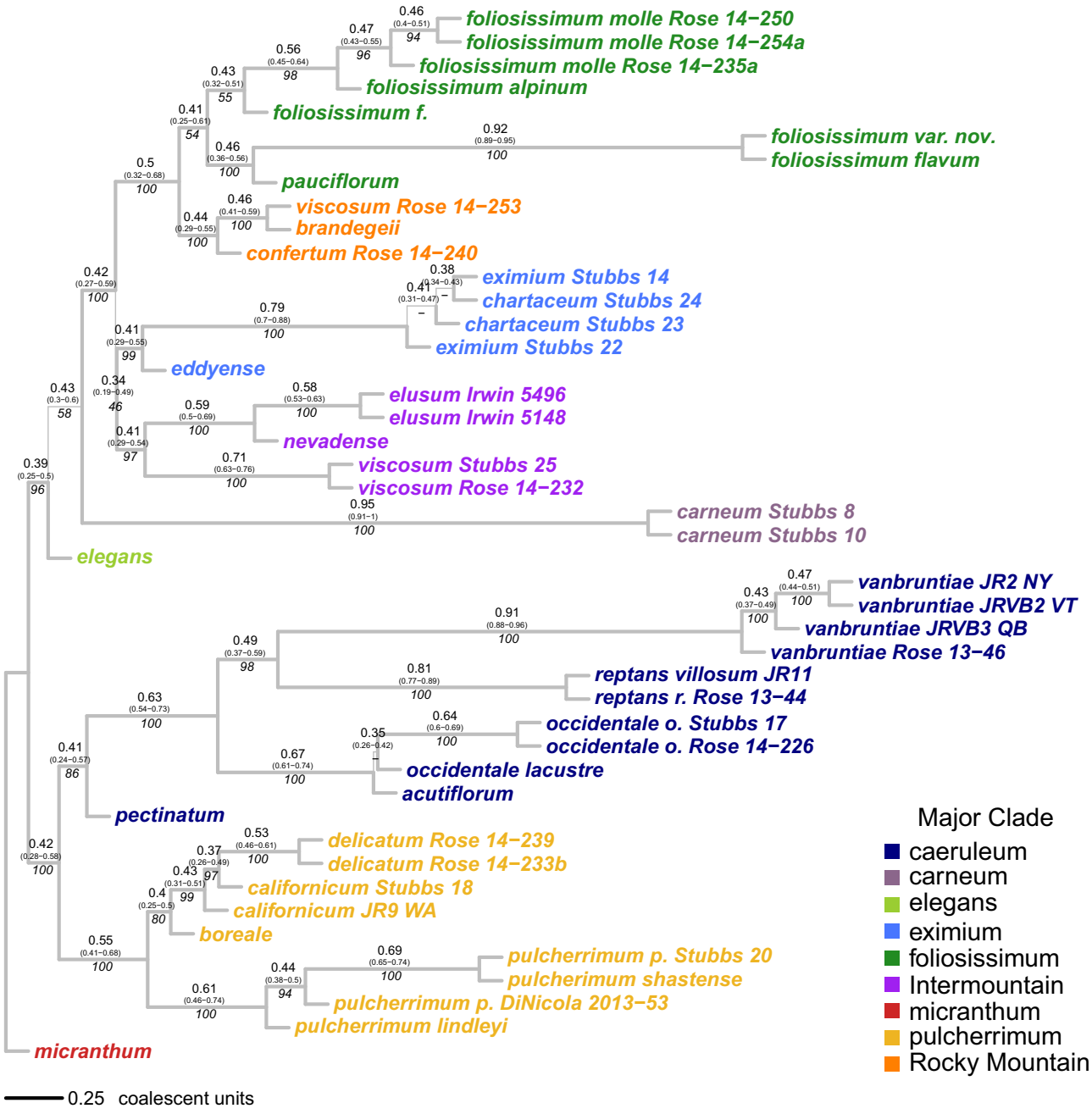

FIGURE 1. Optimal population tree for *Polemonium* inferred using Quartet MaxCut based on CFs derived from 142 nuclear genes without any missing data. The tree is rooted with *P. micranthum* based on the results of other analyses. Internal branch lengths are in coalescent units, terminal branches are diagramed for visualization purposes. Numbers above branches are CFs with their range shown in parentheses. Italicized numbers below branches are BS support values from the RAxML analysis on the concatenated matrix. Thin branches are not found in the most likely tree and branches with support marked by a dash (-) are not present in any BS replicate. Tips are colored according major clades discussed in the text.

to the remainder of the clade excluding *P. elegans* (BS =1.0, QS =0.85, edge best collapsed into a polytomy). Relationships in the remainder of this clade are less clear as backbone relationships receive contrasting BS and QS support. Included in this group are members of the *P. foliosissimum* (foliosissimum clade) and *P. viscosum* species complexes. The *P. viscosum* species complex is recovered as a grade of three clades, the eximium clade (BS =1.0, QS =0.92, edge best collapsed into a polytomy), an Intermountain clade (BS =1.0, QS =0.89, edge best collapsed into a polytomy), and a Rocky Mountain clade (BS =1.0, QS =0.99). The Rocky Mountain clade is sister to members of the foliosissimum clade (BS =1.0, QS =1.0), although the monophyly the foliosissimum

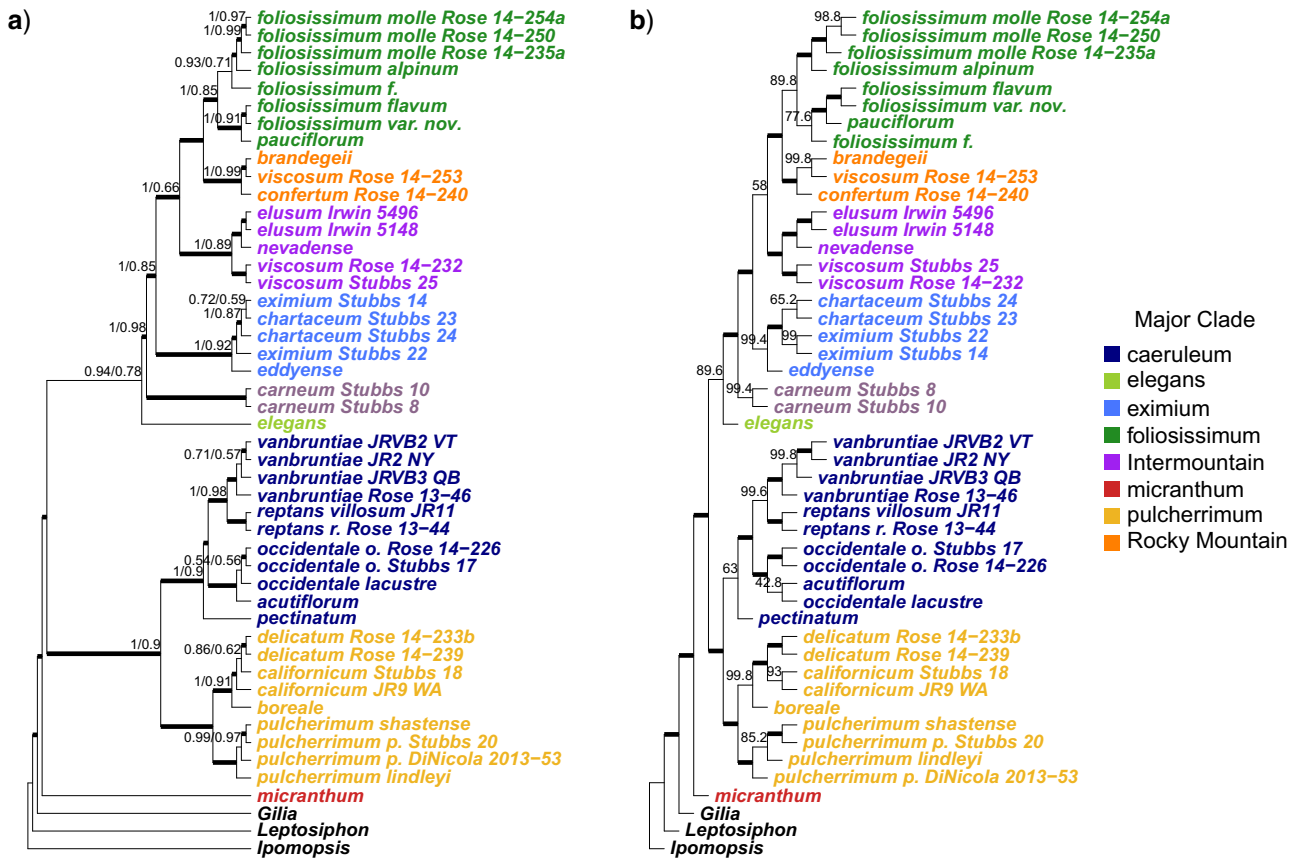

FIGURE 2. Coalescent species tree for *Polemonium* based on 316 nuclear genes inferred using a) ASTRAL and b) SVDquartets. Tips are colored according major clades discussed in the text. Thick edges in the SVDquartets tree are fully supported; values < 100% BS support are indicated. Thick edges in the ASTRAL tree have 100% BS support. Numbers above edges show support based on the BS analysis followed by quartet support. Branches without any numbers above them are fully supported by both metrics.

