Molecular Systematics Large and Small: Patterns of Evolution in *Lamourouxia* and Orobanchaceae

A Thesis
Presented in Partial Fulfillment of the Requirements for the
Degree of Master of Science
with a
Major in Biology
in the
College of Graduate Studies
University of Idaho
by
Sebastian M. E. Mortimer

Major Professor: David C. Tank, Ph.D.
Committee Members: Jack Sullivan, Ph.D.; Luke Harmon, Ph.D.; Eric Roalson, Ph.D.
Department Administrator: James Nagler, Ph.D.

May 2019

Authorization to Submit Thesis

This thesis of Sebastian M. E. Mortimer, submitted for the degree of Master of Science with a Major in Biology and titled "Molecular Systematics Large and Small: Patterns of Evolution in *Lamourouxia* and Orobanchaceae" has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: _____ Date: _____
David C. Tank, Ph.D.

Committee Members: _____ Date: _____
Jack Sullivan, Ph.D.

_____ Date: _____
Luke Harmon, Ph.D.

_____ Date: _____
Eric Roalson, Ph.D.

Department
Administrator: _____ Date: _____
James Nagler, Ph.D.

Abstract

The extent of our current systematic knowledge has grown tremendously in recent years. However, there are still large gaps in our understanding of evolutionary relationships in the tree of life. Here we explore systematics and phylogenetic relationships of angiosperms at two different scales. First, at the genus level, we elucidate species level relationship of the genus *Lamourouxia* and evaluate diagnostic traits for monophyletic clades using phylogenetic half-life. Second, we explore the extent of our systematic knowledge in the Orobanchaceae. We assess where the gaps in phylogenetic, systematic, and geographic knowledge exist. Furthermore, we summarize taxonomic conflicts and inconsistencies, and identify taxonomic and molecular gaps in our current sampling.

Acknowledgements

I would like to thank Dr. David Tank, my major professor, for these past years of mentoring as a researcher, as a field botanist, and as an instructor. Thank you for the opportunities you provided to explore my interests in teaching, botany, systematics, and computational biology. Your excitement for and dedication to botany and systematics is daily inspiration.

I would also like to thank my committee: Dr. Jack Sullivan for challenging systematic conversations and support and encouragement in my scientific career. Dr. Luke Harmon for enthusiasm, ideas, and insightful conversations about my project. Dr. Eric Roalson for insights into the methods and concepts of systematic research concerning my project.

I would like to thank Dr. Sarah Jacobs for incredible support and mentorship throughout my time at the University of Idaho, I could not have done it without you.

A big thank you to all the members of room 343 during my time at the University of Idaho: Dr. Daniel Caetano, Dr. Diego Morales-Briones, Ian Gilman, Megan Ruffley, and Sam McCauley. You all made my time in the office an absolute blast, whether late nights, early mornings, or coffee breaks.

Thank you to Grahm Johnson for reminding me why I love field botany.

Thank you to Dr. Susan Kephart at Willamette University for introducing me to systematics and evolution and helping me on my way.

Dedication

To my parents

For unending support

Table of Contents

List of Tables

List of Figures

Chapter 1: Evolution of morphological traits in the genus *Lamourouxia* (Orobanchaceae): Using phylogenetic comparative models and phylogenetic half-life to inform selection of taxonomic diagnostic characters

Abstract

Evolutionary relationships within a number of genera of the hemi- and holoparasitic lineage Orobanchaceae Vent. have not been studied using modern molecular phylogenetic methods. The genus Lamourouxia Kunth is a moderately sized clade in Orobanchaceae consisting of 28 described species. Morphological evidence for hypothesized evolutionary relationships between species and general morphological groupings in Lamourouxia require further study. Here we sampled 63 individuals representing 18 species of Lamourouxia, inferred evolutionary relationships using maximum likelihood phylogenetic methods and quartet-based species tree methods. Finally, we determined the usefulness of common taxonomic characters using phylogenetic half-life and present a new taxonomy based on molecular phylogenetic evidence.

Introduction

The cosmopolitan angiosperm family Orobanchaceae Vent. displays several notable neotropical radiations. These radiations are primarily hemiparasitic lineages, and several have recently been studied with molecular phylogenetic techniques (*Castilleja* Mutis ex L.f., Tank and Olmstead, 2008; *Neobartsia* Benth., Uribe-Convers and Tank, 2015; *Pedicularis* L., Eaton and Ree, 2013, Yu et al., 2015). However, there are a number of medium to large clades of Orobanchaceae that have only been included as representative taxa in higher level molecular phylogenetic studies to date, and have yet to be the subject of focused molecular systematic work (Mortimer, Ch. 2). Among these, only three of the ~28 species (Ernst, 1972; Turner, 1993) of the neotropical genus *Lamourouxia* Kunth have been sampled to date (Bennett and Mathews, 2006; Tank and Olmstead, 2008; McNeal et al., 2013). *Lamourouxia* is the fourth largest genus of the Pedicularideae clade, and has been placed in a subclade sister to Castillejinae along with *Agalinis* Raf. (third largest), and *Seymeria* Pursh (fifth largest). The most recent circumscription of *Lamourouxia* (Ernst, 1972) hypothesized three taxonomic sections, based largely on floral morphology and their corresponding pollination syndromes; two of these sections are hummingbird pollinated, while the third is bee pollinated. Pollination syndrome and associated floral traits such as flower color, staminate morphology, and anther fertility are the primary traits used for taxonomic circumscription (Ernst, 1972). Vegetative

characters were deemed "difficult to correlate" with taxonomy due to their variability across the clade.

Hypothesized relationships between the three sections of *Lamourouxia* are largely driven by a series of reductions in staminate morphology (Fig. 1 A & B). The hummingbird pollinated section Lamourouxia displays red flowers and four fertile anthers born on equal to sub-equal filaments, and was considered by Ernst (1972) to be the earliest diverging clade in the genus. Section Adelphidion is thought to be primarily bee-pollinated, and displays classic bee pollinated flowers that are pink to purple with. In contrast to section Lamourouxia, section Adelphidion displays four strongly dimorphic stamens with four fertile anthers. This section is comprised of only four species, but they are geographically disjunct, with two occurring in Mexico and two occurring in Andean South America (Fig. 1 D). Lastly, section Hemispadon, which Ernst (1972) hypothesized to be more closely related to Adelphidion than to Lamourouxia, based on a further reduction in staminate morphology with only two fertile anthers and strongly dimorphic stamens, displays primarily red flowers, and is thought to be hummingbird pollinated. Although placing species within these sections is relatively straightforward based on these floral morphologies, these putatively pollinator driven traits focused on by Ernst (1972) may not reflect evolutionarily distinct clades (Smith et al., 2008), making them less than ideal for characterizing monophyletic groups within the clade. For example, in the sister species *Erythranthe lewisii* and *Erythranthe cardinalis*, pollinator-driven morphological changes involve a few genes of large effect (Bradshaw and Schemske, 2003).

In this paper, we infer a multi-locus nuclear and chloroplast species phylogeny of the genus *Lamourouxia*. For this, we sample across the three described taxonomic sections of the genus using a high-throughput, amplicon-based data collection strategy (Uribe-Convers et al. 2016). This systematic evaluation of the three taxonomic sections will allow us to determine whether Ernst's hypothesized taxonomic sections are monophyletic, and evaluate the suitability of floral morphological characters for classification. Furthermore, a focused phylogenetic hypothesis of the clade allows us to determine evolutionary relationships and patterns of morphological diversity within *Lamourouxia.* Finally, to determine which floral traits are evolutionarily conserved, we use phylogenetic half-life (Parins-Fukuchi, 2017) as a measure of lability of discrete traits (e.g., flower color, anther fertility, leaf shape), to ensure that diagnostic traits are slowly evolving traits.

Methods and Materials

Molecular methods: We sampled 63 individuals of *Lamourouxia* representing 18 of 26 species described by W.R. Ernst (1972). Samples were taken from herbarium specimens collected between 1965 and 2014 (Table 1). Additionally, we sampled three outgroup species: two from the

Castillejineae, *Cordylanthus molle* (A. Gray) A. Heller subsp. *molle*, *Chloropyron tecopense* (Munz & J.C. Roos) Tank & J.M. Egger, and *Paulownia fortunei* (Seem.) Hemsl. (Paulowniaceae). Total genomic DNA was extracted using a modified 2x CTAB method (adapted from Doyle and Doyle, 1987). A total of 11 chloroplast regions (primers from Latvis et al., 2017a), two nuclear ribosomal regions (ITS and ETS; Baldwin et al., 1995; Baldwin and Markos, 1998), and 18 nuclear regions (primers from Latvis et al. 2017b) were amplified using microfluidic PCR on the Fluidigm Juno Array System (Fluidigm, San Francisco, CA, USA) at the University of Idaho Genomics Resources Core facility following Uribe-Convers et. al (2016) and Latvis et al. (2017a, 2017b). Resulting amplicons were sequenced in an Illumina MiSeq Sequencing v3 (600 cycles).

The resulting reads were annotated and demultiplexed using a dual barcoding system and the R-script dbcAmplicons (Uribe-Convers et al., 2016). The annotated reads were then filtered and merged using the *fluidigm2purc* python script (Blischak et al., 2018), as paired reads in a FASTQ format. Briefly, the *fluidigm2purc* script filters sequences using Sickle (Joshi and Fass, 2011), merges the paired reads using FLASH2 (Magoc and Salzberg, 2011), and then converts the sequences into FASTA formatted files for each locus or region that was amplified. Settings for the *fluidigm2purc* re-clustering step using the Pipeline for Untangling Reticulate Complexes (PURC; Rothfels et al., 2017) to clean up PCR error, sequencing error, and chimeras were -c 0.925 0.90 0.875 0.925 -s 2 5. Chloroplast and nuclear ribosomal DNA regions were processed using the script *crunch_clusters.py*, with a cleaning threshold of 0.4 (Blishak et al., 2018) under the assumption of a haploid genome. Nuclear loci were processed using the *crunch_clusters.py* script allowing for multiple haplotypes (i.e., unknown ploidy) with the same cleaning threshold.

Phylogenetic methods: Initial alignments were used to build gene trees with RAxML v.8.2.9 (Stamatakis, 2014) using the GTRGAMMA model of sequence evolution. The presence of paralogs was determined if the resulting tree contained separate strongly supported clades containing haplotypes from multiple species. If a locus was found to have paralogous sequences, these sequences were then sorted into a new alignment and analyzed as a separately evolving locus. Once paralog splitting was completed, loci were aligned using MAFFT v.7.307 (Katoh and Standley, 2013) and adjusted by eye. Paralog splitting resulted in 27 single copy nuclear loci that were again aligned using MAFFT v.7.307, and adjusted by eye. Finally, alignments of each locus were cleaned using Gblocks v.0.91b (Castresana J., 2000) to minimize missing data in gappy alignments, where the minimum length of a block was 2, allowing for gap positions in sequences and non-conserved blocks.

For each single copy nuclear locus, chloroplast region, and nuclear ribosomal region we estimated an appropriate model of sequence evolution using decision theory (Minin et al., 2003), as

implemented in PAUP* v.4.0a (Swofford, 2002). Using the best-fit models of sequence evolution for each locus, we estimated maximum likelihood gene trees using Garli v.2.01 (Zwickl, 2006). A preliminary tree was estimated for each locus or region using a stepwise addition starting tree with the number of attachments per taxon equal to twice the number of tips represented in that particular locus. Ten independent searches were conducted in this way to find the most supported topology. The best tree found in the initial 10 searches was then used as a starting tree for another 25 independent searches. This tree was then used as the starting tree for 1000 bootstrap replicates, which were summarized on the starting topology using *sumtrees.py* (Sukumaran and Holder, 2010). These alignments and gene trees were later used in a concatenated maximum likelihood analysis, as well as in species tree estimation methods (Appendix A).

Nuclear, chloroplast, and nuclear ribosomal alignments were concatenated so that each sample was represented by a single sequence (29 loci, 18,170bp, 66 tips). As with gene tree estimates, a maximum likelihood tree was estimated using Garli v.2.01 (Zwickl, 2006). Twenty-five independent searches were run to find an initial best tree, using stepwise addition to find a starting tree. The initial best tree was used as a starting tree for 25 additional independent searches, to find a secondary best tree. The secondary best tree was used as the starting tree for 1000 bootstrap replicates.

Species tree topologies were estimated using two quartet-based methods. First, using ASTRAL-III (Zhang et al., 2018), we estimated a species topology based on quartets found in 29 previously estimated maximum likelihood gene trees (27 nuclear gene trees, one nrDNA tree, and one cpDNA tree). Each gene tree was collapsed at nodes where bootstrap support was lower than 25%, and tips from every gene tree were mapped to hypothesized species assignments based on current taxonomy. The second method used for species tree estimation was SVDquartets (Chifman and Kubatko, 2014), as implemented in the PAUP* v.4.0a (Swofford, 2002). The input dataset for SVDquartets was the same fully concatenated dataset used in the maximum likelihood estimation described above. Here, each sample was represented by a single sequence (29 loci, 18,170bp, 66 tips). We searched all possible quartets, and mapped each tip to hypothesized species assignment. We subsequently ran 100 bootstrap replicates in SVDquartets, and summarized them using *sumtrees.py* (Sukumaran and Holder, 2010).

Ancestral state reconstruction: For downstream comparative analyses, we created ultrametric species tree estimates with branch lengths scaled using sequences of the nrDNA ITS region, as this was the only gene region that had complete sampling for all species. A representative accession for each the 18 species of *Lamourouxia* included in this study was selected from the nrDNA ITS alignment. This

abbreviated ITS alignment was tested to see if it conformed to a global molecular clock, as well as fit to a model of nucleotide substitution using decision theory (Minin et al. 2003), as implemented in PAUP* v.4.0a (Swofford, 2002). We held the root node constant at 1.0, using a narrow normal distribution with a mean of 1.0 and a standard deviation equal to 0.001, to ensure that branch lengths between the two resulting species tree topologies would be comparable. We ran three iterations of 10,000,000 generations and checked each analysis for convergence in Tracer v1.6 (Rambaut et al., 2013).

Measurements from eleven binary characters, representing both vegetative and floral morphologies in the clade, were gathered from descriptions in Ernst's (1972) monograph. These traits were then verified using herbarium specimens where measurements were adjusted to reflect variation seen in the specimens. This was especially important for where Ernst's measurements were incomplete (Appendix A). Characters were selected based on their use in the most current taxonomic keys, indicating their utility in diagnosing taxonomic groups in *Lamourouxia* (Ernst, 1972). We reconstructed the ancestral states of each character on each species topology using two Mk models, equal rates (ER) and all rates different (ARD), in the R package Phytools (Revell, 2012), tested model fit between the two models using a likelihood ratio test, and proceeded with best fit model.

