

Resolving Phylogenetic Relationships of a Recently and Rapidly Evolving Lineage from Western North America (*Mentzelia* Section *Bartonia*, Loasaceae)

Paige P. Fabre,^{1,5,6} Joshua M. Brokaw,² Larry D. Hufford,³ Matthew G. Johnson,⁴ and John J. Schenk¹

¹Department of Environmental and Plant Biology, Ohio University, 22 Richland Avenue, 401 Porter Hall, Athens, Ohio 45701, USA; pf271621@ohio.edu; schenk@ohio.edu

²Department of Biology, Abilene Christian University, Abilene, Texas 79699, USA; josh.brokaw@acu.edu

³Department of Biological Sciences, Washington State University, Pullman, Washington 99164, USA; hufford@wsu.edu

⁴Department of Biological Sciences, Texas Tech University, Box 43131, Lubbock, Texas 79409, USA; matt.johnson@ttu.edu

⁵Department of Environmental and Plant Biology, Ohio University, 22 Richland Avenue, 401 Porter Hall, Athens, Ohio 45701, USA

⁶Author for correspondence (pf271621@ohio.edu)

Communicating Editor: James I. Cohen

Abstract—The landscape of western North America has dramatically transformed since the Miocene to become increasingly heterogeneous, in turn promoting the evolution of many rapidly radiating angiosperm lineages. Phylogenetic relationships of these recently and rapidly radiating groups are difficult to resolve as there is little genetic variation among species and a high degree of noise from incomplete lineage sorting and hybridization. *Mentzelia* section *Bartonia* (51 species; Loasaceae) exemplifies this problem well. The clade has been investigated with Sanger sequencing, RADSeq, and genome skimming methods, however, most species relationships remain elusive due to low genetic variability. To better infer species relationships, we applied a hybrid enrichment approach with the Angiosperms353 probe set and implemented a novel bioinformatics workflow that aimed to maximize phylogenetic signal and minimize noise from low-quality sequences, paralogy, and incomplete lineage sorting. Our phylogenomic approach increased phylogenetic resolution of species relationships compared to previous studies based on nrDNA loci. Although a few species relationships still lack strong support, our results indicate that our methods were effective in phylogenetic inference of this recently and rapidly evolving lineage from western North America. To better characterize major groups in the Section, we propose the formal designation of three subsections: *Decapetala*, *Multicaulis*, and *Multiflora*.

Keywords—Angiosperms353, bioinformatics pipeline, high throughput sequencing, HybPiper, phylogenomics, rapid radiation.

INTRODUCTION

The heterogeneous landscape of western North America has been shaped by dynamic geological and climatic processes, including surface uplift (Foster et al. 2010), glaciation cycles (Thorne 1993), erosion (Cather et al. 2012), aridification (Stebbins 1952; Axelrod 1972), and the formation of biologically restrictive soils (e.g., serpentine, gypsum, limestone, deep sand; Kruckeberg 2004; Moore et al. 2014), all of which have been hypothesized to promote diversification of flowering plants within this region. As the western North American landscape became increasingly heterogeneous with isolated habitats during the Miocene (23.03–5.33 Ma), newly-opened areas of niche space promoted the rapid diversification of endemic genera (Qian et al. 2024), particularly in arid and semi-arid regions (Axelrod 1958). The Great Basin, for example, serves as a center of diversity for *Eriogonum* Michx. (Polygonaceae; ca. 250 spp.) and *Penstemon* Schmidel (Plantaginaceae; ca. 270 spp.), the two largest angiosperm genera endemic to North America (Reveal 1978; Wolfe et al. 2006), both of which underwent rapid diversification beginning 6 Ma and 2.5 Ma, respectively (Kostikova et al. 2014; Stone and Wolfe 2021; Wolfe et al. 2021).

Indeed, many of the rapid species radiations in western North America have occurred relatively recently (i.e., within the last ~10 Ma), including some of the region's most recognizable and charismatic genera: *Arctostaphylos* Adans. (Ericaceae; Wahlert et al. 2009), *Atriplex* L. (Amaranthaceae; Brignone et al. 2019), *Castilleja* Mutis ex L.f. (Orobanchaceae; Tank and Olmstead 2008), *Ceanothus* L. (Rhamnaceae; Burge et al. 2011), *Eriogonum* (Polygonaceae; Kostikova et al. 2014), *Lomatium* Raf. (Apiaceae; George et al. 2014), *Lupinus* L. (Fabaceae; Drummond 2008), *Mentzelia* L. (Loasaceae; Schenk and Hufford 2010), *Oenothera* L. (Onagraceae; Patsis et al. 2021;

Cooper et al. 2023), *Opuntia* (L.) Mill. (Cactaceae; Majure et al. 2012), *Penstemon* (Plantaginaceae; Stone and Wolfe 2021; Wolfe et al. 2021), *Phacelia* Juss. (Gilbert et al. 2005), *Polemonium* L. (Worley et al. 2009), and *Yucca* L. (Asparagaceae; Pellmyr et al. 2007). While these lineages are compelling, their biology introduces major challenges to taxonomy and phylogenetic reconstruction, including cryptic speciation, low morphological variability, limited genetic variation, conflicting phylogenetic signal (cytonuclear discordance, homoplasy, ancient and recent hybridization, introgression), and high rates of incomplete lineage sorting (ILS). *Mentzelia* section *Bartonia* (Torrey & A. Gray) Benth. & Hooker f. is no exception to these taxonomic and phylogenetic challenges, however, a recent monograph by Schenk and Hufford (2020) has provided a deeper understanding of species diversity compared with other groups, making it more amenable to phylogenomic studies of taxonomically well-understood species.

Mentzelia section *Bartonia* is a clade of 51 extant species (Schenk and Hufford 2020) that diversified relatively recently (4.9–9.2 Ma crown age; Schenk et al. 2025) out of the Colorado Plateau and expanded throughout the western United States and northern Mexico (Schenk 2013a). The plants are herbaceous biennials or short-lived perennials with actinomorphic flowers and non-stinging trichomes used in herbivory defense (Eisner et al. 1998; Schenk and Hufford 2020). Much of the clade's species diversification was driven by movement onto nutrient-poor and disturbed soils, such as talus slopes or gypsum and limestone outcrops (Schenk 2013a; Schenk and Hufford 2020). Shifts in aneuploidy also commonly coincided with diversification (Schenk and Hufford 2020). Species are diploid with a base chromosome count of either $n = 10$ or $n = 11$ (Schenk and Hufford 2020), however, there is contention over whether *M. leucophylla* Brandege is a polyploid ($n = 18$; Reveal

and Styer 1973; Schenk and Hufford 2020). Most extant species are narrowly distributed endemics confined to a single floristic province (Schenk 2013a), with 31 out of 51 species currently listed as either threatened or endangered by state or federal agencies (Schenk and Hufford 2020). Like many other recently and rapidly radiating groups in western North America, *Mentzelia* section *Bartonia* is in considerable need of a robust phylogeny to inform future conservation efforts. Much like those other groups, however, taxonomic and phylogenetic work in *Mentzelia* section *Bartonia* has been challenging.

The first molecular phylogenetic study to include section *Bartonia* was Hufford et al. (2003) who sampled 12 out of 51 species as part of a larger study of Loasaceae. Using two chloroplast markers (*matK* and *trnL-trnF*), the authors found little resolution among species, with 11 out of 12 species belonging to a single large polytomy, demonstrating that a limited number of chloroplast markers are insufficient for phylogenetic inference of the section. Subsequently, Schenk and Hufford (2011) reconstructed the first complete molecular phylogeny of section *Bartonia* using two nuclear ribosomal (nrDNA) regions (ITS and ETS), which was later updated to include additional sampling (Schenk and Hufford 2020). While the Schenk and Hufford (2011) phylogeny offered significant insights into species relationships and informed ancestral character estimations of reproductive traits (Schenk 2013b; Botnaru and Schenk 2019) and biogeography (Schenk 2013a), the tree had 45% unresolved relationships with 62% of all relationships being weakly supported (Bayesian posterior probabilities < 0.95).

High throughput sequencing methods have not necessarily been more effective at equivocally resolving relationships. Moore et al. (2023) implemented genome skimming of plastid genomes and the nuclear ribosomal cistron region for 20 out of 51 species and found low phylogenetic signal and high levels of homoplasy. One striking result of the study was that out of ~159,000 bp recovered for the chloroplast genome, only 76 sites were phylogenetically informative, illuminating how little genetic variation exists amongst species. To address these limitations, Cohen and Schenk (2022) used ddRADseq to conduct a population-level study of five species endemic to the Mojave Desert. While their approach obtained resolution among tips, their study also revealed evidence of putative hybridization and ILS.

While high-throughput sequencing has enhanced our ability to detect genetic variation, the abundance of genetic data introduces further complications (Moore et al. 2023). Incomplete lineage sorting and paralogy pose substantial challenges for phylogenetic inference of recently and rapidly diversifying lineages. The likelihood of ILS, which is characterized by the persistence of ancestral polymorphisms after speciation events (Maddison 1997), increases with younger lineages, resulting in incongruent gene tree topologies and diminished support values in species trees. Gene duplications in large genomic data sets also obscure homologous relationships that are essential for accurate phylogenetic inference. Paralogous genes are genes that evolve from duplication events within a lineage and do not coincide with duplication events during speciation. In some situations, a random paralogous gene may be lost, resulting in a single-copy gene that can be mistaken for an ortholog. Leaving these “pseudoorthologs” in a data set can lead to an overestimation of branch lengths and discordant gene tree topologies (Smith and Hahn 2022). In addition, recently and rapidly evolved lineages have not

had time for species to accrue synapomorphies before speciation events occurred, hindering our ability to detect phylogenetic signal (Moore et al. 2023). Therefore, we must implement methods that maximize our ability to detect phylogenetic signal while minimizing common issues that confound relationships, such as paralogy and ILS.

Based on the results of Cohen and Schenk (2022), a robust phylogeny of *Mentzelia* section *Bartonia* subclades using RADseq is plausible, however, we chose to implement a HybSeq approach for four reasons. First, RADSeq and HybSeq methods can find congruent phylogenies and divergence times (e.g., Zhou and Xiang 2022). Second, HybSeq can outperform RADSeq when DNA quality and concentration are low (Zhou and Xiang 2022), which is often true for genomic extractants of *Mentzelia* section *Bartonia* because of secondary metabolite contamination (Carey et al. 2023). Third, RADSeq approaches often fail to recover orthologous SNPs among distantly related species (Leaché and Oaks 2017), and we aim for the compatibility of our data with future studies of older lineages, including *Mentzelia* (70–35.3 Ma; Schenk and Hufford 2010), Loasaceae (91.60–57.96 Ma; Schenk and Hufford 2010), and tree of life projects (e.g., Baker et al. 2022). Finally, the recent development of the Angiosperms353 probe set (Johnson et al. 2019) provides a relatively easy and cost-effective HybSeq method that can be as equally phylogenetically informative as lineage-specific probes (Ufimov et al. 2021). Angiosperms353 probes have been successfully implemented in recent and rapidly radiating lineages outside of western North America, including *Burmeisteria* H.Karst. & Triana (Campanulaceae; Bagley et al. 2020), *Cyperus* L. (Cyperaceae; Larridon et al. 2020), *Tillandsia* L. (Bromeliaceae; Yardeni et al. 2022), and *Veronica* L. (Plantaginaceae; Thomas et al. 2021). Taken together, HybSeq with Angiosperms353 is a promising method for resolving species relationships in *Mentzelia* section *Bartonia*.

The goal of the present study is to elucidate species relationships in *Mentzelia* section *Bartonia*, a recently and rapidly evolving angiosperm clade from western North America. To achieve our goal, we applied a HybSeq approach with Angiosperms353 using a novel bioinformatics pipeline that aimed to minimize noise in our data while maximizing phylogenetic signal.

MATERIALS AND METHODS

To robustly resolve phylogenetic relationships among species of *Mentzelia* section *Bartonia*, we employed methods that address the myriad challenges of recent and rapidly radiating lineages, including having a low phylogenetic signal-to-noise ratio, a high chance of ILS, and the potential for hybridization.

Sample Preparation and Sequencing—We sampled 50 of 51 species in *Mentzelia* section *Bartonia*, including five infraspecific taxa, and four of six species from its sister clade, section *Bicuspidaria* S.Watson (Hufford et al. 2003; Brokaw et al. 2020). Most ingroup taxa were represented by a single individual; however, eleven taxa were represented by up to three individuals to reflect both phenotypic and geographic variation. We prioritized sampling the same vouchers as Schenk and Hufford (2011, 2020) to make direct comparisons of our results to theirs. The complete list of 76 total vouchers is detailed in Appendix 1.

Genomic DNA was obtained from previously published work (Schenk and Hufford 2011, 2020; Brokaw et al. 2020; Moore et al. 2023) and by fresh extraction of 5–10 mg of silica-dried leaves or those preserved as herbarium specimens using a modified CTAB protocol (Schenk et al. 2023, protocol v4). Genomic DNA was quantified with gel electrophoresis and a Qubit fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Per sample, ~500 ng of genomic DNA was used as input for genomic libraries that were constructed with the NEBNext Ultra II FS DNA Prep Kit for

Illumina (New England BioLabs, Ipswich, Massachusetts, USA), which also enzymatically fragmented DNA to 200–450 bp fragments, following manufacturer instructions for inputs ≥ 100 ng. Size selection was performed for a targeted library size of ~ 300 bp. Fragment analyses of the prepared libraries were performed with a TapeStation (Agilent Technologies, Santa Clara, CA, USA). Libraries were indexed with NEBNext Multiplex Oligos for Illumina (New England BioLabs, Ipswich, Massachusetts, USA) and either enriched with Angiosperms353 v5 probe set (myBaits, Daicel Arbor Biosciences, Ann Arbor, MI, USA; Johnson et al. 2019) or left unenriched to increase sequence complexity during the sequencing reaction. Hybrid enrichment was performed following the manufacturer's standard protocol for plants with the following modifications: hybridization temperature was 62°C, library amplification ran for 14 PCR cycles, and wash temperature was 65°C. Enriched and unenriched libraries were pooled to a final concentration of 2.25 nM in 220 μ l in accordance with sequencing facility recommendations and sequenced together at a 1:10 ratio of unenriched to enriched libraries.

