

Phylogenomic analyses using genomes and transcriptomes do not “resolve” relationships among major clades in Phrymaceae

Diego F. Morales-Briones^{1,2}, Nan Lin^{1,3}, Eileen Y. Huang¹, Dena L. Grossenbacher⁴, James M. Sobel⁵,
Caroline D. Gilmore⁵, David C. Tank⁶, Ya Yang^{1,7}

¹ Department of Plant and Microbial Biology, University of Minnesota-Twin Cities, 1445 Gortner
Avenue, St. Paul, MN 55108-1095, USA

² Systematics, Biodiversity and Evolution of Plants, Department of Biology I, Ludwig-Maximilians-
Universität München, Menzinger Strasse 67, 80638, Munich, Germany

³ College of Life Science, Henan Agricultural University, 63 Nongye Road, Zhengzhou, Henan 450002,
China

⁴ Biological Sciences Department, California Polytechnic State University, 1 Grand Avenue, San Luis
Obispo, CA 93407 USA

⁵ Department of Biological Sciences, Binghamton University (State University of New York), 4400
Vestal Parkway E, Binghamton, NY 13902, USA

⁶ Department of Botany & Rocky Mountain Herbarium, University of Wyoming, 1000 E. University
Avenue, Laramie, WY 82071, USA

⁷ Author for correspondence: Ya Yang yangya@umn.edu

Manuscript received _____; revision accepted _____.

Running head: Phylogenomics of Phrymaceae

Abstract

• **Premise of the study:** Phylogenomic datasets using genomes and transcriptomes provide rich opportunities beyond resolving bifurcating phylogenetic relationships. Monkeyflower (Phrymaceae) is a model system for evolutionary ecology. However, it lacks a well-supported phylogeny for a stable taxonomy and for macroevolutionary comparisons.

• **Methods:** We sampled 24 genomes and transcriptomes in Phrymaceae and closely related families, including eight newly sequenced transcriptomes. We reconstructed the phylogeny using IQ-TREE and ASTRAL, evaluated gene tree discordance using PhyParts, Quartet Sampling, and cloudogram, and carried out phylogenetic network analyses using PhyloNet and HyDe. We searched for whole genome duplication (WGD) events using chromosome numbers, synonymous distance, and gene duplication events.

• **Key results:** Most gene trees support the monophyly of Phrymaceae and each of its tribes. Most gene trees also support the tribe Mimuleae being sister to Phrymeae + Diplaceae + Leucocarpeae, with extensive gene tree discordance among the latter three. Despite the discordance, polyphyly of *Mimulus* s.l. is strongly supported, and no particular reticulation event among the Phrymaceae tribes is well supported. Reticulation likely occurred among *Erythranthe bicolor* and close relatives. No ancient WGD event was detected in Phrymaceae. Instead, small-scale duplications are among potential drivers of macroevolutionary diversification of Phrymaceae.

• **Conclusions:** We show that analysis of reticulate evolution is sensitive to taxon sampling and methods used. We also demonstrate that genome-scale data do not always fully “resolve” phylogenetic relationships. They present rich opportunities to investigate reticulate evolution, and gene and genome evolution involved in lineage diversification and adaptation.

49

50 **Key Words:** *Diplacus*, *Erythranthe*, Lamiales; *Mimulus*; monkeyflowers; Phrymaceae; phylogenomics;

51 reticulate evolution; transcriptome; whole genome duplication

52

53

INTRODUCTION

Phylogenomic analyses using genomes and transcriptomes often focus on “resolving” relationships that previously show conflict among markers or low statistical support using Sanger sequencing. With thousands of genes, these datasets are rich in insights beyond providing a bifurcating phylogenetic relationship, pointing to reticulate evolution, gene/genome duplications, and adaptation. However, detecting these events are difficult due to computational limitations, and studies often do not fully interrogate the data.

Monkeyflowers, as part of Phrymaceae, are a model system for evolutionary ecology (Wu et al., 2008; Twyford et al., 2015). With ca. 200 species, a primarily North American distribution, a rich history of ecological studies, and accumulating genomic resources, monkeyflower research has provided insights in speciation (Schemske and Bradshaw, 1999; Sobel, 2014), local adaptation (MacNair, 1983; Hall et al., 2010), pigment evolution (Streisfeld et al., 2013; Yuan et al., 2016; Ding et al., 2020), and development (Yuan, 2019). However, previous studies often focused on a single species or clades of closely related species (Stankowski and Streisfeld, 2015; Chase et al., 2017; Stankowski et al., 2019; Nelson et al., 2021), and we still lack a robust phylogenetic framework for the family. In addition, with phylogenetic uncertainty across Phrymaceae from previous analyses, the circumscription of *Mimulus* L. is under debate (Lowry et al., 2019; Nesom et al., 2019). Previous phylogenetic studies have established the polyphyly of *Mimulus* in its broad sense (Barker et al., 2012), and therefore we follow the narrow definition of *Mimulus* sensu Barker et al. (2012) that includes only seven species as part of the tribe Mimuleae, with the remaining species distributed into the genera *Diplacus* (tribe Diplaceae) and *Erythranthe* (tribe Leucocarpeae).

Previous molecular phylogenetic studies using Sanger sequencing have consistently supported five well-supported tribes within Phrymaceae [Fig. 1; the tropical Asian tribe Cyrtandromoeae was only included in the sampling by Liu et al. (2020)]. However, phylogenetic relationships in Phrymaceae have been problematic due to discordance among analyses (Fig. 1) that used 1) different taxon sampling; 2)

nuclear vs. plastome (cpDNA) markers, or even among cpDNA regions [*trnL-F* only (Beardsley and Olmstead, 2002) vs. six cpDNA regions (Liu et al., 2020)]; and 3) different analytical approaches (maximum likelihood vs. Bayesian, ITS + ETS, Liu et al., 2020). In addition, despite previous and ongoing whole genome and exome sequencing efforts (Hellsten et al., 2013; Edger et al., 2017; Nelson et al., 2021), there yet lacks multi-locus analyses using nuclear genes across major clades of the family.

Macroevolutionary analyses using a large number of nuclear genes are powerful not only for inferring the phylogenetic relationships and history of reticulate evolution, but also investigating gene and genome evolution associated with lineage diversification and adaptation. Previous investigation of patterns of chromosome number changes across the North American members of Phrymaceae suggested extensive polyploidy events across the family (Beardsley et al., 2004). However, comparison of linkage maps established that the higher chromosome base number in *Erythranthe lewisii* compared to *E. guttata* is due to chromosome fission and fusion instead of whole genome duplication (WGD; Fishman et al., 2014). To what extent WGD events are distributed, or the lack of, along the backbone of the family is still unexplored. Sampling of genomes and transcriptomes across major clades of Phrymaceae and analyzing both nuclear and cpDNA regions are needed to investigate the backbone structure in Phrymaceae and the genomic basis of the macroevolutionary diversification of Phrymaceae.