To rank taxonomic traits in relative order of potential taxonomic utility, we calculated the phylogentic half-life of each discrete character based on the Mk rates associated with each topology (Parins-Fukuchi, 2017). For example, a trait that has a relatively short phylogenetic half-life means that the trait has evolved quickly, or multiple times independently, and thus, may not be taxonomically informative in a certain clade. Conversely, if the phylogenetic half-life of a trait is high, it suggests that this trait may be more taxonomically informative or diagnostic for the clade.

Results

Molecular methods: Figure 3 summarizes the results of amplicon sequencing in our study. Of the original 19 nuclear loci sampled, eight contained paralogous sequences. These sequences were aligned separately and treated as independently evolving loci (Table 2), resulting in a total of 27 nuclear loci, 11 chloroplast DNA regions, and two nuclear ribosomal DNA regions. The smallest representation of samples in a locus was six in locus CS1_At4g24830_948F_CS2_At4g24830_1447R_t1, which was originally combined with CS1_At4g24830_948F_CS2_At4g24830_1447R_t2 to make a single locus. The maximum number of samples represented in a single locus was 66 in ITS5*_CS1_ITS2*_CS2 & ITS3*_CS1_ITS4_CS2. The minimum number of gene regions represented in an individual sample was three in *Lamourouxia_viscosa*_16725, and the maximum number gene regions represented in an individual

sample was 33 in *Lamourouxia_xalapensis*_17088 (Fig. 3). Initial alignment lengths of each locus or region was generally reduced after realignment using MAFFT v.7.307 (Katoh and Standley, 2013), and cleaning using Gblocks v.0.91b (Castresana J., 2000) (Table 2). Models of nucleotide substitution estimated using decision theory, as implemented in PAUP* v.4.0a (Swofford, 2002). Alignments and gene trees used in downstream analyses can be found at https://github.com/mortimersebastian/Lamourouxia_Supp.

Concatenated analysis: Maximum likelihood estimation of our concatenated dataset consisted of 29 loci, 18,170bp, and 66 tips, representing 18 species of *Lamourouxia* and three outgroups. We found strong support for the monophyly of the genus *Lamourouxia* (crown support: 100% BS, stem support: 100% BS; Fig. 2). Several clades within *Lamourouxia* were well supported at the crown node, but we found moderate to low support along the backbone of the phylogeny. We found strong support (crown and stem support:100% BS) for the clade containing *L. pringlei* Robinson & Greenman ex Pringle*, L. longiflora* Benth.*, L. dispar* Ernst*,* and *L. multifida* H.B.K. as the sister group to the rest of the genus. *Lamourouxia dependens* Bentham was found sister to all other species except those in the earliest diverging clade. Three species, originally belonging to section Lamourouxia, *L. xalapensis* H.B.K.*, L. stenoglossa* Hunnewell & Smith*,* and *L. macrantha* Martens & Galeotti were found to form a separate clade with moderate support (crown support: 83% BS). All but two samples of the taxonomic section Adelphidion, *L. dasyantha* (Chamisso & Schlechtendal) Ernst*.* and *L. brachyantha* Greenman*,* were found to be monophyletic in this analysis. A sample of *L. dasyantha* included in this study was found on a solitary branch sister to a clade of species primarily belonging to Hemispadon, and a sample representing *L. brachyantha* was found nested with the clade containing species of section Hemispadon. In sum, the concatenated approach found relatively strong support for paraphyly of section *Lamourouxia*, strong support for the monophyly of most species, and relatively strong support for a clade containing primarily samples of section Hemispadon. However, the low bootstrap values along the backbone of the phylogeny provide little support for relationships between these clades.

Species Tree Analyses: We found conflicting topologies between the two species tree methods employed here (Fig. 4). However, both methods recovered a paraphyletic section Lamourouxia, and there was consistent support for a clade contained the four species *Lamourouxia pringlei, L. longiflora, L. dispar,* and *L. multifida* as the sistergroup to the rest of the genus. Although, relationships between *L. longiflora, L. dispar,* and *L. multifida* differed between analyses, *L. pringlei* was found sister to the other species in both analyses. Our ASTRAL-III analysis found the subsequently diverging lineage to be a clade of three species belonging to section Lamourouxia,

including *L. xalapensis, L. stenoglossa,* and *L. macrantha.* However, our SVDquartets analysis found *L. sylvatica* to be a subsequently diverging lineage with relatively high support (stem support: 100% BS). Furthermore*,* our SVDquartets analysis found *L. xalapensis, L. stenoglossa,* and *L. macrantha* sister to *L. dependens* as part of the clade diverging subsequently to *L. sylvatica.* Both analyses found the clades (*L. dasyantha* + *L. virgate H.B.K.*) and (*L. integerrima J.D. Smith* + *L. lanceolate Bentham*) to be the subsequently diverging lineages in *Lamourouxia* with variable support at crown and stem nodes (ASTRAL quartet support = 46-68% and SVDquartets 39%-100% BS). The remaining five species sampled in our analysis were recovered as monophyletic by both analyses. However, relationships between species were not congruent between analyses, and support for relationships within the clade was relatively low (quartet support = 56-34%, 45-34% BS). To visualize the conflict between species tree methods and assess the evolution of morphological characters in *Lamourouxia* with respect to taxonomic conclusions, we collapsed conflicting nodes in our analyses into a consensus tree of our two species tree results (Fig. 6).

Ancestral State Reconstruction: We found that one taxonomic character was best fit under an asymmetric model of trait evolution on the SVDquartets topology (Table 3). All other discrete traits were best fit under a symmetric model of trait evolution on both trees. Ancestral state reconstructions were similar for taxonomic traits between species tree topologies (Fig. 5). We found evidence for three transitions from red flower color to pink or violet flower color in both trees. Furthermore, we found evidence for two transitions from equal stamens to dimorphic stamens in both trees, and finally, we found evidence for a transition from four fertile stamens to two fertile stamens at the crown node of section Hemispadon, with a reversal in the *L. dasyantha* lineage. The three taxonomic characters with the highest phylogenetic half-life for ASTRAL-III were leaf division (2.8004), branching of corolla hairs (1.3227), and anther fertility (1.2023; Table 3). The same three taxonomic characters were found to have the highest phylogenetic half life in the SVDquartets tree (Table 3; 2.8170, 1.2862, 1.2406, respectively).

Discussion

Our phylogenetic results highlight several important taxonomic issues in *Lamourouxia*, as the current infrageneric circumscription (Ernst, 1972) does not reflect our current understanding of phylogenetic relationships. Ernst (1972) circumscribed three taxonomic sections, section Lamourouxia, section Adelphidion, and section Hemispadon, but none of these were monophyletic in our analyses. However, we did find support for resurrecting Bentham's (1846) section Euphrasioides, and incorporating it into a modified circumscription of Ernst's infrageneric classification. Furthermore, several of the floral traits used for circumscription of Ernst's (1972) sections are still

taxonomically informative for major clades within *Lamourouxia*, and these traits often have a high phylogenetic half-life (Figs. 5 and 6).

Here, we propose a new infrageneric classification of *Lamourouxia* that includes four sections: 1) section Lamourouxia, containing seven species, each with four fertile anthers, equal to subequal stamens, and red corollas with simple, glandular corolla hairs, 2) section Euphrasioides with four species that also have equal to subequal stamens with four fertile anthers and red corollas, but whose corolla hairs are branched and not glandular, 3) a reduced section Adelphidion with only two species that have lavender, pink, or magenta, corollas, and strongly dimorphic stamens with four fertile anthers, and 4) an expanded section Hemispadon comprising 12 species, again with strongly dimorphic stamens, but with infertile anthers on the reduced stamens (except in *L. brachyantha*), and typically red corollas that may have flecks of yellow/orange throughout. Two species, *L. dependens* and *L. sylvatica* remain unplaced in our phylogenetic classification. Our phylogenetic analyses, evaluation of morphological characters across the clade, and the resulting revised infrageneric classification will facilitate the placement of taxa not sampled in this study, as well as taxa discovered in future work.

Uncertainty in the placement of L. dependens and L. sylvatica: Phylogenetic results from individual gene trees (Appendix A), concatenated analyses (Fig. 2), and both species tree analyses (Fig. 4) are in conflict with respect to the placement of *L. dependens* and *L. sylvatica*. In general, these analyses are characterized by low bootstrap support, and an almost random placement of these taxa in the gene trees (Appendix A). While the Andean species *L. sylvatica* was represented by several independent accessions in our analyses, only one sample of *L. dependens* was included here. In our concatenated analysis, all accessions of *L. sylvatica* were monophyletic, and these were recovered as the sister group of *L. dasyantha* and *L. virgata*, albeit with limited statistical support (Fig. 2). These three represent three of the four species of Ernst's (1972) bee-pollinated section Adelphidion, and *L. sylvatica* and *L. virgata* are the only two species of *Lamourouxia* that are distributed in Andean South America. However, our species tree analyses do not recover this same relationship, placing *L. sylvatica* on a lone branch separate from other bee-pollinated species, again without much statistical support (Fig. 4). Likewise, *L. dependens* was recovered on a lone branch in both the concatenated analysis (Fig. 2), and the ASTRAL-III species tree analysis (Fig. 4A), but was recovered as sister to *L. macrantha, L. stenoglossa,* and *L. xalapensis* (our resurrected section Euphrasiodes) in the SVDquartets species tree analysis (Fig. 4B). Interestingly, Bentham (1846) included *L. dependens* in section *Euphrasioides* based on morphological similarities to *L. macrantha, L. stenoglossa,* and *L. xalapensis*. Given the conflicting relationships across individual gene trees, variable placements of *L.*

sylvatica and *L. dependens* in the concatenated and coalescent-based species tree analyses, and overall lack of bootstrap support for the placement of these taxa in our analyses, we leave these two species as currently unplaced in our infrageneric classification. Future phylogenetic work using more variable genomic data from a broader sampling of the geographic breadth of the species will help to understand the evolutionary processes resulting in this conflicting phylogenetic signal (e.g., reticulate histories).

Sections Lamourouxia and Euphrasiodes: Ernst's (1972) section Lamourouxia was paraphyletic in all of our analyses (Figs. 2 and 4). Nearly all independent gene trees, as well as the concatenated analysis, and both species tree analyses, recover a strongly supported clade containing the type species, *L. multifida*, along with *L. dispar, L. longiflora*, and *L. pringlei* as the sister clade to the rest of the genus. However, the three species *L. macrantha, L. stenoglossa,* and *L. xalapensis,* considered by Ernst (1972) to belong to section Lamourouxia, were found to form a separate clade in both species trees (Fig. 4) and the concatenated tree (Fig. 2). These species were first described by Bentham (1846) as belonging to a separate section he named Euphrasioides, and for this reason, we suggest resurrecting section Euphrasioides to resolve the paraphyly of Ernst's (1972) section Lamourouxia. *Lamourouxia dependens,* which was also included in Euphrasioides by Bentham (1846), was resolved as sister to our Euphrasioides clade in the SVDquartets analysis (Fig. 4B), but with little statistical support, and as mentioned above, phylogenetic signal for the placement of this species conflicted across all of our analyses, so we hesitate to include it here until future phylogenetic work can resolve the evolutionary history of this species. The morphological characters separating these clades are not conspicuous. However, species of section Lamourouxia have simple, glandular corolla hairs, while species belonging to section Euphrasioides often have branched, eglandular corolla hairs (Fig. 6). Based on these morphological characteristics, it is likely that the unsampled species, *L. ovata*, which Ernst (1972) considered to be closely allied with *L. macrantha*, belongs in this clade as well (Fig. 6).

Species in sections Lamourouxia and Euphrasiodes are found primarily in Mexico, with largely overlapping distributions. The Isthmus of Tehuantepec separates the distributions of most species (Ernst, 1972). Two species, *L. multifida* (section Lamourouxia) and *L. macrantha* (section Euphrasioides), have distributions that cross the Isthmus of Tehuantepec, distributed both throughout central Mexico, and with disjunct populations in southern Mexico and Guatemala. With the separation of sections Lamourouxia and Euphrasioides based on our phylogenetic results, most species of the narrowly circumscribed section Lamourouxia are found north of the Isthmus of Tehuantepec, while putative species pairs in section Euphrasioides are found on both sides of the

isthmus. *Lamourouxia xalapensis* and *L. ovata* (unsampled) are distributed north of the isthmus, and *L. stenoglossa* and *L. dependens* are found south of the isthmus. A better understanding of interspecific relationships in section Euphrasioides would allow for more precise biogeographic and phylogeographic analyses investigating the influence of the Isthsmus of Tehuantepec on diversification in this group.

Section Adelphidion: Ernst (1972) united species in section Adelphidion based primarily on their shared bee pollination syndrome, and included four species, *L. sylvatica* and *L. virgata*, both distributed in Andean South America (Fig. 1), and the two central Mexican species *L. dasyantha* and *L. brachyantha*. Interestingly, Ernst (1972) noted this large geographic disjunction, and discussed the possibility that morphological similarities of these species may be the result of the lability of floral traits and convergence on this pollination syndrome in Mexico and South America. Here, we consistently recover a clade containing *L. virgata* (distributed in northern Ecuador and Colombia) and *L. dasyantha* (central Mexico), suggesting a long distance dispersal, mountain hopping, or vicariant biogeographic history of this clade. *Lamourouxia sylvatica*, found throughout the mountains of southern Ecuador and Peru, is also likely a member of this clade, however, uncertainty in its phylogenetic placement in our analyses prevents us from including this species in Adelphidion pending further study. Given its South American distribution, the resolution of how *L. sylvatica* is related to *L. virgata* and *L. dasyantha* is key to understanding the disjunct distribution of this clade.

The final member of section Adelphidion sensu Ernst (1972) is *L. brachyantha*, which was strongly placed with species of section Hemispadon in our analyses (Figs. 2 and 4). Ernst (1972) noted that this species is easily distinguishable from other *Lamourouxia* species based on its uniquely galeate corollas. Ernst (1972) went further to state that *L. brachyantha* is probably not particularly closely related to any other species of *Lamourouxia,* and was only placed with section Adelphidion because of the shared characteristics of the development of the stamens and the non-red, short corolla. Previous classifications (Bentham, 1846; Greenman 1905) included *L. brachyantha* in section Hemispadon based on mostly vegetative similarities to *L. rhinanthifolia* H.B.K., and in both our concatenated tree and ASTRAL-III species tree, *L. brachyantha* is recovered as the sister species to *L. rhinanthifolia*, albeit with weak support.

