Ancestral Gene Flow and Parallel Organellar Genome Capture Result in Extreme Phylogenetic Discord in a Lineage of Angiosperms

RYAN A. FOLK1,*, JENNIFER R. MANDEL2, and JOHN V. FREUDENSTEIN1

1Herbarium, Department of Evolution, Ecology and Organismal Biology, The Ohio State University, Columbus, OH 43212, USA and
2Department of Biological Sciences, University of Memphis, Memphis, TN 38152, USA

*Correspondence to be sent to: Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA;
E-mail: rfolk@fsuohio.ufl.edu

Received 27 August 2015; reviews returned 24 August 2016; accepted 4 September 2016

Associate Editor: Mark Fishbein

Abstract.—While hybridization has recently received a resurgence of attention from systematists and evolutionary biologists, there remains a dearth of case studies on ancient, diversified hybrid lineages—clades of organisms that originated through reticulation. Studies on these groups are valuable in that they would speak to the long-term phylogenetic success of lineages following gene flow between species. We present a phylogenomic view of Heuchera, long known for frequent hybridization, incorporating all three independent genomes: targeted nuclear (~400,000 bp), plastid (~160,000 bp), and mitochondrial (~470,000 bp) data. We analyze these data using multiple concatenation and coalescence strategies. The nuclear phylogeny is consistent with previous work and with morphology, confidently suggesting a monophyletic Heuchera. By contrast, analyses of both organellar genomes recover a grossly polyphyletic Heuchera, consisting of three primary clades with relationships extensively rearranged within these as well. A minority of nuclear loci also exhibit phylogenetic discord; yet these topologies remarkably never resemble the pattern of organellar loci and largely present low levels of discord interspecifically. Two independent estimates of the coalescent branch length of the ancestor of Heuchera using nuclear data suggest rare or nonexistent incomplete lineage sorting with related clades, inconsistent with the observed gross polyphyly of organellar genomes (confirmed by simulation of gene trees under the coalescent). These observations, in combination with previous work, strongly suggest hybridization as the cause of this phylogenetic discord. [Ancient hybridization, chloroplast capture, incongruence; phylogenetics; reticulation.]

Since Darwin’s introduction of the theory of evolution by natural selection, biologists have debated exactly how biodiversity arises in nature and which processes are most important. Hybrids, the product of sexual reproduction between different species, varieties, or races, have long held a contentious position, at least since Lotsy’s (1916) well-known treatise arguing it is the most important of all causes, surpassing or replacing natural selection. Though this position is probably extreme, and modern authors do not see hybridization and selection as mutually exclusive, botanists in particular have long suspected a significant role of hybridization in diversification (Anderson 1949; Stebbins 1950; Grant 1981), where it serves as a potential source of new variation through introgression as well as a direct source of novel lineages through hybrid speciation. Molecular studies from a variety of taxa across the tree of life have increasingly acknowledged that hybridization is an important source of evolutionary novelty (Mallet 2007; Solis and Solis 2009). Genomic data, while confirming earlier hypotheses in the land plants, have revealed numerous novel instances in unexpected taxa (Cui et al. 2013; Martin et al. 2013; Garrigan et al. 2015; Liu et al. 2015; Sullivan et al. 2014). These findings are eroding foundational notions reinforced by the neo-Darwinian synthesis, which considered the sterile mule, clearly not the beginning of an evolutionary radiation, as representative of all biological hybrids (Mayr 1963).

The most well-understood plant systems in which hybridization has been studied are very recent lineages, such as Spartina anglica (Ainouche et al. 2004), Tragopogon miscellas (Soltis et al. 2004), desert Helianthus (Rieseberg and Willis 2007), and Louisiana Iris (Arnold et al. 1990). A more complete understanding of the long-term success of hybrids, and, therefore, their contribution to biodiversity, should incorporate hybrid lineages at a variety of ages, including instances of organismal lineages that share a single homoploid hybridization event and subsequently experienced cladogenetic events. Yet hybrid lineages, particularly at the diploid level, remain difficult to detect and are rarely reported (reviewed by Solis and Solis 2009), such that identifying suitable test cases remains challenging. Study of these ancient diversified clades—clades whose ancestors originated by hybridization—would advance the dialog about hybridization from whether it occurs or persists in nature to whether and how hybrids are capable of passing from the tokogenetic relationships of a population to the phylogenetic relationships of a species complex (sensu Henning 1966, fig. 4; contra Wagner 1970, fig. 1).

The genomic consequences of interspecific gene flow are a major point of interest. It is well understood that hybridization creates a mosaic genome, composed of a recombined patchwork of chromosomal stretches having fundamentally different underlying histories, inherited from each parent. Yet given a general lack of knowledge about the commonality and success of different pathways to interspecific gene flow, it is uncertain whether hybrids usually contain approximately equal contributions from each parent or whether most hybridization events involve more or less asymmetrical contributions. Smaller data sets now dominate the hybridization literature, which
Methods

Materials

We included 42 of 43 recognized species of Heuchera and samples of most varieties recognized in recent monographic treatments (Rosendahl et al. 1936; Wells 1984; Wells and Shipes 2009; Folk and Freudenstein 2014b). Several suspected hybrids (on the basis of morphology, cf. Rosendahl et al. 1936; Wells 1984; Wells and Shipes 2009) were included as well. On the basis of recent multiclonal work with Sanger-sequenced loci (Folk and Freudenstein 2014a), four species were chosen as outliers to represent other major lineages in the Heuchera group of genera: Mitella diphylla (Mitella sensu stricto), Mitella stauropetala (Ozonelis group, which also includes Elmera, Tiarella, and Coniminitella), Mitella pentandra (Pictantia group, which also may include Tolmiea and Tellima), and Bensoniella oregona. DNAs were extracted from fresh, silica-dried, or herbarium material (up to ~70 years old, as low as 115 ng total) by a CTAB method (Doyle and Doyle 1987) or by a DNEasy kit (Qiagen) as described previously (Folk and Freudenstein 2014a). The methods that follow are summarized from a detailed protocol presented in (Folk et al. 2015).

Library Preparation

DNAs were quantified with a Qubit (dsDNA BR assays: Thermo Fisher Scientific, Waltham, MA) and assessed for quality with a NanoDrop 2000 (Thermo Fisher Scientific). Any DNAs too dilute for library preparation were concentrated via vacuum centrifugation. We prepared 70 genomic libraries using a NEBNext Ultra kit (New England Biolabs, Ipswich, MA), with the following modifications: we used 1 μg input DNA sheared for 500–700 bp fragments with a Covaris sonicator (model S220; Covaris, Woburn, MA), bead-based size selection aimed for 500–700 bp, using bead cleanup protocol with 0.8 volumes of Ampure XP beads (Beckman Coulter, Brea, CA) for the first step and 0.115 volumes for the second; PCR enrichment used six cycles (most samples) or eight cycles (herbarium and low-yield samples). Libraries were barcoded with 24 NEBNext multiplex oligos (New England Biolabs). Libraries were quantified with a Qubit (any libraries that were of low yield were again vacuum centrifuged) and quality was assessed with an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA).

Target Enrichment

We used in-solution biotinylated probes for a previously developed panel of 277 low-copy nuclear loci based on genomic sequences from Heuchera parviflora var. saurensis (mean length 1362 bp; range 555–3672 bp; total target length 378,553 bp; Folk et al. 2015; synthesized for us by MYcroarray, An Arbor, MI) that enrich for intron-containing loci. We followed the MYcroarray MyBaits v. 2 protocol, with the following modifications: seven to eight libraries were pooled per enrichment reaction (60 ng each for 420–480 ng total input DNA in 6 μL; nine reactions total); nucleotide hybridization was performed for 36 h; the thermocycler profile used 10 μL of template, an annealing temperature of 60°C, an elongation time of 45 s, and 15 total cycles. All enriched library pools were cleaned by repeating the NEBNext magnetic bead cleanup protocol with 0.8 volumes of Ampure
XP to 1 volume of library pool. The enriched library pools were again assessed for quality with an Agilent Bioanalyzer, and quantified with a KAPA Illumina library quantification kit (Kapa Biosystems, Wilmington, MA). The success of enrichment was verified with a qPCR protocol described elsewhere (Folk et al. 2015).