clade is poorly supported (BS = 1.0, QS = 0.71, edge best collapsed into a polytomy). Most species are recovered as monophyletic if multiple accessions exist, but several cases of paraphyletic/polyphyletic species are recovered including 1) *P. delicatum* embedded in *P. californicum*, 2) species polyphyly in the *P. eximium* (eximium) clade, and 3) two clades of *P. viscosum*: one in the Rocky Mountain clade sister to *P. brandegeei* and another in the Intermountain clade sister to *P. elusum* + *P. nevadense*. The topology and support for relationships in the SVDquartets tree are largely consistent with that in the ASTRAL topology along the backbone (Fig. 2b) with only 58% BS support for the eximium clade as sister to the Rocky Mountain, Intermountain, and foliosissimum clades. Other differences in the SVDquartets topology are the placement of *P. foliosissimum* var. *foliosissimum*, the placement of *P. occidentale* subsp. *lacustre*, and the monophyly of *P. californicum P. chartaceum*, and *P. eximium*, although these topological conflicts are poorly supported in one or both analyses.

The analysis of the concatenated matrix of 45-terminals and 142 loci produced a mostly well-supported topology (BS > 80) largely consistent with the topologies of the optimal population tree of the same data set as well as the analyses on the larger data set of 48 terminals and 316 loci, with the exception of several edges. These edges include the positions of *P. carneum* and *P. elegans*, relationships among the Intermountain, eximium, and Rocky Mountain + foliosissimum clades, and relationships within the caeruleum, eximium, and foliosissimum clades (Figs. 1 and Fig. 2; Supplementary Fig. 2 available on Dryad).

Estimating a Phylogenetic Network

Filters for taxonomic and genetic completeness in the 360 recovered loci from AHE sequencing resulted in 325 loci analyzed for the 17 target taxa. As with the 45-taxon optimal population tree, CFs mapped on the 17-taxon optimal population tree are relatively low (0.37–0.61, mean = 0.45), suggesting high discordance between gene trees (Supplementary Fig. 3 available on Dryad). The TICR test suggests that ILS does not fully explain the observed quartet CFs ($X^2 = 7.88, P = 0.049$). Furthermore, panmixia is strongly rejected ($X^2 = 551.06, P = 4.09 \times 10^{-119}$). Stepwise addition of edges suggests that all edges should be retained. A plot of

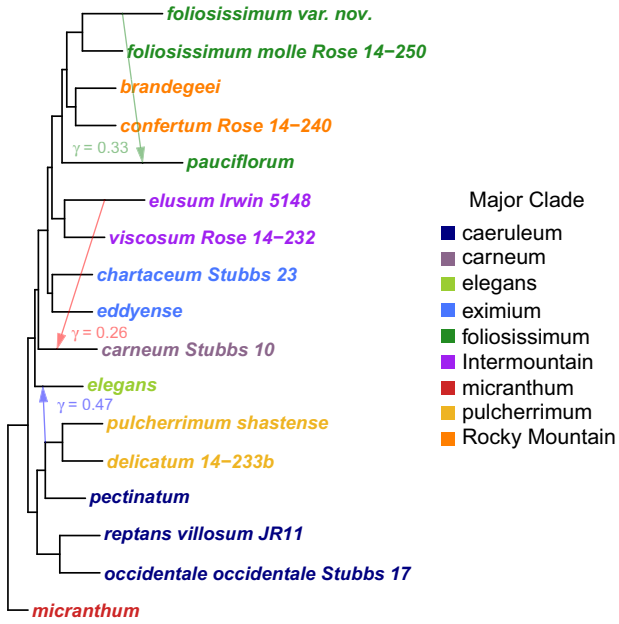

FIGURE 3. Best SNaQ phylogenetic network from our analysis of 17 exemplars of *Polemonium*. Tip color corresponds to clades recovered in the coalescence-only species tree approaches. Arrows indicate inferred direction of gene flow in the three detected reticulation events, with the parameter γ indicating the percentage of the genome involved in the reticulation. All major topology edges and all hybrid edges receive 100% BS support.

pseudologlikelihood scores (Supplementary Fig. 4 available on Dryad) suggests the best network model is one with a maximum number of four hybridization events allowed (-ploglik = 1446.48). The best network has three hybridization events (Fig. 3). The parameter gamma (γ) is a metric of the proportion of the genome contributed by one parental population in the reticulation event, with 1 - γ being the proportion of the other parent (stem lineage in a level-1 network). In each of the three reticulation events, large portions of the genome have been exchanged. Detected reticulations are: 1) between the pulcherrimum clade and *P. elegans* (γ = 0.47), 2) between *P. foliosissimum* var. nov. and *P. pauciflorum* (γ = 0.33), and 3) between *P. elusum* and *P. carneum* (γ = 0.26). The BS analysis found full support (BS = 100%) for all edges of the major topology as well as for all hybrid edges. The range of estimated γ values of the major hybrid edge from the BS analysis was 0.45–0.48 for the contribution of the pulcherrimum clade to *P. elegans*, 0.30–0.44 for the contribution of *P. foliosissimum* var. nov. to *P. pauciflorum*, and 0.24–0.27 for the contribution of *P. elusum* to *P. carneum*. The major topology of the phylogenetic network contrasts with that recovered from the coalescence-only approaches in two ways. First, *P. pectinatum* is recovered as sister to the pulcherrimum clade instead of sister to the caeruleum clade. Second, *P. pauciflorum* is not recovered within the foliosissimum clade. Instead, it is sister to the foliosissimum clade + the Rocky Mountain clade.

Cyto-nuclear Discordance

We were able to recover nearly the entire plastid genome from the raw reads of the AHE sequencing. After removing ambiguously aligned portions, the ptDNA data set of 48 terminals contains 132 389 aligned nucleotides. The maximum clade credibility (MCC) tree has most edges supported by a PP < 0.95 (Supplementary Fig. 5 available on Dryad). The ptDNA topology differs markedly from those obtained from the nuclear loci in terms of clade composition, clade placement, and species monophyly (Fig. 4). The ptDNA genealogy suggests the eximium clade is sister to the remainder of the genus. *Polemonium micranthum* is in turn sister to the remainder of the genus. Within this larger clade, morphological species are scattered throughout the tree but do form several smaller clades, including two clades of members of the foliosissimum clade and a close relationship among many (but not all) members of the pulcherrimum clade. The nrDNA data set of 48 terminals contains 6342 aligned nucleotides. A number of sites are clearly polymorphic. We interpret this as a result of sequencing paralogous copies of the tandem repeat, and these are therefore treated as ambiguous characters in our analyses. While several areas of our nrDNA genealogy are strongly supported, many branches in the MCC tree are supported by low PP (Supplementary Fig. 6 available on Dryad). *Polemonium micranthum* is sister to the rest of *Polemonium*. Within *Polemonium* excluding *P. micranthum*, there are two major subclades. The first subclade contains members of the *P. carneum*, *P. viscosum*, and foliosissimum clades, although none of the polytypic clades are recovered as monophyletic. *Polemonium eddyense* is sister to the Intermountain clade + *P. carneum* (PP = 1.0). *Polemonium carneum* is recovered as sister to *P. elusum* + *P. nevadense* (PP = 1.0). The *P. eddyense* + Intermountain clade + *P. carneum* clade is sister to the foliosissimum clade + Rocky Mountain clade, although the monophyly of this clade is poorly supported (PP = 0.47). The Rocky Mountain clade is not monophyletic and clearly nested within the foliosissimum clade. The second major nrDNA clade contains members of the caeruleum, *P. elegans*, eximium, and pulcherrimum clades. The backbone of this clade is unsupported (PP < 0.50), but the pulcherrimum clade is polyphyletic, with *P. elegans* nested within a clade of *P. pulcherrimum* (PP = 1.0). Though poorly supported, the eximium clade (excluding *P. eddyense*) is nested within the caeruleum clade (PP = 0.86). Overall, the well-supported areas of the nrDNA genealogy are more congruent with the species trees based on nuclear DNA than to the ptDNA genealogy (Fig. 5).