Section Hemispadon: Section Hemispadon is characterized by a further reduction in two of the four stamens, in which the reduced stamens lack fertile anthers (Ernst, 1972). Floral color in this section is primarily red, although, species are known to have flecks of other colors such as oranges and yellows in their corollas, and they are thought to be primarily hummingbird pollinated (Ernst, 1972). *Lamourouxia brachyantha* is the exception here with pink corollas and four fertile anthers, suggesting

that floral traits associated with pollination syndrome in *Lamourouxia* are more labile than previously thought, and may be best viewed in combination with other characteristics (e.g., vegetative morphology, in the case of *L. brachyantha*).

Although, our phylogenetic analyses provide little support for interspecific relationships within section Hemispadon, we do consistently recover two clades in our concatenated (Fig. 2) and coalescent-based species tree (Fig. 4) analyses (Fig. 6; Hemispadon I and II). These clades are separated primarily by vegetative characteristics, where Hemispadon I contains species with ovate to cordate leaves with dentate to crentate margins, and species comprising Hemispadon II have leaves that are lanceolate to subulate (Ernst, 1972). We sampled four species of Hemispadon I here (*L. rhinanthifolia, L. viscosa* H.B.K.*, L. colimae* Ernst & Baad*, L. nelsonii* Robinson & Greenman), and place an additional four Hemispadon species in this clade based on similar leaf morphologies (Fig. 6; *L. gracilis* Robinson & Greenman*, L. guiterrezii* Oersted in Bentham & Oersted*, L. microphylla* Martens & Galeotti*, L. smithii* Robinson & Greenman). Likewise, we sampled *L. lanceolata* and *intergerrima* here, and include the unsampled species *L. tenuifolia* Martens & Galeotti and *L. jaliscana* Ernst & Baad in Hemispadon II (Fig. 6).

Section Hemispadon is the most widespread section of *Lamourouxia*, distributed throughout Mexico on either side of the Isthmus of Tehuantepec, as well as in Guatemala, Honduras, and Panama. Hemispadon I contains the two widespread species *L. viscosa* and *L. rhinanthifolia*. *Lamourouxia viscosa* displays the largest geographic range of any species in the genus, and is found from northern Mexico to Panama, while *L. rhinanthifolia* is found in Mexico north of the Isthmus of Tehuantepec. In contrast to these widespread species, Hemispadon I also contains several rare, narrow endemics, such as *L. nelsonii, L. smithii, L colimae, L gracilis, L. gutierrezii,* and *L. microphylla.* Similarly, Hemispadon II contains three narrow endemics, *L. integerrima, L. tenuifolia,* and *L. jaliscana* (Fig. 6), and *L. lanceolata*, which is distributed south of the Isthmus of Tehuantepec through Central American to Panama. Hemispadon I appears to be more species rich than Hemispadon II, and in contrast to sections Lamourouxia and Euphrasioides, the Isthmus of Tehuantepec does not seem to play as large a role as a biogeographic driver of diversification. Rather, much of the described variation is found in narrowly distributed species. Increased sampling in this clade will be necessary to investigate species boundaries, and patterns of diversification.

Phylogenetic half-life: Ernst's (1972) detailed monograph of *Lamourouxia* exhaustively documents the morphological diversity found across the clade, and provides a wealth of information that may be useful in light of a phylogenetic hypothesis. We analyzed a suite of putatively diagnostic characters that were heavily relied on by Ernst (1972) in his classification of *Lamourouxia* (Table 3), and

investigated the phylogenetic half-life of these traits across our species tree hypotheses. Phylogenetic half-life has been used to model the evolution of continuous characters where it represents the time that it takes to move halfway from the ancestral state to the primary optimum (Parins-Fukuchi, 2017). Here, we use a modified version of this concept for discrete morphological characters under an Mk model, and use transition rates to calculate the phylogenetic half-life (Hansen and Orzack, 2005) (Table 3). Under this modification, labile characters with high transition rates have relatively low phylogenetic half-lives (e.g., leaf shape, pubescence; Table 3), and those with low transition rates (e.g., stamen morphology, corolla hair morphology; Table 3) exhibit a high phylogenetic half-life.

We found four characters with high phylogenetic half-lives ($\lambda > 1.0$; Table 3), and these characters are among those used here to delineate the major clades, including stamen morphology and anther fertility (Fig. 6). By contrast, floral color, which was relied on heavily by Ernst (1972) with respect to section Adelphidion, was shown to have an intermediate phylogenetic half-life, and in isolation, is likely not a character that should be relied on for diagnosing major clades. While some vegetative characters have relatively high phylogenetic half-lives (e.g., leaf division, leaf margin), others are very low (e.g., leaf shape). Thus, we found this quantification of evolutionary rates of character change to be useful in identifying characters that will be more likely to diagnose named clades.

Literature Cited

Baldwin, B. G., M. J. Sanderson, J. M. Porter, M. F. Wojciechowski, C. S. Campbell, and M. J. Donoghue. (1995). The ITS region of nuclear ribosomal DNA: A valuable source of evidence on angio- sperm phylogeny. Annals of the Missouri Botanical Garden 82: 247–277.

Baldwin, B. G., and S. Markos. (1998). Phylogenetic utility of the external transcribed spacer (ETS) of 18S-26S rDNA: Congruence of ETS and ITS trees of Calycadenia (Compositae). Molecular Phylogenetics

and Evolution 10: 449–463.

Bentham, G. (1846). Scrophulariaceae. Prodromus systematis naturalis regni vegetabilis, 10, 186-586.

Blischak, P. D., Latvis, M., Morales-Briones, D. F., Johnson, J. C., Di Stilio, V. S., Wolfe, A. D., & Tank, D. C. (2018). Fluidigm2PURC: automated processing and haplotype inference for double-barcoded PCR amplicons. *bioRxiv*, 242677.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., ... & Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS computational biology*, *10*(4), e1003537.

Bradshaw Jr, H. D., & Schemske, D. W. (2003). Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. Nature, 426(6963), 176.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution*, *17*(4), 540-552.

Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, *30*(23), 3317-3324.

Doyle, J. J. and J. L. Doyle. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19:11–15.

Eaton, D. A., & Ree, R. H. (2013). Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). Systematic biology, 62(5), 689-706.

Ernst, W. R. (1972). Floral morphology and systematics of Lamourouxia (Scrophulariaceae: Rhinanthoideae). 63 pp. – 2

Gutiérrez-García, T. A., & Vázquez-Domínguez, E. (2013). Consensus between genes and stones in the biogeographic and evolutionary history of Central America. *Quaternary Research*, *79*(3), 311-324.

Hansen, T. F., & Orzack, S. H. (2005). Assessing current adaptation and phylogenetic inertia as explanations of trait evolution: the need for controlled comparisons. *Evolution*, *59*(10), 2063-2072.

Joshi, N. A., & Fass, J. N. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33)[Software].

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution, 30(4), 772-780.

Kay, K. M., Reeves, P. A., Olmstead, R. G., & Schemske, D. W. (2005). Rapid speciation and the evolution of hummingbird pollination in neotropical Costus subgenus Costus (Costaceae): evidence from nrDNA ITS and ETS sequences. American Journal of Botany, 92(11), 1899-1910.

Latvis, M., Jacobs, S. J., Mortimer, S. M., Richards, M., Blischak, P. D., Mathews, S., & Tank, D. C. (2017). Primers for Castilleja and their utility across Orobanchaceae: II. Single-copy nuclear loci 1. *Applications in plant sciences*, *5*(9), 1700038.[a]

Latvis, M., Mortimer, S. M., Morales-Briones, D. F., Torpey, S., Uribe-Convers, S., Jacobs, S. J., ... & Tank, D. C. (2017). Primers for Castilleja and their utility across Orobanchaceae: I. Chloroplast primers 1. *Applications in plant sciences*, *5*(9), 1700020.[b]

Magoč, T. and S. L. Salzberg. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27: 2957–2963.

Minin, V., Abdo, Z., Joyce P., Sullivan, J., (2003). Performance-Based Selection of Likelihood Models for Phylogeny Estimation, *Systematic Biology*, 52(5), 674-683

Parins-Fukuchi, C. (2017). Use of Continuous Traits Can Improve Morphological Phylogenetics. *Systematic biology*, *67*(2), 328-339.

Rambaut, A., Drummond, A. J., & Suchard, M. (2013). Tracer v1. 6—MCMC trace analysis package. *Institute of Evolutionary Biology, University of Edinburgh, UK*.

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217-223.

Rothfels, C. J., F.-W. Li, AND K. M. Pryer. (2017). Next-generation polyploid phylogenetics: rapid resolution of hybrid polyploid complexes using PacBio single-molecule sequencing. New Phytologist 213: 413–429.

Smith, S. A., & Dunn, C. W. (2008). Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics*, *24*(5), 715-716.

Smith, S. D., Ané, C., & Baum, D. A. (2008). The role of pollinator shifts in the floral diversification of Iochroma (Solanaceae). Evolution: International Journal of Organic Evolution, 62(4), 793-806.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312-1313.

Sukumaran, J., & Holder, M. T. (2010). DendroPy: a Python library for phylogenetic computing. Bioinformatics, 26(12), 1569-1571.

Swofford, D. (2000). PAUP*: phylogenetic analysis using parsimony and other methods (software). Sunderland, MA: Sinauer Associates.

Tank, D. C., & Olmstead, R. G. (2008). From annuals to perennials: phylogeny of subtribe Castillejinae (Orobanchaceae). American Journal of Botany, 95(5), 608-625.

Turner, B. L. (1993). Two new species of Lamourouxia (Scrophulariaceae) from Mexico. Phytologia (USA).

Uribe-Convers, S., Settles, M. L., & Tank, D. C. (2016). A phylogenomic approach based on PCR target enrichment and high throughput sequencing: Resolving the diversity within the South American species of Bartsia L.(Orobanchaceae). *PLoS one*, *11*(2), e0148203.

Uribe-Convers, S., & Tank, D. C. (2015). Shifts in diversification rates linked to biogeographic movement into new areas: An example of a recent radiation in the Andes. *American Journal of Botany*, *102*(11), 1854-1869.

Yu, W. B., Liu, M. L., Wang, H., Mill, R. R., Ree, R. H., Yang, J. B., & Li, D. Z. (2015). Towards a comprehensive phylogeny of the large temperate genus Pedicularis (Orobanchaceae), with an

emphasis on species from the Himalaya-Hengduan Mountains. *BMC plant biology*, *15*(1), 176.

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC bioinformatics*, *19*(6), 153.

Zwickl, D. J. (2006). GARLI: genetic algorithm for rapid likelihood inference. http://www.bio.utexas.edu/faculty/antisense/garli/Garli.html.

Table 1.1. Voucher information for Lamourouxia samples included in this study. Localities denoted with a * were estimated from the locality description on the herbarium label. Sample IDs denoted with # are samples used in scaling the branches of species topology using the nrDNA ITS region.

Sample ID	Accession	Year Collected	Herbarium ID	latitude	longitude
*Lamourouxia_brachyantha*_17031#	Rzedowski 20951	1965	CAS	19.53437	-99.129962*
*Lamourouxia_colimae*_17051#	Santana Cervantes 900	1984	MEXU	19.106297	-103.773915*
*Lamourouxia_dasyantha*_16732	Ventura y Lopez 8822	1990	FMNH	21.244013	-100.520677*
*Lamourouxia_dasyantha*_16733#	Salinas 7628	1993	CAS	17.816596	-97.5334
*Lamourouxia_dasyantha*_16735	Tenorio 21521	2001	CAS	17.96661	-97.366486
*Lamourouxia_dependens*_16632#	Quedensley 2794	2006	CAS	14.76105	-91.4697
*Lamourouxia_dispar*_17052#	Torres 7759	1985	MEXU	18.027055	-100.55351*
*Lamourouxia_integerrima*_17027	Santiz Ruiz 550	1988	CAS	18.027055	-100.55351*
*Lamourouxia_integerrima*_17039#	Breedlove 42086	1976	MEXU	16.139204	-91.913967*
*Lamourouxia_integerrima*_17040	Breedlove 29232	1972	MEXU	16.907591	-92.788673*
*Lamourouxia_integerrima*_17075	Breedlove 23115	1971	FMNH	17.214984	-92.968731*
*Lamourouxia_integerrima*_17093	Breedlove 49696	1981	MEXU	16.101494	-91.677022*
*Lamourouxia_lanceolata*_16639#	Williams 40329	1969	CAS	15.343533	-90.282636*
*Lamourouxia_lanceolata*_16678	Montalvo 6433	1996	CAS	13.737164	-89.289611*
*Lamourouxia_lanceolata*_17054	Martinez 8604	1984	MEXU	16.828793	-93.09439*
*Lamourouxia_lanceolata*_17055	Veliz 7748	2000	MEXU	14.567782	-90.690169*
*Lamourouxia_lanceolata*_17090	Ventura y Lopez 4285	1987	MEXU	15.086805	-92.082388*
*Lamourouxia_lanceolata*_17091	Lavin 4519	1984	MEXU	16.694207	-92.765396*
*Lamourouxia_lanceolata*_17096	Schwalbe 1976	1976	MEXU	14.66777	-91.191188*
*Lamourouxia_longiflora*_16684#	Breedlove 61265	1984	CAS	27.353559	-107.505314*
*Lamourouxia_macrantha*_17061	Rzedowski 54019	2002	MEXU	21.093516	-100.134308*
*Lamourouxia_macrantha*_17066	Rzedowski 44452	1987	MEXU	20.923438	-100.17604*
*Lamourouxia_macrantha*_17077#	Gonzalez 207	2008	MEXU	19.594444	-103.587222
*Lamourouxia_multifida*_16691	Ventura A 5656	1972	CAS	19.898256	-96.590751*
*Lamourouxia_multifida*_16692	Garcia y Perez 3248	1990	FMNH	19.614585	-102.043485*
*Lamourouxia_multifida*_16698	Mejia E 695	1995	CAS	16.801503	-92.335917
*Lamourouxia_multifida*_16712	Brett 454	1991	CAS	16.885578	-92.371716*
*Lamourouxia_multifida*_16746	Breedlove 60079	1983	CAS	16.948283	-96.846936*
*Lamourouxia_multifida*_17056#	Corral 2089	2011	MEXU	19.121389	-100.113833
*Lamourouxia_multifida*_17078	Calderon (32) fb 33	2003	MEXU	14.729522	-91.29719*
*Lamourouxia_multifida*_17081	Gonzalez-Espinosa 1647	1991	MEXU	16.727361	-92.78077*
*Lamourouxia_multifida*_17097	Duran 4380	1999	MEXU	17.416667	-100.083889
*Lamourouxia_nelsonii*_17033	Tenorio 14151	1988	FMNH	18.716667	-97.383333
*Lamourouxia_pringlei*_16737#	Tovar 6407	1991	CAS	18.436811	-97.164731