Finally, three enriched library pools (seven to eight samples each) were pooled to give 22–24 samples per each of three final pools; each of these was sequenced on an Illumina MiSeq (W. Harry Feinstone Center for Genomic Research, University of Memphis, TN). The first pool of samples was sequenced twice using a 600-cycle MiSeq kit (i.e., paired-end 300 bp reads). For the subsequent two sample pools, we used a 500-cycle kit since the insert sizes were found to be shorter than anticipated; we sequenced each of these sample pools once since locus coverage from each of the first runs was found to be quite high (see Online Appendix 1 available on Dryad at http://dx.doi.org/10.5061/dryad.cd546). Prior to analysis, libraries were demultiplexed through the MiSeq controller software and all reads were trimmed using Trimomatic v. 0.33 (Bolger et al. 2014) with a sliding window of 20 bp, accepting quality scores of Q20 or better.

Marker Assembly

As references for short read assembly, we used the same low-copy nuclear genomic sequences from *H. parviflora* var. *saurensis* that were used for probe design. We used the Burrows-Wheeler Alignment tool (BWA v. 0.7.12; Li and Durbin 2009; Li 2013), which for this NGS protocol performed better than de novo methods (see Folk et al. 2015). Since organellar phylogenies are of particular interest in *Heuchera* due to the presence of chloroplast capture (Soltis et al. 1991b; Soltis and Kozloff 1995), we also assembled off-target organellar genomic data. We used BWA to assemble these data as well, using as references three previously assembled complete plastid genomes (*H. parviflora* var. *saurensis*, *H. sanguinea*, and *H. parviflora*; Folk et al. 2015) and one mitochondrial genome (*H. parviflora* var. *saurensis*; Folk et al. 2015). For the plastid phylogeny, all reads from each individual were mapped to the closest of the three available *Heuchera* plastid genomes based on a Sanger-sequenced plastid phylogeny (see Online Appendix 2 available on Dryad). For the mitochondrial phylogeny, all reads from each individual were mapped to the mitochondrial genome of *H. parviflora* var. *saurensis*. While the *Heuchera* plastid genome does not appear to have foreign DNA from the mitochondrion (determined using BLAST searches), the mitochondrion has significant amounts of horizontally transferred plastid DNA as in other plants (Goremykin et al. 2008). To ensure that the mitochondrial analysis only reflects the history of uniquely mitochondrial DNA, all BLAST hits of the mitochondrial reference genome to the *H. parviflora* var. *saurensis* plastid genome were deleted before assembly. BWA contigs were imported into Geneious 7 and consensus sequences were extracted using default settings (in particular, trimming the consensus to the length of the reference sequence, calling regions with no coverage as ‘?’, and calling bases by per-base quality, which includes ambiguity code calls). We did not perform phasing on the data (consistent with other phylogenomic studies, e.g., Mirarab et al. 2014b), since methods like PHASE (Stephens et al. 2001) are not tractable for large data sets and highly diverged samples. Hence, the consensus sequences produced are referred to here as genotypic sequences rather than allelic sequences.

Since extensive chloroplast capture is known from the *Heuchera* group of genera, to the extent that *Heuchera* is polyphyletic for plastid markers (Soltis et al. 1991b), outgroup choice is nontrivial; we mapped publicly available reads of *Saxifraga granulata* (van der Meer et al. 2014) to the *H. parviflora* plastid and mitochondrial genomes to develop an appropriate outgroup well outside of the *Heuchera* group of genera, and used this sequence to root these trees.

Low-Copy Nuclear Phylogenetic Analysis

A number of issues surrounding phylogenomic analysis, such as intralocus recombination, lineage sorting, and reticulation, are likely for this data set, meaning that all assumptions inherent in any single phylogenetic analysis method are unlikely to be met. In particular, the choice between coalescent and concatenation methods may partly depend on the *a priori* expected frequency of incomplete lineage sorting (ILS; Bayzid and Warnow 2013; Mirarab et al. 2014b), which is generally unknown for empirical systems. Therefore, we performed phylogenetic estimation using a plurality of approaches.

As an initial estimate of phylogenetic relationships, all 277 retained loci were aligned individually in MAFFT (v. 7.017; Katoh et al. 2009) implemented in Geneious 7 using default settings (gap opening penalty 1.53; offset 1223) and concatenated in Geneious for a maximum likelihood analysis in RAxML (command-line version, v. 8.1.3). Alignments were individually examined but not edited, as they typically appeared unambiguous; gaps were treated as missing data. We used 500 rapid bootstrap replicates combined with an ML tree search from every fifth bootstrap tree (option -fa) to assess clade support; for computational efficiency, we initially used an unpartitioned GTR+I model. Partitioning schemes were not straightforward to address. For these loci, introns have 4-fold the pairwise divergence of exons (Folk et al. 2015), motivating the incorporation of splicing patterns in a partitioned analysis. Fully partitioning by both locus and splice site yields 1961 partitions; an attempt to agglomerate these using PartitionFinder with strict hierarchical clustering (v. 1.0.1; Lanfear et al. 2014) resulted in 1850 partitions, which still remains intractable in terms of runtimes (data not shown).
Therefore, to incorporate the most likely source of site substitution rate variation without an excessive number of partitions, we ran the concatenated analysis with two partitions, containing (i) all exons, and (ii) all introns or intronic regions (the vast majority were intronic), using 1000 fast bootstraps. Additionally, we ran a concatenated analysis with the loci treated as separate partitions ($n = 277$) with other analysis options kept the same as the splice-site partitioned analysis, to address potential gene heterogeneity.

For taxa of section Rhodoheuchera, we observed behavior consistent with the presence of rogue taxa; this conflict could arise from a variety of phenomena such as homoplasys or hemiplasy, the latter possibly due either to ILS or to hybridization. Reanalysis of a data set excluding rogues is justified, since the characteristic ambiguity of taxon placement suggests that we cannot confidently place these taxa given the data. Removing them would reveal consistent placement of related nonrogue taxa that is obscured due to information lost by using consensus techniques such as bootstrap summaries. We, therefore, employed RogueNaRok (Aberer et al. 2012, implemented as a webserver at http://rnr.h-its.org/), to identify and remove these OTUs. RogueNaRok was run with a maximum dropset size of five, optimizing for support using a majority rule consensus threshold. Four taxa were identified for removal: H. woodsiaphila, H. pulchella, H. caespitosa, and H. brevistamnma. The resultant OTU set ($n = 66$, reduced from 70) was reanalyzed in RAxML using the same settings as the full analysis.

We measured the tree-likeness of the concatenated nuclear alignment using SplitsTree employing the average s score (Holland et al. 2002), a metric based on quartet distances that ranges from 0 (entirely tree-like) to 1 (not tree-like). We summarized patterns of conflict among nuclear markers by inferring a split network on the concatenated data set with the NeighborNet algorithm and uncorrected $P$ distances in SplitsTree4 (Huson 1998). To address the possibility of lineage sorting confounding concatenation-generated estimates of the species tree (while still not addressing hybridization concerns), we used coalescent phylogenetic inference. To infer a coalescent tree on gene trees based on entire loci, we used ASTRAL-II (multi-individual version), which is accurate in simulations compared to other coalescent methods based on point estimates of gene trees (Mirarab et al. 2014b; Mirarab and Warnow 2015). It also has the advantage of utilizing unrooted gene trees, which is important given that the correct gene tree rooting was ambiguous for some gene trees and that random rooting can mimic the coalescent (Rosenfeld et al. 2012; Tam and Kubatko 2014). All 277 locus alignments were used to infer gene trees with a custom batch script, using RAxML with the GTR-$\Gamma$ model and 1000 fast bootstrap replicates, with M. stiarentilea treated as the root (cf. Online Appendices 3 and 4 available on Dryad), based on previous work (Folk and Freudenstein 2014a). The ML trees were used in ASTRAL-II along with the bootstrap samples and run with 1000 coalescent bootstrap replicates; bootstrap frequencies were plotted on the tree inferred from ML gene trees. We used the most up-to-date taxonomic sources (Wells and Shipes 2009; Folk 2013; Folk and Freudenstein 2014b; Folk and Alexander 2015) to assign our consensus sequences to putative species for species tree analysis (see Online Appendix 5 available on Dryad), with varieties and putative hybrids treated as subspecific samples of recognized species according to current taxonomic assignment in the literature. The inclusion of putative hybrids is argued as reasonable as a total-evidence analysis of the data (i.e., no taxon exclusion; since their hybrid origin is under investigation), while not treating them as separate taxa, given that they have not been considered hybrid species in recent taxonomic works (Rosendahl et al. 1996; Wells 1984; Wells and Shipes 2009). Input sequences were unphased genotypic consensus sequences. Support values from the coalescent bootstrap were plotted on the species tree estimated using the ML trees. ASTRAL-II was originally designed to handle one sequence per species and the implementation with multiple sequences per species is currently experimental (Mirarab S., personal communication); however, the estimated topology was highly congruent with an ASTRAL-II analysis that we ran with individual sequences treated as species (Online Appendix 6 available on Dryad) and also with an MP-EST run with the same sequence-to-species associations (below this section). It is important to note this similarity in results between treating individual sequences as species and assigning multiple sequences to species (also observed elsewhere, e.g., Stephens et al. 2015). While assigning sequences to taxonomic species hypotheses is considered the best practice, strong incongruence between the two data treatments should be examined and may indicate species delimitation problems.