The results of the pipeline of Olave et al. (2017) suggest that the number of extra lineages required to reconcile the random trees generated under the multispecies coalescent with the major topology of our 17-tip phylogenetic network range from 0 to 15 with a median of 6 for nuclear genes, and vary from 0 to 4 with a median of 0 for organellar genomes (Fig. 6a).

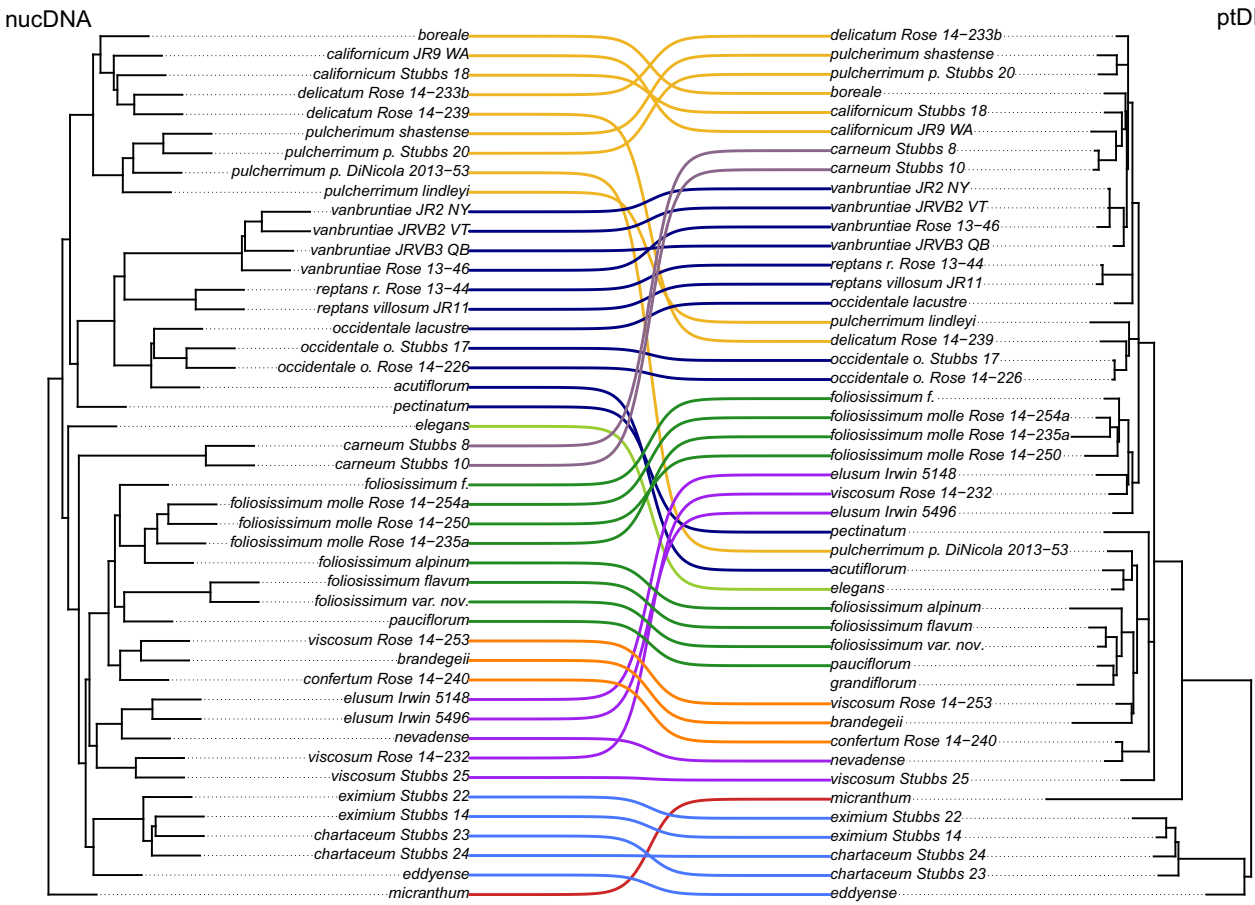

FIGURE 4. Tanglegram illustrating discordance between the ASTRAL nuclear species tree (left) and the plastome tree (right). Trees have been rotated to maximize tip matching with minimal link overlap. Link color corresponds to major clades discussed in the text and links connect identical tips. Clades in the plastome tree with PP < 0.95 have been collapsed.

For our 45-tip optimal population tree, the number of extra lineages in the nuclear genome ranges from 2 to 31 with a median value of 16 and from 0 to 10 for organellar genomes with a median of 3. For observed gene trees, the number of extra lineages required in the 17-tip data set is 16 for the nrDNA gene tree and 27 for the ptDNA gene tree. For the 45-tip data set, these numbers are 24 for the nrDNA gene tree and 80 for the ptDNA gene tree (Fig. 6b). For both data sets, the number of extra lineages for the nrDNA gene tree falls within the upper tail of the distribution expected for the nuclear genome under a strictly coalescent process, whereas the number of extra lineages for the ptDNA genome greatly exceeds the expected distribution (Fig. 6). Assessing support for observed ptDNA bipartitions based on the 5000 organellar trees simulated under the multispecies coalescent suggests the backbone of the ptDNA gene tree does not appear in any simulated gene trees, and that many other clades are only present in a small percentage of gene trees. The most commonly recovered clades in the simulated gene trees are found towards the tips and usually support the monophyly of species (Fig. 7a). By contrast, support for bipartitions of nrDNA clades is much higher overall, with many backbone bipartitions appearing in a significant number (>33%, within the expected distribution of splits under a polytomy) of simulated coalescent trees (Fig. 7b).

DISCUSSION

A major challenge for inferring phylogenetic networks are the computational limitations associated with searching "network space" on data sets that contain more than a dozen taxa, especially when the number of possible hybridization events is high. By contrast, methods that incorporate only ILS in inferring species trees are popular because they account for some sources of discordance and are fast and tractable with data sets of dozens to hundreds of terminals. However, species trees inferred using these methods are suboptimal models of relationships if gene flow has occurred between lineages, even if the major topology of the phylogenetic network is identical, as branch lengths will be underestimated