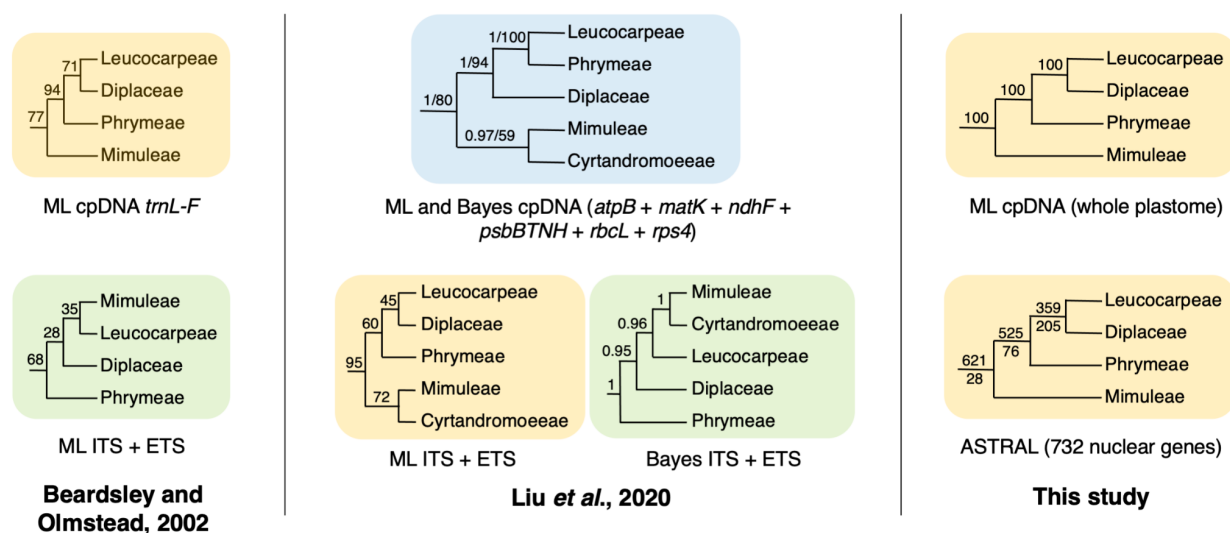


Figure 1. Summary of the topologies among the five tribes of Phrymaceae among studies focusing on the backbone relationship of the family. Numbers above branches are maximum likelihood (ML) bootstrap support (BS), Bayesian posterior probabilities (PP), or BS/PP. For the ASTRAL analysis, numbers above and below branches are the number of genes with concordant vs. discordant topologies that had MLBS >50 in each individual gene tree. Trees shaded by the same color share compatible backbone topologies.

In this study we sampled transcriptomes and genomes covering four of the five tribes in Phrymaceae, including eight newly generated transcriptomes, to 1) provide a phylogenetic backbone and examine gene tree discordance and 2) investigate patterns of gene and genome duplication in the family. We found that most gene trees support Mimuleae being sister to Phrymeae+Diplaceae+Leucocarpeae, with the relationship among the latter three showing extensive gene tree discordance. However, no particular reticulation event among the Phrymaceae tribes was strongly supported. Instead, we found evidence for reticulation events among closely related species contributing to *Erythranthe bicolor*. Our analyses did not support any ancient WGD events in Phrymaceae; instead, small-scale gene duplications in defense, stress response, growth and development, and certain biochemistry pathways are candidates for potential drivers underlie macroevolutionary diversification of Phrymaceae.

MATERIALS AND METHODS

Taxon sampling—Transcriptomes were newly generated from eight accessions representing seven ingroup Phrymaceae species for this study (Appendix S1). Seeds were collected from natural populations and were cold-treated in soil for a week at 4°C in the dark before growing with 15-hour daylight in a greenhouse. Young leaves and flower buds were flash frozen in liquid nitrogen. RNA extraction, library preparation (rRNA removal or poly-A enrichment), and sequencing procedures are detailed in Appendix S1. In addition, we included three genomes and five transcriptomes from Phrymaceae that are publicly available (Appendix S2). Together our taxon sampling included 16 datasets representing 15 Phrymaceae species in four of the five tribes (missing the tropical Asian tribe Cyrtandromoeae) and five of the 14

genera (Barker et al., 2012; Liu et al., 2020). We also included four genomes and four transcriptomes in closely related Lamiales families (Zhang et al., 2020).

Data processing for nuclear genes — Read processing, assembly, and translation were carried out following Morales-Briones et al. (2021). Homology inference started with an all-by-all BLASTN search on coding sequences (CDS) with an *E* value cutoff of 10. Hits were filtered with a minimal hit coverage of 40%. Homolog groups were clustered using MCL v. 14-137 (van Dongen, 2000) with a minimum minus log-transformed *E* value cutoff of 5 and an inflation value of 1.4. Only clusters with at least 20 out of 24 taxa were retained. Sequences from each cluster were aligned using the OMM_MACSE pipeline v. 10.02 (Scornavacca et al., 2019), which pre-filters non-homologous sequence fragments with HMMCleaner (Di Franco et al., 2019) before translation accounting for frameshifts using MACSE v. 2.03 (Ranwez et al., 2018). The resulting CDS alignments were trimmed to remove columns with more than 90% missing data using Phyx (Brown et al., 2017). Homolog trees were built with RAxML v. 8.2.11 (Stamatakis, 2014) using the GTRCAT model and 200 rapid bootstrap (BS) replicates. Monophyletic and paraphyletic tips of the same dataset were removed, keeping the tip with the highest number of characters in the trimmed alignment. Spurious tips were detected and removed with TreeShrink v. 1.3.2 (Mai and Mirarab, 2018) with the 'per-gene' mode and a false positive error rate threshold (α) of 0.001. The resulting trees were visually inspected, and deep paralogs producing internal branch lengths longer than 0.25 were cut apart retaining subclades with a minimum of 20 taxa to obtain final homolog trees.

Orthology inference was carried out using the 'monophyletic outgroup' approach from Yang and Smith (2014). The approach filters unrooted homolog trees, requiring outgroups to be monophyletic and single-copy. It then traverses the ingroup clade rooted by outgroups from root to tip and trims the side with less taxa each time a gene duplication event was detected, until every taxon is represented by a single sequence. We set the three Lamiaceae genomes as outgroups (Zhang et al., 2020), keeping ortholog groups that included at least 15 taxa.

Species tree inference and evaluation of support — Sequences from individual ortholog groups

were aligned using OMM_MACSE. Columns with more than 20% missing data were trimmed with Phyx, and only alignments with at least 1000 characters and all 24 taxa were retained. We first estimated a maximum likelihood (ML) tree of the concatenated matrix with IQ-TREE v. 2.1.13 (Minh et al., 2020) searching for the best partition scheme (Lanfear et al., 2012), followed by ML tree inference and 1000 ultrafast bootstrap replicates. To estimate a coalescent-based species tree, we first inferred individual gene trees with IQ-TREE using extended model selection (Kalyaanamoorthy et al., 2017) and 200 non-parametric bootstrap replicates for clade support. Gene trees were then used to infer a species tree with ASTRAL-III v. 5.6.3 (Zhang et al., 2018) using local posterior probabilities (LPP; Sayyari and Mirarab, 2016) to assess clade support.

To explore discordance among gene trees, we calculated the number of concordant and discordant bipartitions on each node of the species tree using PhyParts (Smith et al., 2015). We mapped bipartitions from gene trees with local BS support of at least 50% against the IQ-TREE tree from the concatenated supermatrix (identical to the ASTRAL tree; see results). Next, to distinguish conflict from poorly supported branches, we carried out a Quartet Sampling (QS; Pease et al., 2018) analysis using the concatenated supermatrix, the IQ-TREE tree, and 1000 replicates. Lastly, to visualize gene tree conflict, we built a cloudogram using the DensiTree function of phangorn v. 2.7.1 (Schliep et al., 2017). Individual ortholog gene trees were time-calibrated for visualization purposes with TreePL v. 1.0 (Smith and O'Meara, 2012). The root was fixed to 70.3 MYA, the most recent common ancestor (MRCA) of Lamiaceae was fixed to 57.69 MYA, and the MRCA of all remaining species was fixed to 68.8 MYA based on Zhang et al. (2020).