As an alternative approach, we used a second coalescent method, SVDquartets (Chifman and Kubatko 2014; implemented in PAUP v. 4.0a142). This is a method designed for SNP data that treats each nucleotide polymorphism as having been drawn from an independent gene tree distribution. It is also known to perform well on large multilocus data sets in simulation, despite violation of the assumption of independence among sites (Chifman and Kubatko 2014). As with ASTRAL-II, this algorithm uses unrooted phylogenetic quartets and, therefore, does use gene tree rooting information. The input matrix consisted of all 277 concatenated low-copy nuclear loci, from which all invariant sites were stripped to simulate SNP data (SVDquartets can be run with or without invariant sites; Chifman and Kubatko 2014; Kubatko L., personal communication). Support was assessed using 1000 bootstrap replicates and sampling 100,000 random quartets per replicate. The assignment of sequences to putative species was the same as that used in the ASTRAL-II analysis. Bootstrap support (BS) values were plotted on a majority-rule bootstrap consensus (as in
Chifman and Kubatko 2014). Topology comparisons between the two coalescent methods were performed through the tanglegram tool of Dendroscope (v. 3.2.10; Scornavacca et al. 2011; Huson and Scornavacca 2012).

Coalescent methods typically assume that input genes have not undergone intralocus recombination during the phylogenetic history of the taxa being analyzed, an assumption that is open to criticism (Gatesy and Springer 2014). To explore the possibility of recombination in our data set, we tested each locus using the permutation-based test (see Online Appendix 7 available on Dryad), which does not make assumptions concerning the relationships of the samples (Bruen 2005), as implemented in SplitsTree4 (Huson 1998; see Online Appendix 8 available on Dryad). We then ran the ASTRAL-II analysis only on loci that showed no significant recombination, with all other analysis options identical to those in the analysis of all loci.

Concordance Analysis

To summarize phylogenetic conflict across the genome, gene tree distributions were used to calculate concordance factors and coalescent branch lengths using BUCKy (v. 1.4.3; Ane et al. 2007; Baum 2007; Larget et al. 2010), which incorporates gene tree estimation error via Bayesian MCMC samples. As with the coalescent methods, BUCKy does not use rooted gene trees. The 277 nuclear locus alignments were used to infer gene tree distributions using MrBayes 3.2.5 (Ronquist and Huelsenbeck 2003), called by a custom batch script, sampling each every 1000 generations for 1,000,000 total generations under a GTR+I+G1 model, but otherwise using default settings (e.g., four MCMC chains of which three are heated, discarding the first 25% of samples as burn-in; these gene tree distributions are summarized on 50% majority rule consensus trees labeled with clade posterior probabilities in Online Appendix 4 available on Dryad). This number of generations was enough to return standard deviations of bipartitions that were almost always ~0.01 and always <0.02. For a data set of this scale, computation in BUCKy was found to be intractable, likely requiring time on the order of months; we, therefore, subsampled the gene tree distributions to reduce computational time significantly (which is expected given the use of a reference to assemble the data) but otherwise using default settings. Whereas plastid sequences were essentially complete, mitochondrial assemblies had more missing data. We addressed this by stripping all alignment columns with >75% missing data (analyses were also attempted with 50% removed and no sites removed, with few differences, not shown). Four samples that had less than 40% mitochondrial genome completeness (H. caroliniana, H. merriami, Heuchera micrantha var. erubescens, and the putative hybrid H. rubescens × H. cylindrica) were not included in the alignment, leaving 66 taxa.

While MP-EST may be less accurate than ASTRAL-II for estimating topology based on simulated data (Mirarab et al. 2014b), and it assumes rooted gene trees, the current implementation of ASTRAL-II (multi-individual version) does not estimate branch length; hence, we used it to have an independent estimate of coalescent branch lengths to corroborate the results from BUCKy (see results of both in Online Appendix 9 available on Dryad). The same ML trees used for the ASTRAL-II analysis were used in the STRA W webserver implementation of MP-EST (Shaw et al. 2013; available at http://bioinformatics.publichealth.uga.edu/SpeciesTreeAnalysis/index.php). BUCKy cannot use multiallelic data, but the map used for MP-EST was the same as that used for ASTRAL-II and SVDquartets.

Organellar Phylogenetic Analysis

Whole plastid genomes were aligned in MAFFT using default settings and the entire plastid matrix was analyzed in RAxML using an unpartitioned GTR+I+G1 model with 5000 bootstrap replicates. Additionally, we accounted for potentially differing evolutionary rates by running an analysis with two partitions: (i) all coding DNA (exons, tRNAs, and rDNA) and (ii) all noncoding DNA (introns and intergeneric regions). This analysis was otherwise run with the same settings as the unpartitioned analysis. Conventional plant plastid genomes contain a large inverted repeat (IR), such that a significant minority of plastid DNA is present in two copies. Sites from one of the IR region copies (133,984–159,735 bp) were excluded from the alignment.

The use of MAFFT to align the longer mitochondrial sequences (~550,000 bp) was found to be computationally excessive, so we used Mauve (Darling 2004; v. 2.3.1, implemented in Geneious 7) to align the genomes, enforcing collinear genomes (which is expected given the use of a reference to assemble the data) but otherwise using default settings. Whereas plastid sequences were essentially complete, mitochondrial assemblies had more missing data. We addressed this by stripping all alignment columns with >75% missing data (analyses were also attempted with 50% removed and no sites removed, with few differences, not shown). Four samples that had less than 40% mitochondrial genome completeness (H. caroliniana, H. merriami, Heuchera micrantha var. erubescens, and the putative hybrid H. rubescens × H. cylindrica) were not included in the alignment, leaving 66 taxa. Additionally, to test the effect of other low-coverage samples on the mitochondrial topology, a reduced matrix was constructed consisting only of the upper 50th percentile of individuals in terms of coverage. Both mitochondrial matrices were analyzed in RAxML using an unpartitioned GTR+I+G1 model; this analysis was more computationally intensive than the nuclear and plastid analyses, so we used only 1000 bootstrap replicates. Additionally, as with the plastid analysis, a two-partition analysis with coding and noncoding
DNA was run on the 66-taxon mitochondrial data set with the same settings as the unpartitioned analysis. Tanglegrams were again generated in Dendroscope to compare the plastid and mitochondrial trees to the low-copy nuclear tree. We also inferred a cluster network (Huson and Rupp 2008) using the plastid tree and the 70-taxa concatenated nuclear tree, also in Dendroscope.

Sanger Sequencing

To increase the sampling of genera outside of Heuchera for the plastid genome, we performed Sanger sequencing of three noncoding markers: trnL-F, rps2-trnL, and rps16-trnK (for thermocycler conditions and primer pairs, see Online Appendix 10 available on Dryad; for voucher and GenBank information, see Online Appendix 11 available on Dryad). Outgroup taxa added were Conomitella, Tiarella, Elmera, Mitella acerina, Mitella caulescens, Mitella ovalis, Tellima, Lithophragma, Tolmiea, Darmera, Rodgersia, and Peltoboykinia; this analysis was rooted with the latter three taxa. These loci were aligned independently in MAFFT, concatenated, and analyzed in RAxML with the GTR/Gamma1 model, with 5000 fast bootstrap replicates.