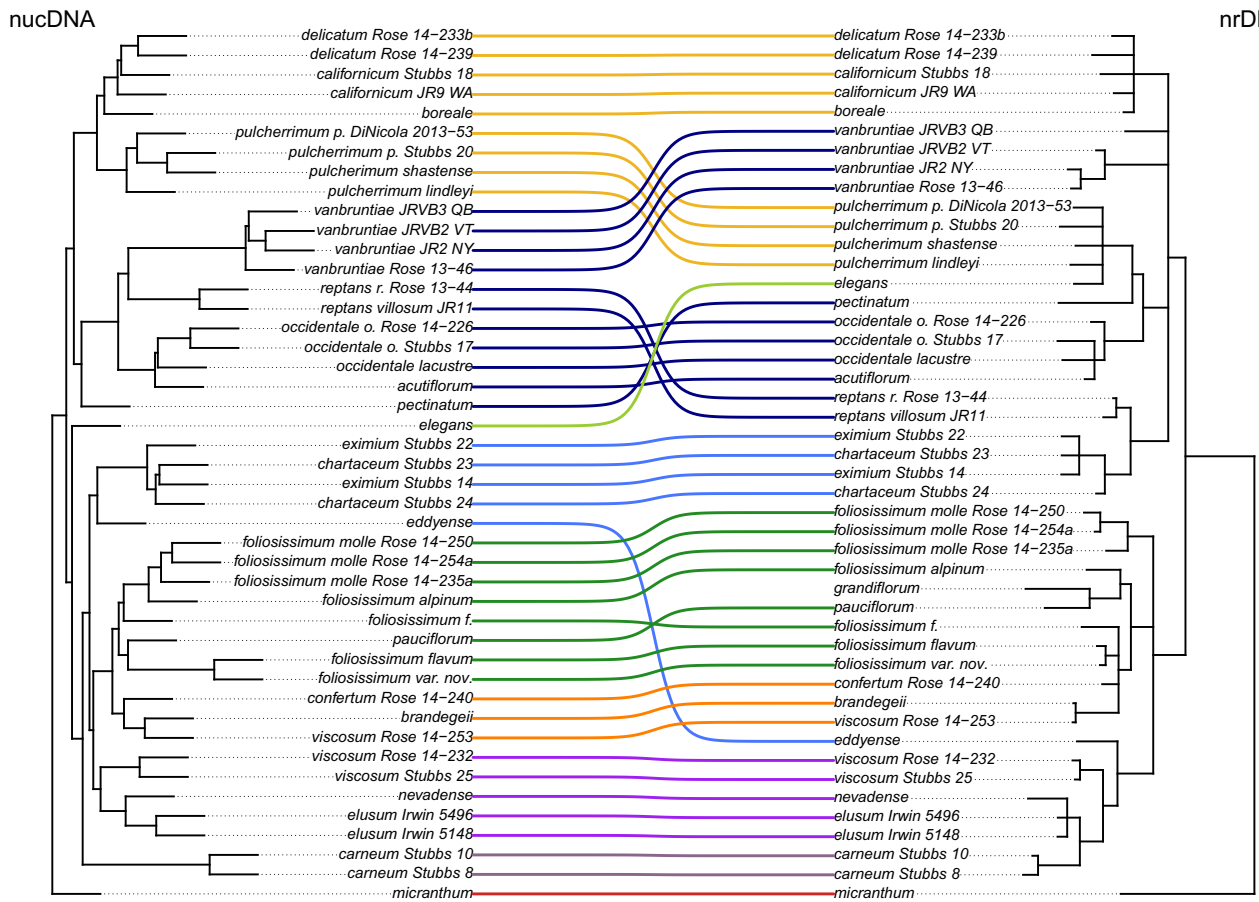

FIGURE 5. Tanglegram illustrating discordance between the ASTRAL nuclear species tree (left) and the nuclear ribosomal DNA gene tree (right). Trees have been rotated to maximize tip matching with minimal link overlap. Link color corresponds to major clades discussed in the text and links connect identical tips. Clades in the nuclear ribosomal DNA tree with PP < 0.70 have been collapsed.

if all genomic discordance is explained by ILS alone (Leaché et al. 2014). In this study, we document that morphological species of *Polemonium* are often reciprocally monophyletic based on nuclear data. Despite evidence for widespread gene flow within *Polemonium*, there is considerable topological congruence among species trees inferred using disparate methods. Furthermore, evidence from the plastid genome suggests that there has been more gene flow than is evident solely in the nuclear genome.

Support for Relationships in Polemonium *Despite Strong Nuclear Discordance*

Our results suggest that there is a large amount of discordance between nuclear gene trees as evidenced by low CF < 0.50 in the optimal population tree, yet the vast majority of these edges are statistically supported by all analyses, strongly suggesting an underlying tree-like structure. One possibility for some of this discordance is error in gene tree estimation. While this may be true to some extent for the ASTRAL-III analysis which relies on maximum likelihood gene trees estimated using RAxML, uncertainty in gene tree estimation is tempered somewhat by using BS replicates. Other analyses should be robust to this error as they either estimate a species tree directly (SVDquartets) or rely on CFs derived from BUCKy, which accounts for gene tree estimation error. However, low but significant CFs are best explained as the result of conflict caused by ILS and reticulation. Widespread ILS is not unexpected in the cases of *Polemonium*, given the apparent rapid radiation of the genus in the last 10 My (Landis et al. 2018; Rose et al. 2018).

Our topologies are consistent with the conclusions of the only previous multilocus phylogenetic study of *Polemonium* based on a concatenated parsimony analysis of AFLP data (Worley et al. 2009). In cases where Worley et al. (2009) found strong support for relationships, we likewise find strong support. As in Worley et al. (2009), we recover a sister relationship of *P. reptans* and *P. vanbruntiae*, as well as two subclades in the pulcherrimum clade: one of *P. pulcherrimum s.s.* and another of *P. boreale, P. californicum*, and *P. delicatum*. The major difference between our estimate and

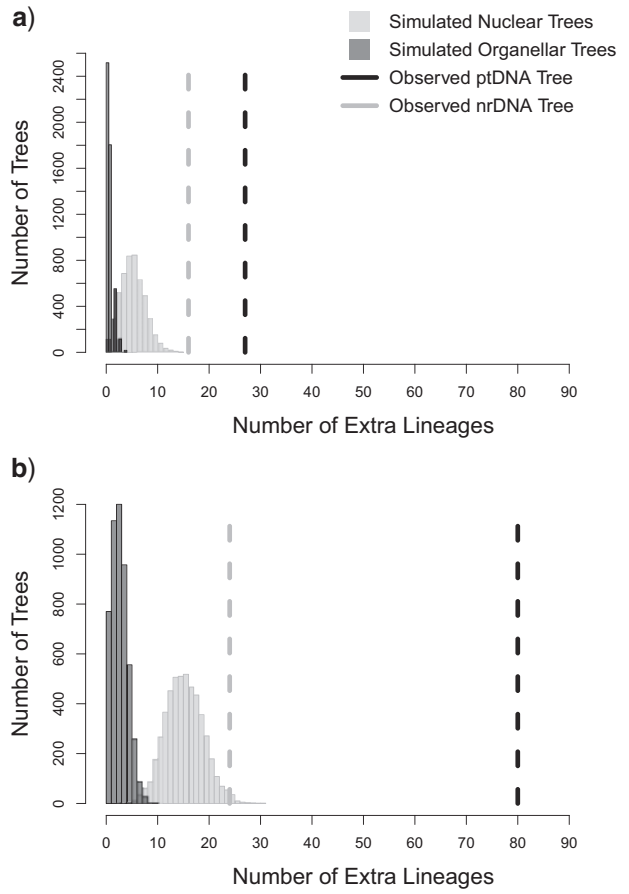

FIGURE 6. Histogram of number of extra lineages required to reconcile gene trees simulated under the multispecies coalescent in MS using a) the major topology of our 17-tip phylogenetic network or b) the 45-tip optimal population tree. Extra lineages were counted using deep_coal_count in Phylonet. Light grey bars show the distribution of nuclear genome trees and dark grey bars show the distribution of organellar genome trees. Dashed vertical lines represent the number of extra lineages observed in the nuclear ribosomal (light grey) and plastid (dark grey) gene trees.

Worley et al. (2009) is that they suggest that *P. viscosum* is either monophyletic or *P. brandegeei* is nested within it, instead of polyphyletic as suggested in this study. This discrepancy, however, is based on their geographic sampling of *P. viscosum*, which was restricted to members of the Rocky Mountain clade.

Despite the large amount of nuclear gene tree discordance detected in our study, major topological relationships are largely consistent across species tree approaches. Between our coalescent species tree approaches, topological relationships along the backbone are consistent, although the placement of the eximium clade is poorly supported across analyses, and the optimal population tree suggests the split between the eximium, Rocky Mountain, Intermountain, and foliosissimum clades occurred rapidly. Additionally, as noted above, error in gene tree estimation may explain some of these differences between these coalescent analyses (Fig. 2).

The major topological difference among our results is the placement of *P. pauciflorum* and *P. pectinatum*. In our optimal population trees and in our coalescent analyses, *P. pauciflorum* falls sister to southern members of the foliosissimum clade: *P. foliosissimum* var. *flavum* and an undescribed variety of *P. foliosissimum*. In our SNaQ analysis, *P. foliosissimum* is recovered as monophyletic. In addition, in our concordance and coalescent analyses *P. pectinatum* falls sister to the remainder of the caeruleum clade whereas in the SNaQ analysis the species is placed sister to the pulcherrimum clade. Our results are also consistent across methods in suggesting that *P. viscosum* is polyphyletic, and that these lineages have not exchanged genes. Only 19% of the nuclear genome supports a sister relationship of the two clades of *P. viscosum*. Our finding of polyphyly in *P. viscosum* might be expected given that the species exhibits a wide range of morphotypes and that species circumscription based on morphology has resulted in various schemes for delimiting species (Brand 1907; Davidson 1950; Grant 1989). At present, it is unclear what (if any) morphological characters distinguish Intermountain *P. viscosum* from Rocky Mountain *P. viscosum*, but it appears that these clades represent a case of extreme convergence in alpine habitats. Preliminary study suggests that length and deciduousness of leaves as well as corolla length may be important characters (J. Rose, unpublished data).