Plastome assembly and tree inference—We obtained nine reference plastomes from RefSeq

(Appendix S3). For the remaining species, we carried out plastome assemblies from either transcriptomic or genomic reads (Appendix S3) with Fast-Plast v. 1.2.8 (McKain and Wilson, n.d.). In four cases, those assemblies resulted in low plastome coverage and were redone using alternative transcriptomic or

genomic libraries (Appendix S3). When the resulting plastome were incomplete (7 out of 14 accessions), filtered contigs from Spades v. 3.9.0 (Bankevich et al., 2012) were mapped to the closest available reference plastome using Geneious v. 11.1.5 (Kearse et al., 2012) to produce oriented and contiguous contigs with missing regions masked with ‘N’. The assembly of *Striga asiatica* had many contigs that were poorly mapped even to congeneric plastomes, likely due to major structural rearrangements in this hemiparasitic species of Orobanchaceae (Frailey et al., 2018). We replaced it with the published plastome of *Striga forbesii* (Appendix S3) for downstream analyses.

The resulting plastomes with one Inverted Repeat removed were aligned with MAFFT and columns with more than 50% missing data were trimmed with Phyx. An ML tree was inferred with IQ-TREE with automated extended model selection and 1000 rapid BS replicates. Additionally, we used QS with 1,000 replicates to evaluate branch support.

Testing potential reticulation at nodes with elevated levels of gene tree conflict — We investigated two areas with elevated gene tree conflict: 1) the backbone of Phrymaceae using one species for each well-supported clade corresponding to a tribe; and 2) among *Erythranthe cardinalis*, *E. lewisii*, and *E. bicolor*.

For each of the two areas, we first ran PhyParts using all taxa from the reduced dataset. We then removed one taxon at a time to determine which taxon produced the highest gene tree conflict. We inferred species networks using ML (Yu et al., 2014) in PhyloNet v. 3.6.9. (Than et al., 2008) with the command “InferNetworks_ML” from individual ML gene trees. Network searches were performed allowing for up to three reticulation events, and optimizing the branch lengths and inheritance probabilities of the inferred species networks. To estimate the optimal number of reticulations and to test whether a species network fits our gene trees better than a strictly bifurcating tree, we computed the likelihood scores of the nuclear and plastid trees given the individual gene trees using the command ‘CalGTProb’ (Yu et al., 2012). We performed model selection using the Akaike information criterion (AIC; Akaike, 1973), bias-corrected AIC (AICc; Sugiura, 1978) and the Bayesian information criterion

(BIC; Schwarz, 1978). Next, we performed a more thorough Phylonet analysis using a Bayesian inference of species networks approach (Wen et al., 2016) with the command “MCMC_GT”, full likelihood, and allowing up to three reticulations events. Analyses consisted of four independent runs with four reversible-jump Markov chain Monte Carlo (RJMCMC) chains, temperatures set to one cold and two hot chains (1.0, 2.0 and 3.0 respectively), 30 million generations, sampling every 1000 generation, and a burn-in of 500,000 generations. The four MCMC runs were summarized with the command “MCMC_GT-sum”. Convergence was assessed once the posterior sampling reached ESS (effective sample size) ≥ 200 .

In addition to PhyloNet, we also tested for hybridization with HyDe (Blischak et al., 2018) that uses site pattern frequencies (Kubatko and Chifman, 2019) to quantify admixture (γ) between two parental lineages that form a hybrid lineage. We tested all triplet combinations in all directions using ‘run_hyde.py’, the concatenated nuclear alignment, and a mapping file to assign individuals to species. Test significance was assessed with a Bonferroni correction ($\alpha = 0.05$) for the number of tests conducted with estimates of γ between 0 and 1 (Blischak et al., 2018).

Gene and whole genome duplication events — We employed three approaches (Yang et al., 2018) to detect WGD events in Phrymaceae. 1) We summarized chromosome counts from Nesom (2012) and the Chromosome Counts Database (Rice et al., 2015). 2) We mapped gene duplication events onto the inferred species tree by extracting rooted ingroup clades from the final homolog trees by requiring an average BS ≥ 50 and at least 15 taxa. Gene duplication events were then mapped onto the MRCA on the species tree when two or more taxa overlapped between the two daughter clades on the rooted ingroup clade (“extract_clades.py” and “map_dups_mrca.py” from <https://bitbucket.org/blackrim/clustering>). 3) We analyzed the distribution of synonymous distances (Ks) from RNA-seq (<https://bitbucket.org/blackrim/clustering>; “ks_plots.py”). Ks peaks were identified using a mixture model in mixtools v. 1.2.0 (Benaglia et al., 2009).

To explore genes with elevated instances of gene duplication within Phrymaceae, we extracted Phrymaceae clades from the final homologs. We then obtained functional annotation for the top 10

Phrymaceae clades with the highest number of genes using the *Erythranthe guttata* genome annotation (Hellsten et al., 2013).

RESULTS

Sequence processing — Organellar reads represented 30–57% of quality-filtered read pairs in RNA-seq libraries prepared using RiboZero, compared to 0.13–0.3% in libraries prepared using Poly-A enrichment (Appendix S1). Of the eight newly generated transcriptomes, we retained 13.7–21.4 million nuclear read pairs after quality filtering and separating organellar reads; each CDS set represented 49–62% of nuclear genes when compared against the *Erythranthe guttata* reference genome (Appendix S2). Although libraries prepared with RiboZero produced lower numbers of nuclear reads, they produced more contiguous assemblies and some of the highest numbers of genes in our final nuclear ortholog set (Appendix S2). This can be due to ribosomal depletion resulting in more even coverage in slightly degraded transcripts compared to poly-A enrichment.

Assemblies from each of the six genomic libraries produced full plastomes in a single contig (Appendix S3). Despite the large numbers of plastid reads from the six RiboZero RNA libraries, only one complete plastome was assembled due to uneven coverage. Still, RiboZero RNA libraries produced contiguous contigs that covered most of the plastomes and recovered similar numbers of CDS compared to the full plastomes (Appendix S3).