We performed three separate runs of Illumina sequencing: Illumina NovaSeq 6000 (San Diego, California, USA) on a single lane with a target length of 250-bp paired-end reads (Nationwide Children's Hospital, Columbus, OH, USA), Illumina MiSeq on two lanes with a target length of 300-bp paired-end reads (Psmagen Inc., Rockville, MD, USA), and Illumina NextSeq 2000 on two lanes with a target length of 250-bp paired-end reads (Nationwide Children's Hospital, Columbus, OH, USA). Both enriched and unenriched libraries were simultaneously sequenced. MiSeq and NextSeq runs were initially performed as exploratory methods for a small proportion of samples before we ultimately chose NovaSeq for the remainder of our sequencing. Identical vouchers that were sequenced on more than one platform had their raw reads bioinformatically combined prior to locus assembly.

Phylogenetic Inference—The complete bioinformatics pipeline is illustrated in Fig. 1. Select data and command-line code is available on Dryad (Fabre et al. 2025).

Quality control of raw, paired-end reads was performed with fastp v0.23.4 (Chen et al. 2018) using the following adjustments to default parameters: minimum qualifying phred score per base ≥ 20 , minimum read

length = 50, maximum number of unqualifying bases per read = 10, and removal of PCR duplicates (fastp code: -q 20 -n 10 -l 50 -D). We used fastp over Trimmomatic (Bolger et al. 2014) to remove homodimers from the ends of Illumina NextSeq and Illumina NovaSeq reads (<https://github.com/mossmatters/HybPiper/wiki>). The post-fastp reads are hereafter called “clean” reads.

HybPiper v2.1.6 (Johnson et al. 2016) was used with BWA (Li and Durbin 2009) to map clean, paired-end reads to a customized nucleotide target file and assemble supercontigs (i.e., exons with flanking introns). Our target file contained Cornales sequences filtered from Mega353 v1 (McLay et al. 2021) in addition to published exons from *Mentzelia decapetala* (Pursh) Urb. and *Mentzelia involucrata* S.Watson (<https://treeoflife.kew.org>), which increased our recovery rates in preliminary analyses. We checked the target file for low-complexity regions using the “check_targetfile” function in HybPiper with default settings and removed the offending sequences before locus assembly. After locus assembly, we used the “stats” command to summarize gene recovery statistics, and the “paralog_retriever” command to summarize putative paralog recovery and retrieve paralogous sequences.

Although the Angiosperms353 probes were designed to capture single-copy nuclear loci (Johnson et al. 2019), they can still recover paralogs, which HybPiper (Johnson et al. 2016) (the most popular pipeline for assembly of Angiosperms353 loci) can fail to detect (Nauheimer et al. 2021; Zhou et al. 2022). In an exploratory analysis, we reviewed the paralogs recovered by HybPiper, selected a locus with only one specimen flagged as paralogous, and assembled a new target file that contained only the paralogous sequences from that specimen. When we re-assembled the locus based on the paralog-only target file, we discovered additional paralogous copies that were not initially reported. Based on this finding, we decided against removing individual paralogous copies and instead conservatively removed entire loci from downstream analyses if they were flagged with one or more paralogous sequences.

After supercontig assembly with HybPiper, we ran HybPhaser v2.1 (Nauheimer et al. 2021) to evaluate genomic samples for additional hidden-paralogy, hybridization, and contamination. In short, HybPhaser uses the supercontigs generated by HybPiper as a reference to re-map clean reads, discover SNPs, and remove low-quality or outlying individuals and loci.

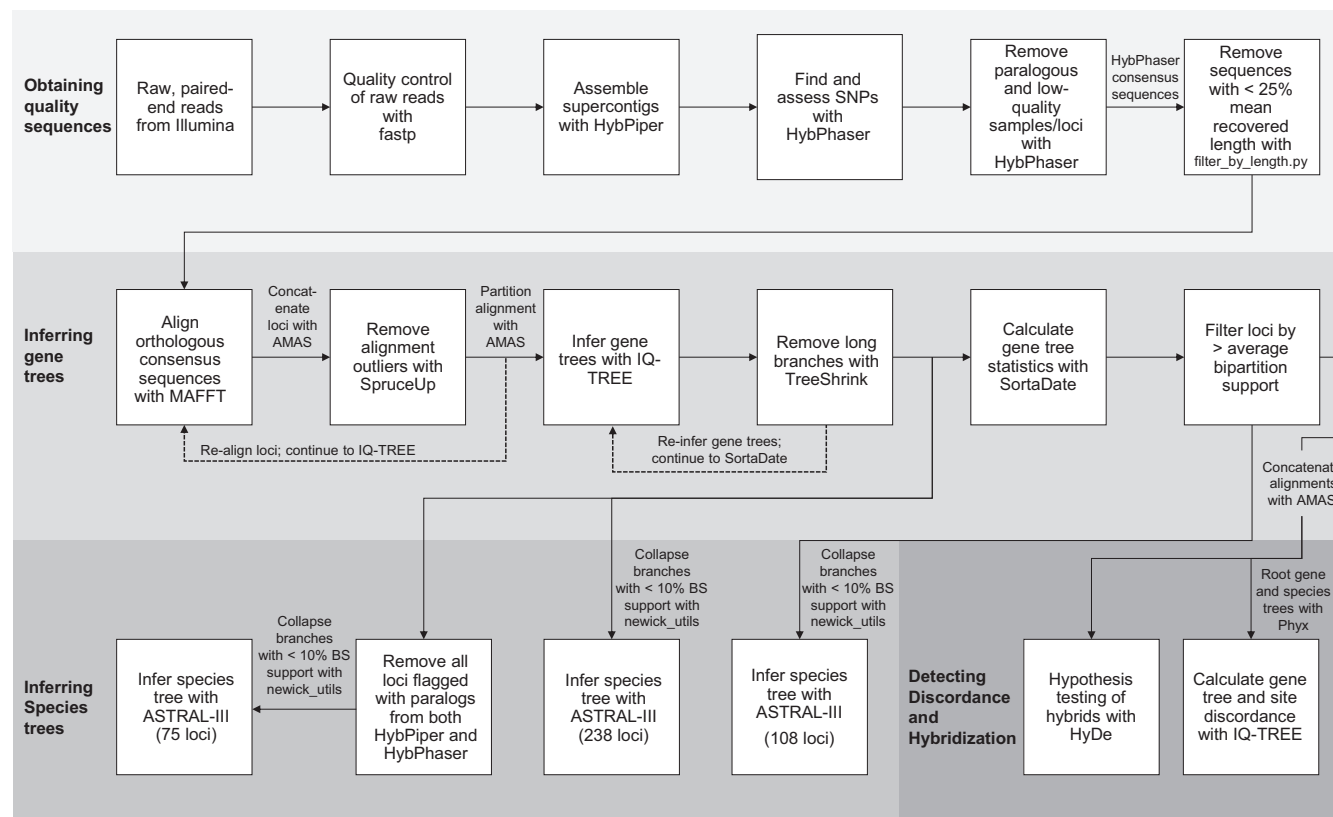


FIG. 1. Bioinformatics pipeline that was implemented in our study to minimize noise while maximizing phylogenetic signal. To maximize the low signal-to-noise ratio present in recently and rapidly evolving lineages, we implemented a rigorous bioinformatics pipeline to address paralogy, noise from low-quality sequences, species, and loci, incomplete lineage sorting, and hybridization. Bootstrap support is abbreviated as BS and single nucleotide polymorphisms are abbreviated as SNPs.

HybPhaser flags both samples and loci as statistical outliers if they have a relatively high proportion of SNPs (i.e., more than $1.5 \times \text{IQR}$ [interquartile range] above the 3rd quartile of mean), indicating putative paralogy. Loci and samples flagged as statistical outliers/paralogs were removed from downstream analyses. Loci that contained $\leq 15\%$ of samples, and samples that were found in $\leq 15\%$ loci, were also removed. When completed, HybPhaser generated consensus sequences with ambiguity codes for heterozygous sites, which reduces conflicting phylogenetic signal from hybrids (Nauheimer et al. 2021).

To reduce the deleterious effects of fragmentary data on downstream analyses of gene and species trees (Sayyari et al. 2017; Mirarab 2019), we ran the Python script `filter_by_length.py` (<https://github.com/mossmatters/phyloscripts/tree/master/HybPiperUtils>) on the HybPhaser “consensus_length” file to flag samples with sequences shorter than 25% of the mean length recovered for a particular locus. To do this, we modified the “consensus_length” file to match the format of the HybPiper “seq_lengths” file and calculated the mean length of recovered sequences (i.e., sequences with recovered length > 0) for each locus. We used the mean recovered length rather than the default mean targeted length to ensure that loci with low recovery across all taxa were still retained. Samples that were flagged as being too short per locus were then removed from the data set.

Putative orthologous consensus sequences were aligned with MAFFT v7.520 (Katoh and Standley 2013) using 1000 maximum iterations. Resulting gene alignments were concatenated with AMAS v1.0 (Borowiec 2016) and analyzed with spruceup v2022.2.4 (Borowiec 2019), utilizing default settings to identify and remove outlier sequences (i.e., individual poorly aligned sequences or sequence fragments that were outside of the 95% confidence interval) using a preliminary ASTRAL (Zhang et al. 2018) tree as a guide. The preliminary tree was comprised of 247 putative orthologous supercontigs from HybPiper and inferred under identical conditions as described below. Following spruceup, the concatenated alignment was partitioned with AMAS (Borowiec 2016) utilizing the “remove-empty” command to remove samples that were characterized by only gaps or ambiguous nucleotides. To reduce regions with excessive gaps, gene alignments were re-aligned with MAFFT under identical conditions outlined above.

After the above conditions were met, individual-locus alignments were used to infer maximum-likelihood gene trees with 1000 ultrafast bootstrap approximations in IQ-TREE v2.2.2.3 (Nguyen et al. 2015; Minh et al. 2022) using the ModelFinder command (Kalyaanamoorthy et al. 2017) to determine the best-fit substitution-model for each locus. Unusually long branches were detected with TreeShrink (Mai and Mirarab 2018) and offending tips were removed from gene trees and corresponding alignments for the offending locus. To re-estimate bootstrap support after TreeShrink, gene trees were then re-estimated with IQ-TREE under identical conditions as above. Gene tree nodes with $< 10\%$ bootstrap support were collapsed with newick_utils v1.6 (Junier and Zdobnov 2010; Mirarab 2019), and resulting gene trees were used to infer a multispecies coalescent tree with ASTRAL-III v5.7.8 (Minh et al. 2022) utilizing default parameters. We did not implement a concatenation-based approach because it underperforms in the face of ILS and is therefore an inferior method for groups that have undergone recent and rapid speciation (Mirarab 2019). The final ASTRAL tree was visualized in FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/>).

Discordance and Hybridization—To evaluate the consistency among gene trees, we rooted the gene trees using `pxrr` in `PhyX` (Brown et al. 2017) and used `SortaDate` (Smith et al. 2018) to calculate tip-to-root variation, tree length, and bipartition support. Tip-to-root variation indicates clock-likeness, tree length is the total sum of branch lengths, and bipartition support measures the similarity between the gene trees and species trees (Smith et al. 2018). Bipartition support has also been used as a metric to remove paralogs that are otherwise only detectable by eye (Frost et al. 2024). Based on this, we subset our data to include only loci with higher than average bootstrap support (Frost et al. 2024; Fig. 1) to remove additional putative paralogs. Simultaneously, we also generated a data subset to include only loci that were not flagged as paralogous by either HybPiper or HybPhaser (Fig. 1). Alignment statistics, including alignment length, percent missing data, proportion of invariable sites, and proportion of parsimony informative sites, were calculated with AMAS (Borowiec 2016) for all three data sets.

To better understand gene tree discordance in our three data sets, gene concordance factors (gCF) and site concordance factors (sCF) were calculated with IQ-TREE (Minh et al. 2022), with the latter utilizing the ModelFinder function (Kalyaanamoorthy et al. 2017) and maximum likelihood (Mo et al. 2023) with 100 quartets. Gene concordance factors are the

proportion of gene trees that are concordant with the species tree at a given node and signify the proportion of the genome that shares this evolutionary history (Baum 2007). Site concordance factors represent the proportion of alignment sites that are decisive for a particular branch (Minh et al. 2020).

Putative hybrids were detected by HybPhaser if they had a high proportion of loci containing SNPs (indicative of high locus heterozygosity [LH]) in conjunction with a high proportion of SNPs distributed across all loci (indicative of high allele divergence [AD]). We designated taxa as putative hybrids worthy of further exploration if they (i) had a high LH/AD ratio based on HybPhaser analyses, (ii) were recovered as non-monophyletic in one or more of our ASTRAL trees, or (iii) had discordant relationships between our ASTRAL trees. Using only the data subset containing loci with above-average bipartition support, we then tested explicit hypotheses of hybridization with HyDe v0.4.3 (Blischak et al. 2018; Fig. 1) for the following specimens: *M. lagarosa* (Hufford 4332 [WS]), *M. longiloba* var. *longiloba* (Schenk 1225 [WS]), *M. perennis* (Schenk 2642 [WS]), *M. perennis* (Spellenberg 10586 [NMC]), *M. saxicola* (Schenk 1256 [WS]), and *M. humilis* var. *humilis* (Schenk 1264 [WS]). Parental crosses were chosen based on the most likely specimens (Schenk and Hufford 2020) and specimen availability. Specific tests are detailed in the Supplementary Materials.

RESULTS

We produced genomic sequence data for 77 individuals, comprising 96% of taxa in *Mentzelia* section *Bartonia* and 67% of taxa in section *Bicuspidaria*. In total, we recovered 498 billion clean, paired-end reads, ranging from 23 thousand to 26 million reads per sample (median = 5.4 million; Table S1).