Organelar Gene Tree Simulations

To differentiate chloroplast capture from ILS, we simulated ILS under the model of an organelar gene tree, generating a distribution of 1,000 gene trees under the coalescent with the DendroPy package (executed with the gene tree simulation pipeline of Mirarab et al. 2014a), using the 46-taxon MP-EST result as a guide tree, with all branch lengths scaled by a factor of four to account for organellar inheritance. Matrilineal inheritance has been demonstrated for both the chloroplast genome (Soltis et al. 1990) and the mitochondrial genome (full-length assembly) was 146.4 kbp. The aligned length of the plastid genome matrix (one IR excluded) was 137,758 bp, of which 2119 characters (1.5%) were parsimony informative. The plastid matrix had 1.2% undetermined characters and 6.0% gap characters. For ingroup taxa, 1948 characters (1.4%) were parsimony informative, 11% undetermined, and 6.0% were gap characters. The average coverage across individuals for the mitochondrial genome was 13.9 x. The aligned length of the mitochondrial genome matrix (with columns >75% missing data stripped) was 467,783 bp, of which 6133 characters (1.3%) were parsimony informative (5601 or 1.2% considering only ingroup taxa). The mitochondrial matrix had 60.6% missing data or 47.1% considering only ingroup taxa. The q tests indicated that 116/277 genes showed significant signs of recombination (see Online Appendix 7 available on Dryad); hence, recombination is common for this data set. Subsampling the BUCKy analysis was successful; concordance factors calculated on fully sampled and subsampled posterior tree distributions differed on average by 0.2% (see Online Appendix 16 available on Dryad), demonstrating that subsamples gave adequate estimates of concordance factors. Additionally, all concordance factors on the maximum concordance tree were identical across all three α values (Online Appendix 17).

Concatenated Phylogenetic Inference

The partitioned and unpartitioned concatenated analyses (Fig. 1, Online Appendices 18 and 19 available on Dryad) corroborate the monophyly of Heuchera, with strong support (BS 100), as observed in earlier studies (Soltis and Kuzoff 1995; Okuyama and Pellmyr 2008; Folk and Freudenstein 2014a). Most other relationships in the tree were similarly strongly supported. The tree contains three major clades recovered previously (Folk and Freudenstein 2014a): section Holochloa (Fig. 1, orange; BS 100); section Heuchera (blue; BS 100), and section Rhodoheuchera (magenta; BS 50). A fourth clade, section Bracteatae (green), tentatively suggested by (Folk and Freudenstein 2014a) based on combined ribosomal data and morphology, was reconstructed here as paraphyletic with strong support for the relevant nodes (BS 100 and 75 for two nodes contradicting group monophyly).

Most nodal support values were high (BS ≥ 95). However, nodal support values within section Rhodoheuchera were lower; the deletion of four rogue taxa (Online Appendix 20 available on Dryad) considerably

RESULTS

The average sequencing coverage for each nuclear locus, averaged across OTUs, was 573.9 x; the least-covered locus had 237.7 x average coverage. The aligned length of the concatenated 70-taxa, 277-locus low-copy nuclear matrix was 404,475 bp, of which 26,305 characters (6.5%) were parsimony informative. The nuclear matrix had 0.4% undetermined characters (no coverage of target sequence or gap heterozygosity) and 70% gap characters (only indels). Considering only the ingroup taxa in the nuclear matrix, 22,090 characters (5.4%) were parsimony informative, 0.2% were undetermined, and 70% were gap characters. The average coverage across individuals for the plastid genome (full-length assembly) was 146.4 x. The aligned length of the plastid genome matrix (one IR excluded) was 137,758 bp, of which 2119 characters (1.5%) were parsimony informative. The plastid matrix had 1.2% undetermined characters and 6.0% gap characters. For ingroup taxa, 1948 characters (1.4%) were parsimony informative, 11% undetermined, and 6.0% were gap characters. The average coverage across individuals for the mitochondrial genome was 13.9 x. The aligned length of the mitochondrial genome matrix (with columns >75% missing data stripped) was 467,783 bp, of which 6133 characters (1.3%) were parsimony informative (5601 or 1.2% considering only ingroup taxa). The mitochondrial matrix had 60.6% missing data or 47.1% considering only ingroup taxa. The q tests indicated that 116/277 genes showed significant signs of recombination (see Online Appendix 7 available on Dryad); hence, recombination is common for this data set. Subsampling the BUCKy analysis was successful; concordance factors calculated on fully sampled and subsampled posterior tree distributions differed on average by 0.2% (see Online Appendix 16 available on Dryad), demonstrating that subsamples gave adequate estimates of concordance factors. Additionally, all concordance factors on the maximum concordance tree were identical across all three α values (Online Appendix 17).

Concatenated Phylogenetic Inference

The partitioned and unpartitioned concatenated analyses (Fig. 1, Online Appendices 18 and 19 available on Dryad) corroborate the monophyly of Heuchera, with strong support (BS 100), as observed in earlier studies (Soltis and Kuzoff 1995; Okuyama and Pellmyr 2008; Folk and Freudenstein 2014a). Most other relationships in the tree were similarly strongly supported. The tree contains three major clades recovered previously (Folk and Freudenstein 2014a): section Holochloa (Fig. 1, orange; BS 100); section Heuchera (blue; BS 100), and section Rhodoheuchera (magenta; BS 50). A fourth clade, section Bracteatae (green), tentatively suggested by (Folk and Freudenstein 2014a) based on combined ribosomal data and morphology, was reconstructed here as paraphyletic with strong support for the relevant nodes (BS 100 and 75 for two nodes contradicting group monophyly).

Most nodal support values were high (BS ≥ 95). However, nodal support values within section Rhodoheuchera were lower; the deletion of four rogue taxa (Online Appendix 20 available on Dryad) considerably
**FIGURE 1.** Optimal likelihood tree inferred from concatenated low-copy nuclear loci with all 70 taxa, using an exon/intron partitioning scheme. Support values $\geq 50$ are plotted on branches; all very strongly supported clades (BS $\geq 95\%$) are represented by a black diamond. Branches are shown proportional to ML branch lengths. Taxonomic sections as currently recognized are indicated by branch coloration as a visual reference to relationships. The scale represents branch lengths measured in per-site substitution rate.

- **Sect. Holochloa**
- **Sect. Heuchera**
- **Sect. Rhodoheuchera**
- **Sect. Bracteatae**
- **Incertae sedis**
- **Outgroup**
improved support values for the Rhodoheuchera clade. The remaining low support values, other than for section Rhodoheuchera, largely concern subspecific samples and, therefore, were less relevant to our focal questions. The nuclear concatenated analysis was minimally affected by partitioning scheme. The analysis partitioned by splice site (Fig. 1) was topologically identical to the unpartitioned analysis (Online Appendix 18 available on Dryad), with support values that differed by single digits. There were a few topological differences in the analysis partitioned by genetic locus (Online Appendix 19 available on Dryad); however, all of these received low support.

Coalescent Phylogenetic Inference

Relationships in the coalescent analyses (Fig. 2) were largely similar to those recovered by concatenation: the monophyly of Heuchera was decisively supported (BS 100 in ASTRAL-II, 98 in SVDquartets). As with concatenation, both coalescent methods found monophyletic sections Heuchera, Holochloa, and Rhodoheuchera, while finding section Bracteatae clearly paraphyletic. Interestingly, section Rhodoheuchera received higher support in the coalescent analysis than in the 70-taxon concatenation analysis (ASTRAL-II, BS 100; SVDquartets, BS 98). Hence, backbone relationships were nearly identical for both coalescent methods and for concatenation. Among relationships closer to the tips of the phylogeny, there were more differences between coalescence and concatenation, but these were never consistent among the coalescent methods, and generally not well-supported, suggesting a lack of significant disagreement with concatenation results (Table 1, Fig. 2). Two exceptions to this observation of low support for disagreement occurred: for SVDquartets, M. pentandra was sister to Heuchera rather than to Bensoniella and M. diphylla (BS 89). However, ASTRAL-II found the latter relationship (BS 78), in agreement with concatenation. This difference among outgroups is largely a rooting issue and may be caused by our relatively sparse sampling of outgroup taxa. Additionally, ASTRAL-II supported a sister relationship of H. woodsiaphila with section Rhodoheuchera (BS 93), whereas concatenation and SVDquartets (the latter weakly, BS 54) suggest that section Rhodoheuchera is sister to H. bracteata and H. hallii.