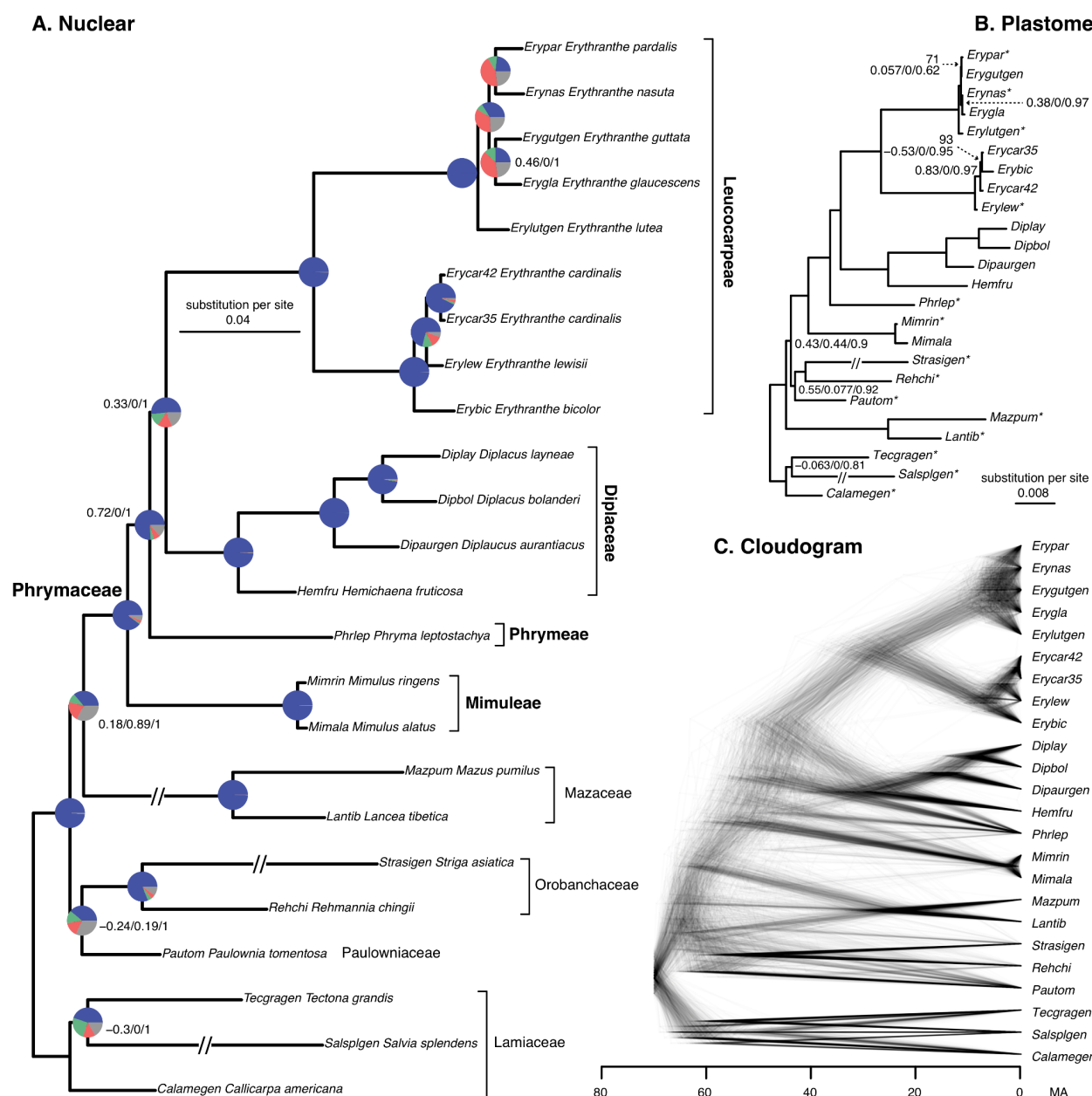


Figure 2. Maximum likelihood phylogeny of Phrymaceae inferred with IQ-TREE from the concatenated 732-nuclear gene supermatrix. Quartet Sampling (QS) scores are shown below branches. QS scores: Quartet concordance/Quartet differential/Quartet informativeness. Nodes with maximum QS support (1/—/1) not shown. All nodes have full bootstrap support (BS =100) and local posterior probability (LLP =1). See Appendix S4. Pie charts represent the proportion of ortholog trees that support that clade (blue), the main alternative bifurcation (green), the remaining alternatives (red), and the remaining with (conflict or support) <50% bootstrap support (gray). Branch lengths are in number of substitutions per site (scale bar). Exceptionally long branches are shortened with a broken

segment (//) for illustration purposes (See Appendix S5 for original branch lengths); b) Maximum likelihood phylogeny inferred with IQ-TREE from plastomes. Bootstrap support shown above branches and QS scores below the branches. Full BS and QS support values are not shown. Branch lengths are in number of substitutions per site (scale bar). Longest branches are shortened with a broken segment (//) for illustration purposes (See Appendix S5 for original branch lengths); c) Cloudogram inferred from 732 nuclear ortholog trees. Scale in millions of years ago (MA).

Orthology inference and phylogenetic analysis — The final set of nuclear orthologs included 732 genes, and the concatenated matrix consisted of 1,246,075 aligned columns with a character occupancy of 93.1% (Appendix S2). The topologies from the IQ-TREE and ASTRAL trees were identical and all nodes had maximum support (BS = 100, LPP = 100; Figs. 1, 2A, & Appendix S4). The monophyly of Phrymaceae and each tribe of Phrymaceae was strongly supported by most informative gene trees and full QS support (1/-/1; i.e., all sampled quartets supported that branch; blue in Fig. 2A and Appendix S5). The sister relationship of Leucocarpeae and Diplaceae was supported by 359/564 informative gene trees and strong Quartet Concordance (QC = 0.3), but the Quartet Differential (QD = 0) indicates the presence of a single alternative topology (Diplaceae sister to Phrymeae). Similarly, the placement of Phrymeae was supported by 525/601 informative gene trees, strong QC (0.72) and signal of a single potential alternative topology (QD = 0; Mimulaceae sister to Phrymeae). The cloudogram (Fig. 2C) showed discordance in the backbone of Phrymaceae, especially on the placement of Phrymeae (“Phrlep”), consistent with the concordance analyses and QS results.

The final cpDNA alignment included 128,056 characters with a character occupancy of 89%. The plastome topology recovered the monophyly of Phrymaceae and each of its tribes with maximum support (BS = 100, QS 1/-/1; Fig. 2B) and the backbone relationships were identical to the nuclear results (Fig. 2A, Appendix S5). However, relationships among closely related ingroup taxa differ in two places (Fig. 2, A vs. B): 1) The two *Erythranthe cardinalis* accessions were sister to each other in the nuclear tree but were paraphyletic with *E. bicolor* nested among them in the cpDNA tree; 2) relationships among

Erythranthe pardalis, *E. nasuta*, *E. guttata*, and *E. glaucescens* showed extensive gene tree conflict among nuclear genes, low quartet concordance, and dominant secondary topologies in the cpDNA tree, and conflicting topology between cpDNA and nuclear trees. In addition, extensive nuclear gene tree conflict and discordance between nuclear and cpDNA trees are present among other Lamiales families sampled.

Phylogenetic network analyses — Phylogenetic network analyses focused on two ingroup areas with elevated levels of conflict. To investigate the backbone of Phrymaceae, we used one taxon to represent each tribe. The cloudogram (Fig. 2C) showed *Phryma leptostachya* (Phrymeae) shifted its placement among other Phrymaceae tribes. When removing one tip at a time, conflict among nuclear gene trees (red and green in Fig. 3A) reduced the most when removing *Phryma*, followed by removing *Erythranthe guttata* (Leucocarpeae). Visual inspection of individual gene trees and alignments confirmed that *Phryma*'s placement shifted among gene trees with short internal branches attached to the backbone of Phrymaceae. PhyloNet ML searches (Appendix S6) recovered three networks with small amounts of gene flow towards *Mimulus ringens*. Model selection (Appendix S7) using AIC and AICc both preferred three reticulations while BIC did not support significant differences among the three networks. MPP from the MCMC PhyloNet searches (Fig. 3A) recovered the same network as the 1-reticulation network from ML, and estimated that 9.17% of *M. ringens* genes had contribution from *E. guttata*, with the 95% credible set consisted of a single network. HyDe (Appendix S8) results indicated *E. guttata* and *Diplacus aurantiacus* were each derived from reticulation between the stem lineage of Mimuleae and Phrymeae, and *E. guttata* from reticulation between Mimuleae and Diplaceae. As almost all informative gene trees supported the monophyly of each Phrymaceae tribe, any potential hybridization events would have occurred among stem branches of tribes. Given the disagreement among PhyloNet and HyDe analyses, none of the putative hybridization events were well supported. Phylogenetic uncertainty, ILS, and population structure likely contributed to the high levels of gene tree discordance among Phrymaceae tribes.