HybPiper recovered sequence data for 351 total loci with an average of 336 loci per specimen (Table S1). HybPiper flagged 104 loci as containing at least one paralog (Table S1). Based on both HybPiper and HybPhaser analyses, no individuals recovered entirely paralogous loci, suggesting there are no allopolyploids in our data set (including *M. leucophylla*, the only hypothesized polyploid). HybPhaser failed to detect SNPs for 20 individuals; however, of the 57 individuals with SNP data, 236 loci were flagged as containing at least one individual with a paralog (Table S1), and 16 loci were flagged as paralogous across all individuals. Based on quality filtering by HybPhaser, two loci were removed from the data set for containing fewer than 15% of samples, and one specimen (*M. canyonensis*) was removed from the data set completely for being recovered in fewer than 15% of loci (Table S1).

Our most inclusive data set contained 238 loci, the data set based on above-average bipartition support contained 108 loci, and our strict ortholog data set contained 75 loci (Table S2). On average, the three datasets exhibited similar proportions of missing characters and variable sites; however, the 75-locus dataset contained 70.8% fewer parsimony informative sites than the 238-locus dataset and 38.5% fewer than the 108-locus dataset (Table S2). In general, support values decreased with the reduction of loci, but the major relationships were consistent across all three phylogenies (Fig. 2; Figs. S1, S2). The greatest relationship differences we observed between the phylogenies included the placement of *M. decapetala*, *M. pumila* Nutt., *M. densa* Greene, *M. springeri* (Standl.) Tidestr., and *M. longiloba* var. *chihuahuensis* J.J.Schenk & L.Hufford. We will base the remaining Results and Discussion on the above-average bipartition data set of 108 loci unless noted otherwise. We focus on the 108-locus data set to maintain a balance between retaining phylogenetic signal and removing noise associated with paralogy.

Each locus contained an average of 69 taxa and had an average alignment length of 6397 bp (Table S2). A concatenated alignment was 690,863 bp long with 60% undetermined

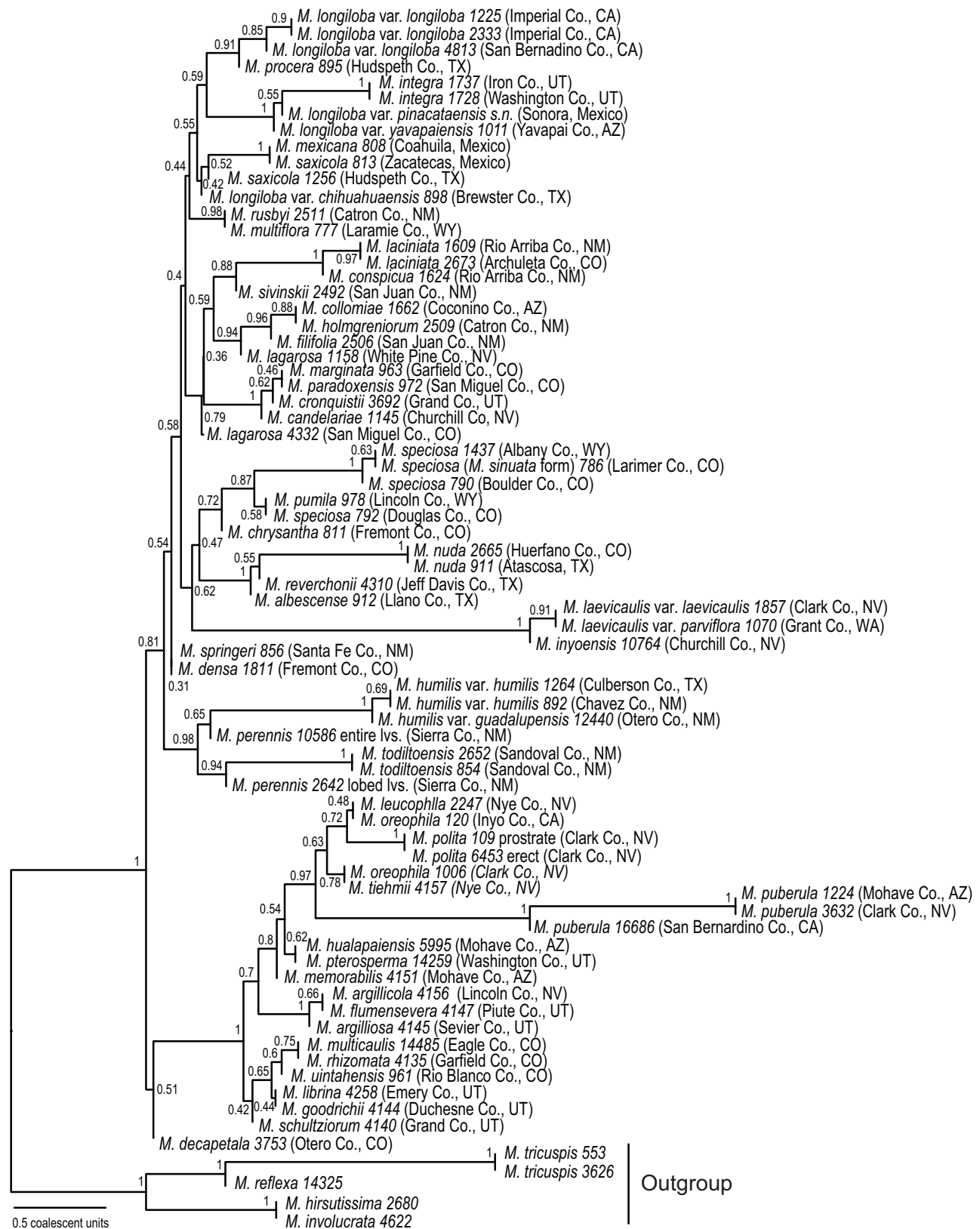


FIG. 2. ASTRAL species tree of *Mentzelia* section *Bartonia* based on 108 nuclear loci. Taxa are designated with a voucher number and place of origin in USA (county, state) or Mexico (state, country). Values at nodes indicate local posterior probability.

characters, 23% variable sites, and 8% parsimony informative sites (Table S2). Gene trees had an average root-to-tip variation of 0.0002, treelength of 0.62, and bipartition support of 0.19 (Table S2). Gene concordance factors (gCF) ranged 0–99.1

(avg. 22.8) and site concordance factors (sCF) ranged 17.51–92.6 (avg. 40.92; Fig. 3; Table S2).

The ASTRAL-III analysis recovered a monophyletic *Mentzelia* section *Bartonia* that is comprised of three major lineages, each

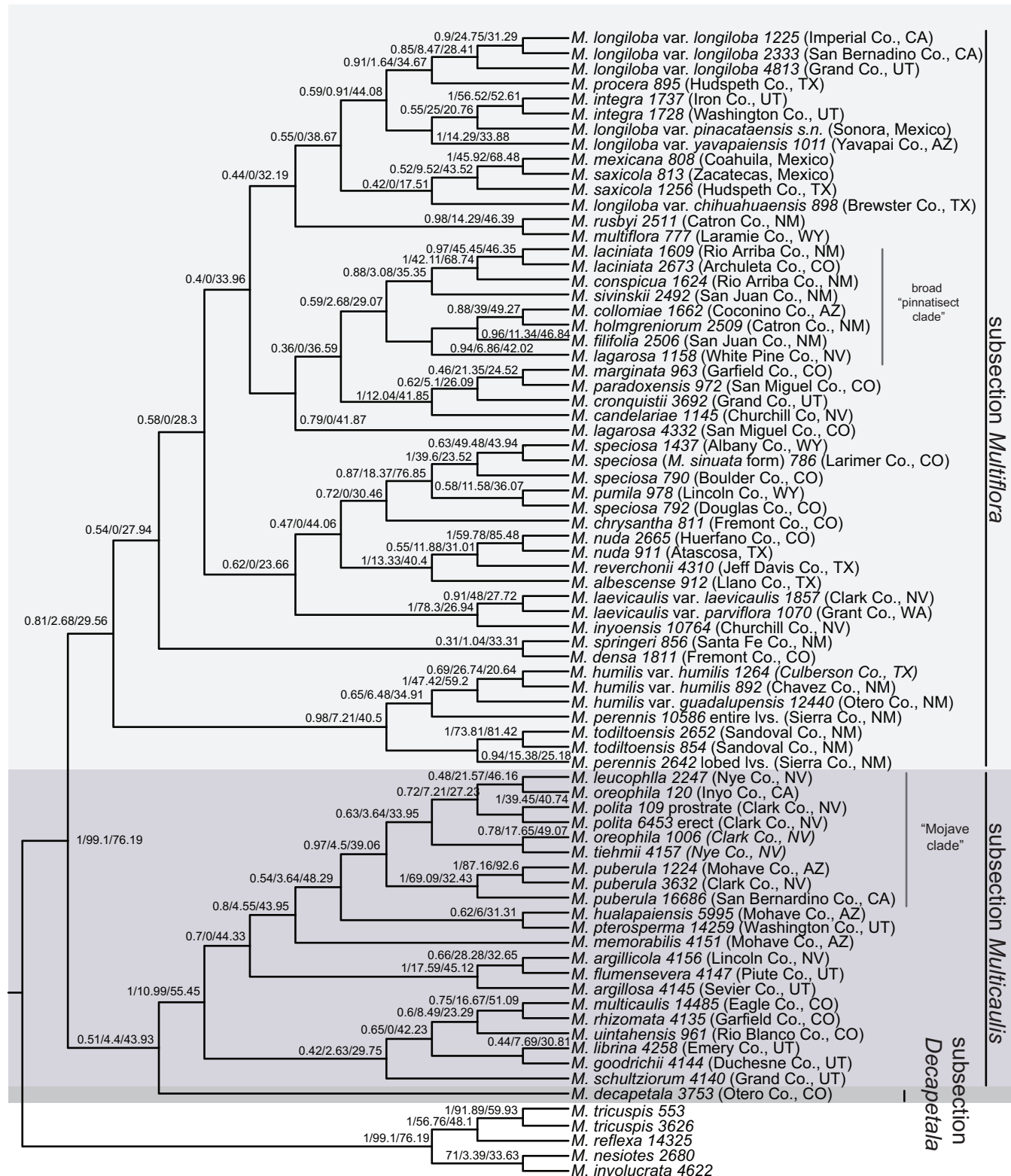


FIG. 3. Cladogram of *Mentzelia* section *Bartoniana* based on an ASTRAL analysis of 108 nuclear loci. Colored boxes designate newly named subsections. Taxa are designated with a voucher number and place of origin in USA (county, state) or Mexico (state, country). Values at nodes indicate local posterior probability (LPP), gene concordance factors (gCF), and site concordance factors (sCF) and are in the order of LPP/gCF/sCF.

with varying levels of support (Fig. 3). The largest clade was supported by a local posterior probability (LPP) of 0.81 with nodes along the backbone averaging an LPP of 0.56, a gCF of 0.51, and an sCF of 33.53 (Fig. 3). In contrast, the second-largest lineage was supported by an LPP of 1 with nodes along the

backbone averaging an LPP of 0.76, a gCF of 4.93, and an sCF of 41.75 (Fig. 3). The monotypic lineage comprised of *M. decapetala* was supported by an LPP of 0.51, a gCF of 4.4, and an sCF of 43.93 (Fig. 3). In general, support values tended to increase towards the tips (Fig. 3). Only 7% of total relationships across

the tree remain unresolved (Fig. 2). Of the 14 taxa that were sampled with multiple individuals, our analyses revealed six species that are not monophyletic as currently circumscribed: *M. longiloba* J.Darl., *M. lagarosa* (K.H.Thorne) J.J.Schenk & L.Hufford, *M. perennis* Wooton, *M. oreophila* J.Darl., *M. speciosa* Osterh., and *M. saxicola* H.J.Thomps. & Zavort (Figs. 2, 3).

Allele divergence varied between 0.04% and 2.36% (avg. 0.58%) and LH varied between 8.17% and 99.7% (avg. 67%; Table S1). Three ingroup individuals had a high LH/AD ratio compared to other samples, indicating possible hybridization: *M. longiloba* var. *longiloba* (Schenk 1225), *M. humilis* var. *humilis* (Urb. & Gilg) J.Darl. (Schenk 1264), and *M. saxicola* (Schenk 1256; Table S1). Individual hypothesis testing with HyDe inferred significant hybridization in *M. longiloba* var. *longiloba* 1225 with both *M. procera* (Wooton & Standl.) J.J.Schenk & L.Hufford ($p < 0.001$, $\gamma = 0.50$) and *M. longiloba* var. *yavapaiensis* J.J.Schenk & L.Hufford ($p < 0.001$, $\gamma = 0.74$), and in *M. saxicola* with both *M. mexicana* H.J.Thomps. & Zavort. ($p < 0.001$, $\gamma = 0.71$) and *M. longiloba chihuahuensis* ($p < 0.001$, $\gamma = 0.47$; Table S3). No significant hybridization was detected in *M. humilis* var. *humilis* 1264 among any of the hypotheses we tested (Table S3).

TAXONOMIC TREATMENT

Below, we propose to formally name subsections of *Mentzelia* section *Bartonia*. We will use these newly designated subsection names throughout the Discussion. Please see Schenk and Hufford (2020) for further details of the characters and distribution maps of the species.

Mentzelia* section *Bartonia* subsection *Decapetala P.Fabre, J.Brokaw, L.Hufford, M.Johnson, & J.Schenk subsection nova—*Mentzelia decapetala* (Pursh) Urb., Ber. Deutsch. Bot. Ges. 10: 263. 1892. TYPE: UNITED STATES. On the banks of the Missouri, from the river Platt to the Andes, on arid volcanic soil, 1811, T. Nuttall s.n. (lectotype: Bot. Mag. 36: pl. 1487, 1812; by Reveal, Moulton & Schuyler, Proc. Acad. Nat. Sci. Philadelphia 149: 32. 1999).

Biennial to short-lived perennials (as rosettes only). Subterranean caudex unbranched. Ovary five–seven carpellate. Stigma lobes short. Seed coat anticlinal cell walls straight.

Mentzelia* section *Bartonia* subsection *Multicaulis P.Fabre, J.Brokaw, L.Hufford, M.Johnson, & J.Schenk subsection nova—*Mentzelia multicaulis* (Osterh.) J.Darl., Ann. Missouri Bot. Gard. 21: 156. 1934. TYPE: UNITED STATES. Colorado. Eagle Co.: Wolcott, 11 Jul 1902, G. E. Osterhout 2663 (lectotype: RM!; isolectotypes: CAS!, COLO! [2], NY!, POM! [2], RM!, RSA! [2]).