Nuclear Discordance is Explained by ILS and Gene Flow

Our SNaQ results demonstrate that reticulate evolution has played a large role in the evolutionary history of the genus. In fact, the presence of reticulation events in *Polemonium* is one possible explanation for deeper edges rejected as "true" by the ASTRAL polytomy test. The phylogenetic network analysis suggests at least three exchanges of genetic material have occurred between unrelated lineages, several of which occur along deeper (internal) edges of the phylogeny. Estimates of the amounts of the genome exchanged between lineages during these events are high (26, 33, and 47%), especially when considering animal systems (Solís-Lemus and Ané 2016; Blair et al. 2019; Morando et al. 2020; Pyron et al. 2020) but are not atypical for plant systems (Crowl et al. 2017; Morales-Briones et al. 2018; Roberts and Roalson 2018). Given high CFs for sister relationships not analyzed by SNaQ (Fig. 1), it is likely that the *P. carneum*–*P. elusum* reticulation event occurred before the origin of *P. carneum* (*P. carneum* CF = 0.95) and that the *P. foliosissimum* var. nov.–*P. pauciflorum* reticulation event also involved *P. foliosissimum* var. *flavum* (*P. foliosissimum* var. nov. + *P. foliosissimum* var. *flavum* CF = 0.92). One of the most commonly used markers for Sanger-based phylogenetic inference in land plants has been nrITS sequences. Despite lack of variability in nrDNA and weak support for relationships in *Polemonium* (Irwin et al. 2012; Supplementary Fig. 6 available on Dryad), this nuclear region contains evidence of the history of reticulate evolution in the clade through the strongly supported sister relationship of *P. pulcherrimum* and

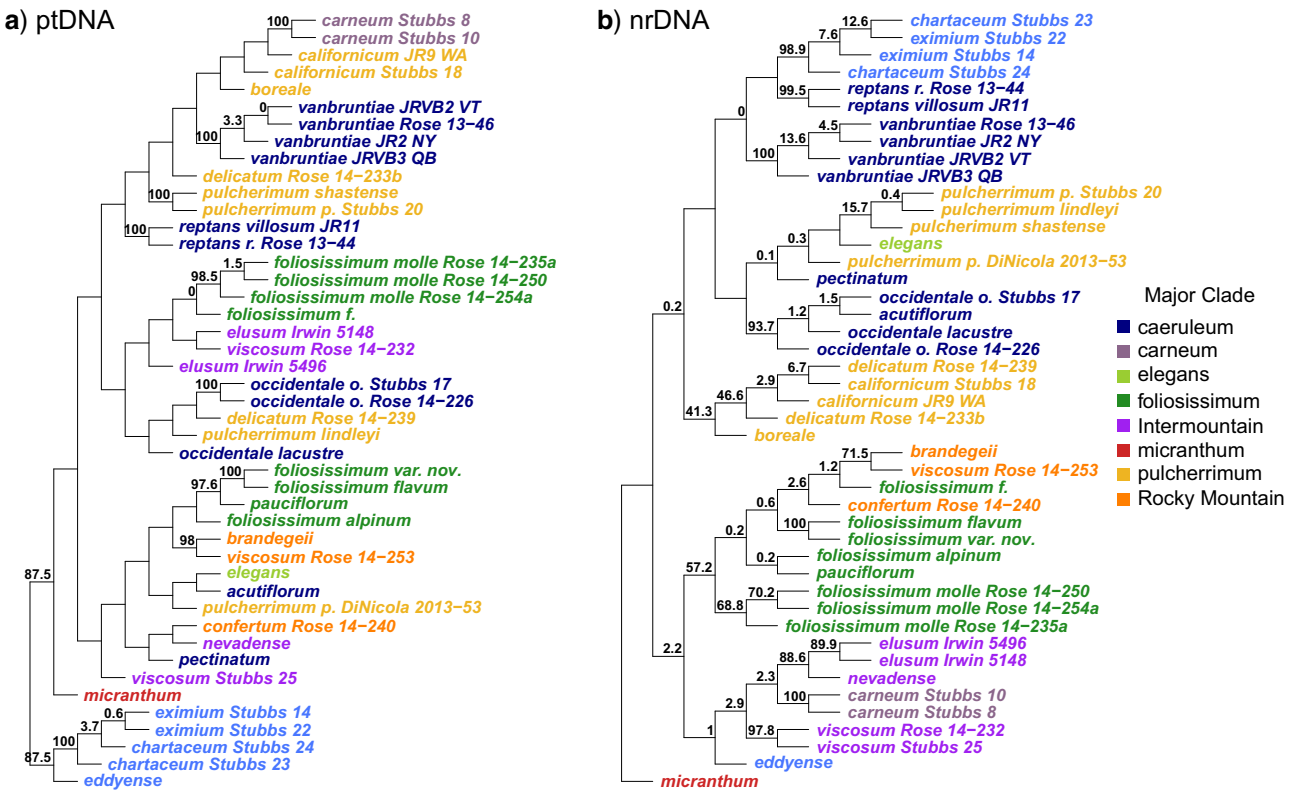

FIGURE 7. Percent support for bipartitions in the maximum clade credibility tree in 5000 simulated multispecies coalescent trees from MS assuming the a) the plastid gene tree or b) nuclear ribosomal DNA gene tree represents the true species tree. Numbers are rounded. If no number is given above the edge, that bipartition does not occur in any simulated coalescent trees. Tips are colored according major clades discussed in the text.

P. elegans (PP = 1.0) as well as the sister relationship of *P. carenum* and *P. elusum* (PP = 1.0). Based on the distribution of extra lineages as well as support for the topology provided by nrDNA in the simulated gene trees, discordance for the nrDNA genealogy can be explained by 1) a lack of phylogenetically informative substitutions for some relationships, 2) ILS, and 3) detected reticulate evolution events. While it is clear that species of *Polemonium* have exchanged genes, it is unclear which phenotypic traits, if any, were exchanged in these reticulation events. Candidate traits demonstrating such reticulation events include pollen color and leaflet size. Future study into character evolution within *Polemonium* should explicitly test for this phenomenon in the context of a phylogenetic network (Bastide et al. 2018).

Strong Discordance Between Plastid and Nuclear Genomes

From our analyses, it is clear that the well-supported tree obtained with the plastid genome of *Polemonium* is in strong conflict with the species tree of the genus inferred from nuclear data. Based on data from the nuclear genome, when multiple individuals of a given morphological species have been sampled, the species are either monophyletic or paraphyletic. By contrast, most morphological species appear polyphyletic in the ptDNA genealogy. Moreover, this topology is not only inexplicable based on ILS, but tests for ILS as the sole cause of discordance in the ptDNA tree demonstrate that the amount of discordance in the ptDNA tree is on exceeds the amount of discordance expected under a strictly coalescent process (Fig. 6). This is further emphasized by the small number of bipartitions present in the plastid genealogy also present in bipartitions in the 5000 simulated organellar gene trees (Fig. 7a). The finding of strong plastid-nuclear discordance is contrary to what would be expected based on the smaller effective population sizes (and thus shorter coalescence times) expected for organellar genomes (Moore 1995). Of the three reticulation events supported by our nuclear data, only the genetic exchange between *P. foliosissimum* var. nov. and *P. pauciflorum* is also reflected in the plastid genome. Plastid discordance not matching hybridization events has also been reported recently in *Pinus* (Gernandt et al. 2018).

Causes of Cyto-nuclear Discordance

How do we reconcile these results? First, it is possible that we simply did not sample loci in the nuclear genome that show evidence for additional reticulation events. This may come about from not sampling loci showing reticulation in our 316-gene subsample of the nuclear

genome. Our sample of loci may come from one or a few chromosomes or within a specific area of each chromosome and not capture all evidence of horizontal gene flow. However, the 499 target loci are found throughout the *Arabidopsis* genome (Buddenhagen et al. 2016), so it is unlikely that our 316 genes are strongly biased towards certain chromosomes. Second, our subsample of 17-taxa for inferring a phylogenetic network may not have included most of the ptDNA/nucDNA discordances and we may have not sampled individuals with these reticulation events in their history. This could be addressed by generating a phylogenetic network for all 45 samples of *Polemonium* if this were computationally feasible. ABBA-BABA statistics are widely used to investigate patterns of gene flow (Patterson et al. 2012; Eaton and Ree 2013; Martin et al. 2014; Pease and Hahn 2015; Blischak et al. 2018) but are limited in that they only treat four or five terminals at a time and cannot detect gene flow along internal edges of the phylogeny in the four-terminal case (Pease and Hahn 2015, p. 654). Third, there may be no sign of these reticulation events in the nuclear genome. In order to detect introgression between different lineages, crossing over between the two parental nuclear genomes in the F_1 generation must occur, and extensive backcrossing must not occur. This could be caused by prezygotic or postzygotic barriers preventing crossing between F_1 hybrids, or nuclear signal being swamped out by extensive backcrossing with one parent. We propose that this third possibility best explains the seemingly incompatible ptDNA/nucDNA results. That is, the unexplained discordance between genomes reflects a pattern of gene flow followed by immediate backcrossing with one parent, swamping out any signal of gene flow in the nuclear genome. There is experimental evidence for this scenario in *Polemonium*. While many species of *Polemonium* lack prezygotic barriers to reproduction (Ostenfeld 1929), interspecific F_1 hybrids are often sterile but may be backcrossed with one of the parents to produce a fertile hybrid. Clausen (1967) investigated the chromosomes of these *Polemonium* hybrids and found a large number of irregularities during meiosis in the anthers. These chromosomal rearrangements may provide a postzygotic barrier to reproduction in F_1 hybrids (Rieseberg 2001). In other cases, interspecific hybrids show few chromosomal irregularities (Clausen 1931) and may be able to produce viable seed (Anway 1968). In the case of *Polemonium*, both chromosomal incompatibilities and/or backcrossing with one parent may have erased a signal of gene flow in the nuclear genome, but this signal has been retained in the plastid genome by chance or by selection on particular interspecific ptDNA haplotypes.