Organelar Phylogenetic Relationships

The plastid analysis recovered a polyphyletic Heuchera, in contrast with all nuclear analyses, with species distributed in three strongly supported clades. We will refer to these as clades A, B, and C (see labels in Fig. 3, Online Appendices 21 and 22 available on Dryad). Two of these clades have been recovered previously, though with less taxon sampling. Clade A contains species of Heuchera that Soltis et al. (1991b)
TABLE 1. A summary of conflicting support for major clades among data sets

<table>
<thead>
<tr>
<th>Clade monophyly</th>
<th>C</th>
<th>Cr</th>
<th>Csp</th>
<th>Csp*</th>
<th>A*</th>
<th>S*</th>
<th>C*</th>
<th>M*</th>
<th>CF\text{_nuc}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heuchera</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>98</td>
<td>100</td>
<td>100</td>
<td>0.680</td>
</tr>
<tr>
<td>Section Holochloa*</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>95</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0.048</td>
</tr>
<tr>
<td>Subsection Villosa</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99</td>
<td>100</td>
<td>100</td>
<td>0.519</td>
</tr>
<tr>
<td>Subsection Cylindricalialae</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99</td>
<td>99</td>
<td>100</td>
<td>100</td>
<td>0.241</td>
</tr>
<tr>
<td>Subsection Micrantha</td>
<td>94</td>
<td>92</td>
<td>86</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>100</td>
<td>0.231</td>
</tr>
<tr>
<td>H. glabra + Micrantha</td>
<td>94</td>
<td>92</td>
<td>86</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>100</td>
<td>0.024</td>
</tr>
<tr>
<td>Section Heuchera</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0.217</td>
</tr>
<tr>
<td>Eastern S. Heuchera</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>97</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0.091</td>
</tr>
<tr>
<td>Western S. Heuchera</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>98</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0.079</td>
</tr>
<tr>
<td>Section Rhodeoheuchera</td>
<td>58</td>
<td>99</td>
<td>62</td>
<td>97</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>100</td>
<td>0.004</td>
</tr>
<tr>
<td>S. Rhodeoheuchera + H. brachata + H. hallii</td>
<td>75</td>
<td>100</td>
<td>100</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>100</td>
<td>0.004</td>
</tr>
<tr>
<td>S. Rhodeoheuchera + H. woodii</td>
<td>75</td>
<td>n/a</td>
<td>75</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>100</td>
<td>0.003</td>
</tr>
<tr>
<td>Subsection Hensleyana</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>0.030</td>
</tr>
<tr>
<td>Subsection Rosendahlii</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>0.222</td>
</tr>
<tr>
<td>Subsection Rosendahlii + H. sanguinea</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>100</td>
<td>0.020</td>
</tr>
<tr>
<td>H. elegans group (Chloro. clade C)</td>
<td>99</td>
<td>100</td>
<td>99</td>
<td>99</td>
<td>96</td>
<td>97</td>
<td>97</td>
<td>100</td>
<td>0.002</td>
</tr>
<tr>
<td>Mitochondrial Heuchera + H. elegans group</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0.000</td>
</tr>
<tr>
<td>Plastid clade A</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0.000</td>
</tr>
<tr>
<td>Plastid clade A + M. pentandra</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0.000</td>
</tr>
<tr>
<td>Plastid clade B</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Notes: Green cell = clade supported. Red cell = clade rejected. Yellow cell = clade rejected, but a similar clade was present that differed only by four OTUs (see text, Fig. 3). C = concatenated analysis; Cr = concatenated analysis with rogue sequences removed; Csp = concatenated with splice-site partition; Csp* = concatenated with by-locus partition. A* = ASTRAL-II analysis; S* = SVDquartets analysis. C* = plastid analysis; M* = mitochondrial analysis. For organellar analyses, partitioning gave essentially identical results, so only the partitioned results are shown. Numbers given in columns 2–7 are BS values in each analysis, either in support of the clade (for green cells) or the highest support value among incompatible clades (for red cells). CF\text{\_nuc} (last column) = concordance factor for the clade in 277 nuclear genes.

a Section Holochloa here taken to include H. glabra.

Hypothesized had undergone ancestral introgression of plastid DNA with the related genus Tiarella (sampled in the Sanger analysis, Online Appendix 2 available on Dryad); clade B contains species hypothesized to have retained the ancestral, nonintrogressed Heuchera plastid genome. Clade C, sister to M. diphylla, consisting largely of taxa unsampled in the earlier study, has not been recovered previously. Heuchera is also polyphyletic with respect to the mitochondrial genome. In particular, a clade corresponding in taxon content exactly to plastid clade C is present in the mitochondrial phylogeny and is likewise sister to M. diphylla. Clades mostly corresponding to the content of plastid clades A and B [differing by four and three OTUs, respectively] are also present in the mitochondrial analysis (Fig. 3, Online Appendices 23 and 24 available on Dryad; the similar clades are denoted A' and B', while clade C had identical taxon content in both analyses). One important difference arose: the relationships within mitochondrial clade A are extensively rearranged as compared to those in the plastid tree. Sister to M. pentandra is a clade more restricted in content, containing mostly species of section Holochloa and M. stauropetala (embedded in clade A'; Fig. 3).

The imposition of a coding versus noncoding partitioning scheme resulted in single-digit differences in support values on identical topologies for the mitochondrial analysis (cf. plastid, Online Appendix 22 available on Dryad). Notably, Bensoniella, a longer branch, occupied a more basal position in the unpartitioned analysis, but given the lack of support for this relationship, the well-supported partitioned placement appears optimal. It is important to acknowledge that our mitochondrial genomic assemblies have more missing data and lower coverage than the plastid assemblies; high repeat content and horizontally transferred chloroplast DNA represent further challenges. The analysis of only the upper 50% of mitochondrial genomes in terms of coverage still resulted in a tree congruent with the full result (see Online Appendix 25 available on Dryad), suggesting that the backbone topology is not sensitive to lower coverage in mitochondrial regions.

Organellar Gene Tree Simulations

The branch subtending the clade of all species of Heuchera was estimated in BUCKy to be 1.803 coalescent units long, and it was 1.815 units long in MP-EST (see Online Appendix 9 available on Dryad; estimates use nuclear data; branch length congruence among these two methods indicates that branch lengths were insensitive to the treatment of sequences with respect to putative species, since BUCKy treats individual sequences as tip taxa). Of the simulated plastid gene trees, 100% contained a monophyletic Heuchera. Monophyly was also observed >50% of the time for 17
FIGURE 3. Tanglegram comparing the plastid and mitochondrial phylogenies inferred in partitioned RAxML analyses, optimized in Dendroscope to minimize line crossings. To minimize spurious disagreement between phylogenies due to estimation error, all clades with <50% BS have been collapsed. Nodal labels are clade names, as referred to in the text; mitochondrial clades A' and B' are similar but not identical to chloroplast clades A and B.

of the 40 subclades within Heuchera that were contained in the species tree estimate (Online Appendices 12–13 available on Dryad). Hence, gene tree discordance with the species tree is expected to be moderate within Heuchera for organellar loci, and essentially nonexistent between Heuchera and related genera. Examining the frequency of observed plastid clades resulted in clade frequencies of 0% for the plastid clades (A + M. pentandra), B, and (C + M. diphylla) in the simulated gene tree set, and similarly all other backbone clades and most other clades at shallower levels had very low clade frequencies (Online Appendices 14–15 available on Dryad). When the two simulations were rerun with the original branch lengths, that is, under a diploid, biparentally inherited model, 70% of gene trees still contained a monophyletic Heuchera (congruent with the empirical nuclear estimate in BUCKy at 68%), and the plastid clades remained at 0% frequency (results not shown).