Overwintering annual/biennial (in *M. pterosperma* only) or short-lived perennials.

Subterranean caudex branched if perennial. Ovaries three–carpellate. Stigma lobes short. Seed coat anticlinal cell walls straight.

Mentzelia* section *Bartonia* subsection *Multiflora P.Fabre, J.Brokaw, L.Hufford, M.Johnson, & J.Schenk subsection nova—*Mentzelia multiflora* (Nutt.) A.Gray, Mem. Amer. Acad. Arts ser. 2 [Pl. Fendl.] 4:48. 1849. TYPE: UNITED STATES. New Mexico. Santa Fe, sandy hills along the borders of the Rio del Norte, Aug 1841, W. Gambel s.n. (lectotype: K! [K000810481]).

Biennials or short-lived perennials. Subterranean caudex unbranched. Ovaries three–carpellate. Stigma lobes long, slender, and tapering. Seed coat anticlinal cell walls wavy, sinuate, or straight.

DISCUSSION

Mentzelia section *Bartonia* is one of many lineages that have recently and rapidly diversified in western North America. To date, phylogenetic relationships of recently and rapidly radiating groups from western North America have mostly been investigated with Sanger sequencing (e.g., Gilbert et al. 2005; Drummond 2008; Brignone et al. 2019) or amplified fragment-length polymorphism (AFLP; e.g., Worley et al. 2009), but few have been investigated with high-throughput sequencing methods (e.g., Ottenlips et al. 2021; Wenzell et al. 2021; Overson et al. 2023). Although larger data sets might promise resolution among species trees over traditional sequencing, they do not necessarily resolve all relationships. Phylogenetic studies of *Penstemon*, for example, highlight the challenges of resolving evolutionary relationships in recently and rapidly radiating groups of western North America despite extensive data sets. While both RADseq (Wessinger et al. 2016, 2019) and HybSeq with lineage-specific probes (Wolfe et al. 2021) recovered greater phylogenetic resolution over Sanger sequencing-based approaches in *Penstemon*, more recently diverged relationships remained unsupported due to substantial discordance caused by ILS, hybridization, and/or introgression (Wolfe et al. 2006).

Of the 14 aforementioned lineages that have recently and rapidly evolved in western North America, we are aware of only two that have been investigated with the Angiosperms353 probe set: *Castilleja* (Wenzell et al. 2021) and *Lomatium* (Ottenlips et al. 2021), in which both studies examined shallow-level relationships within species complexes. Wenzell et al. (2021) recovered too low genetic variation to implement tree-based methods, but Ottenlips et al. (2021) recovered enough variation to infer well-supported phylogenies that improved resolution over Sanger sequencing-based trees, although they still found high gene-tree discordance likely caused by ILS or hidden paralogy. Both Wenzell et al. (2021) and Ottenlips et al. (2021) did not rigorously address paralogy beyond removing those flagged by HybPiper. As we discovered in our exploratory HybPiper analyses and subsequent HybPhaser analyses (Table S1), HybPiper does not detect all paralogs, and data sets that remove loci flagged only by HybPiper likely maintain hidden paralogs that can obscure species relationships by increasing discordance and lowering support values (Smith and Hahn 2022).

To elucidate species relationships of *Mentzelia* section *Bartonia*, we implemented a hybrid-enrichment approach with the Angiosperms353 bait-set and inferred a phylogenetic tree with only 7% unresolved relationships. In contrast, Schenk and Hufford (2011), who utilized Sanger sequencing approaches, inferred a phylogeny with 45% unresolved relationships. We believe that our ability to resolve most phylogenetic relationships was due to our rigorous bioinformatics pipeline, especially regarding paralogs. In lineages with low phylogenetic signal, relationships are substantially impacted by paralogy, hybridization, and ILS. We detected paralogs because we looked for them in multiple steps of our bioinformatics pipeline (Fig. 1). We also further reduced conflicting

phylogenetic signal by utilizing the HybPhaser consensus sequences to account for hybridization. Furthermore, we accounted for ILS by implementing a multispecies coalescent approach with ASTRAL-III rather than implementing a concatenated supermatrix approach that is insufficient for recently and rapidly radiating groups (Mirarab 2019). One consequence of rigorous paralog detection, however, is that fewer loci decrease support values and alter species tree topology among poorly supported taxa. A custom bait approach could have alleviated this problem by enabling the retention of more loci for phylogenetic analyses following our rigorous quality control steps.

By successfully resolving most relationships in *Mentzelia* section *Bartonia*, our study demonstrated that Angiosperms353 holds considerable promise for elucidating relationships in recently and rapidly radiating lineages of western North America, provided a rigorous bioinformatics pipeline is also implemented to account for paralogy, ILS, and other noise. The robust phylogenetic hypothesis of *Mentzelia* section *Bartonia* we reconstructed now allows us to explore patterns among species relationships at a deeper level than previous analyses allowed.

Subsectional Taxonomy—Across our data sets, we consistently recovered two well-supported major clades (Fig. 3) that were concordant with what Schenk and Hufford (2011, 2020) called clade-1 (now *Mentzelia* subsection *Multiflora*) and clade-2 (now *Mentzelia* subsections *Multicaulis* and *Decapetala*). Both clades represent major groupings of species that are consistent with geographic distribution and morphological variation. Subsection *Multiflora* (clade-1) is especially diverse in the Chihuahuan, Great Plains, Rocky Mountains, and Sonoran floristic provinces (Schenk 2013a), while *M.* subsection *Multicaulis* is diverse in the Colorado Plateau, Great Basin, Sonoran Desert, and Mojave Desert. *Mentzelia* subsection *Multicaulis* also contains many species with highly limited ranges that are confined to biologically restricted soils, putting them at great risk of extinction (Schenk 2013a; Schenk and Hufford 2020). Most species in *M.* subsection *Multicaulis* have adaptations for steep slope habitats with loose substrates, such as having “pseudo-rhizomes” (i.e., aerial shoots that become buried) and a sub-shrubby habit (Schenk and Hufford 2009, 2020), with a notable exception in the overwintering annual or biennial *M. pterosperma* Eastw. *Mentzelia pterosperma* is the only species in *M.* subsection *Multicaulis* that has converged on the *M.* subsection *Multiflora* form, but multiple studies (Schenk and Hufford 2011, 2020; this study) confirm its position in *M.* subsection *Multicaulis*, and this is further supported by the geographic distribution of species (Schenk and Hufford 2020). *Mentzelia decapetala* was recovered as sister to the remaining members of clade-2 in some, but not all analyses (Fig. 3; Figs. S1, S2), however, the species is morphologically and geographically distinct from *M.* subsection *Multicaulis*. Based on the inconsistency of our results as well as the results of previous studies (Hufford et al. 2003; Schenk and Hufford 2011, 2020), we recognize *M. decapetala* as its own clade, although it remains possible that the species might have evolved as sister to either subsections. Additional sampling across the distribution of *M. decapetala* is needed to fully understand its relationship with the rest of *M.* section *Bartonia*. Despite this, the distinct morphology and distribution of *M. decapetala* encourages us to recognize the species as a monotypic subsection. The recognition of the subsections as natural groups brings clarity to the diversity of *Mentzelia*

section *Bartonia*, and for this reason, we propose to formally name them as subsections.

Mojave and Pinnatisect Clades—Phylogenomic studies of *Mentzelia* section *Bartonia* have been conducted with both RADSeq (Cohen and Schenk 2022) and genome skimming approaches (Cohen and Schenk 2022; Moore et al. 2023), which explored phylogenetic relationships among the “Mojave clade” and “pinnatisect clade,” respectively. Below, we compare the relationships we recovered with Angiosperms353 to those recovered in previous studies.

MOJAVE CLADE—In *Mentzelia* subsection *Multicaulis*, Schenk and Hufford (2020) recovered a polytomy of seven species from in and around the Mojave Desert that they colloquially named the “Mojave clade” (their clade 2c), which included: *M. canyonensis*, *M. hualapaiensis* J.J.Schenk, W.C.Hodgs. & L.Hufford, *M. leucophylla*, *M. oreophila*, *M. polita* A.Nelson, *M. puberula* J.Darl, and *M. tiehmii* N.H.Holmgren & P.K.Holmgren. Cohen and Schenk (2022) brought resolution to this clade with RADseq and redefined the “Mojave clade” to include all the above species except *M. hualapaiensis* and *M. canyonensis*, which consistently grouped with *M. memorabilis* N.H.Holmgren & P.K.Holmgren in a sister clade that the authors referred to as the “Colorado Plateau clade.” Consistent with Cohen and Schenk (2022), we recovered a monophyletic “Mojave clade” in all our analyses (Fig. 3; Figs. S1, S2). We did not, however, recover a “Colorado Plateau clade” in any of our analyses (Fig. 3; Figs. S1, S2), although this lack of resolution may be due to the absence of *M. canyonensis* in our datasets. Like Cohen and Schenk (2022), the relative placement of *M. pterosperma* Eastw. to the Mojave and Colorado Plateau clades was also inconsistent.

We recovered a polyphyletic *M. oreophila* that is consistent with the findings of both Cohen and Schenk (2022) and Schenk and Hufford (2011, 2022). We sampled both Nevada and California populations of *M. oreophila*, including the type locality of Inyo County, California. *Schenk 1006* (WS) from Clark County, Nevada, was recovered as sister to *M. tiehmii* from Nye County, Nevada. *Mentzelia tiehmii* is a state-listed gypsum endemic that closely resembles small forms of *M. oreophila* and occurs in the Great Basin near the northernmost edge of the *M. oreophila* distribution (Schenk and Hufford 2020). The specimen *Cohen 120* (RSA) from Inyo County, California, on the other hand, was recovered as sister to the federally endangered *M. leucophylla* from Nye County, Nevada, near the California border. *Mentzelia oreophila* and *M. leucophylla* are morphologically similar but distinguished by geographical distribution and leaf trichome density, with *M. leucophylla* being confined to Ash Meadows National Wildlife Preserve and having dense hairs that form a white cast (Schenk and Hufford 2020). Previous hypotheses have suggested that *M. oreophila* and *M. leucophylla* may be the same species (Prigge 1993; Holmgren and Holmgren 2002; Holmgren et al. 2005), however, Schenk and Hufford (2020) argued that phylogenetic, morphological, distribution, and ecological differences are substantial enough to continue recognizing *M. leucophylla* at the specific level. Our analyses corroborate that the specimen *Schenk 1006* represents a new species, which we are currently describing.

Cohen and Schenk (2022) also investigated the two growth forms of *M. polita*, a state-listed narrow endemic with an affinity for limestone soils. Populations near the type locality of Clark County, Nevada, have erect stems while populations away from the type locality have prostrate stems. We sampled

both growth forms of *M. polita* and found the species to be monophyletic and closely related to the California *M. oreophila* (Figs. 2, 3), which is consistent with Cohen and Schenk (2022). Our Angiosperms353 phylogeny is largely consistent with the relationships found in Cohen and Schenk (2022) and solidifies the monophyly of a core “Mojave clade,” *M. polita*, and the polyphyly of *M. oreophila*. The results of both studies support that Angiosperms353 is comparable with RADseq for recently and rapidly radiating clades.

PINNATISECT CLADE—Moore et al. (2023) used genome skimming to investigate relationships among 20 species in *M.* subsections *Multiflora* and *Decapetala* and tested the monophyly of a “pinnatisect clade”, a hypothesized group of five species with pinnatisect leaves, including: *M. conspicua* Todsén, *M. filifolia* J.J.Schenk & L.Hufford, *M. holmgreniorum* J.J.Schenk & L.Hufford, *M. laciniata* (Rydb.) J.Darl., and *M. sivinskii* J.J.Schenk & L.Hufford. They recovered three data sets (nrDNA, an anonymous nuclear locus, and chloroplast genomes) all with highly discordant topologies, and recovered them as a clade or failed to reject them as a monophyletic group. Although we did not recover a strict pinnatisect clade, we did recover a clade with all pinnatisect members in it, with the addition of *M. collomiae* Christy from Coconino County, Arizona, and *M. lagarosa* from White Pine County, Nevada (Schenk 1158 [WS]; Figs. 2, 3). Like the nrDNA data set from Moore et al. (2023), we recovered *M. collomiae* and *M. lagarosa* as closely related to *M. holmgreniorum*. The inclusion of *M. lagarosa* in a broadened “pinnatisect clade” makes sense based on the species having nearly pinnatifid leaves. *Mentzelia collomiae*, on the other hand, has toothed leaves, but this may represent a character reversal in response to the volcanic cinder soil it inhabits, although this hypothesis would need formal testing.

While Schenk and Hufford (2020) recovered *M. lagarosa* as monophyletic in their Bayesian analysis, we recovered no such relationship in any of our analyses (Figs. 2, 3; Figs. S1, S2). Our sampling, however, was more widespread; we sampled Schenk 1158 (WS) from White Pine County, Nevada, and Hufford 4332 from San Miguel County, Colorado (WS; Figs. 2, 3), while Schenk and Hufford sampled from White Pine County, Nevada and Wayne County, Utah. Our results indicate that there may be geographic isolation in the Colorado *M. lagarosa*, but population sampling across the species range should be performed to elucidate the presence of any hidden species or hybridization. Although we found no evidence of hybridization of *M. lagarosa* with *M. cronquistii* (Table S3), hybridization or introgression may also explain why we recovered a polyphyletic *M. lagarosa* and why the placement of the species was so discordant between data sets in Moore et al. (2023).

Discordance and Hybridization—While HybPhaser analyses flagged *M. longiloba* var. *longiloba* (Schenk 1225 [WS]), *M. humilis* var. *humilis* (Schenk 1264 [WS]), and *M. saxicola* (Schenk 1256 [WS]) as putative hybrids, discordance between data sets and the recovery of polyphyletic species may also indicate potential hybridization. We discuss the relationships of these putative hybrids and their close relatives below.