Future work should examine whether the mitochondrial genome is concordant with the plastid genome, as would be expected if the cytoplasm is inherited as a single unit. Inferring the mitochondrial genealogy of *Polemonium* is hampered by having no closely related reference genome and the tendency of plant mitochondrial genomes to rearrange (Palmer and Herbon 1988). However, preliminary results with ca. 30 kb of highly conserved mitochondrial sequences, where strongly supported, suggest a similar topology to the plastid gene tree (J. Rose, unpublished data). Future work also should examine if there is a signal of natural selection in the coding regions of certain ptDNA haplotypes which maintain the persistence of interspecific and intraspecific haplotypes in populations of *Polemonium*. Phylogenies inferred from the plastid genome may reflect the biogeographic history of *Polemonium* where relationships are indicators of present or past geographic distance between species rather than phylogenetic distance (Whittemore and Schaal 1991; Petit et al. 1993; Wolf et al. 1997; Rautenberg et al. 2010; Pham et al. 2017; Schuster et al. 2018).

A Workflow for Analyzing Introgression in AHE Data Sets

This study presents a case study for examining the evolution of recently radiated plant clades using AHE and teasing apart ILS and horizontal gene flow. We recommend a procedure such as the following be implemented, with some caveats relating to the number of tips being analyzed (Supplementary Fig. 7 available on Dryad).

1. Generate gene trees for all loci, paying attention to the informativeness of the loci at your particular taxonomic scale and the potential impact of missing data in the phylogenetic methods used. Loci may be highly conserved (e.g., transcriptomes) or short reads (e.g., restriction site methods) where there may be few or no potentially phylogenetically informative SNPs per locus. If a reference genome is available, consider combining loci into a "superlocus" using a sliding window approach.
2. Generate a preliminary species tree from these loci.
3. Test for evidence for all or some branches in the species tree and if gene tree discordance has a simple evolutionary explanation (ILS alone) or is best explained by ILS and one or more cases of horizontal gene transfer, as suggested by outlier quartets.
4. If ILS alone is a sufficient explanation for the discordance, ILS is low, or there is no conflict among loci, coalescent-only or concatenation approaches are appropriate for analyzing the data set. If not, phylogenetic network approaches are appropriate.
5. If the number of terminals is <20, analyze all samples using currently available methods for inferring phylogenetic networks, allowing for models with reticulation events varying from 0 to n. If there are >20 terminals, carefully consider subsetting the data set and avoid if at all possible. If necessary, subsetting should focus on the specific questions being addressed in a given

study, informed by the results of (2) and (3) (e.g., major clades recovered in a coalescent only or concatenation approaches, known/suspected hybrids and their parental taxa, or composition of outlier quartets), perhaps with several tiers of subsampling and tests for the robustness of results in particularly complex systems.

6. Consider follow-up analyses to investigate gene flow in outlier quartets, including investigating signal from organellar genome(s) and evidence from pattern-based statistics, if appropriate.

Given the rapid progress and interest in the field of phylogenetic network inference, we are confident that soon, new algorithms will drastically decrease the computational cost associated with modeling gene flow with a larger number of tips and reticulation events, allowing for ease of analysis for ever larger data sets.

Conclusions

We have presented a case study for examining the evolution of recently radiated plant clades using AHE. This has allowed us to robustly tease apart the contribution of ILS and reticulation in the history of *Polemonium*, as well as explain patterns of discordance between the nuclear genome and additional genomic compartments. Our results suggest that reticulation has been an integral factor in the evolutionary history of angiosperms and that caution should be applied when making claims about the evolutionary history and taxonomic circumscription of recently radiated groups based on models that do not or only account for ILS, and/or overly rely upon drawing conclusions based on the inferred history of organellar genomes. Furthermore, given the contrasting patterns of species or species complex monophyly in the plastid genome compared with species and species complex polyphyly in the nuclear genome revealed by this study, we urge caution against describing new, cryptic species based solely on variation in the plastid genome.

Supplementary Material

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.5hqbzkh2r.

Funding

This work was funded in part through a NSF doctoral dissertation improvement grant [DEB-1501867 to K.J.S. and J.P.R], the O.N. and E.K. Allen Fellowship awarded to J.P.R. through the University of Wisconsin-Madison Department of Botany, the University of Wisconsin Botany Department Hofmeister Endowment, a NSF-DOB [DOB-1046355 to K.J.S], a NSF-DEB [DEB-1655606 to K.J.S.], and from small research grants to J.P.R. through the Rocky Mountain Biological Laboratory Idaho Native Plant Society, Nevada Native Plant Society, Native Plant Society of New Mexico, and Utah Native Plant Society.

Acknowledgments

We would like to thank Rebecca Stubbs for providing tissue collected during her M.S. Thesis, Cecilé Ané and Nisa Karimi for discussion of SNAq, Steve Hunter for discussion of plastomes, associate editors Bryan Carstens and David Tank, Ryan Folk, and four anonymous reviewers for their helpful suggestions which greatly improved the manuscript.

References

Ané C., Larget B., Baum D.A., Smith S.D., Rokas A. 2007. Bayesian estimation of concordance among gene trees. Mol. Biol. Evol. 24:412–426.

Anway J.C. 1968. The systematic botany and taxonomy of *Polemonium foliosissimum* A. Gray (Polemoniaceae). Am. Midl. Nat. 79:458–475.

Bastide P., Solís-Lemus C., Kriebel R., Sparks K.W., Ané C. 2018. Phylogenetic comparative methods on phylogenetic networks with reticulations. Syst. Biol. 67:800–820.

Blair C., Bryson Jr. R.W., Linkem C.W., Lazcano D., Klicka J., McCormack J.E. 2019. Cryptic diversity in the Mexican highlands: thousands of UCE loci help illuminate phylogenetic relationships, species limits and divergence times of montane rattlesnakes (Viperidae: *Crotalus*). Mol. Ecol. Resour. 19:349–365.

Blischak P.D., Chifman J., Wolfe A.D., Kubatko L.S. 2018. HyDe: a Python package for genome-scale hybridization detection. Syst. Biol. 67:821–829.

Bock D.G., Andrew R.L. Rieseberg L.H. 2014. On the adaptive value of cytoplasmic genomes in plants. Mol. Ecol. 23:4899–4911.

Brand A. 1907. Polemoniaceae. In: Engler A., editor. Das Pflanzenreich IV(250). Leipzig: Engelmann. p. 1–203.

Buckler E.S., Ippolito A., Holtsford T.P. 1997. The evolution of ribosomal DNA divergent paralogues and phylogenetic implications. Genetics 145:821–832.

Buddenhagen C., Lemmon A.R., Lemmon E.M., Bruhl J., Cappa J., Clement W.L., Donoghue M., Edwards E.J., Hipp A.L., Kortyna M., Mitchell N., Moore A., Prychid C.J., Segovia-Salcedo M.C., Simmons M.P., Soltis P.S., Wanke S., Mast A. 2016. Anchored phylogenomics of angiosperms I: assessing the robustness of phylogenetic estimates. bioRxiv:086298. doi:10.1101/086298.

Cardillo M., Weston P.H., Reynolds Z.K.M., Olde P.M., Mast A.R., Lemmon E.M., Lemmon A.R., Bromham L. 2017. The phylogeny and biogeography of *Hakea* (Proteaceae) reveals the role of biome shifts in a continental plant radiation. Evolution 71:1928-1943.

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

Clausen J. 1931. Genetic studies in *Polemonium*. III. Hereditas 15:62–66.

Clausen J. 1967. Biosystematic consequences of ecotypic and chromosomal differentiation. Taxon 16:271–279.

Cranston K.A., Hurwitz B., Ware D., Stein L., Wing R.A. 2009. Species trees from highly incongruent gene trees in rice. Syst. Biol. 58:489–500.

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

Cui R., Schumer M., Kruesi K., Walter R., Andolfatto P., Rosenthal G.G. 2013. Phylogenomics reveals extensive reticulate evolution in *Xiphophorus* fishes. Evolution. 67:2166-2179.

Davidson J.F. 1950. The genus *Polemonium* [Tournefort] L. Univ. Calif. Publ. Bot. 23:209–282.