The second instance of conflict we focused on was among *Erythranthe bicolor*, *E. cardinalis*, and *E. lewisii*. Consistent with the cloudogram (Fig. 2C) where some gene trees supported *E. bicolor* being sister to *E. lewisii*, removing either *E. bicolor* or *E. lewisii* (Fig. 3B) removed most of the gene tree conflicts. PhyloNet ML analysis (Appendix S6) recovered networks with 38–48% of *E. bicolor* genes from *E. lewisii* or its close relatives. AIC or AICc did not prefer any network, while BIC preferred the 1-reticulation network (Appendix S7). PhyloNet MCMC searches recovered a 95% credible set of three 1-reticulation networks with gene flow towards *E. bicolor*, but the source of gene flow varied among networks (Appendix S9). The MPP (54% of the credible set) showed that *E. bicolor* had 26% genes from *E. lewisii*, similar to the 1-reticulation ML network (Fig. 3B). HyDe, on the other hand, did not recover any significant hybridization events.

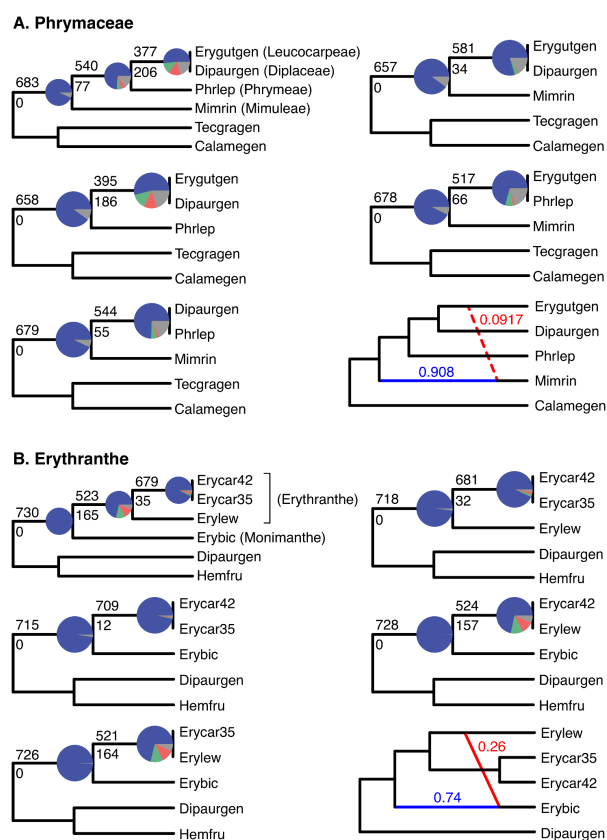


Figure 3. Conflict and network analyses using reduced datasets. Cladograms showing relationships among the reduced dataset, removing one tip at a time, and Maximum posterior probability (MPP) network from the PhyloNet

Bayesian inference. A) Phrymaceae backbone with one representative species for each tribe. B) *Erythranthe cardinalis*, *E. lewisii*, and *E. bicolor*. Pie charts on cladograms represent the proportion of gene trees that support that clade (blue), the main alternative bipartition (green), the remaining alternatives (red), and conflict or support with <50% bootstrap support (gray). Number above and below branches represent the number of concordant and discordant gene trees, respectively. Red and blue branches in networks indicate the minor and major edges respectively, of hybrid nodes, with the inheritance probabilities next to each branch. Dotted edge indicates uncertainty in inference (see Results).

Although we also detected extensive gene tree conflict among *Erythranthe pardalis*, *E. nasuta*, *E. guttata*, and *E. glaucescens*, as branches among them were short and we lacked any intraspecific sampling, we did not carry out additional analyses on reticulate evolution.

Gene and Genome Duplications — Mapping gene duplication events did not reveal any node with more than 4.3% of gene duplications in Phrymaceae (Appendix S10). Similarly, the Ks plots (Appendix S11) did not support any Phrymaceae-specific WGD that involved more than one species sampled. All 24 datasets included in this study shared two optimal mixing components (i.e. Ks peaks). The first component had a Ks mean of 1.8–2.2, corresponding to a whole-genome triplication event early in the core eudicots (Jiao et al., 2012). The second component had Ks means of 0.3–0.9 (lower in woody species and higher in herbaceous species), corresponding to a WGD at the MRCA of the core Lamiales (Zhang et al., 2020). A third component at Ks~0.1 was found only in *Erythranthe lutea*, corresponding to a previously reported WGD event (Edger et al., 2018). In addition, *Diplacus layneae* and *Erythranthe guttata* each showed a putative Ks peak at ~0.04 and 0.09 respectively. However, chromosome counts (Appendix S10) did not support a WGD in either species. Overall, all sampled Phrymaceae species except *Erythranthe lutea* had low chromosome counts compared to outgroups, and the uptick in Ks density below 0.1 was likely due to recent small-scale duplications. Among non-Phrymaceae species, *Salvia splendens* showed a Ks peak ~0.1, consistent with a WGD in *Salvia* and relatives within Lamiaceae

(Godden et al., 2019). Both *Mazus pumilus* and *Lancea tibetica* showed a Ks peak ~0.07, which could be due to WGD or small-scale duplications.

The ten Phrymaceae genes (Appendix S12) with the highest numbers of copies were involved in defense/immune response (Serine protease inhibitor, aspartyl protease, MLP-like protein), stress response (HSP20-like chaperones, Ribosomal protein L10 family protein), mitochondria organization (prohibitin 2), regulating plant growth (small auxin up-regulated RNA-like auxin-responsive protein family), cell wall architecture (Glycosyl hydrolase), and various other biochemical processes (S-adenosyl-L-methionine-dependent methyltransferases, hydroxymethyltransferase 4). Since we reduced sequences from the same sample that formed monophyletic or paraphyletic groups, we ignored isoforms from alternative splicing, assembly artifacts, and recent copy number increase involving only a single sample, as these are difficult to quantify using *de novo* assembled transcriptomes. Therefore, only gene duplication events involving more than two taxa in our sampling contributed to our copy number counts. Given our much denser taxon sampling in Leucocarpeae and Diplaceae, the top ten are heavily influenced by genes that had multiple rounds of gene duplications in these two tribes.

DISCUSSION

Extensive gene tree discordance and potential hybridization events in Phrymaceae — Our phylogenomic analyses recovered strong support of the monophyly of Phrymaceae and each of its tribes sampled. We also recovered extensive and well-supported gene tree discordance along the backbone of Phrymaceae. The discordance is not an artifact of gene and genome duplications; nor is any particular reticulation event well supported by phylogenetic network analyses and hypothesis testing. Therefore, phylogenetic uncertainty, ILS, and population structure likely contributed to the extensive gene tree discordance among Phrymaceae tribes.