MENTZELIA LONGILOBA—We found evidence that *M. longiloba* var. *longiloba* from Imperial County, California (Schenk 1225 [WS]) may represent an F1 hybrid between *M. procera* from Hudspeth County, Texas (Schenk 895 [WS]) and *M. longiloba* var. *longiloba* from San Bernardino County, CA (De Groot 2333 [WS]; Table S3). Introgression from *M. procera* may explain why Schenk and Hufford (2020) recovered only Texas *M. procera*, and no other populations, in a polytomy with *M. longiloba*

var. *longiloba*. Hybridization between *M. procera* and *M. longiloba* var. *longiloba* may be limited to west Texas, however, proper population-level studies will be necessary to confirm this, especially in eastern Arizona where the ranges of the two species overlap (Schenk and Hufford 2020). Additional sampling of *M. procera* populations outside of Texas will be needed to confirm the relative placement of species in the phylogeny.

Although not flagged by HybPhaser as a putative hybrid, *Mentzelia longiloba* var. *chihuahuensis* had discordant relationships between data sets that may be indicative of hybridization. *Mentzelia longiloba* var. *chihuahuensis* (Schenk 898 [WS]) and *M. procera* (Schenk 895 [WS]) have overlapping distributions in west Texas where they have the potential for hybridization (Schenk and Hufford 2020); however, we did not have the sampling to formally test this hypothesis with HyDe. We recovered *Mentzelia longiloba* var. *chihuahuensis* as sister to *M. longiloba* var. *longiloba* in our strictest data set of 75 loci (Fig. S2), suggesting that it is likely the true sister to *M. longiloba* var. *longiloba*, which does not contradict previous studies (Schenk and Hufford 2011, 2020); however, additional population sampling of *M. longiloba* var. *chihuahuensis* across its distribution will be necessary to fully understand its relationship to *M. longiloba* var. *longiloba*.

MENTZELIA HUMILIS—Populations of *Mentzelia humilis* var. *humilis* in northern New Mexico (Schenk 1264 [WS]) are sometimes mistaken for *M. perennis* and *M. todiltoensis* N.D. Atwood & S.L. Welsh due to similarities in floral, leaf, and growth forms (Schenk and Hufford 2020). Despite morphological similarities and overlapping distributions, HyDe analyses found no signs of hybridization among these three taxa (Table S3).

We recovered a monophyletic *Mentzelia humilis* var. *humilis* in our 108-locus analysis, however, we recovered *M. humilis* var. *guadalupensis* as sister to *M. humilis* var. *humilis* from Culberson County, Texas (Schenk 1264 [WS]) in our 75- and 238-locus analyses, which may indicate gene flow between the taxa. We investigated *M. humilis* var. *humilis* 1264 as a putative hybrid with *M. humilis* var. *guadalupensis* but found no significant signs of hybridization (Table S3). Since *M. humilis* var. *humilis* 1264 was flagged as a putative hybrid by HybPhaser, it remains possible there is an unknown parent we did not investigate, although we sampled the most likely parents, *M. perennis* and *M. todiltoensis* (Schenk and Hufford 2020). More comprehensive sampling of the *M. humilis* varieties and its close relatives will be necessary to elucidate the presence of any cryptic taxa or hybridization.

MENTZELIA PERENNIS—We sampled two populations of *Mentzelia perennis* to represent two distinct laminal forms. The plants from the type locality (White Sands, NM) have mostly entire to slightly lobed leaves, while plants outside the type locality have deeply lobed to pinnatisect leaves that are similar to *M. humilis* (Schenk and Hufford 2020). In this study, Spellenberg 10586 (ID, NMC) from Sierra County, New Mexico represents the type locality and entire leaves, while Schenk 2642 (UNM) from Sierra County, New Mexico represents pinnately lobed leaves. In previous studies, populations representing both laminar types of *M. perennis* have been recovered as closely related to *M. todiltoensis* (Schenk and Hufford 2011, 2020). Our results, however, suggest that only the pinnately lobed *M. perennis* (Schenk 2642) is closely related to *M. todiltoensis* while the entire-margined *M. perennis* (Spellenberg 10586) is more closely related to *M. humilis* (Figs. 2, 3). Future studies should focus on deeper sampling of *M. perennis*,

M. humilis, and *M. todiltoensis* to investigate geographic isolation and hybridization among these species.

MENTZELIA SAXICOLA—The close relationship recovered between *Mentzelia saxicola* (Fig. 2), *M. mexicana*, and *M. longiloba* var. *chihuahuensis* was not found in previous studies (Schenk and Hufford 2011, 2020); however, the taxa do have overlapping ranges in the Chihuahuan Desert, and *M. mexicana* and *M. saxicola* are notably morphologically similar (Thompson and Powell 1981; Schenk and Hufford 2020). We found substantial evidence that *M. saxicola* from Hudspeth County, Texas (Schenk 1256 [WS]) had large genomic contribution from both *M. mexicana* from Coahuila, Mexico (Granados & Schenk 808 [MEXU]) and *M. longiloba* var. *chihuahuensis* from Brewster County, Texas (Table S3). A hybrid origin would explain why, in only the 108-locus tree, *M. longiloba chihuahuensis* is recovered in a clade with *M. saxicola* and *M. mexicana* and not with *M. longiloba* var. *longiloba*. *Mentzelia saxicola* and *M. procera* are also known to hybridize in Hudspeth County, Texas (Schenk and Hufford 2020), although we sampled both species from this locality and we did not find evidence of hybridization between them (Table S3).

MENTZELIA PUMILA—*Mentzelia pumila* has a history of confounding taxonomy (Schenk and Hufford 2011) and was recovered as a monotypic lineage with no clearly defined close relatives in Schenk and Hufford (2020). We recovered *M. pumila* within (Figs. 2, 3) or sister to (Fig. S2) *M. speciosa* in our 108- and 75-locus analyses, respectively, but sister to a clade comprised of *M. laevicaulis* (Douglas) Torr. & A. Gray and *M. inyoensis* Prigge in our 238-locus analysis (Fig. S1). Recovery of *M. pumila* within *M. speciosa* was unexpected, as Schenk and Hufford (2020) recovered *M. speciosa* as monophyletic; however, this close relationship does make sense geographically because both *M. pumila* and *M. speciosa* are distributed in the middle Rocky Mountains (Schenk and Hufford 2020). On the other hand, *M. pumila* has notably similar seed morphology and chromosome numbers as *M. laevicaulis* (Hill 1975). We did not have the population sampling to test for hybridization in *M. pumila* with HyDe, but future studies should explore potential hybridization with *M. speciosa* and/or *M. laevicaulis*.

The goal of our study was to elucidate evolutionary relationships within *Mentzelia* section *Bartonia* using Angiosperms353 target-capture. Like Schenk and Hufford (2011), the phylogeny we recovered was well-supported at the base and the tips but lacked resolution or strong support along the backbone. Relationships between tips were mostly consistent across data sets, with most relationships being well-supported, consistent with previous studies, and resolving previous polytomies, but some topological differences warrant further investigation (Figs. 2, 3; Figs. S1, S2). We also recovered several nonmonophyletic species that illuminated potential cryptic speciation or hybridization that should be investigated in future studies (Figs. 2, 3; Figs. S1, S2). Overall, the Angiosperms353 approach, in conjunction with our rigorous bioinformatics pipeline, provided species-level resolution to this recent and rapidly evolving lineage from western North America, significantly progressing our understanding of its evolution. To provide stronger support for relationships across all clades, however, lineage-specific baits may be necessary to resolve relationships in areas of the phylogeny with little to no support.

ACKNOWLEDGMENTS

Funding for this research was generously provided to JJS by NSF DEB award 2117446, and to PPF by the Graduate Student Research Grant from the Ohio University College of Arts and Sciences. We thank undergraduate researchers E. Ivory and J. McClure for their assistance with extracting and quantifying genomic DNA. We are grateful to herbarium curator E. H. Roalson for permitting our destructive sampling of specimens at WS, N. Daniels at the Ohio University Genomics Facility for bioanalyzing samples, and D. Cohen for donation of leaf material. We are thankful to L. E. Becklund, T. C. Ling, and M. Vis for their generous feedback on earlier drafts of this manuscript.

AUTHOR CONTRIBUTIONS

PPF wrote the manuscript with edits contributed by JMB, LDH, MGJ, and JJS. JMB, LDH, and JJS contributed specimen vouchers and DNA. PPF and JJS performed laboratory procedures and generated specimen samples. MGJ set up the sequencing procedure and designed the HybPiper pipeline. PPF performed bioinformatics and data analyses.

LITERATURE CITED

- Axelrod, D. I. 1958. Evolution of the Madro-Tertiary geoflora. *The Botanical Review* 24: 433–509.
- Axelrod, D. I. 1972. Edaphic aridity as a factor in angiosperm evolution. *The American Naturalist* 106: 311–320.
- Bagley, J. C., S. Uribe-Convers, M. M. Carlsen, and N. Muchhala. 2020. Utility of targeted sequence capture for phylogenomics in rapid, recent angiosperm radiations: Neotropical *Burmeistera* bellflowers as a case study. *Molecular Phylogenetics and Evolution* 152: 106769.
- Baker, W. J., P. Bailey, V. Barber, A. Barker, S. Bellot, D. Bishop, L. R. Botigué, et al. 2022. A comprehensive phylogenomic platform for exploring the Angiosperm tree of life. *Systematic Biology* 71: 301–319.
- Baum, D. A. 2007. Concordance trees, concordance factors, and the exploration of reticulate genealogy. *TAXON* 56: 417–426.
- Blischak, P. D., J. Chifman, A. D. Wolfe, and L. S. Kubatko. 2018. HyDe: A Python package for genome-scale hybridization detection. *Systematic Biology* 67: 821–829.
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.
- Borowiec, M. L. 2016. AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4: e1660.
- Borowiec, M. L. 2019. Spruceup: Fast and flexible identification, visualization, and removal of outliers from large multiple sequence alignments. *Journal of Open Source Software* 4: 1635.
- Botnaru, L. and J. J. Schenk. 2019. Staminate evolution in *Mentzelia* section *Bartonia* (Loasaceae) and its impact on insect visitation rates. *Botanical Journal of the Linnean Society* 190: 151–164.
- Brignone, N. F., R. E. Pozner, and S. S. Denham. 2019. Origin and evolution of *Atriplex* (Amaranthaceae s.l.) in the Americas: Unexpected insights from South American species. *TAXON* 68: 1021–1036.
- Brokaw, J. M., J. J. Schenk, J. K. Devitt, and D. J. Brokaw. 2020. Phylogeny and floral character evolution of *Mentzelia* section *Bicuspidaria* (Loasaceae). *Systematic Botany* 45: 306–314.
- Brown, J. W., J. F. Walker, and S. A. Smith. 2017. Phyx: Phylogenetic tools for unix. *Bioinformatics* 33: 1886–1888.
- Burge, D. O., D. M. Erwin, M. B. Islam, J. Kellermann, S. W. Kembel, D. H. Wilken, and P. S. Manos. 2011. Diversification of *Ceanothus* (Rhamnaceae) in the California Floristic Province. *International Journal of Plant Sciences* 172: 1137–1164.
- Carey, S. J., L. E. Becklund, P. P. Fabre, and J. J. Schenk. 2023. Optimizing the lysis step in CTAB DNA extractions of silica-dried and herbarium leaf tissues. *Applications in Plant Sciences* 11: e11522.
- Cather, S. M., C. E. Chapin, and S. A. Kelley. 2012. Diachronous episodes of Cenozoic erosion in southwestern North America and their relationship to surface uplift, paleoclimate, paleodrainage, and paleoaltimetry. *Geosphere* 8: 1177–1206.
- Chen, S., Y. Zhou, Y. Chen, and J. Gu. 2018. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34: i884–i890.
- Cohen, D. H. and J. J. Schenk. 2022. Investigating phylogenetic placement and species-level relationships in a recent radiation of *Mentzelia* section *Bartonia* (Loasaceae) from the Mojave Desert. *Systematic Botany* 47: 140–151.
- Cooper, B. J., M. J. Moore, N. A. Douglas, W. L. Wagner, M. G. Johnson, R. P. Overson, S. P. Kinsian, A. J. McDonnell, R. A. Levin, R. A.