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

Everson K.M., Soarimalala V., Goodman S.M., Olson L.E. 2016. Multiple loci and complete taxonomic sampling resolve the phylogeny and biogeographic history of tenrecs (Mammalia: Tenrecidae) and reveal higher speciation rates in Madagascar's humid forests. Syst. Biol. 65:890–909.

Faircloth B., McCormack J.E., Crawford N.G., Harvey M.G., Brumfield R.T., Glenn T.C. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61:717–726.

Fehrer J., Gemeinholzer B., Chrtek Jr. J., Bräutigam S. 2007. Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in *Pilosella* hawkweeds (*Hieracium*, Cichorieae, Asteraceae). Mol. Phylogenet. Evol. 42:347–361.

Feulner M., Möseler B.M., Nezadal W. 2001. Introgression und morphologische variabilität bei der Blauen Himmelsleiter, *Polemonium caeruleum* L. in Nordbayern, Deutschland. Feddes Repertorium 112:231–246.

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

Folk R.A., Soltis P.S., Soltis D.E., Guralnick R. 2018. New prospects in the detection and comparative analysis of hybridization in the tree of life. Am. J. Bot. 105:364–375.

Galen C., Kevan P.G. 1980. Scent and color, floral polymorphisms and pollination biology in *Polemonium viscosum* Nutt. Am. Midl. Nat. 104:281–289.

Galen C., Newport M.E.A. 1987. Bumble bee behavior and selection on flower size in the sky pilot, *Polemonium viscosum*. Oecologia 74:20–23.

Galen C., Zimmer K.A., Newport M.E. 1987. Pollination in floral scent morphs of *Polemonium viscosum*: a mechanism for disruptive selection on flower size. Evolution 41:599–606.

Galen C., Shore J.S., Deyoe H. 1991. Ecotypic divergence in alpine *Polemonium viscosum*: genetic structure, quantitative variation, and local adaptation. Evolution 45:1218–1228.

de Geofroy I. 1998. Phylogeny and biogeography of the high-elevation species of *Polemonium* (Polemoniaceae) [M.S. Thesis]. San Francisco State University.

Gernandt D.S., Aguirre Dugua X., Vázquez-Lobo A., Willyard A., Moreno Letelier A., Pérez de la Rosa J.A., Piñero D., Liston A. 2018. Multi-locus phylogenetics, lineage sorting, and reticulation in *Pinus* subsection *Australes* Am. J. Bot 105:711–725.

Grant V. 1959. Natural history of the phlox family. The Hague: Martinus Nijhoff.

Grant V. 1989. Taxonomy of the tufted alpine and subalpine polemoniums (Polemoniaceae). Bot. Gaz. 150:158–169.

Grummer J.A., Morando M.M., Avila L.J., Sites Jr. J.W., Leaché A.D. 2018. Phylogenomic evidence for a recent and rapid radiation of lizards in the Patagonian *Liolaemus fitzingerii* species group. Mol. Phylogenet. Evol. 125:243–254.

Hamilton C.A., Lemmon A.R., Lemmon E.M., Bond J.M. 2016. Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. BMC Evol. Biol. 16:212.

Heled J., Drummond A.J. 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol 27:570–580.

Ho L.S.T., Ané C. 2014. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. Syst. Biol. 63:397–408.

Hudson R.R. 2002. Generating samples under a Wright-Fisher neutral model. Bioinformatics 2:337–338.

Hultén E. 1971. The circumpolar plants. 2. Dicotyledons. Stockholm: Almqvist and Wiksell.

Irwin J.J., Stubbs R., Hartman R.L. 2012. *Polemonium elusum* (Polemoniaceae), a new species from east central Idaho, USA. J. Bot. Res. Inst. Texas 6:331–338.

Johnson L.A., Porter J.M. 2017. Fates of angiosperm species following long-distance dispersal: Examples from American amphitropical Polemoniaceae. Am. J. Bot. 104:1729–1744.

Johnson L.A., Chan L.M., Weese T.L., Busby L.D., McMurry S. 2008. Nuclear and cpDNA sequences combined provide strong inference of higher phylogenetic relationships in the phlox family (Polemoniaceae). Mol. Phylogenet. Evol. 48:997–1012.

Johnson L.A., Schultz J.L., Soltis D.E., Soltis P.S. 1996. Monophyly and generic relationships of Polemoniaceae based on matK sequences. Am. J. Bot. 83:1207–1224.

Kamneva O.K., Rosenberg N.A. 2017. Simulation-based evaluation of hybridization network reconstruction methods in the presence of incomplete lineage sorting. Evol. Bioinform. 13:1176934317691935.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

Kriebel R., Drew B.T., Drummond C.P., González-Gallegos J.G., Celep F., Mahdjoub M.M., Rose J.P., Xiang C.-L., Hu G.-X., Walker J.B., Lemmon E.M., Lemmon A.R., Sytsma K.J. 2019. Tracking the temporal shifts in area, biomes, and pollinators in the radiation of *Salvia* (sages, Lamiaceae) across continents: leveraging Anchored Hybrid Enrichment and targeted sequence data. Am. J. Bot. 106:573–597.

Kulbaba M.W., Worley A.C. 2013. Selection on *Polemonium brandegeei* (Polemoniaceae) flowers under hummingbird pollination: in opposition, parallel, or independent of selection by hawkmoths? Evolution 67:2194–2206.

Landis J.B., O'Toole R.D., Ventura K.L., Gitzendanner M.A., Oppenheimer D.G., Soltis, D.E. Soltis P.S. 2016. The phenotypic and genetic underpinnings of flower size in Polemoniaceae. Front. Plant Sci. 6:1144.

Landis J.B., Bell C.D., Hernandez M., Zenil-Ferguson R., McCarthy E.W., Soltis D.E., Soltis P.S. 2018. Evolution of floral traits and impact of reproductive mode on diversification in the phlox family (Polemoniaceae) Mol. Phylogenet. Evol. 127:878–890.

Larget B.R., Kotha S.K., Dewey C.N., Ané C. 2010. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. Bioinformatics 26:2910–2911.

Leaché A.D., Harris R.B., Rannala B., Yang Z. 2014. The influence of gene flow on species tree estimation: a simulation study. Syst. Biol. 63:17–30.

Lee-Yaw J.A., Grassa C.J., Joly S., Andrew R.L., Rieseberg L.H. 2019. An evaluation of alternative explanations for widespread cytonuclear discordance in annual sunflowers (*Helianthus*). New Phytol. 221:515–526.

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

Léveillé-Bourret E., Starr J.R., Ford B.A., Lemmon E.M., Lemmon A.R. 2017. Resolving rapidly radiations within angiosperm families using anchored phylogenomics. Syst. Biol. 67:94–112.

Levin D.A. 2003. The cytoplasmic factor in plant speciation. Syst. Bot. 28:5–11.

Liu L., Yu L., Edwards S.V. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol. Biol. 10:302.

Martin S.H., Davey J.W., Jiggins C.D. 2014. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. Mol. Biol. Evol. 32:244–257.

Meyer B.S., Matschiner M., Salzburger W. 2015. A tribal level phylogeny of Lake Tanganyika cichlid fishes based on a genomic multi-marker approach. Mol. Phylogenet. Evol. 83:56–71.

McVay J.D., Hauser D., Hipp A.L., Manos P.S. 2017a. Phylogenomics reveals a complex evolutionary history of lobed-leaf white oaks in Western North America. Genome 60:733–742.

McVay J.D., Hauser D., Hipp A.L., Manos P.S. 2017b. A genetic legacy of introgression confounds phylogeny and biogeography in oaks. Proc. R. Soc. Lond. 284:20170300.

Meyer M., Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols 2010:pdb.prot5448.

Mirarab S., Reaz R., Bayzid M.S., Zimmermann T., Swenson M.S., Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree. Bioinformatics 30:i541–i548.

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

Mitchell N., Lewis P.O., Lemmon E.M., Lemmon A.R., Holsinger K.E. 2017. Anchored phylogenomics improves the resolution of evolutionary relationships in the rapid radiation of *Protea* L. (Proteaceae). Am. J. Bot. 104:102–115.

Moore W.S. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution 49:718–726.

Morales-Briones D.F., Liston A., Tank D.C. 2018. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). New Phytol. 218:1668–1684.

Morando M., Olave M., Avila L.J., Sites Jr J.W., Leaché A.D., 2020. Phylogenomic data resolve higher-level relationships within South American *Liolaemus* lizards. Mol. Phylogenet. Evol. 147:106781.

Murray D.F., Elven R. 2011. *Polemonium villosissimum* (Polemoniaceae), an overlooked species in Alaska and Yukon Territory. J. Bot. Res. Inst. Texas. 5:19–24.