Among closely related species, our phylogenetic network analyses support introgression from *E. lewisii* or close relatives towards *E. bicolor*. In addition, our plastome analysis (Fig. 2B) recovered *E.*

bicolor being nested among accessions of *E. cardinalis*, suggesting that *E. cardinalis* is also involved in the reticulation. However, without sampling of other closely related species or additional within-species sampling, the timing, source, and prevalence of the introgression is unclear. Nelson et al. (2021) analyzed over 8,000 nuclear gene trees and recovered extensive reticulation among *E. lewisii*, *E. cardinalis*, and *E. parishii* (not sampled in our study), with *E. bicolor* set as the outgroup. Our analyses suggest that additional clades are involved in reticulations involving *E. lewisii*, *E. cardinalis*, and/or other close relatives.

In summary, our phylogenetic analyses suggest that: 1) network inferences are sensitive to the methods used, sources of data, and taxon sampling, including the choice of both ingroups and outgroups. 2) Despite phylogenetic uncertainty along the Phrymaceae backbone, most gene trees support Mimuleae (likely together with the unsampled Cyrtandromoeae) being sister to a strongly supported clade of Phrymeae + Diplaceae + Leucocarpeae, consistent with previous Sanger-based studies (summarized in Barker et al. 2012). 3) Our results strongly support the polyphyly of the monkeyflower genus *Mimulus* s.l. (= part of Mimuleae + part of Leucocarpeae + part of Diplaceae).

Genomic drivers of macroevolution in Phrymaceae — Analyses of chromosome counts, gene tree mapping, and Ks plots did not find any evidence for ancient WGD in Phrymaceae that involves more than one species in our taxon sampling. Our results are consistent with the previous comparisons using linkage maps between *E. lewisii* and *E. guttata* (Fishman et al., 2014) and whole genome sequences *E. guttata* and *E. lutea* (Edger et al., 2018), both primarily focused on species in the tribe Leucocarpeae. Our analysis broadened the genome-wide sampling to four of the five tribes in Phrymaceae and found that ancient WGD is not a driving force in macroevolution of Phrymaceae. Instead, reticulate evolution, small-scale gene duplication in genes involved in defense, stress response, growth and development, and additional biochemical pathways are some of the potential drivers of macroevolutionary diversification in Phrymaceae. Our study provides initial insights into the gene space of species across Phrymaceae and

potential genomic drivers of macroevolution in the family. In addition, our newly generated transcriptome datasets using RiboMinus provide data for future studies looking into non-coding RNAs.

CONCLUSIONS

Our phylogenomic analysis evaluated the support (or the lack of) in the backbone of Phrymaceae, confirmed the polyphyly of *Mimulus* s.l., and detected an area of reticulate evolution among closely related species. We show that analysis of reticulate evolution is sensitive to taxon sampling and methods used. We also show a lack of ancient WGD events in Phrymaceae; instead, small-scale gene duplications are potential drivers underlie macroevolutionary diversification of Phrymaceae.

Our analyses demonstrate that genome-scale data do not always “resolving” phylogenetic relationships. Instead, they provide resolution for some areas, but also recover “clouds” and “networks” that point to future opportunities for investigating their significance in adaptation and lineage diversification.

Acknowledgments

The authors thank Rahul Roy for assistance with greenhouse work, and for Aaron Lee and Benjamin Cooper for commenting on the draft. Funding was provided by the University of Minnesota to DFM-B, YH, and YY; China Scholarship Council to NL (CSC: 201904910676), National Science Foundation grant to JMS (DEB-1856158), and the Frost Fund by the California Polytechnic State University to DLG.

Author Contributions

YY and DLG designed the study; YH, DLG, JMS, CDG, and YY generated the data; DFM-B, NL, and YY analyzed data and drafted the manuscript. All authors edited the manuscript and approved the final version.

420

421 Data Availability Statement

422 Raw reads of newly sequenced transcriptomes were deposited in the NCBI Sequence Reads Archive
423 (BioProject: PRJNA770153). Analysis files are available from Dryad (to be deposited).

424

425 Supporting Information

426 Additional supporting information may be found online in the Supporting Information section at the end
427 of the article.

428

429 **Appendix S1.** Collection, plant growth, and sequencing information for the eight newly generated
430 transcriptomes.

431

432 **Appendix S2.** Taxon sampling, source of data, and nuclear matrix statistics. Naming authorities above
433 species level (Stevens, n.d.): 1) Order: Lamiales Bromhead. 2) Lamiales families: Phrymaceae Schauer,
434 Orobanchaceae Ventenat, Mazaceae Reveal, Paulowniaceae Nakai, Orobanchaceae Ventenat. 3)
435 Phrymaceae tribes: Diplaceae Bo Li, B. Liu, S. Liu & Y. H. Tan; Phrymeae Hogg; Leucocarpeae
436 Conzatti; Mimuleae Dumortier; Cyrtandromoeae Bo Li, B. Liu, S. Liu & Y. H. Tan. 4) Phrymaceae
437 genera: *Diplacus* Nuttall, *Hemichaena* Benthham, *Erythranthe* Spach, *Mimulus* L., *Phryma* L.

438

439 **Appendix S3.** Plastome assembly and sources information.

440

441 **Appendix S4.** A. Maximum likelihood phylogeny of Phrymaceae inferred with IQ-TREE from the
442 concatenated 732-nuclear gene supermatrix. Numbers above branches represent bootstrap support (BS).

443 Branch lengths are in number of substitutions per site (scale bar on the bottom). B. ASTRAL-Pro tree of

Phrymaceae inferred from the 732 nuclear gene trees. Local posterior probabilities (LLP) are shown next to nodes. Internal branch lengths are in coalescent units (scale bar on the bottom). C. Maximum likelihood phylogeny of Phrymaceae inferred with IQ-TREE from plastomes. BS values are shown above branches. Branch lengths are in number of substitutions per site (scale bar on the bottom)

Appendix S5. Maximum likelihood cladogram of Phrymaceae inferred with IQ-TREE from the concatenated 732-nuclear gene supermatrix. Pie charts represent the proportion of gene trees that support that clade (blue), the main alternative bifurcation (green), the remaining alternatives (red), and conflict or support that have <50% bootstrap support (gray). Number above and below branches represent the number of concordant and discordant informative gene trees, respectively.

Appendix S6. Species network inferred from PhyloNet maximum likelihood analyses with one to three maximum reticulations of the reduced data sets. A) Phrymaceae backbone. B) *Erythranthe cardinalis*, *E. lewisii*, and *E. bicolor*. Red and blue branches indicate the minor and major edges, respectively, of hybrid nodes. Numbers next to curved branches indicate inheritance probabilities for each hybrid node.

Appendix S7. Model testing between trees and PhyloNet networks for the reduced data sets of Phrymaceae and *Erythranthe*. The number of parameters were set to equal the number of branch lengths plus the number of inheritance probabilities. The number of gene trees was used to correct for finite sample size.

Appendix S8. HyDe tests for hybridization events along the backbone of Phrymaceae.