- Raguso, H. F., Olvera, H., Ochoterena, J. B., Fant, K. A., Skogen, and N. J. Wickett. 2023. Target enrichment and extensive population sampling help untangle the recent, rapid radiation of *Oenothera* sect. *Calylophus*. *Systematic Biology* 72: 249–263.
- Drummond, C. S. 2008. Diversification of *Lupinus* (Leguminosae) in the western New World: Derived evolution of perennial life history and colonization of montane habitats. *Molecular Phylogenetics and Evolution* 48: 408–421.
- Eisner, T., M. Eisner, and E. R. Hoebeke. 1998. When defense backfires: Detrimental effect of a plant's protective trichomes on an insect beneficial to the plant. *Proceedings of the National Academy of Sciences* 95: 4410–4414.
- Fabre, P. P., J. M. Brokaw, L. D. Hufford, M. G. Johnson, and J. J. Schenk. 2025. Resolving phylogenetic relationships of a recently and rapidly evolving lineage from western North America (*Mentzelia* section *Bartonia*, Loasaceae). Dryad Digital Repository. <https://doi.org/10.5061/dryad.sf7m0cghb>.
- Foster, G. L., D. J. Lunt, and R. R. Parrish. 2010. Mountain uplift and the glaciation of North America – a sensitivity study. *Climate of the Past* 6: 707–717.
- Frost, L. A., A. M. Bedoya, and L. P. Lagomarsino. 2024. Artifactual orthologs and the need for diligent data exploration in complex phylogenomic data sets: A museomic case study from the Andean flora. *Systematic Biology* 73: 308–322.
- George, E. E., D. H. Mansfield, J. F. Smith, R. L. Hartman, S. R. Downie, and C. E. Hinchliff. 2014. Phylogenetic analysis reveals multiple cases of morphological parallelism and taxonomic polyphyly in *Lomatium* (Apiaceae). *Systematic Botany* 39: 662–675.
- Gilbert, C., J. Dempcy, C. Ganong, R. Patterson, and G. S. Spicer. 2005. Phylogenetic relationships within *Phacelia* subgenus *Phacelia* (Hydrophyllaceae) inferred from nuclear rDNA ITS sequence data. *Systematic Botany* 30: 627–634.
- Holmgren, N. H. and P. K. Holmgren. 2002. New *Mentzelias* (Loasaceae) from the Intermountain Region of western United States. *Systematic Botany* 27: 747–762.
- Holmgren, N. H., P. K. Holmgren, and A. Cronquist. 2005. *Intermountain Flora: Vascular Plants of the Intermountain West, USA. Volume two, Part B: Subclass Dilleniidae*. Bronx: New York Botanical Garden.
- Hufford, L., M. M. McMahon, A. M. Sherwood, G. Reeves, and M. W. Chase. 2003. The major clades of Loasaceae: Phylogenetic analysis using the plastid *matK* and *trnL-trnF* regions. *American Journal of Botany* 90: 1215–1228.
- Johnson, M. G., E. M. Gardner, Y. Liu, R. Medina, B. Goffinet, A. J. Shaw, N. J. C. Zerega, and N. J. Wickett. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* 4: 1600016.
- Johnson, M. G., L. Pokorny, S. Dodsworth, L. R. Botigué, R. S. Cowan, A. Devault, W. L. Eiserhardt, N. Epitawalage, F. Forest, J. T. Kim, J. H. Leebens-Mack, I. J. Leitch, O. Maurin, D. E. Soltis, P. S. Soltis, G. K. S. Wong, W. J. Baker, and N. J. Wickett. 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology* 68: 594–606.
- Junier, T. and E. M. Zdobnov. 2010. The Newick utilities: High-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* 26: 1669–1670.
- Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler, and L. S. Jermiin. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* 14: 587–589.
- Katoh, K. and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.
- Kostikova, A., G. Litsios, S. Burgý, L. Milani, P. B. Pearman, and N. Salamin. 2014. Scale-dependent adaptive evolution and morphological convergence to climatic niche in Californian eriogonoids (Polygonaceae). *Journal of Biogeography* 41: 1326–1337.
- Kruckeberg, A. R. 2004. *Geology and Plant Life: The Effects of Landforms and Rock Types on Plants*. Seattle: University of Washington Press.
- Larridon, I., T. Villaverde, A. R. Zuntini, L. Pokorny, G. E. Brewer, N. Epitawalage, I. Fairlie, M. Hahn, J. Kim, E. Maguilla, O. Maurin, M. Xanthos, A. L. Hipp, F. Forest, and W. J. Baker. 2020. Tackling rapid radiations with targeted sequencing. *Frontiers in Plant Science* 10: 1655.
- Leaché, A. D. and J. R. Oaks. 2017. The utility of single nucleotide polymorphism (SNP) data in Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 48: 69–84.
- Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25: 1754–1760.
- Maddison, W. P. 1997. Gene trees in species trees. *Systematic Biology* 46: 523–536.
- Mai, U. and S. Mirarab. 2018. TreeShrink: Fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics* 19: 272.
- Majure, L. C., R. Puente, M. P. Griffith, W. S. Judd, P. S. Soltis, and D. E. Soltis. 2012. Phylogeny of *Opuntia* s.s. (Cactaceae): Clade delineation, geographic origins, and reticulate evolution. *American Journal of Botany* 99: 847–864.
- McLay, T. G. B., J. L. Birch, B. F. Gunn, W. Ning, J. A. Tate, L. Nauheimer, E. M. Joyce, L. Simpson, A. N. Schmidt-Lebuhn, W. J. Baker, F. Forest, and C. J. Jackson. 2021. New targets acquired: Improving locus recovery from the Angiosperms353 probe set. *Applications in Plant Sciences* 9: e11420.
- Minh, B. Q., M. W. Hahn, and R. Lanfear. 2020. New methods to calculate concordance factors for phylogenomic data sets. *Molecular Biology and Evolution* 37: 2727–2733.
- Minh, B. Q., R. Lanfear, N. Ly-Trong, J. Trifinopoulos, D. Schrempf, and H. A. Schmidt. 2022. IQ-TREE version 2.2.0: Tutorials and manual phylogenomic software by maximum likelihood. <http://www.iqtree.org/doc/iqtree-doc.pdf> (last accessed 03 January 2025).
- Mirarab, S. 2019. Species tree estimation using ASTRAL: Practical considerations. <http://arxiv.org/abs/1904.03826> (last accessed 03 January 2025).
- Mo, Y. K., R. Lanfear, M. W. Hahn, and B. Q. Minh. 2023. Updated site concordance factors minimize effects of homoplasy and taxon sampling. *Bioinformatics* 39: btac741.
- Moore, A. G., K. Kelly, and J. J. Schenk. 2023. Resolving relationships in *Mentzelia* section *Bartonia* (Loasaceae) in the face of cytonuclear discordance. *Systematic Botany* 48: 386–394.
- Moore, M. J., J. F. Mota, N. A. Douglas, H. Flores-Olvera, and H. Ochoterena. 2014. The ecology, assembly, and evolution of gypsophile floras. Pp. 97–128 in *Plant Ecology and Evolution in Harsh Environments*, eds. N. Rajakaruna, R. S. Boyd, and T. B. Harris. New York: Nova Science Publishers.
- Nauheimer, L., N. Weigner, E. Joyce, D. Crayn, C. Clarke, and K. Nargar. 2021. HybPhaser: A workflow for the detection and phasing of hybrids in target capture data sets. *Applications in Plant Sciences* 9: e11441.
- Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274.
- Ottenslips, M. V., D. H. Mansfield, S. Buerki, M. A. E. Feist, S. R. Downie, S. Dodsworth, F. Forest, G. M. Plunkett, and J. F. Smith. 2021. Resolving species boundaries in a recent radiation with the Angiosperms353 probe set: The *Lomatium packardiae*/*L. anomalum* clade of the *L. triternatum* (Apiaceae) complex. *American Journal of Botany* 108: 1217–1233.
- Overson, R. P., M. G. Johnson, L. L. Bechen, S. P. Kinoshian, N. A. Douglas, J. B. Fant, P. C. Hoch, R. A. Levin, M. J. Moore, R. A. Raguso, W. L. Wagner, K. A. Skogen, and N. J. Wickett. 2023. A phylogeny of the evening primrose family (Onagraceae) using a target enrichment approach with 303 nuclear loci. *BMC Ecology and Evolution* 23: 66.
- Patsis, A., R. P. Overson, K. A. Skogen, N. J. Wickett, M. G. Johnson, W. L. Wagner, R. A. Raguso, J. B. Fant, and R. A. Levin. 2021. Elucidating the evolutionary history of *Oenothera* Sect. *Pachylophus* (Onagraceae): A phylogenomic approach. *Systematic Botany* 46: 799–811.
- Pellmyr, O., K. A. Segraves, D. M. Althoff, M. Balcázar-Lara, and J. Leebens-Mack. 2007. The phylogeny of yuccas. *Molecular Phylogenetics and Evolution* 43: 493–501.
- Prigge, B. 1993. Loasaceae. Pp. 741–744 in *The Jepson Manual: Higher Plants of California*, ed. J. C. Hickman. Berkeley: Univ of California Press.
- Qian, H., B. D. Mishler, J. Zhang, and S. Qian. 2024. Global patterns and ecological drivers of taxonomic and phylogenetic endemism in angiosperm genera. *Plant Diversity* 46: 149–157.
- Reveal, J. L. 1978. Distribution and phylogeny of Eriogonoideae (Polygonaceae). *Great Basin Naturalist Memoirs*: 169–190.
- Reveal, J. L. and E. L. Styer. 1973. Miscellaneous chromosome counts of western American plants—II. *Great Basin Naturalist* 33: 19–25.
- Sayyari, E., J. B. Whitfield, and S. Mirarab. 2017. Fragmentary gene sequences negatively impact gene tree and species tree reconstruction. *Molecular Biology and Evolution* 34: 3279–3291.

- Schenk, J. J. 2013a. Biogeographical diversification of *Mentzelia* section *Bartonia* in western North America. *Journal of Biogeography* 40: 455–465.
- Schenk, J. J. 2013b. Evolution of limited seed dispersal ability on gypsum islands. *American Journal of Botany* 100: 1811–1822.
- Schenk, J. J. and L. Hufford. 2009. Name changes in the *Mentzelia multicaulis* Complex (Loasaceae). *Novon: A Journal for Botanical Nomenclature* 19: 117–121.
- Schenk, J. J. and L. Hufford. 2010. Effects of substitution models on divergence time estimates: Simulations and an empirical study of model uncertainty using Cornales. *Systematic Botany* 35: 578–592.
- Schenk, J. J. and L. Hufford. 2011. Phylogeny and taxonomy of *Mentzelia* section *Bartonia* (Loasaceae). *Systematic Botany* 36: 711–720.
- Schenk, J. J. and L. Hufford. 2020. *A Phylogenetic Monograph of Mentzelia section Bartonia* (Loasaceae). Ann Arbor: American Society of Plant Taxonomists.
- Schenk, J. J., L. E. Becklund, S. J. Carey, and P. P. Fabre. 2023. What is the “modified” CTAB protocol? Characterizing modifications to the CTAB DNA extraction protocol. *Applications in Plant Sciences* 11: e11517.
- Schenk, J. J., S. J. Jacobs, and L. Hufford. 2025. Comparative diversification analyses of Hydrangeaceae and Loasaceae reveal complex evolutionary history as species disperse out of Mesoamerica. *American Journal of Botany* : e16455.
- Smith, M. L. and M. W. Hahn. 2022. The frequency and topology of pseudorthologs. *Systematic Biology* 71: 649–659.
- Smith, S. A., J. W. Brown, and J. F. Walker. 2018. So many genes, so little time: A practical approach to divergence-time estimation in the genomic era. *Plos One* 13: e0197433.
- Stebbins, G. L. 1952. Aridity as a stimulus to plant evolution. *The American Naturalist* 86: 33–44.
- Stone, B. W. and A. D. Wolfe. 2021. Asynchronous rates of lineage, phenotype, and niche diversification in a continental-scale adaptive radiation. *bioRxiv* <https://doi.org/10.1101/2021.06.14.448393> [Preprint] (last accessed 1 April 2024).
- Tank, D. C. and R. G. Olmstead. 2008. From annuals to perennials: Phylogeny of subtribe Castillejinae (Orobanchaceae). *American Journal of Botany* 95: 608–625.
- Thiers, B. M. 2021 (updated continuously). Index Herbariorum. <http://sweetgum.nybg.org/science/ih/> (last accessed August 2024).
- Thomas, A. E., J. Igea, H. M. Meudt, D. C. Albach, W. G. Lee, and A. J. Tanentzap. 2021. Using target sequence capture to improve the phylogenetic resolution of a rapid radiation in New Zealand *Veronica*. *American Journal of Botany* 108: 1289–1306.
- Thompson, H. J. and A. M. Powell. 1981. Loasaceae of the Chihuahuan Desert region. *Phytologia* 49: 16–33.
- Thorne, R. 1993. Phytogeography. Pp. 132–153 in *Flora of North America; Volume 1: Introduction*, eds. Flora of North America Editorial Committee. New York: Oxford University Press.
- Ufimov, R., V. Zeisek, S. Pišová, W. J. Baker, T. Fér, M. van Loo, C. Dobeš, and R. Schmickl. 2021. Relative performance of customized and universal probe sets in target enrichment: A case study in subtribe Malinae. *Applications in Plant Sciences* 9: e11442.
- Wahlert, G. A., V. T. Parker, and M. C. Vasey. 2009. A phylogeny of *Arctostaphylos* (Ericaceae) inferred from nuclear ribosomal ITS sequences. *Journal of the Botanical Research Institute of Texas* 3: 673–682.
- Wenzell, K. E., A. J. McDonnell, N. J. Wickett, J. B. Fant, and K. A. Skogen. 2021. Incomplete reproductive isolation and low genetic differentiation despite floral divergence across varying geographic scales in *Castilleja*. *American Journal of Botany* 108: 1270–1288.
- Wessinger, C. A., C. C. Freeman, M. E. Mort, M. D. Rausher, and L. C. Hileman. 2016. Multiplexed shotgun genotyping resolves species relationships within the North American genus *Penstemon*. *American Journal of Botany* 103: 912–922.
- Wessinger, C. A., M. D. Rausher, and L. C. Hileman. 2019. Adaptation to hummingbird pollination is associated with reduced diversification in *Penstemon*. *Evolution Letters* 3: 521–533.
- Wolfe, A. D., C. P. Randle, S. L. Datwyler, J. J. Morawetz, N. Arguedas, and J. Diaz. 2006. Phylogeny, taxonomic affinities, and biogeography of *Penstemon* (Plantaginaceae) based on ITS and cpDNA sequence data. *American Journal of Botany* 93: 1699–1713.
- Wolfe, A. D., P. D. Blischak, and L. S. Kubatko. 2021. Phylogenetics of a rapid, continental radiation: Diversification, biogeography, and circumscription of the Beardtongues (*Penstemon*; Plantaginaceae). *bioRxiv* <https://doi.org/10.1101/2021.04.20.440652> [Preprint] (last accessed 30 May 2024).
- Worley, A. C., H. Ghazvini, and D. W. Schemske. 2009. A phylogeny of the genus *Polemonium* based on amplified fragment length polymorphism (AFLP) markers. *Systematic Botany* 34: 149–161.
- Yardeni, G., J. Viruel, M. Paris, J. Hess, C. Groot Crego, M. de La Harpe, N. Rivera, M. H. J. Barfuss, W. Till, V. Guzmán-Jacob, T. Krömer, C. Lexer, O. Paun, and T. Leroy. 2022. Taxon-specific or universal? Using target capture to study the evolutionary history of rapid radiations. *Molecular Ecology Resources* 22: 927–945.
- Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19: 15–30.
- Zhou, W. and Q.-Y. Xiang. 2022. Phylogenomics AND biogeography of *Castanea* (chestnut) and *Hamamelis* (witch-hazel) – Choosing between RAD-seq and Hyb-Seq approaches. *Molecular Phylogenetics and Evolution* 176: 107592.
- Zhou, W., J. Soghigian, and Q.-Y. Xiang. 2022. A new pipeline for removing paralogs in target enrichment data. *Systematic Biology* 71: 410–425.