Olave M., Avila L.J., Sites J.W., Morando M. 2017. Detecting hybridization by likelihood calculation of gene tree extra lineages given explicit models. Methods Ecol. Evol. 9:121–133.

Ostenfeld C.H. 1929. Genetic studies in *Polemonium*. Hereditas 12:33–40.

Patterson N., Moorjani P., Luo Y., Swapan M., Rohland N., Zhan Y., Genschoreck T., Webster T., Reich D. 2012. Ancient admixture in human history. Genetics 192:1065–1093.

Palmer J.D., Herbon L.A. 1988. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. J. Mol. Evol. 28:87–97.

Pease J.B., Hahn M.W. 2015. Detection and polarization of introgression in a five-taxon phylogeny. Syst. Biol. 64:651–662.

Petit R.J., Kremer A., Wagner D.B. 1993. Geographic structure of chloroplast DNA polymorphisms in European oaks. Theor. Appl. Genet. 87:122-128.

Pham K.K., Hipp A.L., Manos P.S., Cronn R.C. 2017. A time and a place for everything: phylogenetic history and geography as joint predictors of oak plastome phylogeny. Genome 60:720–732.

Poczai P., Hyvönen J. 2010. Nuclear ribosomal spacer regions in plant phylogenetics: problems and prospects. Mol. Biol. Rep. 37:1897–1912.

Porter J.M. 1996. Phylogeny of Polemoniaceae based on nuclear ribosomal internal transcribed spacer DNA sequences. Aliso 15:57–77.

Porter J.M., Johnson L.A. 1998. Phylogenetic relationships of Polemoniaceae: inferences from mitochondrial *nad1* intron sequences. Aliso 17:157–188.

Prather L.A., Ferguson C.J., Jansen R.K. 2000. Polemoniaceae phylogeny and classification: Implications of sequence data from the chloroplast gene *ndhF*. Am. J. Bot. 87:1300–1308.

Prum R.O., Berv J.S., Dornburg A., Field D.J., Townsend J.P., Lemmon E.M., Lemmon A.R. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 526:569–573.

Pyron R.A., O'Connell K.A., Lemmon E.M., Lemmon A.R., Beamer D.A., 2020. Phylogenomic data reveal reticulation and incongruence among mitochondrial candidate species in Dusky Salamanders (*Desmognathus*). Mol. Phylogenet. Evol. 146:106751.

Rautenberg A., Hathaway L., Oxelman B., Prentice H.C. 2010. Geographic and phylogenetic patterns in *Silene* section *Melandrium* (Caryophyllaceae) as inferred from chloroplast and nuclear DNA sequences Mol. Phylogenet. Evol. 57:978–991.

Rieseberg L.H., Soltis D.E. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evol. Trends Plants 5:65–84.

Rieseberg L.H. 2001. Chromosomal rearrangements and speciation. Trends Ecol. Evol. 16:351– 358.

Roberts W.R., Roalson E.H. 2018. Phylogenomic analyses reveal extensive gene flow within the magic flowers (*Achimenes*). Am. J. Bot. 105:726–740.

Rokyta D.R., Lemmon A.R., Margres M., Aronow K. 2012. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). BMC Genomics 13:312.

Ronquist F., Teslenko M., Van Der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.

Rose J.P., Kleist T.J., Löfstrand S.D., Drew B.T., Schönenberger J., Sytsma K.J. 2018. Phylogeny, historical biogeography, and diversification of angiosperm order Ericales suggest ancient Neotropical and East Asian connections. Mol. Phylogenet. Evol. 122:59– 79.

Rojas-Andres B.M., Albach D.C., Martinez-Ortega M.M. 2015. Exploring the intricate evolutionary history of the diploid-polyploid complex *Veronica* subsection *Pentasepalae* (Plantaginaceae). Bot. J. Linn. Soc. 179: 670–692.

Sayyari E., Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33:1654–1668.

Sayyari E., Mirarab S. 2018. Testing for polytomies in phylogenetic species trees using quartet frequencies. Genes 9:132.

Schuster T.M., Setaro S.D., Tibbits J.F.G., Batty E.L., Fowler R.M., McLay T.G.B., Wilcox S., Ades P.K., Bayly M.J. 2018. Chloroplast variation is incongruent with classification of the Australian bloodwood eucalypts (genus *Corymbia*, family Myrtaceae). PLoS One 13:e0195034.

Sledge J.L., Anway J.C. 1970. Hybridization between members of *Polemonium delicatum* Ryd. and *P. foliosissimum* A. Gray var. *molle* (Greene) Anway in southern Utah. Am. Midl. Nat. 84:136–143.

Smith R.L., Sytsma K.J. 1990. Evolution of *Populus nigra* (sect. *Aigeiros*): introgressive hybridization and the chloroplast contribution of *Populus alba* (sect. *Populus*). Am. J. Bot. 77:1176–1187.

Snir S., Rao S. 2012. Quartet MaxCut: a fast algorithm for amalgamating quartet trees. Mol. Phylogenet. Evol. 62:1–8.

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

Solís-Lemus C., Bastide P., Ané C. 2017. PhyloNetworks: a package for phylogenetic networks. Mol. Biol. Evol. 34:3292–3298.

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

Steele K.P., Vilgalys R. 1994. Phylogenetic analysis of Polemoniaceae using nucleotide sequences of the plastid gene *matK*. Syst. Bot. 19:126–142.

Stenz N.W., Larget B., Baum D.A., Ané C. 2015. Exploring tree-like and non-tree-like patterns using genome sequences: an example using the inbreeding plant species *Arabidopsis thaliana* (L.) Heynh. Syst. Biol. 64:809–823.

Stubbs R.L., Patterson R. 2013. Revisions in *Polemonium* (Polemoniaceae): a new species and a new variety from California. Madroño 60:243–248.

Stubbs R.L., Folk R.A., Xiang C.-L., Soltis D.E., Cellinese N. 2018. Pseudo-parallel patterns of disjunctions in an arctic-alpine plant lineage. Mol. Phylogenet. Evol. 123:88–100.

Swofford D.L. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). version 4.0a. Sunderland, MA: Sinauer Associates.

Than C.V., Nakhleh L. 2009. Species tree inference by minimizing deep coalescences. PLoS Comp. Biol. 5:e1000501.

Than C., Ruths D., Nakhleh L. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. BMC Bioinformatics 9:322.

Timme R.E. 2001. A molecular phylogeny of the genus *Polemonium* (Polemoniaceae) [M.S. Thesis]. San Francisco State University.

Tsitrone A., Kirkpatrick M., Levin D.A. 2003. A model for chloroplast capture. Evolution 57:1776–1782.

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

Vassiljev V. 1974. *Polemonium* L. In: Shishkin B.K., editor. Flora of the U.S.S.R.,Volume XIX Tubiflorae. Jerusalem: Israel Program for Scientific Translations. p. 58–69.

Wherry E.T. 1942. The genus *Polemonium* in America. Am. Midl. Nat. 27:741–760.
Wherry E.T. 1967. Our temperate tufted polemoniums. Aliso 6:97–101.
Whittemore A.T., Schaal B.A. 1991. Interspecific gene flow in sympatric oaks. Proc. Natl. Acad. Sci. USA 88:2540–2544.
Wickett N.J., Mirarab S., Nguyen N., Warnow T., Carpenter E., Matasci N., Ayyampalayam S., Barker M.S., Burleigh J.G., Gitzendanner M.A., Ruhfel B.R. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl. Acad. Sci. USA 111:E4859–E4868.
Wolf P.G., Murray R.A., Sipes S.D. 1997. Species-independent, geographical structuring of chloroplast DNA haplotypes in a montane herb *Ipomopsis* (Polemoniaceae). Mol. Ecol. 6:283–291.
Worley A.C., Ghazvini H., Schemske D.W. 2009. A phylogeny of the genus *Polemonium* based on amplified fragment length polymorphism (AFLP) markers. Syst. Bot. 34:149–161.
Yi T.-S., Jin G.-H., Wen J. 2015. Chloroplast capture and intra- and intercontinental biogeographic diversification in the Asian-New World disjunct plant genus *Osmorhiza* (Apiaceae). Mol. Phylogenet. Evol. 85:10–21.
Yu Y., Nakhleh L. 2015. A maximum pseudo-likelihood approach for phylogenetic networks. BMC Genomics 16:S10.
Zhang C., Sayyari E., Mirarab S. 2017. ASTRAL-III: increased scalability and impacts of contracting low support branches. In: Meidanis J., Nakhleh L., editors. RECOMB International Workshop on Comparative Genomics. Cham, Switzerland: Springer. p. 53–75.