*Lamourouxia_pringlei*_17050	Reyes S 2691	1990	MEXU	17.436761	-97.443154*
*Lamourouxia_pringlei*_17062	Hunn Oax-599	1996	MEXU	16.30398	-96.006495
*Lamourouxia_rhinanthifolia*_17057#	Panero 4094	1994	MEXU	17.193167	-97.813244
*Lamourouxia_rhinanthifolia*_17058	Campos V 3802	1991	MEXU	16.683333	-95.516667
*Lamourouxia_stenoglossa*_16738#	Chame 525	1994	CAS	16.813056	-92.290833
*Lamourouxia_stenoglossa*_16739	Martinez-Ico 98	1994	CAS	16.745833	-92.696111
*Lamourouxia_stenoglossa*_16748	Breedlove 55585	1981	CAS	16.746335	-92.699138*
*Lamourouxia_sylvatica*_16680	Binder 192	1999	CAS	-7.909628	-78.556757*
*Lamourouxia_sylvatica*_16719#	Dostert 140	1998	CAS	-7.242278	-78.375099*
*Lamourouxia_sylvatica*_16721	Binder 83	1999	CAS	-7.374345	-78.799223*
*Lamourouxia_sylvatica*_16730	Dostert 32	1998	CAS	-6.898907	-78.107774*
*Lamourouxia_sylvatica*_16731	Rodriguez 2223	1999	CAS	-7.37019	-78.806119*
*Lamourouxia_virgata*_16706	Zak and Jaramillo 3409	1988	CAS	0.390006	-77.85804
*Lamourouxia_virgata*_16744#	Zak and Jaramillo 3387	1988	CAS	0.358159	-77.83277
*Lamourouxia_virgata*_17086	Uribe-Convers 088	2011	UI	-1.546138889	-78.50661111
*Lamourouxia_viscosa*_16640#	Espinosa-Jimenez 573	2008	CAS	16.792917	-93.078389
*Lamourouxia_viscosa*_16671	Cascante 1205	1996	CAS	9.911111	-84.283333
*Lamourouxia_viscosa*_16710	Mathewson 2	1974	FMNH	14.755161	-91.178909*
*Lamourouxia_viscosa*_16723	Turner 15458	1983	FMNH	16.697148	-92.856073*
*Lamourouxia_viscosa*_16725	Fryxell and Lott 3256	1980	FMNH	16.736307	-92.959776*
*Lamourouxia_viscosa*_16752	Wilbur 21647	1976	CAS	9.808726	-83.861938*
*Lamourouxia_viscosa*_17049	Fishbein 2206	1995	MEXU	26.983333	-108.683333
*Lamourouxia_viscosa*_17071	Veliz 17984	2006	MEXU	15.280243	-91.100781
*Lamourouxia_xalapensis*_16685	Campos 4931	1993	CAS	17.29778	-98.186836
*Lamourouxia_xalapensis*_17038	Ventura A 19842	1982	MEXU	19.694783	-96.846494*
*Lamourouxia_xalapensis*_17044	Figueroa 293	2014	MEXU	19.466868	-101.809834
*Lamourouxia_xalapensis*_17060	Ventura A 19143	1981	MEXU	19.579184	-97.010403*
*Lamourouxia_xalapensis*_17068#	Miranda-Hernandez 708	1998	MEXU	16.862742	-96.805097*
*Lamourouxia_xalapensis*_17088	Alvarez 14701	2014	MEXU	19.519306	-100.305861

Table 1.2. Nuclear and chloroplast DNA regions used in this study. Low-copy nuclear genes and chloroplast regions are from Latvis et al. (2017a) and Latvis et al. (2017b), respectively. Nuclear regions identified as containing paralogous sequences and subsequently split into separate alignments are denoted by #. Genomic region, raw aligned length following crunch_clusters.py, and final aligned length after manual alignment and cleaning are also indicated.

Locus_Name	Genome	Region	Raw Length (bp)	Aligned Length (bp)
CS1_At1g04200_1908F_CS2_At1g04200_2395R_t1	Nuclear	COSII	473	439
CS1_At1g04200_1908F_CS2_At1g04200_2395R_t2#	Nuclear	COSII	473	446
CS1_At1g09620_1187F_CS2_At1g09620_1688R_t1	Nuclear	COSII	467	462
CS1_At1g09620_1490F_CS2_At1g09620_1989R	Nuclear	COSII	463	458
CS1_At2g26430_2136F_CS2_At2g26430_2632R_t1	Nuclear	COSII	521	457
CS1_At2g26430_2136F_CS2_At2g26430_2632R_t2#	Nuclear	COSII	521	450
CS1_At2g28390_1096F_CS2_At2g28390_1595R	Nuclear	COSII	706	432
CS1_At2g34560_688F__CS2_At2g34560_1186R_t1	Nuclear	COSII	542	451
CS1_At2g34560_688F__CS2_At2g34560_1186R_t2#	Nuclear	COSII	542	414
CS1_At2g38020_1_892F_CS2_At2g38020_1_1391R	Nuclear	COSII	680	434
CS1_At2g38020_1_1797F_CS2_At2g38020_1_2297R_t1	Nuclear	COSII	497	429
CS1_At2g38020_1_1797F_CS2_At2g38020_1_2297R_t2#	Nuclear	COSII	497	439
CS1_At3g04260_147F_CS2_At3g04260_646R_t2	Nuclear	COSII	594	479
CS1_At3g09920_1316F_CS2_At3g09920_1859R_t1	Nuclear	COSII	514	437
CS1_At3g09920_1316F_CS2_At3g09920_1859R_t2#	Nuclear	COSII	514	452
CS1_At3g62010_2_1425F_CS2_At3g62010_2_1926R_t1	Nuclear	COSII	491	457
CS1_At3g62010_2_1425F_CS2_At3g62010_2_1926R_t2#	Nuclear	COSII	491	456
CS1_At4g24190_1278F_CS2_At4g24190_1777R_t1	Nuclear	COSII	659	434
CS1_At4g24830_948F_CS2_At4g24830_1447R_t1	Nuclear	COSII	500	427
CS1_At4g24830_948F_CS2_At4g24830_1447R_t2#	Nuclear	COSII	500	433
CS1_At5g26360_1312F_CS2_At5g26360_1811R_t1	Nuclear	COSII	470	464
CS1_At5g26360_1312F_CS2_At5g26360_1811R_t2#	Nuclear	COSII	470	453
CS1_At5g27620_538F_CS2_At5g27620_1337R	Nuclear	COSII	635	472
CS1_At5g46630_2_851F_CS2_At5g46630_2_1350R_t1	Nuclear	COSII	478	455

CS1_At5g49970_503F_CS2_At5g49970_998R	Nuclear	COSII	490	314
CS1_At5g48300_2_1121F_CS2_At5g48300_2_1621R	Nuclear	COSII	507	458
CS1_At5g52210_456F_CS2_At5g52210_955R	Nuclear	COSII	487	442
CS1_Cas_121734_F_CS2_Cas_122486_R	Chloroplast	ndhA-ndhI	507	443
CS1_Cas_13394_F_CS2_Cas_14062_R	Chloroplast	atpA-atpF	577	469
CS1_Cas_20009_F_CS2_Cas_20813_R	Chloroplast	rpoC2	494	458
CS1_Cas_21290_F_CS2_Cas_22036_R	Chloroplast	rpoC2	481	454
CS1_Cas_21932_F_CS2_Cas_22735_R	Chloroplast	rpoC2-rpoC1	511	470
CS1_Cas_25017_F_CS2_Cas_25720_R	Chloroplast	rpoC1-rpoB	525	498
CS1_Cas_32159_F_CS2_Cas_32745_R	Chloroplast	psbM-trnE	547	468
CS1_Cas_59866_F_CS2_Cas_60624_R	Chloroplast	accD	563	497
CS1_Cas_85146_F_CS2_Cas_85791_R	Chloroplast	rps3-rpl22	552	471
CS1_Cas_90084_F_CS2_Cas_90885_R	Chloroplast	ycf2	506	461
CS1_Cas_94709_F_CS2_Cas_95300_R	Chloroplast	ycf2	556	430
ETSB_CS1_18S_IGS_CS2	rDNA	ETSB-18S_IGS	498	434
ITS5*_CS1_ITS2*_CS2	rDNA	ITS	321	294
ITS3*_CS1_ITS4__CS2	rDNA	ITS	405	379

Table 1.3. Discrete characters used by Ernst (1972) for the infrageneric classification in *Lamourouxia*. Phylogenetic half-life as calculated by Parins-Fukuchi (2017) for discrete traits are presented for two species tree hypotheses (ASTRAL-III and SVDquartets). The best fit Mk model for each trait, p-value from LR test, and transition rates are also indicated (ER = equal rates, ARD = all rates different).

	ASTRAL – III				SVDquartets			
Trait	Phylo. Half Life	mkModel	p-val	Transition rate	Phylo. Half Life	mkModel	p-val	Transition rate(s)
Leaf Division	2.8004	ER	0.3080	0.12376	2.8170	ER	0.3540	0.1230315
Corolla Hairs	1.3227	ER	0.0720	0.2620164	1.2862	ER	0.0651	0.2694629
Branching								
Anther Fertility	1.2023	ER	0.4831	0.2882672	1.2406	ER	0.7280	0.2793707
Stamen Morphology	1.1900	ER	0.7566	0.2912336	1.2202	ER	0.3801	0.2840374
Corolla Hairs	0.8359	ER	0.4822	0.4146161	0.8162	ER	0.6007	0.4246427
Glan/Not								
Leaf Margins	0.8099	ER	0.3154	0.4279323	0.7818	ER	0.2956	0.4433169
Flower Color	0.7302	ER	0.6769	0.4746289	0.7508	ER	0.5318	0.4616172
Glab./Pubes. Calyx	0.3530	ER	0.2516	0.9816914	0.3314	ER	0.1909	1.045658
Leaf Shape	0.0682	ER	0.6323	5.081187	0.0121	ER	0.6370	28.60385
Leaf Pubescence	0.1095	ER	0.6154	3.165384	0.0120	ER	0.6370	28.8237
Inflorescence Tips	0.7948	ER	0.1005	0.4360636	0.0059	ARD	0.0499	98.09226, 19.61849

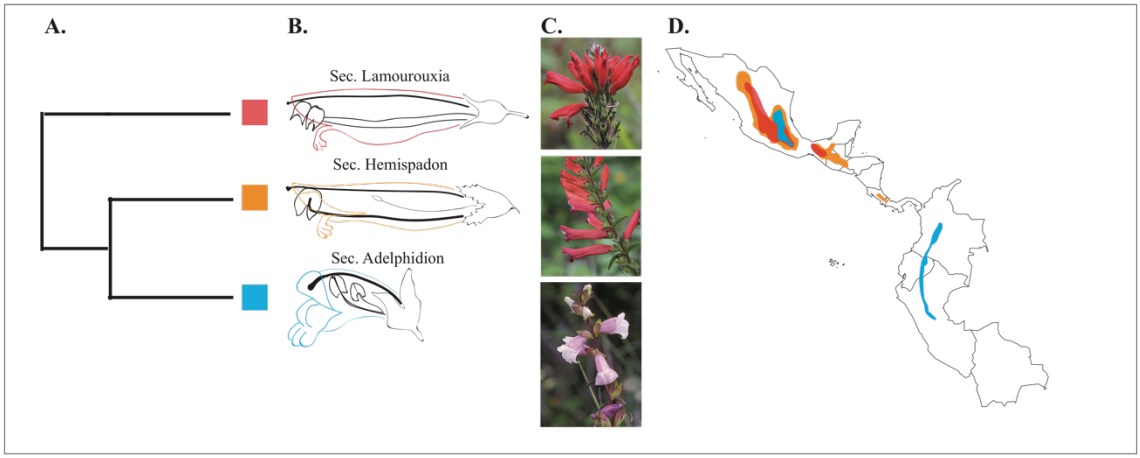

Figure 1.1. A. Evolutionary hypothesis of *Lamourouxia* based on morphological evidence from Ernst 1972. B. General morphology of the three taxonomic sections proposed by W.R. Ernst in Floral Morphology and Systematics of Lamourouxia (Scrophulariaceae: Rhinanthoideae). C. Images of representative taxa of each taxonomic section of *Lamourouxia* (top: *L. multifida,* middle: *L. viscosa,* bottom *L. sylvatica*; photos courtesy of J. M. Egger). D. Geographic ranges of sections Lamourouxia (red), Hemispadon (orange), and Adelpihidion (blue), adapted from Ernst (1972).