APPENDIX 1. Voucher information. Voucher information and GenBank accession numbers for *Mentzelia* section *Bartonia* (ingroup) and section *Bicuspidaria* (outgroup). Entries are listed as: *taxon*, authority, *voucher*, herbarium code, collection location, and GenBank numbers. Herbarium code follows the Index Herbariorum (Thiers, updated continuously).

Ingroup: *M. albescens* (Gill. ex Arn.) Griseb., Schenk 912, WS, Llano Co., TX, SAMN46202335; *M. argillicola* N.H.Holmgren & P.K.Holmgren, Hufford 4156, WS, Lincoln Co., NV, SAMN46202336; *M. argillosa* J.Darl., Hufford 4145, WS, Sevier Co., UT, SAMN46202337; *M. candalariae* H.J.Thomps. & Prigge, Schenk & Tiehm 1145, RENO, WS, Churchill Co., NV, SAMN46202338; *M. canyonensis* J.J.Schenk, W.Hodgson & L.Hufford, *Hodgson et al.* 25716, DES, Coconino Co., AZ, SAMN46202339; *M. chrysantha* Engelm. ex Brandegee, Schenk 811, WS, Fremont Co., CO, SAMN46202340; *M. collomia* Christy, Schenk 1662, WS, Coconino Co., AZ, SAMN46202341; *M. conspicua* Todsén, Schenk 1624, WS, Rio Arriba Co., NM, SAMN46202342; *M. cronquistii* H.J.Thomps. & Prigge, Hufford 3692, WS, Grand Co., UT, SAMN46202343; *M. decapetala* (Pursh) Urb., Hufford 3753, WS, Otero Co., CO, SAMN46202344; *M. densa* Greene, Schenk 1811, WS, Fremont Co., CO, SAMN46202345; *M. filifolia* J.J.Schenk & L.Hufford, Schenk 2506, GAS, San Juan Co., NM, SAMN46202346; *M. flumensevera* (N.H.Holmgren & P.K.Holmgren) J.J.Schenk & L.Hufford, Hufford 4147, WS, Piute Co., UT, SAMN46202347; *M. goodrichii* K.H.Thorne & S.L.Welsh, Hufford 4144, WS, Duchesne Co., UT, PRJNA1209082; *M. holmgreniorum* J.J.Schenk & L.Hufford, Schenk 2509, BHO, Catron Co., NM, SAMN46202349; *M. hualapaiensis* J.J.Schenk, W.C.Hodgson & L.Hufford, Burgess 5995, ARIZ, Mohave Co., AZ, SAMN46202350; *M. humilis* var. *guadalupensis* Spellenberg, Spellenberg 12440, NMC, Otero Co., NM, SAMN46202351; *M. humilis* var. *humilis* (Urb. & Gilg) J.Darl., Schenk 892, WS, Chavez Co., NM, SAMN46202352; *M. humilis* var. *humilis* (Urb. & Gilg) J.Darl., Schenk 1264, WS, Culberson Co., TX, SAMN46202353; *M. integra* (M.E.Jones) Tidestr., Schenk 1728, WS, Washington Co., UT, SAMN46202354; *M. integra* (M.E.Jones) Tidestr., Schenk 1737, WS, Iron Co., UT, SAMN46202355; *M. inyoensis* H.J.Thomps. & Prigge, Tiehm 10764, WS, Churchill Co., NV, SAMN46202356; *M. laciniata* (Rydb.) J.Darl., Schenk 1609, WS, Rio Arriba Co., NM, SAMN46202357; *M. laciniata* (Rydb.) J.Darl., Schenk 2673, BHO, Archuleta Co., CO, SAMN46202358; *M. laevicaulis* var. *laevicaulis* (Douglas ex Hook.) Torr. & A.Gray, Schenk 1857, WS, Clark Co., NV, SAMN46202359; *M. laevicaulis* var. *parviflora* (Dougl. ex Hook.) Torr. & A.Gray, Schenk 1070, WS, Grant Co., WA, SAMN46202360; *M. lagarosa* (K.H.Thorne) J.J.Schenk & L.Hufford, Schenk 1158, WS, White Pine Co., NV, SAMN46202361; *M. lagarosa* H.J.Thomps. & Prigge, Hufford 4332, WS, San Miguel Co., CO, SAMN46202362; *M. leucophylla* Brandegee, *Reveal* 2247, RENO, Nye Co., NV, SAMN46202363; *M. libraria* (K.H.Thorne & F.J.Smith) J.J.Schenk & L.Hufford, Hufford 4258, WS, Emery Co., UT, SAMN46202364; *M. longiloba* var. *chihuahuensis* J.J.Schenk & L.Hufford, Schenk 898, WS, Brewster Co., TX, SAMN46202365; *M. longiloba* var. *longiloba* J.Darl., Schenk 1225, WS, GAS, Imperial Co., CA, SAMN46202366; *M. longiloba* var. *longiloba* J.Darl., *De Groot* 2333, RSA, San Bernardino Co., CA, SAMN46202367; *M. longiloba* var. *longiloba* J.Darl., Hufford 4813, WS, Grand Co., UT, SAMN46202368; *M. longiloba* var. *pinacatensis* J.J.Schenk & L.Hufford, *Dimmit s.n.*, ARIZ, Sonora, Mexico, SAMN46202369; *M. longiloba* var. *yavapaiensis* J.J.Schenk & L.Hufford, Schenk 1011, WS, Yavapai Co., AZ, SAMN46202370; *M. marginata* (Osterh.) H.J.Thomps. & Prigge, Schenk 963, WS, Garfield Co., CO, SAMN46202371; *M. memorabilis* N.H.Holmgren & P.K.Holmgren, Hufford 4151, WS, Mohave Co., AZ, SAMN46202372; *M. mexicana* H.J.Thomps. & Zavort., *Granados & Schenk* 808, MEXU, Coahuila, Mexico, SAMN46202373; *M. multicaulis* (Osterh.) J.Darl., *Holmgren* 14485,

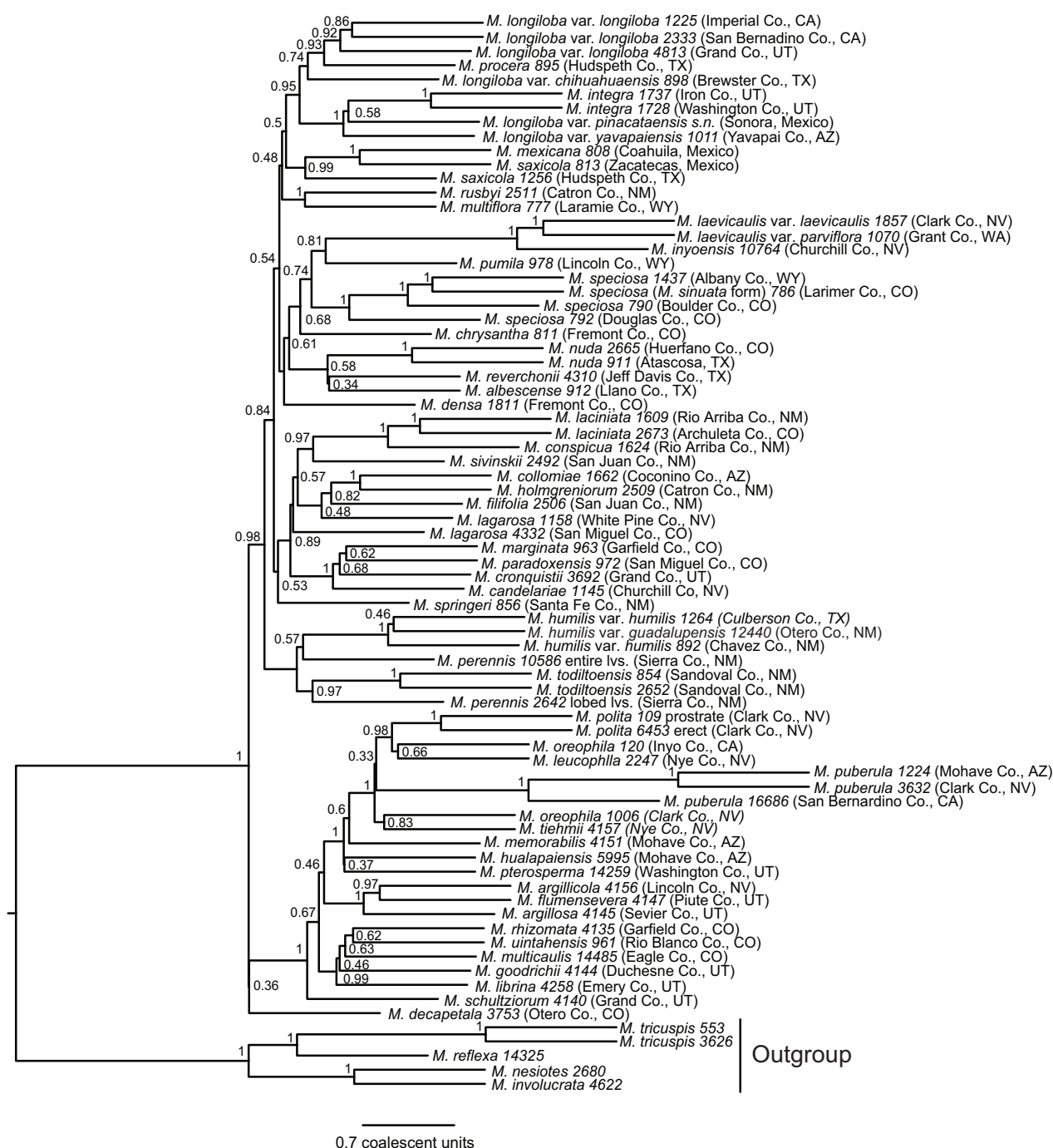


FIG. S1. ASTRAL species tree of *Mentzelia* section *Bartonia* based on the 238 loci not flagged as paralogous by HybPiper. Values at nodes indicate local posterior probability.

WS, Eagle Co., CO, SAMN46202374; *M. multiflora* (Nutt.) A.Gray, Schenk 777, RM, WS, Laramie Co., WY, SAMN46202375; *M. nuda* (Pursh) Torr. & A.Gray, Schenk 911, WS, Atascosa, TX, SAMN46202376; *M. nuda* (Pursh) Torr. & A.Gray, Schenk 2665, BHO, Huerfano Co., CO, SAMN46202377; *M. oreophila* J.Darl., Cohen 120, RSA, Inyo Co., CA, SAMN46202378; *M. oreophila* J.J.Schenk 1006, WS, Clark Co., NV, SAMN46202379; *M. paradoxensis* J.J.Schenk & L.Hufford, Schenk 972, WS, San Miguel Co., CO, SAMN46202380; *M. perennis* (entire leaves) Wootton, Spellenberg 10586, TEX, Sierra Co., NM, SAMN46202381; *M. perennis* (lobed leaves) Wootton, Schenk 2642, BHO, Sierra Co., NM, SAMN46202382; *M. polita* (prostrate) A.Nelson, Cohen 109, RSA, Clark Co., NV, SAMN46202383; *M. polita* (type form, erect) A.Nelson, Niles 6453, WS, Clark Co., NV, SAMN46202384; *M. procera* (Wootton & Standl.) J.J.Schenk & L.Hufford, Schenk 895, WS, Hudspeth Co., TX, SAMN46202385; *M. pterosperma* Eastw., Holmgren 14259, WS, Washington Co., UT, SAMN46202386;

M. puberula J.Darl., Schenk 1224, WS, Mohave Co., AZ, SAMN46202387; *M. puberula* J.Darl., Hufford 3632, WS, Clark Co., NV, SAMN46202388; *M. puberula* J.Darl., Munz 1668A, RSA, WS, San Bernardino Co., CA, SAMN46202389; *M. pumila* Torr. & A.Gray, Schenk 978, WS, Lincoln Co., WY, SAMN46202390; *M. reverchonii* (Urb. & Gilg) H.J.Thomps. & Zavort., Hufford 4310, WS, Jeff Davis Co., TX, SAMN46202391; *M. rhizomata* Reveal, Hufford 4135, WS, Garfield Co., CO, SAMN46202392; *M. rusbyi* Wootton, Schenk 2511, BHO, Catron Co., NM, SAMN46202393; *M. saxicola* H.J.Thomps. & Zavort., Granados & Schenk 813, MEXU, Zacatecas, Mexico, SAMN46202394; *M. saxicola* H.J.Thomps. & Zavort., Schenk 1256, WS, Hudspeth Co., TX, SAMN46202395; *M. shultziorum* Prigge, Hufford 4140, WS, Grand Co., UT, SAMN46202396; *M. sivinskii* J.J.Schenk & L.Hufford, Schenk 2492, GAS, San Juan Co., NM, SAMN46202397; *M. speciosa* (*M. sinuata* form) (Rydb.) R.J.Hill, Schenk 786, WS, Larimer Co., CO, SAMN46202398; *M. speciosa* Osterh., Schenk 790, WS, Boulder Co., CO,



Outgroup: *M. involucrata* S.Wats., De Groot 4622, RSA, Riverside Co., CA, SAMN46202407; *M. nesiotis* (I.M.Johnst.) Brokaw & J.J.Schenk, Hufford 2680, WS, Baja California Sur, Mun. Mulegé, SAMN46202408; *M. reflexa* Coville, Holmgren 14325, NY, Nye Co., NV, SAMN46202409; *M. tricuspis* A.Gray, Hufford 553, WS, Clark Co., NV, SAMN46202410; *M. tricuspis*, Hufford 3626, WS, Clark Co., NV, SAMN46202411.