Appendix S9. Phylonet Bayesian inference 95% credibility set for the reduced data set of *Erythranthe cardinalis*, *E. lewisii*, and *E. bicolor*. Red and blue branches indicate the minor and major edges,

respectively, of hybrid nodes. Numbers next to curved branches indicate inheritance probabilities for each hybrid node.

Appendix S10. Maximum likelihood cladogram of Phrymaceae inferred with IQ-TREE from the concatenated 732-nuclear gene supermatrix. Numbers above branches are gene duplication counts and numbers below branches are gene duplication percentages. Numbers next to species names are gametophytic chromosome counts. All chromosome counts are from the Chromosome Counts Database (Rice et al., 2015), except *Erythranthe pardalis* (Nesom, 2012). When multiple independent counts gave a single consistent chromosome number but different counts were each reported by a single study, we ignored the outlier numbers. Inset: Histogram of percentages of gene duplication per branch.

Appendix S11. Distribution of synonymous distance among gene pairs (Ks plots) for each genome or transcriptome. A) Plots of raw Ks distances between 0 and 3. B) Plots of Ks distances zooming in to values between 0 and 0.5. C) Plots of log-transformed Ks distances. Colored lines indicate components inferred using a mixture model. Blue lines indicate a component from an ancestral whole genome triplication event early in core eudicots; red lines are from more recent whole genome or small-scale duplication events.

Appendix S12. Phrymaceae clades extracted from the final homologs with the highest number of copies.

Literature Cited

- Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N. and Csaki, F., Eds., International Symposium on Information Theory, 267–281.
- Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.

494 *Journal of Computational Biology* 19: 455–477.

495 Barker, W., G. Nesom, P. Beardsley, and N. Fraga. 2012. A taxonomic conspectus of Phrymaceae: a
496 narrowed circumscription for *Mimulus*, new and resurrected genera, and new names and
497 combinations. *Phytoneuron* 39.

498 Beardsley, P. M., and R. G. Olmstead. 2002. Redefining Phrymaceae: the placement of *Mimulus*, tribe
499 Mimuleae, and *Phryma*. *American Journal of Botany* 89: 1093–102.

500 Beardsley, P. M., S. E. Schoenig, J. B. Whittall, and R. G. Olmstead. 2004. Patterns of evolution in
501 western North American *Mimulus* (Phrymaceae). *American Journal of Botany* 91: 474–89.

502 Benaglia, T., D. Chauveau, D. R. Hunter, and D. S. Young. 2009. mixtools: an R package for analyzing
503 mixture models. *Journal of Statistical Software* 32: 1–29.

504 Blischak, P. D., J. Chifman, A. D. Wolfe, and L. S. Kubatko. 2018. HyDe: a Python package for genome-
505 scale hybridization detection. *Systematic Biology* 67: 821–829.

506 Brown, J. W., J. F. Walker, and S. A. Smith. 2017. Phyx: phylogenetic tools for unix. *Bioinformatics* 33:
507 1886–1888.

508 Chase, M. A., S. Stankowski, and M. A. Streisfeld. 2017. Genomewide variation provides insight into
509 evolutionary relationships in a monkeyflower species complex (*Mimulus* sect. *Diplacus*).
510 *American Journal of Botany* 104: 1510–1521.

511 Di Franco, A., R. Poujol, D. Baurain, and H. Philippe. 2019. Evaluating the usefulness of alignment
512 filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evolutionary*
513 *Biology* 19: 21.

514 Ding, B., E. L. Patterson, S. V. Holalu, J. Li, G. A. Johnson, L. E. Stanley, A. B. Greenlee, et al. 2020.
515 Two MYB proteins in a self-organizing activator-inhibitor system produce spotted pigmentation
516 patterns. *Current Biology* 30: 802–814.

517 van Dongen, S. 2000. Graph clustering by flow simulation. PhD Thesis, University of Utrecht, The
518 Netherlands.

519 Edger, P. P., M. R. McKain, K. A. Bird, and R. VanBuren. 2018. Subgenome assignment in

520 allopolyploids: challenges and future directions. *Current Opinion in Plant Biology* 42: 76–80.

521 Edger, P. P., R. Smith, M. R. McKain, A. M. Cooley, M. Vallejo-Marin, Y. Yuan, A. J. Bewick, et al.

522 2017. Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-

523 old naturally established neo-allopolyploid monkeyflower. *The Plant Cell* 29: 2150–2167.

524 Fishman, L., J. H. Willis, C. A. Wu, and Y. W. Lee. 2014. Comparative linkage maps suggest that fission,

525 not polyploidy, underlies near-doubling of chromosome number within monkeyflowers

526 (*Mimulus*; Phrymaceae). *Heredity* 112: 562–568.

527 Frailey, D. C., S. R. Chaluvadi, J. N. Vaughn, C. G. Coatney, and J. L. Bennetzen. 2018. Gene loss and

528 genome rearrangement in the plastids of five hemiparasites in the family Orobanchaceae. *BMC*

529 *Plant Biology* 18: 30.

530 Godden, G. T., T. J. Kinser, P. S. Soltis, and D. E. Soltis. 2019. Phylotranscriptomic analyses reveal

531 asymmetrical gene duplication dynamics and signatures of ancient polyploidy in mints. *Genome*

532 *Biology and Evolution* 11: 3393–3408.

533 Hall, M. C., D. B. Lowry, and J. H. Willis. 2010. Is local adaptation in *Mimulus guttatus* caused by trade-

534 offs at individual loci? *Molecular Ecology* 19: 2739–2753.

535 Hellsten, U., K. M. Wright, J. Jenkins, S. Shu, Y. Yuan, S. R. Wessler, J. Schmutz, et al. 2013. Fine-scale

536 variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing.

537 *Proceedings of the National Academy of Sciences* 110: 19478–19482.

538 Jiao, Y., J. Leebens-Mack, S. Ayyampalayam, J. Bowers, M. McKain, J. McNeal, M. Rolf, et al. 2012. A

539 genome triplication associated with early diversification of the core eudicots. *Genome Biology*

540 13: R3. doi:10.1186/gb-2012-13-1-r3.

541 Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler, and L. S. Jermin. 2017.

542 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14: 587–

543 589.

544 Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, et al. 2012.

545 Geneious Basic: An integrated and extendable desktop software platform for the organization and

546 analysis of sequence data. *Bioinformatics* 28: 1647–1649.

547 Kubatko, L. S., and J. Chifman. 2019. An invariants-based method for efficient identification of hybrid
548 species from large-scale genomic data. *BMC Evolutionary Biology* 19: 112.

549 Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon. 2012. PartitionFinder: Combined Selection of
550 Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Molecular Biology and*
551 *Evolution* 29: 1695–1701.

552 Liu, B., Y.-H. Tan, S. Liu, R. G. Olmstead, D.-Z. Min, Z.-D. Chen, N. Joshee, et al. 2020. Phylogenetic
553 relationships of Cyrtandromoea and Wightia revisited: A new tribe in Phrymaceae and a new
554 family in Lamiales. *Journal of Systematics and Evolution* 58: 1–17.

555 Lowry, D. B., J. M. Sobel, A. L. Angert, T.-L. Ashman, R. L. Baker, B. K. Blackman, Y. Brandvain, et
556 al. 2019. The case for the continued use of the genus name *Mimulus* for all monkeyflowers.
557 *TAXON* 68: 617–623.

558 MacNair, M. R. 1983. The genetic control of copper tolerance in the yellow monkey flower, *Mimulus*
559 *guttatus*. *Heredity* 50: 283–293.