DISCUSSION

Organellar Relationships and Phylogenomic Discord

Both organellar phylogenies were so profoundly discordant from the nuclear phylogeny that points of similarity are not even immediately apparent. As an example, the plastid–nuclear tanglegram (Fig. 4) and cluster network (Online Appendix 26 available on Dryad) demonstrate that while a few small clades are shared among plastid and nuclear phylogenies, the backbones of the two phylogenies have little in common. Similarly, incongruence was very high between mitochondrial and nuclear analyses (Online Appendix 27 available on Dryad). Discordance among genomes has long been observed in plants (Rieseberg and Soltis 1991); hence these results are dramatic, but should not be surprising given the unique population genetic pressures experienced by organellar genomes that lead to rapid haplotype fixation. Importantly, mitochondrial and plastid phylogenies were similar but not identical to each other, a fact to which we return below.

Remarkably, the divergent phylogenetic signal from the two organellar phylogenies was completely unique to them among all gene trees examined. For instance, among the 277 nuclear gene trees, the concordance factor for plastid clades A, (A + M. pentandra), B, and (C + M. diphylla) is 0 (Table 1, rightmost column). Since BUCKy uses posterior distributions of gene trees to incorporate estimation error, this can be taken as indicating that basic aspects of the organellar historical signal are not part of the nuclear data, even considering the possibility of inaccurate gene trees. Manual examination of gene trees (Online Appendices 3 and 4 available on Dryad) confirms the lack of common signal. This contrast among well-sampled genomic compartments has been observed in chipmunks (Good et al. 2015); more studies are needed to explore this issue, which may have broad relevance for explaining cytonuclear discordance. It may be that, for some systems, hybridization was so highly backcrossed or sex-biased as to leave little or no signal in the nuclear genome. Yet, without some biological mechanism to enforce backcrossing or sex-specific hybridization, it seems unparsimonious to rely on such extreme biases in gene flow patterns across multiple lineages in parallel. Another possibility is that hybridization occurred in the context of spontaneous androgenesis (Hedtke and Hillis 2011, where Table 1 provides examples; Alexander
FIGURE 4. Tanglegram comparing plastid and concatenated nuclear phylogenies, optimized in Dendroscope to minimize line crossings. To minimize spurious disagreement between phylogenies due to estimation error, all clades with <70% BS have been collapsed.

et al. 2013), which would create cytonuclear discordance in one generation and avoid the necessity of a sex bias or period of backcrossing among parental taxa. Under one potential scenario (Hedtke and Hillis 2011, fig. 2), an unreduced pollen grain fertilizes an ovule; the nucleus of the egg cell is lost (but not the maternal organelles) so that all nuclear genetic material descends from the paternal parent. Structures in the anther resembling embryo sacs, pollen multinucleation, and other developmental abnormalities have been observed in $H$. micrantha and compared with androgenesis (Vijayaraghavan and Ratnaparkhi 1977), highlighting a role for further developmental studies in this genus.

Causes of Genome Conflict

Both hybridization and ILS cause gene tree discordance, yet expected gene tree distributions under these processes are different. A model-based analysis of the data incorporating both hybridization and the coalescent would ideally complement interpretations of gene tree discord. However, analytical tools available to distinguish between ILS and hybridization with probabilistic methods are currently limited by various constraints in the kinds of hypotheses that can be tested, such as requiring foreknowledge of putative hybrid taxa, population-level sampling of each taxon, a requirement that parents be sister taxa, or requiring that gene flow results in hybrids rather than introgressants (e.g., STEM-by-implementation of Kubatko 2009; Gerard et al. 2011). Additionally, no coalescent-based methods exist that are well suited for phylogenomic data scales and especially large numbers of taxa, so we elected to test hybridization indirectly through examining gene tree distributions expected under ILS alone.

Given a correction of effective population sizes for organellar inheritance (which is matrilineal and effectively haploid), the organellar genomes should experience a branch length of $1.8^4 = 72$ coalescent units in the ancestor of $H$. Heuchera. Coalescent branch lengths ($\lambda = T/2N_e$) above 1.0 generally should result in relatively low levels of ILS (Kubatko and Degnan 2007); hence, this represents a branch length so long that $H$. Heuchera has a high probability of monophyly for the organellar genomes under the assumption that all gene tree discord is due to ILS. The simulations of organellar phylogenies indeed never produced any gene trees that resembled the observed plastid tree, with almost all plastid clade frequencies at or near zero. The calculated branch length values should consistently underestimate the true coalescent branch length, since they are based on attributing gene tree variance (estimation error, gene flow) could be at work and inflate apparent gene tree conflict (e.g., Zhong et al. 2014), and hence this test is a conservative measure because it overestimates true ILS. Depressed branch lengths additionally were clearly observed in the plastid tree as compared to the nuclear tree (see clades A and C, Appendices 21 and 22, available online at Dryad). Gene tree branch length is more difficult to analyze under the coalescent with a data set of this size due to computational requirements (Joly et al. 2009; Joly 2011), but short branch lengths in introgressed genes are an expected outcome of reticulation.
From these observations, it seems highly unlikely that a coalescent process could be responsible for the high degree of intergeneric discord observed. In fact, organellar loci should experience less ILS than nuclear given that they experience ¼ the effective population size (N_e). On the other hand, coalescent branch lengths within the genus were considerably shorter, indicating a potential role for ILS among phylogenetically more proximate species, especially among nuclear loci.

Hypotheses of Organellar Introgression

In agreement with earlier work (Soltis et al. 1991b), Heuchera is here shown to be polyphyletic for the plastid genome, with some species more closely related to taxa from other genera than to other species of Heuchera. Specifically, plastid clades A and B were observed in the earlier study; clade B, with no close relatives, was hypothesized to retain the ancestral Heuchera plastome, with members of clade A having an introgressed plastid genome. Our plastid phylogeny did not contain dense outgroup sampling, so there remains some question as to the sister taxon to plastid clade A and, therefore, the introgressed DNA source. The analysis of Sanger sequences (Online Appendix 2 available on Dryad) has better outgroup sampling and calls into question the previously observed sister relationship between Tiarella and plastid clade A—Tiarella, M. stauropetala, and Conimbriga williamsii were derived members of this clade. Given the depth in the phylogeny at which these putative introgressions would have occurred, we regard an ancestor of M. pentandra and its relatives in the "Pectiantia clade" (Folk and Freudenstein 2014a; see also "clade C" in Okuyama et al. 2012) to be the most likely source of chloroplast DNA, but support values for placing these additional taxa with the Sanger data were generally poor, suggesting that more plastid sequencing of these genera is needed to validate this hypothesis.

Two particularly notable instances of entire clades shifting position in the plastid tree as compared to the nuclear tree were observed. A monophyletic group comprising all eastern US Heuchera species of sect. Heuchera except H. richardsonii (to which they are sister for nuclear data) are members of the introgressed chloroplast clade A, suggesting that plastid DNA introgressed into a single ancestral eastern species of section Heuchera from one of several sympatric members of section Holochla. The most dramatic topological shift in the chloroplast tree involved a monophyletic group of species from Southern California. These are sister to H. brevistamina in most nuclear analyses, but shift to near the base of the tree in both plastid and mitochondrial analyses, being sister to M. diphylla, a widespread species of Eastern North America. For more complete coverage of notable instances of incongruence together with geographic ranges, see Table 2 and Online Appendix 28 available on Dryad.

Almost all inferred instances of gene flow involved diploid hybrids and diploid parental taxa (chromosome counts in Soltis 1980, 1984, 1988; Ness and Soltis 1989; Segraves et al. 1999; Godsoe et al. 2013; Folk and Freudenstein 2014a; see Tables 2–3). Polyploidy in Heuchera is common only in a few widespread species of the northwestern portion of North America. The only clear case of an allopolyploid hybrid is H. rubescens var. truncata (tetraploid; Table 3) for which the inferred parental taxa are tetraploid (H. rubescens var. rubescens) or have a variable cytotype (H. cylindrica). In another instance, the Elegantes group contains a single derived member that is tetraploid (H. caespitosa); yet all other members of this group are diploid, suggesting polyploidy as the ancestral condition at the time of gene flow.