TABLE S1. Recovery statistics for the Angiosperms353 data of *Mentzelia* section *Bartonia* and outgroup. “Clean Reads” indicates the number of reads after quality control with fastp. “Genes With Sequences” is the number of Angiosperms353 loci recovered by HybPiper, regardless of sequence length. “Genes At 50%” is the number of loci with recovered sequences at least 50% of the targeted locus length. Paralog Warnings (Long) and Paralog Warnings (Depth) are flags by HybPiper. “Paralog Warnings (Long)” are the number of loci in which the specimen contains multiple contigs with sequences $\geq 75\%$ of the reference. “Paralog Warnings (Depth)” are the number of loci in which the set of sequences for a specimen is extracted from contigs with a depth greater than 1 across $\geq 75\%$ of the reference. “Paralogs Detected (HybPhaser)” are the number of loci determined to be paralogous for that specimen, as determined by HybPhaser, based on the specimen having a relatively high proportion of SNPs for that locus (more than 1.5*IQR [interquartile range] above the third quartile of mean). “Allele Divergence” (AD) indicates the proportion of SNPs distributed across all loci, and “Locus Heterozygosity” (LH) indicates the proportion of loci containing SNPs. Both AD and LH are calculated by HybPhaser, with an em-dash (—) indicating specimens for which HybPhaser failed to detect SNPs. All specimens were sequenced on Illumina NovaSeq, except those sequenced on MiSeq (*) and NextSeq (+). *Mentzelia collomiae* is the only specimen to be sequenced on both NovaSeq and MiSeq platforms, therefore raw reads from both sequencing runs were bioinformatically combined before quality control steps.

Specimen	Clean Reads	Genes With Sequences	Genes At 50%	Paralog Warnings (Long)	Paralog Warnings (Depth)	Paralogs Detected (HybPhaser)	Allele Divergence	Locus Heterozygosity
<i>M. albescens</i> 912	2,227,842	348	309	0	3	10	0.048	20.73
<i>M. argillicola</i> 4156	11,616,024	352	336	5	17	33	0.504	87.35
<i>M. argillosa</i> 4145	5,889,742	351	330	3	34	0	—	—
<i>M. candelariae</i> 1145	9,806,526	351	331	3	9	9	0.055	23.28
<i>M. canyonensis</i> 25716	23,206	37	4	0	0	0	—	—
<i>M. chrysantha</i> 811	110,866	196	59	0	0	22	0.502	84.29
<i>M. collomiae</i> 1662*	5,584,678	349	326	4	14	16	0.228	55.99
<i>M. conspicua</i> 1624	501,622	325	188	0	0	10	0.452	69.11
<i>M. cronquistii</i> 3692*	2,144,052	331	251	1	3	15	0.773	88.25
<i>M. decapetala</i> 3753	352,110	305	164	0	0	17	0.746	84.29
<i>M. densa</i> 1811	11,501,858	350	331	4	24	0	—	—
<i>M. filifolia</i> 2506	6,196,804	348	307	1	23	0	—	—
<i>M. flumensevera</i> 4147	2,341,312	351	332	7	12	0	—	—
<i>M. goodrichii</i> 4144	8,792,682	351	334	7	62	0	—	—
<i>M. holmgreniorum</i> 2509	4,114,840	351	330	3	31	0	—	—
<i>M. hualapaiensis</i> 5995	1,762,998	350	324	6	15	24	0.291	67.48
<i>M. humilis</i> var. <i>guadalupensis</i> 12440	10,656,326	351	329	4	9	15	0.804	95.21
<i>M. humilis</i> var. <i>humilis</i> 892	5,190,376	349	317	4	11	29	0.635	89.85
<i>M. humilis</i> var. <i>humilis</i> 1264	5,462,922	350	328	3	14	8	0.843	92.94
<i>M. integra</i> 1728	10,410,086	351	336	10	31	5	0.136	26.88
<i>M. integra</i> 1737*	415,368	265	123	0	0	0	—	—
<i>M. inyoensis</i> 10764	4,396,978	348	317	2	8	10	0.190	37.54
<i>M. laciniata</i> 1609	7,096,118	351	334	1	8	7	0.042	17.96
<i>M. laciniata</i> 2673	2,301,896	350	331	0	5	0	—	—
<i>M. laevicaulis</i> var. <i>laevicaulis</i> 1857	3,652,830	351	329	5	14	11	0.166	59.7
<i>M. laevicaulis</i> var. <i>parviflora</i> 1070	5,687,850	351	336	3	12	17	0.072	34.34
<i>M. lagarosa</i> 1158	7,686,418	351	332	4	14	24	0.537	82.34
<i>M. lagarosa</i> 4332	9,572,544	351	335	16	43	0	—	—
<i>M. leucophylla</i> 2247 ⁺	19,360,964	351	325	8	17	0	—	—
<i>M. librina</i> 4258	5,208,592	346	306	1	16	3	1.051	90.36
<i>M. longiloba</i> var. <i>chihuahuensis</i> 898	3,421,310	351	330	12	34	6	1.808	99.7
<i>M. longiloba</i> var. <i>longiloba</i> 1225	4,029,734	351	333	10	39	5	0.989	74.63
<i>M. longiloba</i> var. <i>longiloba</i> 2333	9,671,538	351	334	15	55	16	0.587	83.28
<i>M. longiloba</i> var. <i>longiloba</i> 4813	7,531,206	351	335	7	25	0	—	—
<i>M. longiloba</i> var. <i>pinacatensis</i> s.n.	7,190,992	348	324	5	36	12	0.892	84.73
<i>M. longiloba</i> var. <i>yavapaiensis</i> 1011	9,873,978	350	335	23	52	0	—	—
<i>M. marginata</i> 963	2,559,502	351	328	2	7	18	0.568	90.45
<i>M. memorabilis</i> 4151	8,964,842	352	333	4	20	29	0.754	91.52
<i>M. mexicana</i> 808 ⁺	15,324,096	348	322	8	20	0	—	—
<i>M. multicaulis</i> 14485	16,964,774	344	324	8	147	5	0.931	84.73
<i>M. multiflora</i> 777 ⁺	26,381,858	350	331	20	37	12	0.373	60.00
<i>M. nuda</i> 911	13,129,672	350	333	5	21	18	0.659	68.07
<i>M. nuda</i> 2665	8,563,402	350	333	4	12	14	0.433	64.11
<i>M. oreophila</i> 120	1,635,446	343	307	3	6	14	0.16	39.19
<i>M. oreophila</i> 1006	13,011,326	350	333	14	43	26	0.325	66.97
<i>M. paradoxensis</i> 972*	872,706	312	203	0	0	0	—	—
<i>M. perennis</i> 2642	2,312,220	349	326	9	34	10	0.502	78.38
<i>M. perennis</i> 10586	2,940,612	349	318	3	13	0	—	—
<i>M. polita</i> (prostrate)	7,313,732	351	334	6	13	18	0.339	63.36
<i>M. polita</i> 6453 (type form; erect)	8,770,608	351	330	4	9	9	1.31	80.93
<i>M. procera</i> 895 ⁺	2,633,162	249	105	2	2	0	—	—
<i>M. pterosperma</i> 14259	3,734,858	351	337	3	10	17	0.483	85.37
<i>M. puberula</i> 1224	1,123,438	344	313	2	4	22	0.512	89.82
<i>M. puberula</i> 3632	1,164,588	348	328	3	7	0	0.038	8.17
<i>M. puberula</i> 16686	2,868,834	350	320	2	6	21	0.372	81.63
<i>M. pumila</i> 978*	1,372,364	321	224	0	1	0	—	—
<i>M. reverchonii</i> 4310	5,388,304	351	330	2	16	25	0.692	91.64
<i>M. rhizomata</i> 4135	7,471,098	351	335	10	27	10	0.050	20.25
<i>M. rusbyi</i> 2511	3,148,478	351	331	1	6	11	1.586	94.63
<i>M. saxicola</i> 813	4,131,008	350	333	1	6	0	—	—

(Continued)

TABLE S1. (CONTINUED)

Specimen	Clean Reads	Genes With Sequences	Genes At 50%	Paralog Warnings (Long)	Paralog Warnings (Depth)	Paralogs Detected (HybPhaser)	Allele Divergence	Locus Heterozygosity
<i>M. saxicola</i> 1256	12,074,082	351	336	29	78	22	0.323	55.69
<i>M. shultziorum</i> 4140	81,074	192	53	0	0	14	0.236	38.55
<i>M. sivinskii</i> 2492	2,553,328	351	323	2	10	6	0.529	57.66
<i>M. speciosa</i> (<i>M. sinuata</i> form) 786	1,979,160	350	326	3	8	9	0.813	81.27
<i>M. speciosa</i> 790	12,004,932	341	325	15	39	15	0.798	93.73
<i>M. speciosa</i> 792	8,897,762	351	333	11	27	4	0.325	58.61
<i>M. speciosa</i> 1437	5,416,580	350	329	2	12	6	0.638	58.15
<i>M. springeri</i> 856*	1,036,230	317	217	1	1	10	0.786	89.12
<i>M. tiehmii</i> 4157	13,071,692	349	331	8	35	11	0.751	89.79
<i>M. todiltoensis</i> 854	4,175,890	351	327	3	29	6	0.225	31.4
<i>M. todiltoensis</i> 2652	4,890,634	350	332	9	29	0	—	—
<i>M. uintahensis</i> 961	9,577,866	349	326	7	16	13	0.126	27.46
Outgroup								
<i>M. involuocrata</i> 4622	15,533,990	351	318	15	44	8	0.081	24.18
<i>M. nesiotis</i> 2680	8,838,620	351	328	3	12	27	0.495	71.74
<i>M. reflexa</i> 14325	7,731,924	351	324	1	11	1	2.210	98.81
<i>M. tricuspis</i> 553	15,188,018	351	325	43	113	7	2.358	97.87
<i>M. tricuspis</i> 3626	1,279,880	344	287	9	23	10	0.048	20.73
Average	6,466,205	337	301	6	21	10	0.436	49.78

TABLE S2. Summary statistics for alignments and gene trees from our three data sets: loci with individual paralogous samples removed by HybPiper (238 Loci), loci with higher-than-average bipartition support (108 Loci), and loci with no samples flagged as paralogous by either HybPiper or HybPhaser (75 Loci). Alignment statistics were calculated with AMAS. Gene tree statistics were calculated with SortaDate (root-to-tip variance, tree length, bipartition support) and IQ-TREE (gene concordance factors [gCF] and site concordance factors [sCF]). Gene tree statistics are represented by the average $\pm$ SD. All data sets included 76 total specimens.

	238 Loci	108 Loci	75 Loci
Alignments			
Average number of taxa per locus	74	69	68
Average locus alignment length (bp)	7383	6397	6421
Concatenated alignment length (bp)	1,476,961	690,863	481,597
Total matrix cells in concatenated alignment	112,249,036	52,505,588	36,601,372
Number of undetermined characters	67,349,374	31,513,332	21,558,423
Proportion of missing characters	60.00	60.02	58.90
Proportion of variable sites	0.23	0.23	0.22
Parsimony informative sites	117,955	55,998	34,444
Proportion of parsimony variable sites	0.08	0.08	0.07
GC content (%)	0.38	0.37	0.37
Gene Trees			
root-to-tip variance	0.002 $\pm$ 0.030	0.0002 $\pm$ 0.001	0.007 $\pm$ 0.055
treelength	0.64 $\pm$ 0.28	0.62 $\pm$ 0.24	0.58 $\pm$ 0.32
Bipartition support	0.15 $\pm$ 0.05	0.19 $\pm$ 0.03	0.16 $\pm$ 0.05
gCF	19.00 $\pm$ 24.10	22.88 $\pm$ 26.89	20.35 $\pm$ 25.19
sCF	43.34 $\pm$ 15.61	40.92 $\pm$ 15.50	43.11 $\pm$ 19.15

TABLE S3. Individual hypothesis testing with HyDe based on the 108-loci data set. *Mentzelia longiloba* is abbreviated as *M. lon.*

Parent 1	Hybrid	Parent 2	Z-score	p-value	γ
<i>M. lagarosa</i> 1158	<i>M. lagarosa</i> 4332	<i>M. cronquistii</i> 3692	-3.428	1.000	-2.565
<i>M. lon.</i> var. <i>longiloba</i> 2333	<i>M. lon.</i> var. <i>longiloba</i> 1225	<i>M. procera</i> 895	16.427	< 0.001	0.502
<i>M. lon.</i> var. <i>longiloba</i> 2333	<i>M. lon.</i> var. <i>longiloba</i> 1225	<i>M. lon.</i> var. <i>yavapaiensis</i> 1011	12.438	< 0.001	0.743
<i>M. lon.</i> var. <i>longiloba</i> 2333	<i>M. lon.</i> var. <i>longiloba</i> 1225	<i>M. lon.</i> var. <i>chihuahuensis</i> 898	-99999.900	1.000	10.597
<i>M. perennis</i> 10586	<i>M. perennis</i> 2642	<i>M. todiltoensis</i>	-2.992	0.999	1.347
<i>M. perennis</i> 2642	<i>M. perennis</i> 10586	<i>M. humilis</i> var. <i>humilis</i> 892	-99999.900	1.000	-0.132
<i>M. saxicola</i> 813	<i>M. saxicola</i> 1256	<i>M. reverchonii</i> 4310	-3.021	0.999	-0.379
<i>M. saxicola</i> 813	<i>M. saxicola</i> 1256	<i>M. mexicana</i> 808	16.473	< 0.001	0.714
<i>M. saxicola</i> 813	<i>M. saxicola</i> 1256	<i>M. lon.</i> var. <i>chihuahuensis</i> 898	29.861	< 0.001	0.468
<i>M. saxicola</i> 813	<i>M. saxicola</i> 1256	<i>M. procera</i> 895	-1.449	0.926	-3.425
<i>M. humilis</i> var. <i>humilis</i> 892	<i>M. humilis</i> var. <i>humilis</i> 1264	<i>M. humilis</i> var. <i>guadalupensis</i> 12440	-8.034	1.000	0.614
<i>M. humilis</i> var. <i>humilis</i> 892	<i>M. humilis</i> var. <i>humilis</i> 1264	<i>M. perennis</i> 2642	-0.141	0.556	1.006
<i>M. humilis</i> var. <i>humilis</i> 892	<i>M. humilis</i> var. <i>humilis</i> 1264	<i>M. perennis</i> 10586	-5.192	1.000	1.267
<i>M. humilis</i> var. <i>humilis</i> 892	<i>M. humilis</i> var. <i>humilis</i> 1264	<i>M. todiltoensis</i>	-3.362	1.000	1.299

Copyright (c) American Society for Plant Taxonomists. All rights reserved. Delivered by Ingenta to IP: 76.34.60.88 on: Tue, 02 Sep 2025 03:23:10