560 Mai, U., and S. Mirarab. 2018. TreeShrink: fast and accurate detection of outlier long branches in
561 collections of phylogenetic trees. *BMC genomics* 19: 272.

562 McKain, M. R., and M. Wilson. Fast-Plast: Rapid de novo assembly and finishing for whole chloroplast
563 genomes v.1.2.6 <https://github.com/mrmckain/Fast-Plast>.

564 Minh, B. Q., H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, and R.
565 Lanfear. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the
566 genomic era. *Molecular Biology and Evolution* 37: 1530–1534.

567 Morales-Briones, D. F., G. Kadereit, D. T. Tefarikis, M. J. Moore, S. A. Smith, S. F. Brockington, A.
568 Timoneda, et al. 2021. Disentangling sources of gene tree discordance in phylogenomic data sets:
569 testing ancient hybridizations in Amaranthaceae s.l. *Systematic Biology* 70: 219–235.

570 Nelson, T. C., A. M. Stathos, D. D. Vanderpool, F. R. Finseth, Y. Yuan, and L. Fishman. 2021. Ancient
571 and recent introgression shape the evolutionary history of pollinator adaptation and speciation in

572 a model monkeyflower radiation (*Mimulus* section *Erythranthe*). *PLOS Genetics* 17: e1009095.

573 Nesom, G. 2012. Taxonomy of *Erythranthe* sect. *Simiola* (Phrymaceae) in the USA and Mexico.

574 *Phytoneuron* 40: 1–123.

575 Nesom, G. L., N. S. Fraga, W. R. Barker, P. M. Beardsley, D. C. Tank, B. G. Baldwin, and R. G.

576 Olmstead. 2019. Response to “The case for the continued use of the genus name *Mimulus* for all

577 monkeyflowers”. *TAXON* 68: 624–627.

578 Pease, J. B., J. W. Brown, J. F. Walker, C. E. Hinchliff, and S. A. Smith. 2018. Quartet Sampling

579 distinguishes lack of support from conflicting support in the green plant tree of life. *American*

580 *Journal of Botany* 105: 385–403.

581 Ranwez, V., E. J. P. Douzery, C. Cambon, N. Chantret, and F. Delsuc. 2018. MACSE v2: toolkit for the

582 alignment of coding sequences accounting for frameshifts and stop codons. *Molecular Biology*

583 *and Evolution* 35: 2582–2584.

584 Rice, A., L. Glick, S. Abadi, M. Einhorn, N. M. Kopelman, A. Salman-Minkov, J. Mayzel, et al. 2015.

585 The Chromosome Counts Database (CCDB) – a community resource of plant chromosome

586 numbers. *New Phytologist* 206: 19–26.

587 Sayyari, E., and S. Mirarab. 2016. Fast coalescent-based computation of local branch support from quartet

588 frequencies. *Molecular Biology and Evolution*.

589 Schemske, D. W., and H. D. Bradshaw. 1999. Pollinator preference and the evolution of floral traits in

590 monkeyflowers (*Mimulus*). *Proceedings of the National Academy of Sciences* 96: 11910–11915.

591 Schliep, K., A. J. Potts, D. A. Morrison, and G. W. Grimm. 2017. Intertwining phylogenetic trees and

592 networks. *Methods in Ecology and Evolution* 8: 1212–1220.

593 Schwarz, G. 1978. Estimating the dimension of a model. *The Annals of Statistics* 6: 461–464.

594 Scornavacca, C., K. Belkhir, J. Lopez, R. Dernat, F. Delsuc, E. J. P. Douzery, and V. Ranwez. 2019.

595 OrthoMaM v10: scaling-up orthologous coding sequence and exon alignments with more than

596 one hundred mammalian genomes. *Molecular Biology and Evolution* 36: 861–862.

597 Smith, S. A., M. J. Moore, J. W. Brown, and Y. Yang. 2015. Analysis of phylogenomic datasets reveals

conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evol Biol* 15:150.

Smith, S. A., and B. C. O'Meara. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28: 2689–2690.

Sobel, J. M. 2014. Ecogeographic isolation and speciation in the genus *Mimulus*. *The American Naturalist* 184: 565–579.

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

Stankowski, S., M. A. Chase, A. M. Fuiten, M. F. Rodrigues, P. L. Ralph, and M. A. Streisfeld. 2019. Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers. *PLOS Biology* 17: e3000391.

Stankowski, S., and M. A. Streisfeld. 2015. Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. *Proceedings of the Royal Society B: Biological Sciences* 282: 20151666.

Stevens, P. F. (2001 onwards). Angiosperm Phylogeny Website. Version 14, July 2017 [and more or less continuously updated since]. <http://www.mobot.org/MOBOT/research/APweb/>.

Streisfeld, M. A., W. N. Young, and J. M. Sobel. 2013. Divergent selection drives genetic differentiation in an R2R3-MYB transcription factor that contributes to incipient speciation in *Mimulus aurantiacus*. *PLOS Genetics* 9: e1003385.

Sugiura, N. 1978. Further analysts of the data by Akaike' s information criterion and the finite corrections. *Communications in Statistics - Theory and Methods* 7: 13–26.

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

Twyford, A. D., M. A. Streisfeld, D. B. Lowry, and J. Friedman. 2015. Genomic studies on the nature of species: adaptation and speciation in *Mimulus*. *Molecular Ecology* 24: 2601–2609.

Wen, D., Y. Yu, and L. Nakhleh. 2016. Bayesian Inference of Reticulate Phylogenies under the

- Multispecies Network Coalescent. *PLOS Genetics* 12: e1006006.
- Wu, C. A., D. B. Lowry, A. M. Cooley, K. M. Wright, Y. W. Lee, and J. H. Willis. 2008. *Mimulus* is an emerging model system for the integration of ecological and genomic studies. *Heredity* 100: 220–230.
- Yang, Y., M. J. Moore, S. F. Brockington, J. Mikenas, J. Olivieri, J. F. Walker, and S. A. Smith. 2018. Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy events in Caryophyllales, including two allopolyploidy events. *New Phytol* 217: 855–870.
- Yang, Y., and S. A. Smith. 2014. Orthology inference in non-model organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Molecular Biology and Evolution* 31: 3081–3092.
- Yu, Y., J. H. Degnan, and L. Nakhleh. 2012. The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLOS Genetics* 8: e1002660.
- Yu, Y., J. Dong, K. J. Liu, and L. Nakhleh. 2014. Maximum likelihood inference of reticulate evolutionary histories. *Proceedings of the National Academy of Sciences* 111: 16448–16453.
- Yuan, Y.-W. 2019. Monkeyflowers (*Mimulus*): new model for plant developmental genetics and evo-devo. *New Phytologist* 222: 694–700.
- Yuan, Y.-W., A. B. Rebocho, J. M. Sagawa, L. E. Stanley, and H. D. Bradshaw. 2016. Competition between anthocyanin and flavonol biosynthesis produces spatial pattern variation of floral pigments between *Mimulus* species. *Proceedings of the National Academy of Sciences* 113: 2448–2453.
- 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: 153.
- Zhang, C., T. Zhang, F. Luebert, Y. Xiang, C.-H. Huang, Y. Hu, M. Rees, et al. 2020. Asterid Phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications. *Molecular Biology and Evolution* 37: 3188–3210.