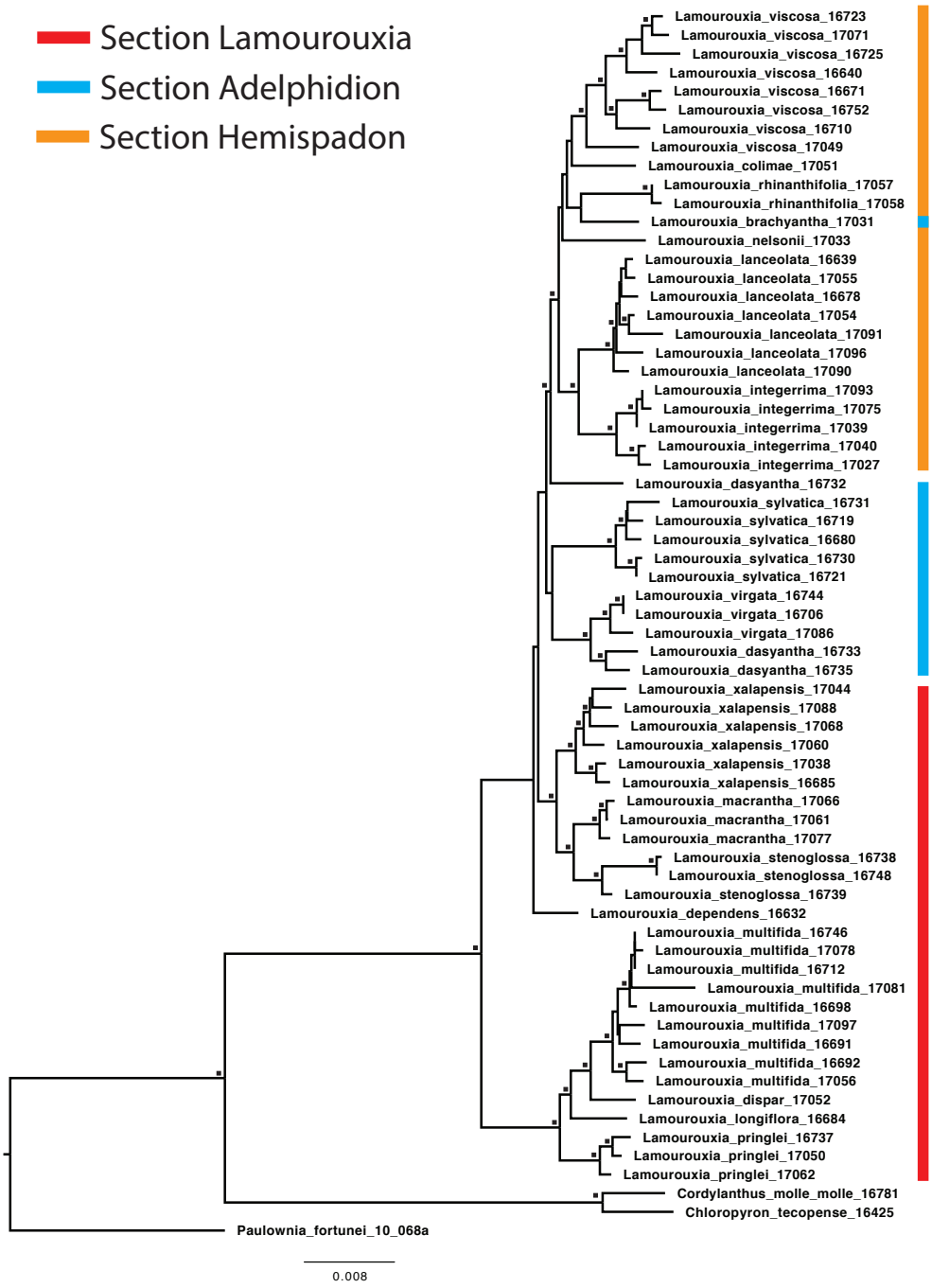

Figure 1.2. Maximum likelihood tree of the concatenated dataset of 27 single-copy nuclear loci, two nuclear ribosomal DNA regions, and 11 chloroplast DNA regions. Branches with squares denote maximum likelihood bootstrap support ≥ 75%. Colored bars represent taxa from each of each of the three taxonomic groups proposed by Ernst (1972). Branch lengths are proportional to substitutions/site, as indicated by the scale bar.

Figure 1.3. Heat map showing the taxon by region coverage for molecular data collected in this study. Rows represent individual samples labeled on the left, and columns represent targeted DNA regions. Dark blue cells represent a sample that was successfully amplified from a given DNA region and used in this study.

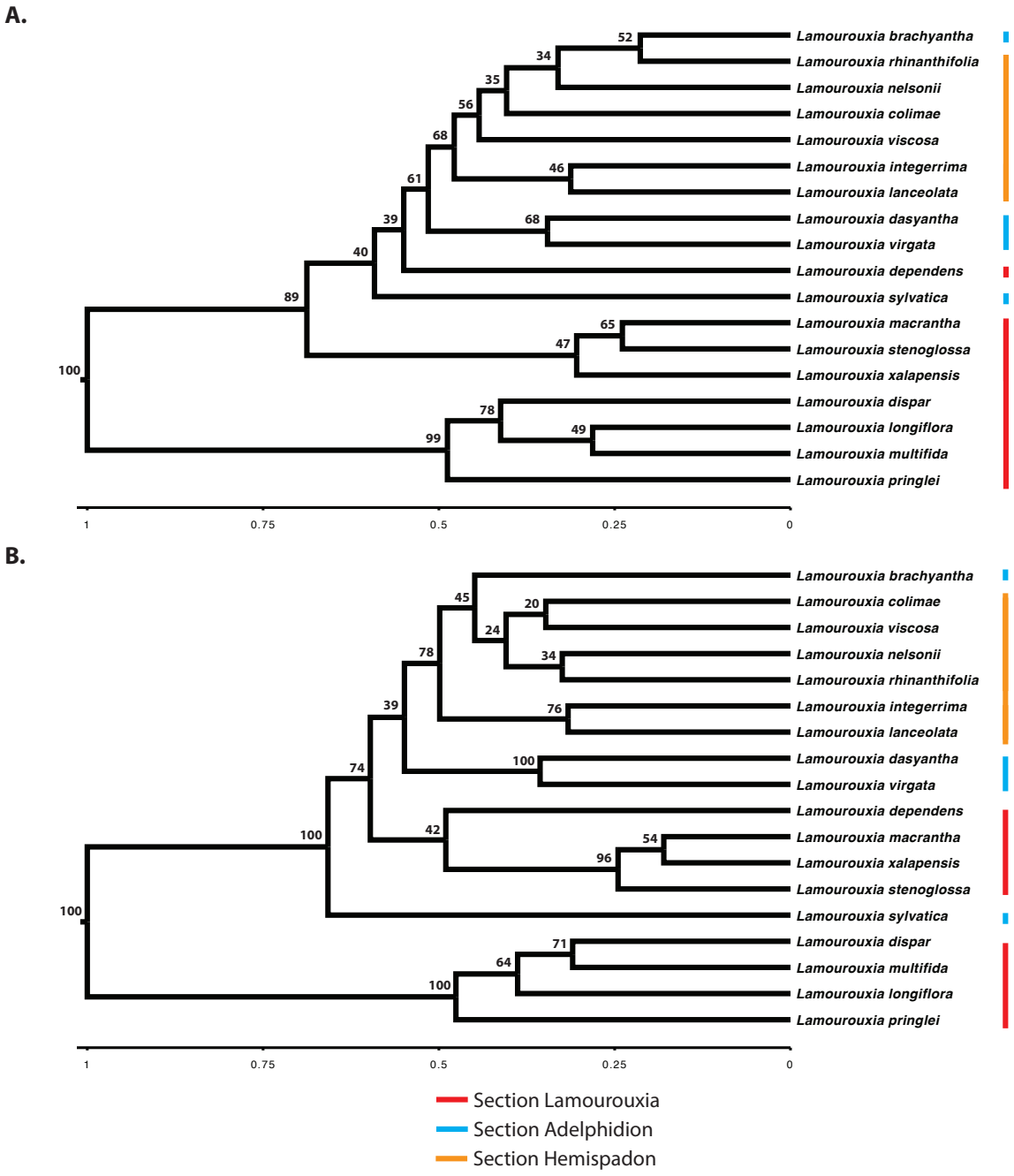

Figure 1.4. Coalescent-based species tree analyses of 18 species of *Lamourouxia*. Relative branch lengths were scaled by setting the crown node of the genus to 1.0 using a narrow normal distribution with mean = 1.0, and standard deviation = 0.001. Colored bars denothe the taxonomic designation of each species based on Ernst (1972): orange = section Hemispadon, blue = section Adelphidion, red = section Lamourouxia. (A) ASTRAL-III analysis. Branch support values are ASTRAL posterior probabilities (quartet support) of the primary topology. (B) SVDquartets analysis. Branch support values are bootstrap proportions from 100 bootstrap replicates.

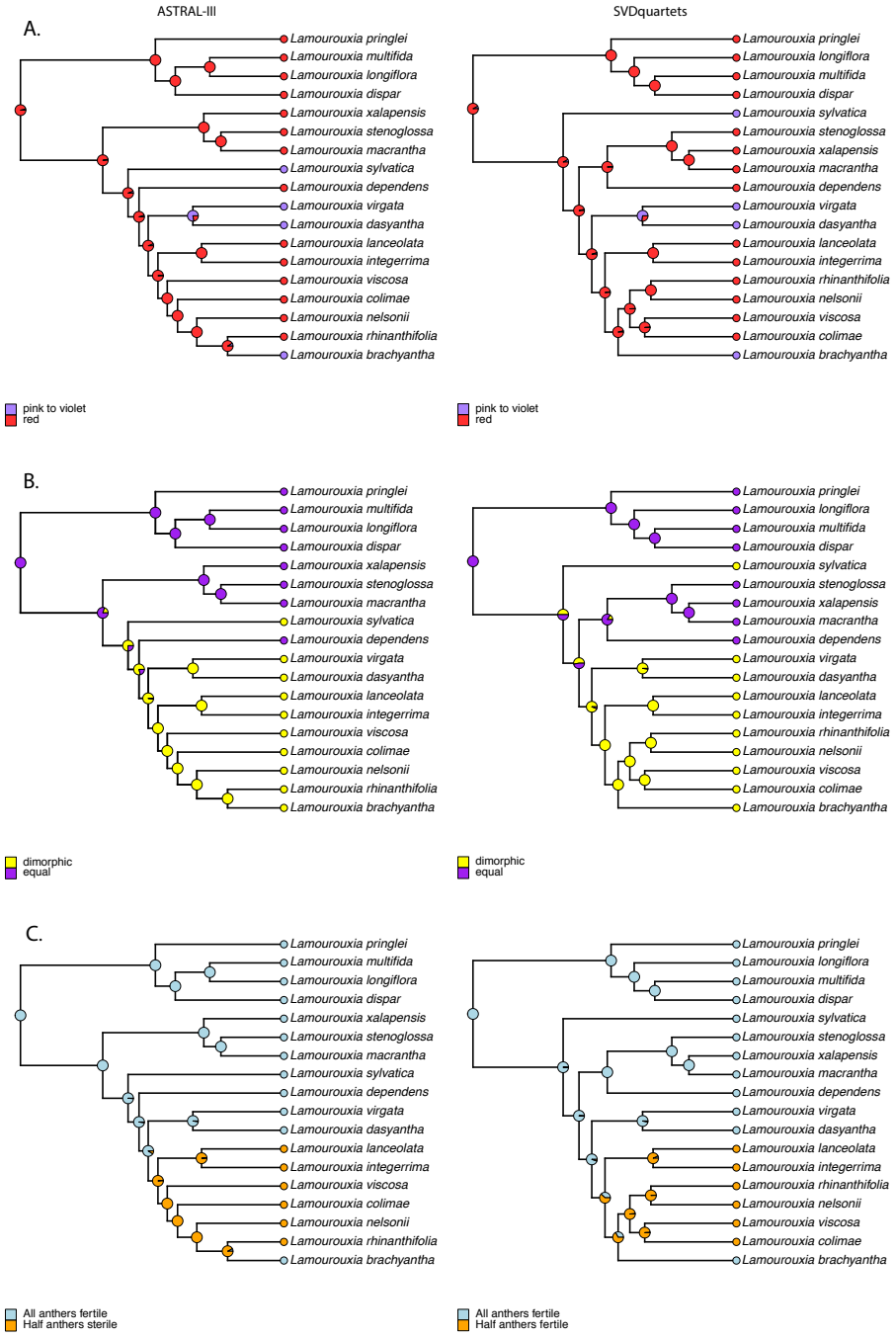

Figure 1.5. Ancestral state reconstructions of three binary taxonomic characters. Pie charts represent certainty of the reconstructed state at a particular node. A. Flower color red or pink to violet, B. staminate morphology pink or violet, C. fertility of staminate structures. Ancestral state reconstructions of all characters evaluated here can be found in Appendix A.

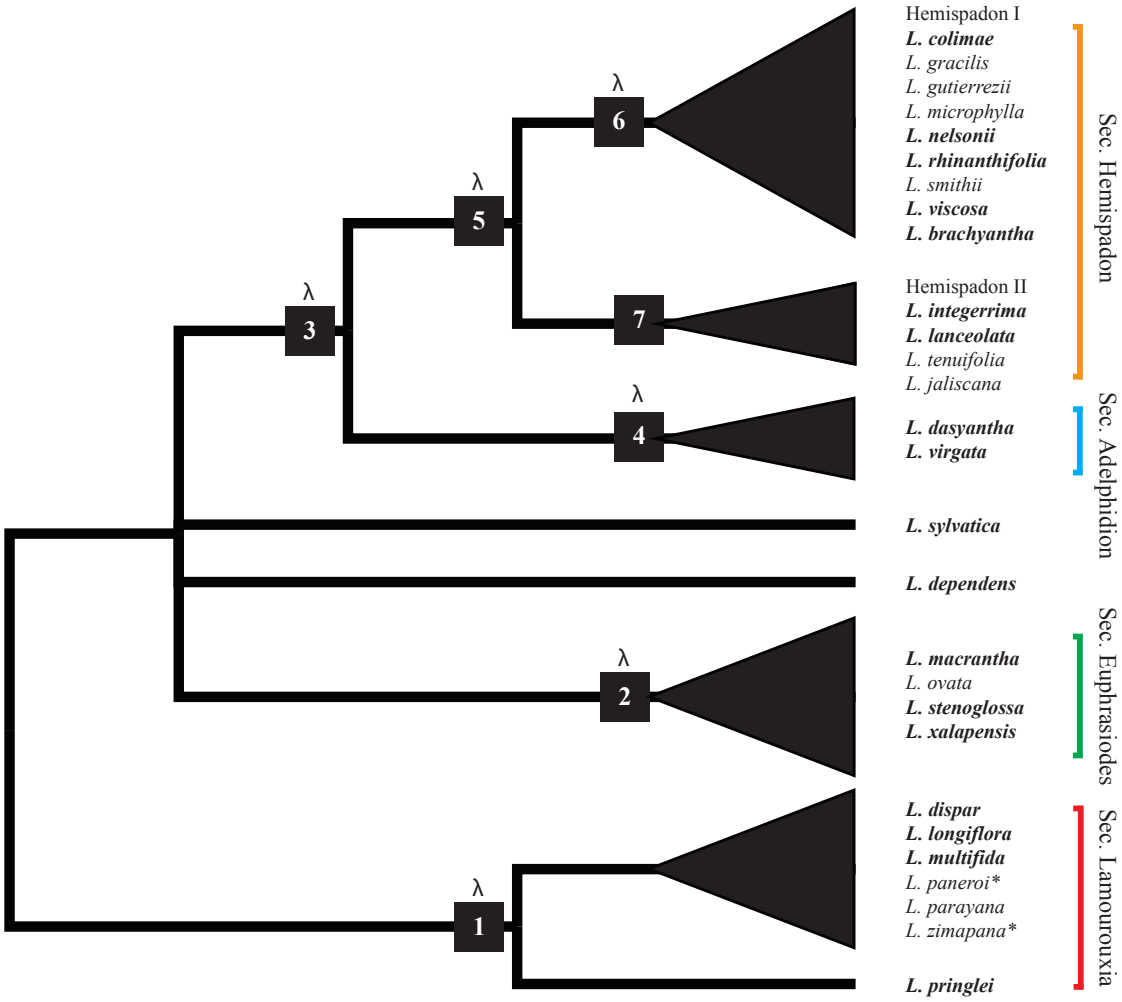

Figure 1.6. Consensus tree of the two coalescent-based species tree analyses. Sampled species (bold) and unsampled species (not bolded) are grouped adjacent to clades represented in our analysis. Species names are bracketed by our revised taxonomic circumscriptions of major clades. Synapomorphies describing major clades of *Lamourouxia* are denoted with numbered squares on branches, and where traits have high phylogenetic half-lives (Table 3), these are preceded by (1) Four fertile anthers, equal to subequal stamens, corolla red, corolla hairs glandular and often simple. (2) Four fertile anthers, equal to subequal stamens, corolla red, corolla hairs eglandular and unbranched. (3) Stamens strongly subequal, anthers strongly dimorphic, longer filaments thickened, corolla red or otherwise. (4) Corolla lavender, pink, or magenta, stamens strongly subequal, anthers all fertile and strongly dimorphic. (5) Stamens strongly subequal, anthers of reduced stamens not fertile (except *L. brachyantha*), corolla red or with flecks of yellow/orange (except *L. brachyantha*). (6) Leaves sessile and/or rounded, truncate, ovate, elliptical to obovate or cordate at base. (7) Leaves lanceolate to subulate. All species are included in the monograph by Ernst (1972) except for *L. paneroi* and *L. zimapana*, which were described by Turner (1993) as allied with *L. pringlei.*