Another unusual element of this taxonomic group is that many apparent instances of introgression involve species that are not closely related; reconciliation of organellar trees with the nuclear tree requires at least two intergeneric hybridization events, and even discord within the genus does not necessarily involve closely related species (as shown dramatically in the cluster network, Online Appendix 26 available on Dryad). In some cases, the hypothesized introgressed taxa are not even geographically close to their parents (summarized in Tables 2 and 3). Heuchera consists largely of montane taxa; since alpine species are thought to have had historically dynamic geographic ranges (e.g., Trollius europaeus; Espíndola et al. 2012), it may not be surprising that a vicariance process could leave vestiges of gene flow among taxa that are restricted in distribution in the present.

Discord among Plastid and Mitochondrial Phylogenies

While we have shown that the plastid and mitochondrial phylogenies have parallel points of discord with respect to the nuclear phylogeny, the plastid and mitochondrial plastids were not completely congruent, especially for the taxa of clade A in each analysis. This is surprising given that plastids and mitochondria are expected to be coinherited in most angiosperms and that there is evidence for maternal inheritance of both organelles in the Saxifragaceae. Support values were high for backbone clades and both organellar topologies were robust to two partitioning schemes. Moreover, misrooting of mitochondrial clade A does not explain the topology; when this subtree is rerooted with M. pentandra to match the plastid analysis, a high amount of conflict remains (Online Appendix 29 available on Dryad). One potential explanation is that the mitochondrial genome has undergone extensive horizontal transfer from other genomes, as is typical for land plants. Chloroplast content is fairly conserved in most angiosperms, to the extent that the BLAST search should have been sufficient to rule out confounding chloroplast sequence, but a full reference genome is unavailable to provide the same guarantee for the nucleus. It is possible that such foreign DNA from the
<table>
<thead>
<tr>
<th>Putative introgressed group</th>
<th>Species tree relationship</th>
<th>Introgressed DNA source: data set (support)</th>
<th>Geographic range/hybrid ploidal level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elegantes group (H. elegans, H. caespitosa, H. hirsutissima, H. abramsii, H. parishii), clade of five species</td>
<td>Section Rhodoheuchera clade</td>
<td>Mitella diphylla: plastid (BS 100), mitochondrion (BS 100)</td>
<td>Allopatric/diploid</td>
</tr>
<tr>
<td>Holochloa group* (H. chlorantha, H. cylindrica, H. micrantha, H. pilosissima, H. parviflora, H. puberula, H. missouriensis, H. villosa, H. glabra), clade of ten species</td>
<td>Sister to all other species of Heuchera</td>
<td>Mitella pentandra clade (&quot;Peziantia&quot;): plastid (BS 100)</td>
<td>Some species sympatric, others allopatric/diploid</td>
</tr>
<tr>
<td>Eastern species of section Heuchera (H. americana, H. alba, H. pubescens, H. longiflora, H. caroliniana), clade of five species</td>
<td>Sister to H. richardsonii</td>
<td>Subsection Villoae (H. parviflora, H. villosa, H. missouriensis, H. puberula): plastid (BS 100); evidence for multiple transfers</td>
<td>Sympatric/diploid</td>
</tr>
<tr>
<td>H. brevistaminea</td>
<td>Sister to Elegantes group (ASTRAL-II, SVDquartets, RAxML unresolved)</td>
<td>Subsection Homalodes (H. acutifolia, H. longipetala, H. mexicana): mitochondrion (BS 100) or only H. acutifolia + H. longipetala: plastid (BS 99)</td>
<td>Allopatric/diploid</td>
</tr>
<tr>
<td>H. parviflora var. utahensis*</td>
<td>Section Heuchera clade</td>
<td>H. micrantha: plastid (BS 100), mitochondrion (BS 100)</td>
<td>Allopatric/diploid</td>
</tr>
<tr>
<td>H. grossulariifolia</td>
<td>Sister to H. merriamii</td>
<td>H. cylindrica: plastid (BS 86/69 for two samples of H. grossulariifolia)</td>
<td>Sympatric/diploid and polyploid</td>
</tr>
<tr>
<td>H. richardsonii</td>
<td>Section Heuchera clade</td>
<td>H. bracteata: plastid (BS 93)</td>
<td>Allopatric/diploid</td>
</tr>
<tr>
<td>Mitella stauropetala clade (&quot;Ozomelis&quot;)</td>
<td>Sister to rest of Heuchera group of genera (Folk and Freudenstein 2014a)</td>
<td>H. micrantha: plastid (BS 92)</td>
<td>Sympatric/diploid</td>
</tr>
<tr>
<td>H. bracteata, H. hallii,</td>
<td>Sister to section Rhodoheuchera</td>
<td>Unresolved member of H. parviflora complex: plastid (BS 99), mitochondrion (BS 91)</td>
<td>Sympatric/diploid</td>
</tr>
<tr>
<td>H. woodiaphila</td>
<td>Sister to H. hallii + H. bracteata + section Rhodoheuchera</td>
<td>Unresolved member of H. parviflora complex: plastid (BS 99), mitochondrion (BS 91)</td>
<td>Sympatric/diploid</td>
</tr>
<tr>
<td>North Mexican clade (H. sanguinea, H. roundiflora, H. wellsi), clade of three species</td>
<td>Section Rhodoheuchera clade</td>
<td>Unresolved member of H. northmexicana-H. soltisi-H. eastwoodiae complex (BS 100), mitochondrion (BS 90)</td>
<td>Allopatric/unknown</td>
</tr>
<tr>
<td>H. wootonii</td>
<td>Section Heuchera clade</td>
<td>Unresolved member of section Holochloa clade: plastid (BS 100), mitochondrion (BS 100)</td>
<td>Allopatric/diploid</td>
</tr>
</tbody>
</table>

Notes: Bootstrap numbers ("BS," partitioned analyses) for introgressed DNA indicate topological support for anomalous placement in the plastid and/or mitochondrial phylogeny.

*Also observed in (Soltis et al. 1991b).

nucleus could be responsible for greatly differing branch lengths subtending taxa in the mitochondrial analysis, especially those with greatly different positions among organellar analyses, such as H. missouriensis, H. bracteata, and H. richardsonii. Inspecting this possibility was not possible in individual genes given low substitution rates and lack of support (data not shown), but we were able to obtain four larger well-aligned regions (9800–26,594 bp; Online Appendix 30 available on Dryad). These topologies conflicted with the main mitochondrial analysis, indicating a possible role for recombination of loci with different histories, but the topologies were also incongruent with plastid and nuclear data, and high topological support was still difficult to achieve. Another potential explanation is that ancestral gene flow events were accompanied by low levels of paternal leakage, likely a typical phenomenon in angiosperms (Azhagiri and Maliga 2007), allowing for differential...
Table 3. Proposed hybrids (with Linnaean names if any exist), with a summary of evidence for their hybrid status;

<table>
<thead>
<tr>
<th>Putative hybrid</th>
<th>Proposed parent one: data set (support)</th>
<th>Proposed parent two: data set (support)</th>
<th>Geographic range/hybrid ploidal level</th>
</tr>
</thead>
<tbody>
<tr>
<td>H. rubescens x H. clyndrica (= H. rubescens var. truncata)*</td>
<td>H. rubescens (nuclear species tree, BS 93; similar overall morphology)</td>
<td>H. clyndrica (plastid, BS 95; short style and truncate leaf base)</td>
<td>Parapatric/polyploid</td>
</tr>
<tr>
<td>H. americana x H. pubescens (= H. americana var. hispida)*</td>
<td>H. pubescens (18.7 nuclear gene trees; anthocyanic rhomboid petals and larger flower size)</td>
<td>H. americana (10.5 nuclear gene trees; similar overall floral morphology)</td>
<td>Parapatric/diploid</td>
</tr>
<tr>
<td>H. americana x H. richardsonii (= H. americana var. hirsuticaulis)*</td>
<td>H. richardsonii (20.7 nuclear genes; hirsute petioles)</td>
<td>H. americana (10.9 nuclear genes; floral morphology)</td>
<td>Parapatric/diploid</td>
</tr>
<tr>
<td>H. micrantha x H. glabra*</td>
<td>H. glabra (65.7 genes; glabrous plant body)</td>
<td>H. americana var. diversifolia (9.2 nuclear genes; plastid, BS 72; pyramidal many-flowered inflorescence, elongate terminal leaf lobe) or var. macropetala (9.4 nuclear genes; mitochondrial, BS 30)</td>
<td>Sympatric/diploid</td>
</tr>
<tr>
<td>&quot;H. eastwoodiae 2&quot; (H. eastwoodiae x H. glomerulata)</td>
<td>H. glomerulata (21.5 genes; 5-merous petalous flowers)</td>
<td>H. eastwoodiae (18.6 nuclear genes; overall floral morphology)</td>
<td>Parapatric/unknown</td>
</tr>
</tbody>
</table>