Chapter 2: What is the current extent of our systematic knowledge? A case study in synthesizing phylogenetic data with large trees

Abstract

What is the extent of our current systematic knowledge of the tree of life? How can we identify the gaps in our understanding of phylogenetic relationships? In recent years, the field of molecular systematics has amassed a significant amount of systematic and molecular data spanning a range of large and small phylogenetic scales. Here, we utilize a recently developed set of tools for compiling and synthesizing publicly available data for molecular phylogenetic studies. These methods assess where the gaps in our phylogenetic knowledge exist in the angiosperm clade Orobanchaceae, summarize taxonomic conflicts and inconsistencies, and identify taxonomic and molecular gaps in our current sampling. Finally, as an example of the types of macroevolutionary questions that are possible given the newly synthesized tree, we perform an exploratory diversification rate analysis and identify several shifts in diversification rate that provide important macroevolutionary hypotheses for future studies.

Introduction

How much do we know about the phylogenetic history of the major branches of the angiosperm tree of life? How can we best synthesize existing knowledge to identify gaps in our understanding of phylogenetic relationships? As more and more sequence data accumulate in databases and methods for the reconstruction of large phylogenetic trees advance, we are at a point where we can begin to carry out comprehensive analyses that span multiple studies and provide a way to address these fundamental questions. Recent advances in molecular systematics on a large scale have established new knowledge about evolutionary relationships and macroevolutionary processes that affect the diversity of angiosperms (e.g., Zanne et al., 2015; Tank et al., 2015; Landis et al., 2018; Magallón et al., 2019) and seed plants (Smith & Brown, 2018). Large scale phylogenetic approaches have gained recent popularity for answering macroevolutionary questions across angiosperms, such as the radiation of angiosperms into freezing environments (Zanne et al., 2015), C4 evolution of grasses (Edwards & Smith, 2010), diversification rates in angiosperms (Tank et al., 2015; Landis et al., 2018; Magallón et al., 2019), and plant habit in campanulid angiosperms (Beaulieu, O'Meara, & Donoghue, 2013). These large-scale studies synthesize non-homogeneous molecular datasets and contribute significantly to our phylogenetic knowledge of extant plant diversity.

Concomitantly, the field of molecular systematics has contributed a significant amount of informative molecular data that span multiple phylogenetic scales and is housed in publicly accessible databases (e.g., NCBI GenBank). Likewise, new tools are continuously being developed for compiling and synthesizing these data for molecular phylogenetic studies (Pearse & Purvis, 2013; Smith et al., 2009; Freyman, 2015; Smith & Walker, 2019). However, there are inherent limitations to the phylogenetic and downstream macroevolutionary inferences that can be made using an incompletely sampled phylogenetic trees (Nabhan & Sarkar, 2012). Shedding a quantitative light on the limits of taxonomic sampling is critical for revealing the limitations of phylogenetic and macroevolutionary inferences about a clade of interest (Folk et al., 2018). Analyses that provide clear assessment of the gaps in our phylogenetic knowledge allow us to strategically direct future phylogenetic efforts with respect to both taxonomic and genetic sampling in specific clades.

To date, little is known about the macroevolutionary processes affecting the strikingly diverse angiosperm family Orobanchaceae. Orobanchaceae (*sensu* APG IV; Angiosperm Phylogeny Group, 2016; Chase et al., 2016; Stevens, 2001) is a clade of 108 genera and an estimated 2310 species of predominately parasitic plants in the Lamiales with a cosmopolitan distribution. Several clade-wide phylogenetic studies have established a broad-scale phylogenetic framework for Orobanchaceae (Young et al., 1999; Bennett & Mathews, 2006; McNeal et al., 2013; Fu et al., 2017). However, these studies have employed representative sampling strategies which include less than 10% of the described diversity in the clade, thus limiting clade-wide macroevolutionary analyses that require (more) complete taxon sampling (e.g., diversification rates). Many individual clades have been the focus of molecular phylogenetic studies primarily aimed at elucidating phylogenetic relationships at the species level. These have greatly increased our knowledge of relationships within the subclade Orobancheae [e.g., Orobanche s.l.:(Manen et al., 2004); (Schneeweiss et al., 2004); (Schneider, 2016)]; Euphrasia: (Gussarova et al., 2008); Pedicularis: (Ree, 2005; Eaton & Ree, 2013); (Eaton et al., 2012); (Tkach et al., 2014); (Yu et al., 2015); Castillejinae: [(Tank & Olmstead, 2008); (Tank et al., 2009); Castilleja: (Tank & Olmstead, 2009); (Jacobs et al., 2018); Chloropyron:(Gilman & Tank, 2018)]; Buchnereae: (Morawetz et al., 2010); Rhinantheae: (Scheunert et al., 2012); (Uribe-Convers & Tank, 2015); Pterygiella complex: (Dong et al., 2013); Neobartsia: (Uribe-Convers & Tank, 2016); (Uribe-Convers et al., 2016). However, no comprehensive phylogenetic analyses have been conducted that include both a wide generic sampling and a large sampling of species. Given our broad-scale phylogenetic knowledge and the recent surge in phylogenetic effort in individual clades, Orobanchaceae provides an excellent opportunity to synthesize taxonomic and phylogenetic

knowledge across the family, with the goal of evaluating where future phylogenetic work should be focused. In addition, with a comprehensive phylogenetic hypothesis for Orobanchaceae, it will be possible to provide a more global assessment of broad macroevolutionary questions including the evolution of hemi- and holoparasitism, the molecular evolutionary consequences of the parasitic habit, the evolution of host-specificity, shifts in the rate of diversification across the clade, and historical biogeography at both broad and fine-scales.

In this paper, we summarize the evolutionary relationships and divergence times between genera and major clades within Orobanchaceae. Using available sequence data from strategically sampled chloroplast and nuclear gene regions that have been employed for phylogenetic analyses at various taxonomic levels in the clade, we assembled a supermatrix of more than 900 species of Orobanchaceae that represents approximately 40% of the known species diversity of the family. Our analyses of this large Orobanchaceae dataset largely confirm phylogenetic relationships revealed in earlier studies, but importantly these analyses provide a clear assessment of where the gaps in our phylogenetic knowledge exist, and allow us to strategically direct future phylogenetic efforts with respect to both taxonomic and genetic sampling in the clade. We summarize taxonomic conflicts and inconsistencies of our tree as compared to previously published phylogenies, and describe these gaps in our phylogenetic knowledge based on publicly available taxonomic data. Finally, we perform an exploratory diversification rate analysis as an example of the types of macroevolutionary questions that are possible given the greatly increased taxonomic sampling.

Methods and Materials

Dataset and phylogeny construction. Following (Beaulieu & O'Meara, 2018) the building of the final super-matrix was an iterative process that included manual curation and addition of ingroup and outgroup sequences, correction of sequence identifiers, and removal of erroneous sequences. First, we built a copy of the GenBank plant division (pln) with phlawd_db_maker (github.com/blackrim/phlawd_db_maker) on September 13, 2018. Subsequently, we ran a clustering analysis of the plant family Orobanchaceae using PyPHLAWD (https://github.com/FePhyFoFum/pyphlawd), making several changes to the default settings of the configuration file: smallest_size = 400, sampsize = 10, length_limit = 0.50, and filternamemismatch = False.

To find individual loci for concatenation, we ran the *find_good_clusters.py* script with settings for smallest_cluster = 10 and cluster_prop = 0.089. We selected 12 clusters based on non-

overlapping gene regions (Table 1). To add an extra-familial outgroup to the analysis we ran PyPHLAWD again to mine sequences of Paulowniaceae, which is a monogeneric family known to be sister to Orobanchaceae (Albach et al., 2009; McNeal et al., 2013; Refulio-Rodriguez and Olmstead, 2017; Fu et al. 2017). We recovered a total of five clusters, two were portions of the nuclear ribosomal (nr) DNA ITS region, a single cluster for the chloroplast (cp) DNA regions *rbcL*, the *trnL* intron and *trnL-trnF* intergenic spacer, and the *rps16* gene. Separate clusters of ITS sequences were viewed in ALIVIEW v1.18.1 (Larsson, 2014), aligned using MUSCLE v.3.8. (Edgar, 2004), as implemented in ALIVIEW, and concatenated manually. Additional sequences from known *Paulownia* (Paulowniaceae) species not found in our PyPHLAWD analysis were manually downloaded from GenBank for *Paulownia tomentosa* for three additional cpDNA gene regions (*matK*, *psbA*, and *rps2*) and two additional nuclear gene regions (PHYA and PHYB). All outgroup sequences were curated and trimmed to length by hand, then aligned using MAFFT v.7.407 (Katoh & Standley, 2013) using the –add flag to keep the existing alignments from PyPHLAWD intact. Lastly, a single species of Orobanchaceae, *Radamaea montana*, that was sampled in the five-gene analysis of McNeal et al., (2013) was conspicuously missing from the matrix. This species was also added to the analysis using manually curated sequences, and aligned with MAFFT v7.407 using the –add option for five gene regions (Table 1). Prior to phylogenetic analyses, missing sites were removed at a threshold of 50% using the –clean function in phyutility v.2.2 (Smith & Dunn, 2008).

We built preliminary gene trees for each cluster, and a preliminary concatenated tree using RAxML v.8.2.9 (Stamatakis, 2014) using the GTRGAMMA option for each by-gene partition and 100 bootstrap replicates with the –fa option. Using the preliminary gene trees and the concatenated tree, we identified and removed erroneous sequences and data-overlap issues in the dataset following criteria from Beaulieu and O'Meara (2018) (Table 2; taxa whose placement was erroneous in the ML estimates due to non-overlapping genetic sampling with congeneric species, paralogous sequences of phytochrome A and phytochrome B, and erroneously gappy sequences or short sequences that caused excessively long branches in the resulting maximum likelihood estimate). Furthermore, to represent each species as a single tip we deleted conspecific sequences of different subspecific taxa (Table 2). Additionally, we created 18 chimeric tips by concatenating conspecific sequences with congruent subspecific taxa – e.g., we combined the five sequences of *Aphyllon californicum* (*ITS, PHYA, PHYB, rbcL*, and *rps2*) with a single *trnL-trnF* sequence from *Aphyllon californicum* ssp. *californicum* (Table 3).

Using this manually curated concatenated alignment, a final tree was built using RAxML v.8.2.9 using the GTRGAMMA option for each by-gene partition and 100 bootstrap replicates with the –fa option (Appendix A). The monophyly of genera in the tree was analyzed using MonoPhy (Schwery & O'Meara, 2016) in R (R Core Team, 2017). We found 13 genera to be non-monophyletic, namely, *Alectra, Aphyllon, Bellardia, Boschniakia, Cycnium, Euphrasia, Graderia, Harveya, Melasma, Nesogenes, Odontites, Orobanche,* and *Phelipanche*, which highlighted multiple tips with deprecated names. We changed names of eight tips to reflect their updated taxonomic status (Table 4), reducing the number of non-monophyletic genera to eight (Table 5).

Taxonomic and Genetic sampling. To quantify sampling effort across the phylogeny, we compared the number of species sampled in our phylogenetic analyses to the most current synoptical classifications of Orobanchaceae Mabberley, 2017; Olmstead, 2016. Where species estimates differed from recent systematic work in a particular clade (e.g., *Orobanche*; (Schneider, 2016)) we updated our species estimates to reflect these publications. In these classifications, species numbers for large genera are often approximated with ranges (e.g., *Pedicularis* is noted to have ~500-750 species); in these cases we chose the larger number to represent the putative species diversity in the clade. To quantify and visualize the extent of genetic sampling in our study, we compared our sampling across the 12 gene regions for all genera and major clades to the known taxonomy of each clade (Tables 1 and 4).

Geographic sampling in Orobancheae, Rhinantheae, Buchnereae, and Pedicularideae. To visualize and quantify geographic gaps in the sampling of the four major Orobanchaceae clades, we compared the range of sampled species to the range of under sampled clades using localities downloaded from GBIF (GBIF.org, Downloaded 19 Oct 2018). We cleaned the data to remove erroneous locality records. We removed all records that were not categorized as "PRESERVED_SPECIMEN" and kept only records of genera represented in our current taxonomy of Orobanchaceae (108 genera). Finally, we adjusted deprecated generic names of sampled species from *Aphyllon, Kopsiopsis, Hedbergia, Bellardia, Parentucellia,* and *Neobartsia* to reflect the current taxonomy. Points associated with sampled species were mapped with points associated with unsampled species names to understand the geographic extent of unrepresented species across these clades in Orobanchaceae (Appendix B).

Divergence time estimation. We used the 'congruification' approach developed by (Eastman et al., 2013) to resolve topological inconsistencies between two previously published Orobanchaceae specific reference time-trees (Fu et al. 2017; Schneider & Moore, 2017) and our unscaled maximum

likelihood target tree from the final, concatenated matrix. Using the 'congruify' function in the R-package Geiger v.2.0 (Pennell et al., 2014), we mapped dates onto concordant nodes of our unscaled phylogeny to use as secondary calibrations. For both reference trees, we used mean node heights from the MCC tree and node heights from a random sample of 500 trees from the posterior distributions to account for topological and temporal heterogeneity and to calculate 95% confidence intervals. Divergence times were estimated using penalized-likelihood as implemented in treePL (Sanderson, 2002); (Smith & O'Meara, 2012) with the penalized-likelihood rate smoothing parameter set to 1000. This parameter was selected after 40 optimizations on the maximum likelihood tree using median node heights from the MCC trees.

While Schneider & Moore (2017) provided posterior distributions of their analyses with their supplemental materials, these were not made available by Fu et al. (2017). Therefore, we reanalysed the Fu et al. (2017) data following their procedures to produce a posterior distribution of trees for this analysis. We ran ten independent analyses; each run consisted of 25 million generations (sampling every 1,000 steps) with the GTR + G model, a Yule tree prior and an uncorrelated lognormal clock. Samples from each run were combined by LogCombiner 1.8.2 (Rambaut & Drummond, 2013). sampling every 5,000 steps. Finally, we assessed convergence using Tracer 1.6.0 (Rambaut & Drummond, 2009).

Diversification rates. To begin to understand diversification dynamics within Orobanchaceae, and as an example of the types of downstream macroevolutionary analyses that could be facilitated by a synthetic phylogeny, we used BAMM (Bayesian Analysis of Macroevolutionary Mixtures) v.2.5.0 (Rabosky, 2014; Rabosky, Mitchell, & Chang, 2017). The BAMM model simulates a posterior distribution of rate shift configurations using reversible jump Markov chain Monte Carlo to explore candidate models. Given the controversy surrounding BAMM (Moore et al., 2016) and to gage the stability of our results following (Magallón et al., 2019), we performed six analyses under different priors of the expected number of shifts (expectedNumberOfShifts = 0.1, 1.0, 1.5, 2.0, 10.0, 25.0). We ran four MCMC chains for 25 million generations sampling every 5000 generations for each of the six analyses. We discarded the first 10% of the MCMC as burn-in. We assessed the convergence of the remaining 4501 samples using BAMMtools v.2.5.0 (Rabosky et al., 2014), compared posteriors and priors, and identified a 95% credible set of rate shift configurations for each analysis. Finally, we present the maximum a posteriori (MAP) probability shift configuration for each analysis.

Results

Dataset construction. PyPHLAWD identified 50 clusters containing between three and 845 taxa (Appendix B). The *find_good_clusters.py* script identified 16 clusters from which we selected 12 gene regions, including the nrDNA ETS and ITS regions, the *psbA-trnH* intergenic spacer, the *rps16* intron, *matK*, *rbcL*, *rpl32*, and *rps2* genes, the *trnL* intron and *trnL-trnF* intergenic spacer from the plastome, and the three nuclear genes GBSSI, *PHYA* (phytochrome A), and *PHYB* (phytochrome B) (Table 1). The four clusters we did not select for downstream analyses contained overlapping or partial gene regions with our selected loci (e.g. "Pedicularis nipponica isolate nf0655 trnK gene, intron; and maturase K (matK) gene, partial cds; chloroplast." [cluster 1216] and "Orobanche ludoviciana var. arenosa isolate Colwell01-93_WTU344844 external transcribed spacer, partial sequence." [cluster 847]). We removed a number of erroneous taxa and sequences based on criteria from Beaulieu & O'Meara (2018) (Table 2).

Taxonomic sampling. Table 4 summarizes our species-level sampling for all genera of Orobanchaceae. Briefly, we sampled 39.7% of the described species diversity of Orobanchaceae, representing all eight primary clades of Orobanchaceae; Rehmannieae, *Lindenbergia*, Cymbarieae, Orobancheae, *Brandisia,* Rhinantheae, Buchnereae, and Pedicularideae. Of these clades Rehmannieae and Cymbarieae are relatively species poor, however both are well sampled in our analysis with 11 of 12 and eight of 14 species sampled, respectively. Similarly, the other small clades *Lindenberia* and *Brandisia* are represented by six of 12 and three of 13 described species, respectively. The four remaining major clades of Orobanchaceae contain ~98% of the species diversity (2,259 species). For these diverse clades we sampled between 20.6% (Buchnereae) and 54.2% (Orobancheae) of the described species diversity. In total, our sampling included 80 of 108 described genera in Orobanchaceae. Generally, the genera not represented by a single species are relatively species poor e.g., *Rhamphicarpa* (six species) and *Seymeriopsis* (one species). Furthermore, 23 of 30 unsampled genera belong to the clade Buchnereae (Table 4, Fig. 1).

Phylogenetic analyses. Relationships among the major clades and most genera were largely the same as in previous studies [e.g., McNeal et al., 2013; Fu et al., 2017; Fig. 2, Appendix B]. Our analyses found strong bootstrap support for the monophyly of the autotrophic Rehmannieae Rouy.. clade (*Rehmannia* + *Triaenophora*), as well as its sister group relationship with the remainder of Orobanchaceae. Furthermore, we found strong bootstrap support along the backbone of the phylogeny (74%-98%; Fig. 2). However, we found 13 of 80 sampled genera to be non-monophyletic (Appendix B). Subsequently, eight species names were changed to reflect updated taxonomic

treatments in Orobanchaceae: *Orobanche bulbosa* to *Aphyllon tuberosum*, *Aphyllon vallicola* to *Aphyllon vallicolum*, *Orobanche latisquama* to *Boulardia latisquama*, *Alectra alba* to *Harveya alba*, *Parastriga alectroides* to *Harveya alectroides*, *Boschniakia strobilacea* to *Kopsiopsis strobilacea*, *Orobanche nowackiana* to *Phelipanche nowackiana*. After these changes were made the number of non-monophyletic genera was reduced to eight (Table 5).

Estimation of the Geographic Extent of unsampled taxa. We downloaded 1,612,982 occurrence records for Orobanchaceae from GBIF. After cleaning the data set, we retained 7,961 records belonging to Orobancheae, 45,219 records belonging to Rhinantheae, 10,249 records belonging to Buchnereae, and 62,306 records belonging to Pedicularideae. These records contained 11 sampled genera in Orobancheae, 17 sampled genera in Rhinantheae, 67 sampled genera in Buchnereae, and 17 sampled genera in Pedicularideae (Appendix B). To estimate where unsampled species and genera are located geographically, we mapped localities of unsampled and sampled species (Fig. 3). For each clade we identified which genera contributed to the unsampled distributions of the four major clades of Orobanchaceae.

Unsampled species in Orobancheae are predominately from old world *Orobanche* and *Phelipanche* clades and the new world clade *Aphyllon*. Unsampled species of Rhinantheae are predominantly from the genera *Euphrasia, Melampyrum, Odontites, and Neobartsia,* with a noticeable gap in Europe. However, the range of *Euphrasia* extends to New Guinea, Australia, and New Zealand. The South American genus *Neobartsia* is found throughout the Andes. The predominantly tropical clade Buchnereae contains the largest number of genera not represented in our phylogeny; the African genera *Baumia*, *Buttonia*, *Cycniopsis*, and *Gerardiina*. The larger African genera of Buchnereae, *Harveya, Cycnium,* and *Cyclocheilon* are represented in our tree, but are considerably under sampled and are found primarily in sub-Saharan Africa. Other under sampled genera of Buchnereae have larger distributions across the paleo- and neotropics (*Buchnera, Striga,* and *Melasma*). The genus *Pedicularis* contains the most unsampled species in Pedicularideae. Given its large species diversity we estimate that unsampled species are still found across the range of the genus. We found that several new world genera of Pedicularideae are also considerably undersampled – especially in South and Central America. The genera *Castilleja, Agalinis, Lamourouxia, Esterhazya,* and *Aureolaria* are predominantly new world genera that are conspicuously undersampled in our phylogeny. *Agalinis* and *Esterhazya* are the primary contributors to the unsampled areas of eastern South America (Fig. 3).

Divergence Time Estimation. Our congruification approach found 87 overlapping nodes with the posterior distribution of trees from Fu et al. (2017) and 74 overlapping nodes with the posterior distribution of trees from Schneider and Moore (2017; Fig. 4). Our cross-validation analysis found our smoothing parameter to be 1000 for 25 of the 40 replicates for the Fu et al. (2017) calibrations and 34 of the 40 replicates for the Schneider and Moore (2017) calibrated trees. Based on secondary calibrations from Fu et al. (2017) we estimated the crown divergence time of Orobanchaceae to be 36.7 mya (q0.025=30.4, q0.975=47.0) while the estimated crown age of Orobanchaceae based on secondary calibrations from Schneider and Moore (2017) was 30.2 mya (q0.025=25.6, q0.975=36.4) (Appendix B). With few exceptions mean stem and crown ages of major clades within Orobanchaceae based on calibrations from Fu et al. (2017) tended to be older than those estimated using calibrations from Schneider and Moore (2017; Table 6).

Diversification rate analyses. We found that for each of our six independent analyses the posterior distribution was decoupled from prior distributions (Appendix B) and that the ESS values were above 200 for each analysis. We found that MAP shift configurations across analyses were relatively similar in their placement of rate changes (Fig. 5). The MAP configuration under the 0.1 expected shifts prior found four speciation rate shifts, while the analyses under 1.0, 1.5, 2.0, 10.0, and 25.0 expected shifts found 11, 10, 11, 11, and 10 speciation rate shifts, respectively (Fig. 5). All analyses found rate increases within the *Castilleja* + *Triphysaria* clade (Pedicularideae), in *Euphrasia* (Rhinantheae), and along the stem branch of *Pedicularis*. All but one analysis (0.1) found another rate shift within *Pedicularis* on the stem branch of a clade containing 32 species. Similarly, all analyses detected rate shifts at the stem nodes of *Phelipanche* and *Orobanche* (Orobancheae) and the clade sister to *Bartsia alpina* (Rhinantheae). Finally, all analyses except for priors 1.5 and 25.0 found a rate shift on the stem branch of (Rhinantheae (Buchnereae, Pedicularideae).

Discussion

Priorities among future sampling efforts. As with most phylogenetic studies of large clades, previous Orobanchaceae-wide phylogenies have been quite limited in the depth of their taxonomic sampling, including less than 10% of the described diversity of Orobanchaceae (e.g., Bennett and Mathews, 2006; McNeal et al., 2013). Here, we have greatly expanded sampling across Orobanchaceae to include ~ 40% of the described diversity of the clade, and although our results do not conflict with our current understanding of relationships among the major clades of Orobanchaceae, they provide an important opportunity to identify where taxonomic and genetic sampling is most needed to increase

our phylogenetic knowledge, including macroevolutionary patterns influencing the diversity of Orobanchaceae.

We are of the opinion that unsampled and undersampled tropical genera in Buchnereae require the most urgent attention in terms of taxonomic and genetic sampling. Increased phylogenetic effort in these areas of the Orobanchaceae tree would aid in the resolution of several currently paraphyletic taxa (e.g., *Cycnium* and *Melasma*), and allow for more robust comparative phylogenetic analyses (Table 5; Fig 2). The next highest priority is additional sampling of missing species-level diversity within the largest genera of Pedicularideae and Rhinantheae (e.g., *Agalinis*, *Castilleja*, and *Pedicularis* of Pedicularideae, and *Euphrasia* and *Odontites* of Rhinantheae). Although a lower priority in terms of species numbers, increased sampling of new world and old world members Orobancheae would allow for a better understanding of the diversity dynamics of the major holoparasitic clade of Orobanchaceae (Fig. 2). Finally, there are several other taxonomic inconsistencies and non-monophyletic genera that will require the attention of systematic experts to resolve. For example, we found *Aureolaria*, *Euphrasia, Odontites,* and *Orobanche* to all have intruding monotypic genera (Table 5), making them paraphyletic.

In addition to the taxonomic and genetic sampling components of this synthesis, we were able to estimate geographic ranges of unsampled taxa using GBIF data. This type of analysis is important for us to understand where, geographically, to direct sampling efforts. In Orobanchaceae, we were able to identify particular geographic areas in which the major clades are currently undersampled. Unsampled geographic diversity in the Orobancheae clade includes both old world species of *Orobanche* and *Phelipanche*, as well as new world species of *Aphyllon* (Table 4; Fig. 3). Notably, a single widespread species, *Orobanche cernua*, largely accounts for the unsampled distribution of Orobancheae in Australia (Fig. 3). The major unsampled diversity of Rhinantheae is found in the neotropical radiation of *Neobartsia* in Andean South America (Fig. 3). However, the recent taxonomic changes to *Neobartsia* and closely related genera, such as *Bartsia, Bellardia,* and *Hedbergia* (Uribe-Convers & Tank, 2016), obscures our ability to estimate geographic ranges from GBIF in some parts of this clade. Pedicularideae is notably undersampled in the neotropics (Fig. 3), and especially in eastern South America where much of the unsampled diversity is found in *Agalinis* and *Esterhazya* and in Mexico (*Castilleja*). In line with our taxonomic estimates of unsampled diversity, *Buchnereae* is undersampled throughout the paleo- and neotropics (Fig. 3), including much of tropical South America (*Buchnera* and *Melasma*), Africa, where several endemic genera are not represented in our phylogeny (*Baumia, Buttonia, Gerardiina,* and *Cycniopsis*), as well as several

widely distributed genera in Buchnereae (*Buchnera* and *Striga*) that are distributed in Africa, but are also undersampled in Indomalay and Australia (Fig. 3). Finally, as with *Orobanche cernua,* two species, *Striga curviflora* (347 localities) and *Buchnera linearis* (607 localities) are the primary contributors to the unsampled Buchnereae distributions in northern Australia (Fig. 3, Appendix B).

Macroevolutionary inferences. Because there are no known fossils in Orobanchaceae available for divergence time estimation, previous studies have relied on a number of different approaches to calibrate nodes including, molecular clock estimates (Wolfe et al., 2005), distant fossils calibrations (Fu et al., 2017), geographic calibrations (Uribe-Convers and Tank, 2015), and calibrations based on the timing of horizontal gene transfer events (Schneider and Moore, 2017). Given the recent clade-wide divergence time estimates of Fu et al. (2017) and Schneider and Moore (2017) we employed the 'congruification' approach of Eastman et al. (2013) which places numerous secondary calibrations from a source tree onto our target tree (Fig. 4) for divergence time estimation using penalized likelihood. In general, divergence time estimates using the Schneider and Moore (2017) calibrations recovered more recent divergence events (noted by Schneider and Moore, 2017). However, the 95% confidence intervals summarized from our 500 independent analyses overlap considerably between analyses (Table 6; Appendix B). Based on the greatly expanded taxonomic sampling of our phylogenetic analysis of Orobanchaceae and the lack of fossil evidence needed for primary calibrations in the clade, we consider our results to be an important resource for use in future macroevolutionary studies where temporal information on crown and/or stem ages within the clade is desired.