Notes: The number of nuclear genes that support a relationship is derived from BUCKy concordance factors (CF; n = CF * total number of genes).
* Hypothesized by Rosendahl et al. (1936).
* Hypothesized by Wells and Shipes (2009).

fixation of parental genomes and, therefore, differing organellar histories. Finally, the exclusion of lower coverage taxa (Online Appendix 25 available on Dryad) decreased the level of discord with the nucleus, also perhaps indicating a role of gene tree error in creating some of this conflict despite high support for their placement in the analysis with more taxa. Despite these concerns, which reflect the general difficulty of using mitochondrial loci in plants, we present the mitochondrial analysis in the spirit of highlighting the use of all three genomic compartments in plant phylogenetics, while discussing primarily the plastid phylogeny and noting mitochondrial relationships where they are congruent.

Nuclear Gene Trees and Hybrid Taxa

The average δ score for the nuclear data set (0.1341) indicated relatively low levels of conflict among nuclear loci, creating a sharp contrast with organellar markers. Nevertheless, the earlier morphological literature provides several hypotheses of hybrid origin for taxa that we were able to test; these are distinguished from possible cases of introgression because they also involve morphological intermediacy. Several of these have been given Linnaean names: H. rubescens x H. clyndrica = H. rubescens var. truncata or H. cuneata (Rosendahl et al. 1936); H. americana x H. pubescens = H. americana var. hispida (Wells 1984); H. americana x H. richardsonii = H. americana var. hirsuticaulis (Wells 1984). Furthermore, we had two putative cases, H. micrantha x H. glabra (Wells and Shipes 2009) and H. eastwoodiae x H. glomerulata (“H. eastwoodiae 2”; not previously hypothesized) that have not been formally named but show morphological intermediacy between distinctive parental taxa. In four of these cases, there was support for hybridization from discordance factors for the nuclear genome (Table 3). For the overall phylogeny, discordance factors were never high. We attribute this to high gene tree uncertainty; discordance factors should be consistently underestimated for gene trees with many taxa and short branches (given a constant gene size) since under these conditions gene trees should have greater topological uncertainty. Since none of the gene trees was perfectly resolved with high support (though the degree of resolution varied and some gene trees were well resolved), this would result in downweighting discordance factors when there is equivocality rather than evidence contradicting particular species relationships. In general, the four cases generally showed more gene trees supporting one putative parent than the other; this could indicate asymmetrical hybridization.

We additionally note that the rogue taxa we identified could potentially be hybrids since hybridization would result in alleles with mixed histories and, therefore, high character conflict. However, all four rogues were in the Rhodoheuchera clade or near it (H. woodsiiaphila). This group was responsible for much of the conflict and low support present in this analysis, suggesting a cautious interpretation of conflicting placements.

Coalescent Analysis

ASTRAL-II and SVDquartets, despite taking completely different approaches to analyzing the data, had broadly similar results and only one major point of incongruence involving outgroup taxa. We
view the ASTRAL-II topology as more likely to be accurate since ASTRAL-II is designed to work with multilocus data and the ASTRAL-II tree was more similar to concatenation estimates with higher support values overall (similar observations in Mirarab et al. 2014b). Furthermore, upon rerunning ASTRAL-II on only those loci that showed no sign of recombination under the ω tests (methods otherwise identical to the main analysis, n = 161; Online Appendix 31 available on Dryad), only a few tip relationships differed. None of these rearrangements had BS greater than 50, a result as likely due to a smaller data set as due to a weak signal of recombination. This suggests that our coalescent analyses may not be sensitive to intralocus recombination, despite its definite presence in the data set, and may be an empirical corroboration of the simulation-based assertion of (Lanier and Knowles 2012) that coalescent methods may be relatively robust to the presence of intralocus recombination, contra (Gatesy and Springer 2014).

**Concatenation Analysis**

There has been recent debate on the empirical importance of ILS at various taxonomic levels (Rokas et al. 2003; Pollard et al. 2006; Rosenfeld et al. 2012; Gatesy and Springer 2013; Wu et al. 2013; Gatesy and Springer 2014; Lanier and Knowles 2015). Despite some differences, the topologies of the concatenation analyses were quite similar to those in the coalescent analyses. Strong similarity in results, despite the great difference in underlying philosophy between these two techniques, has been observed before (Stephens et al. 2015). This is consistent with theoretical results and simulation studies that suggest concatenation can be an accurate inferential tool when ILS occurs with only modest frequency (Bayzid and Warnow 2013; see these issues reviewed in Roch and Warnow 2015).

While congruence between concatenation and coalescence is encouraging, an important issue arose in the placement of hybrid OTUs in concatenation. In agreement with the previous literature (McDade 1990, 1997, especially McDade 1992), our suspected hybrid taxa were typically placed with relatively high support as sister to the most apomorphic putative parent (i.e., the parent with the longest branch length measured in per-site substitutions). For instance, among species in section *Heuchera* (Fig. 1, blue), *H. richardsonii* is sister to all other eastern US species and has the longest branch length; the putative hybrid *H. richardsonii × H. americana* always was sister to both samples of *H. richardsonii* with very high support. While concordance factors showed the hybrid had an affinity with *H. richardsonii* for more loci (Table 3), there was also signal in the gene trees suggesting a relationship with *H. americana*. This is likewise true for the putative hybrid *H. micrantha × H. glabra*, for which the *H. glabra* parent is on a particularly long branch. Concatenation analyses tend to hide conflict in the data set when phylogenetic signal is strong, necessitating the use of other methods to adequately search for this signal. We found that a similar effect occurred with the ASTRAL-II analysis with individuals treated as species (Online Appendix 6 available on Dryad).

**Conclusions**

The *Heuchera* system has been cited frequently as a dramatic system of chloroplast capture; here, phylogenomic evidence has provided a rigorous, probabilistic assessment of the existence of promiscuous organellar gene flow in this lineage. There appear to have been two intergeneric hybridization events in the history of *Heuchera* involving the *M. diphylla* group and the *Pectiantia* group; moreover, we have argued for the ancient introgression of DNA ancestral into several *Heuchera* clades (Table 2). Only 17 out of 43 species were placed solely in plastid clade B, hypothesized to be representative of the ancestral state of the plastid genome, suggesting that in this lineage most taxa have a history of hybridization. Traditionally, hybrids have been seen as ephemeral evolutionary experiments, whereas the thrust of evolution occurred along nonhybrid lines (Wagner 1970). Examples of successful hybrid species and especially hybrids that underwent subsequent cladogenesis call these notions into question, providing living examples of diversification after gene flow. Beyond the identification of ancestral hybridization, we hope this and similar cases will invite future contributions from the perspective of comparative methods, investigating not merely the existence of hybrids but factors that may be associated with their persistence and diversification.

**ACKNOWLEDGMENTS**

We would like to acknowledge the Ohio Supercomputer Center and Ohio Biodiversity Conservation Partnership Cluster for resources to perform analyses, L. Kubatko for discussions on SVQquartets and simulations under the coalescent, and A. Ramsey for facilitating R.A.F. as a visiting scholar; the Soltis Lab for the provision of *Elmira* and *Lithophragma* materials used with supplementary analyses, K. Robis for assistance with assembling Supplementary Materials available on Dryad, B. Stucky for discussions on partitioning schemes, The W. Harry Feinstone Center for Genomic Research (University of Memphis) and the Molecular and Cellular Imaging Center (The Ohio State University) are thanked for sequencing support. D. Soltis and K. Guralnick are thanked for reading the manuscript.

**FUNDING**

This work was supported by the American Society of Plant Taxonomists and the National Science Foundation [DEB 1406721].


Mitella