To demonstrate the usefulness of synthetic studies like ours, we estimated the number and timing of diversification rate shifts in Orobanchaceae. Results from six independent analyses using a range of priors for the number of expected shifts largely agreed with each other, as well as with the few clade-specific diversification rate analyses for Orobanchaceae. For example, in Castillejinae, the clade containing the paintbrushes (*Castilleja*) and related genera, Tank (2006) found an increased diversification rate associated with the evolution of perenniality and polyploidy and all six of our analyses identified significant shifts in this part of the tree (Fig. 5, shift A/A*). Likewise, in Rhinantheae, Uribe-Convers and Tank (2015) identified several shifts in diversification rate across the clade, and these are also recovered here (Fig. 5, shifts D/D* and E). In addition, we found several novel rate shifts (both increases and decreases in speciation rate) involving the large genus *Pedicularis* (Fig. 5, shifts B and C), the clade containing Rhinantheae, Buchnereae, and Pedicularideae (Fig. 5, shift G), Rehmannieae (Fig. 5, shift I), and several lineages within the

holoparasitic Orobancheae clade (Fig. 5, shifts F, H, and L). Although many of these diversification rate shifts were consistent across our six independent analyses, it is important to note that some of these conclusions may change with a completely sampled tree. Most notably, no shifts were recovered within the mostly tropical Buchnereae clade, and this is where our taxonomic sampling is least complete.

Conclusions. We have shown in our study of Orobanchaceae that a phylogenetic synthesis can provide a broad but precise overview of the state of our taxonomic and phylogenetic knowledge of larger clades. Although full automation of the process from sequence databases to phylogenetic hypothesis may be appealing, human intervention and data curation steps will likely always be an important factor in this area of systematics (Beaulieu and O'Meara, 2018). Using multiple resources, we were able to construct a robust and well-sampled phylogenetic hypothesis of Orobanchaceae. This hypothesis reflects a synthesis of data from many individual clade specific studies, allowing for both wide generic sampling and a large sampling of species. Our study reflects the most recent changes to the taxonomy of Orobanchaceae, and provides a new temporal framework for the clade that serves as a stepping-stone for future macroevolutionary studies (e.g., historical biogeography, the evolution of holo- and hemiparasitism, host specificity). Furthermore, in the process, we encountered several themes that may be broadly applicable to future synthetic studies.

Using a synthetic phylogenetic hypothesis in conjunction with a taxonomic framework we were able to quantify where our taxonomic and phylogenetic knowledge is incomplete or conflicting. We predict that this approach will aid in identifying where to focus future systematic efforts (e.g., collecting and sequencing) in any clade of interest. In Orobanchaceae, we found that the majority of missing species from our phylogeny belonged to large, diverse genera (e.g., *Pedicularis* [386 unsampled species], *Euphrasia* [270 unsampled species], *Castilleja* [112 unsampled species]), where representative sampling strategies have been employed to cover their morphological and geographic diversity in higher level phylogenetic studies. Conversely, many small genera (<10 described species) are not represented in our phylogeny by even a single species (Fig. 1, Table 4); in Orobanchaceae, most of these have tropical distributions, and point to a lack of collections and/or molecular data available for species in certain areas of the world. In a few cases, recent phylogenomic studies have developed robust species level phylogenetic hypotheses for several of the undersampled genera of Orobanchaceae that we identified here (e.g., *Pedicularis*, Eaton and Ree, 2013 and *Neobartsia*, Uribe-Convers et al., 2016). Clearly, these genomic studies are valuable for understanding the evolutionary history of these specific clades. However, the anonymity of molecular data employed in these studies

means that they cannot be used for synthetic studies, such as this one. There is the possibility that the use of amplicon-based high throughput sequencing approaches (e.g., Uribe-Convers et al., 2016, Latvis et al., 2017) or universal probe sets for hybridization based target enrichment (e.g., the Angiosperm 353 probe set; Johnson et al., 2018) will result in phylogenetic data that is compatible across individual clade-specific studies. However, this requires that these data be deposited in GenBank as with traditional phylogenetic data, which is not currently the standard practice for high throughput data collection strategies. Rather, these data are most often made available in the NCBI Sequence Read Archive (SRA) or on data repositories such as GitHub or Dryad, where they are largely inaccessible to current phylogenetic data mining pipelines, like PyPHLAWD. As high throughput sequencing approaches replace other data collection strategies for phylogenetic studies, we feel that it will be important to continue to generate – and deposit – standard sets of universal gene regions, including high copy regions of the genome that are already well represented in GenBank across the plant tree of life (e.g., chloroplast regions such as *rbcL*, *trnL-trnF*, *matK, ndhF*, and the nuclear ribosomal ITS and ETS regions), to be able to synthesize with the large amount of previously acquired sequence data.

With a few notable exceptions (e.g., Xi et al., 2014, Ran et al., 2018) phylogenetic relationships among major seed plant lineages are generally well understood. However, relationships at the tips of the spermatophyte tree of life are still largely unknown for many clades. With the development of new tools for large-scale phylogenetic studies, it has become much easier to synthesize publicly available data for systematic and macroevolutionary studies. This kind of clade-wide synthesis of multiple data types from a variety of sources (i.e., molecular, geographic, taxonomic) facilitates a better understanding of phylogenetic relationships and a global assessment of the gaps in our current phylogenetic knowledge. We feel the clade-wide synthesis approach will be a valuable tool to direct future phylogenetic efforts as systematics moves from the backbone of the tree of life to the tips.

Literature Cited

Albach, D. C., Yan, K., Jensen, S. R., & Li, H.-Q. (2009). Phylogenetic placement of Triaenophora (formerly Scrophulariaceae) with some implications for the phylogeny of Lamiales. Taxon, 58(3), 749–756.

Beaulieu, J. M., & O'Meara, B. C. (2018). Can we build it? Yes we can, but should we use it? Assessing the quality and value of a very large phylogeny of campanulid angiosperms. American Journal of Botany, 105(3), 417–432.

Beaulieu, J. M., O'Meara, B. C., & Donoghue, M. J. (2013). Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. Systematic Biology, 62(5), 725–737.

Bennett, J. R., & Mathews, S. (2006). Phylogeny of the parasitic plant family Orobanchaceae inferred from phytochrome A. American Journal of Botany. https://doi.org/10.3732/ajb.93.7.1039

Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., Soltis, D. E., ... & Stevens, P. F. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Botanical Journal of the Linnean Society, 181(1), 1-20.

Dong, L.-N., Wang, H., Wortley, A. H., Lu, L., & Li, D.-Z. (2013). Phylogenetic relationships in the Pterygiella complex (Orobanchaceae) inferred from molecular and morphological evidence. Botanical Journal of the Linnean Society. Linnean Society of London, 171(3), 491–507.

Eastman, J. M., Harmon, L. J., & Tank, D. C. (2013). Congruification: support for time scaling large phylogenetic trees. Methods in Ecology and Evolution. https://doi.org/10.1111/2041-210x.12051

Eaton, D. A. R., Fenster, C. B., Hereford, J., Huang, S.-Q., & Ree, R. H. (2012). Floral diversity and community structure in Pedicularis (Orobanchaceae). Ecology, 93(sp8), S182–S194.

Eaton, D. A. R., & Ree, R. H. (2013). Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (Pedicularis: Orobanchaceae). Systematic Biology. Retrieved from https://academic.oup.com/sysbio/article-abstract/62/5/689/1684460

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. Retrieved from https://academic.oup.com/nar/article-abstract/32/5/1792/2380623

Edwards, E. J., & Smith, S. A. (2010). Phylogenetic analyses reveal the shady history of C4 grasses. Proceedings of the National Academy of Sciences of the United States of America, 107(6), 2532–2537.

Folk, R. A., Sun, M., Soltis, P. S., Smith, S. A., Soltis, D. E., & Guralnick, R. P. (2018). Challenges of comprehensive taxon sampling in comparative biology: Wrestling with rosids. American Journal of Botany, 105(3), 433–445.

Freyman, W. A. (2015). SUMAC: Constructing Phylogenetic Supermatrices and Assessing Partially Decisive Taxon Coverage. Evolutionary Bioinformatics Online, 11, 263–266.

Fu, W., Liu, X., Zhang, N., Song, Z., Zhang, W., Yang, J., & Wang, Y. (2017). Testing the Hypothesis of Multiple Origins of Holoparasitism in Orobanchaceae: Phylogenetic Evidence from the Last Two Unplaced Holoparasitic Genera, Gleadovia and Phacellanthus. Frontiers in Plant Science, 8, 1380.

GBIF.org (19th October 2018) GBIF Occurrence Download https://doi.org/10.15468/dl.be2u2y

Gilman, I. S., & Tank, D. C. (2018). Species Tree Estimation Using ddRADseq Data from Historical Specimens Confirms the Monophyly of Highly Disjunct Species of Chloropyron (Orobanchaceae). Systematic Botany. Retrieved from https://www.ingentaconnect.com/contentone/aspt/sb/2018/00000043/00000003/art00006

Gussarova, G., Popp, M., Vitek, E., & Brochmann, C. (2008). Molecular phylogeny and biogeography of the bipolar Euphrasia (Orobanchaceae): recent radiations in an old genus. Molecular Phylogenetics and Evolution, 48(2), 444–460.

Jacobs, S. J., Kristofferson, C., Uribe-Convers, S., Latvis, M., & Tank, D. C. (2018). Incongruence in molecular species delimitation schemes: What to do when adding more data is difficult. Molecular Ecology, 27(10), 2397–2413.

Johnson, M. G., Pokorny, L., Dodsworth, S., Botigué, L. R., Cowan, R. S., Devault, A., … Wickett, N. J. (2018). A Universal Probe Set for Targeted Sequencing of 353 Nuclear Genes from Any

Flowering Plant Designed Using k-medoids Clustering. Systematic Biology. https://doi.org/10.1093/sysbio/syy086

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution, 30(4), 772–780.

Landis, J. B., Soltis, D. E., Li, Z., Marx, H. E., Barker, M. S., Tank, D. C., & Soltis, P. S. (2018). Impact of whole-genome duplication events on diversification rates in angiosperms. American journal of botany, 105(3), 348-363.

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics , 30(22), 3276–3278.

Latvis, M., Mortimer, S. M. E., Morales-Briones, D. F., Torpey, S., Uribe-Convers, S., Jacobs, S. J., … Tank, D. C. (2017). Primers for Castilleja and their utility across Orobanchaceae: I. Chloroplast primers 1. Applications in Plant Sciences, 5(9), 1700020.

Mabberley, D. J. (2017). Mabberley's Plant-book. https://doi.org/10.1017/9781316335581

Magallón, S., Sánchez-Reyes, L. L., & Gómez-Acevedo, S. L. (2019). Thirty clues to the exceptional diversification of flowering plants. Annals of Botany. https://doi.org/10.1093/aob/mcy182

Manen, J.-F., Habashi, C., Jeanmonod, D., Park, J.-M., & Schneeweiss, G. M. (2004). Phylogeny and intraspecific variability of holoparasitic Orobanche (Orobanchaceae) inferred from plastid rbcL sequences. Molecular Phylogenetics and Evolution, 33(2), 482–500.

McNeal, J. R., Bennett, J. R., Wolfe, A. D., & Mathews, S. (2013). Phylogeny and origins of holoparasitism in Orobanchaceae. American Journal of Botany, 100(5), 971–983.

Moore, B. R., Höhna, S., May, M. R., Rannala, B., & Huelsenbeck, J. P. (2016). Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures. Proceedings of the National Academy of Sciences. https://doi.org/10.1073/pnas.1518659113

Morawetz, J. J., Randle, C. P., & Wolfe, A. D. (2010). Phylogenetic relationships within the tropical clade of Orobanchaceae. TAXON. https://doi.org/10.1002/tax.592007

Nabhan, A. R., & Sarkar, I. N. (2012). The impact of taxon sampling on phylogenetic inference: a review of two decades of controversy. Briefings in Bioinformatics. https://doi.org/10.1093/bib/bbr014

Olmstead R.G. (2016). A Synoptical Classification of the Lamiales. Version 2.6.2. https://depts.washington.edu/phylo/Classification.pdf

Pearse, W. D., & Purvis, A. (2013). phyloGenerator: an automated phylogeny generation tool for ecologists. Methods in Ecology and Evolution. https://doi.org/10.1111/2041-210x.12055

Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn, R. G., … Harmon, L. J. (2014). geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics. https://doi.org/10.1093/bioinformatics/btu181

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rabosky, D. L. (2014). Automatic Detection of Key Innovations, Rate Shifts, and Diversity-Dependence on Phylogenetic Trees. PLoS ONE. https://doi.org/10.1371/journal.pone.0089543

Rabosky, D. L., Grundler, M., Anderson, C., Title, P., Shi, J. J., Brown, J. W., … Larson, J. G. (2014). BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. Methods in Ecology and Evolution. https://doi.org/10.1111/2041-210x.12199

Rabosky, D. L., Mitchell, J. S., & Chang, J. (2017). Is BAMM Flawed? Theoretical and Practical Concerns in the Analysis of Multi-Rate Diversification Models. Systematic Biology. https://doi.org/10.1093/sysbio/syx037

Rambaut, A., & Drummond, A. J. (2009). Tracer v1. 5.0.

Rambaut, A., & Drummond, A. J. (2013). LogCombiner v1. 8.2.

Ran Jin-Hua, Shen Ting-Ting, Wang Ming-Ming, & Wang Xiao-Quan. (2018). Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. Proceedings of the Royal Society B: Biological Sciences, 285(1881), 20181012.

Ree, R. H. (2005). Phylogeny and the Evolution of Floral Diversity in Pedicularis (Orobanchaceae). International Journal of Plant Sciences. https://doi.org/10.1086/430191

Refulio-Rodriguez, N. F., & Olmstead, R. G. (2014). Phylogeny of Lamiidae. American Journal of Botany. https://doi.org/10.3732/ajb.1300394

Sanderson, M. J. (2002). Estimating Absolute Rates of Molecular Evolution and Divergence Times: A Penalized Likelihood Approach. Molecular Biology and Evolution. https://doi.org/10.1093/oxfordjournals.molbev.a003974

Scheunert, A., Fleischmann, A., Olano-Marín, C., Bräuchler, C., & Heubl, G. (2012). Phylogeny of tribe Rhinantheae (Orobanchaceae) with a focus on biogeography, cytology and re-examination of generic concepts. TAXON. https://doi.org/10.1002/tax.616008

Schneeweiss, G. M., Colwell, A., Park, J.-M., Jang, C.-G., & Stuessy, T. F. (2004). Phylogeny of holoparasitic Orobanche (Orobanchaceae) inferred from nuclear ITS sequences. Molecular Phylogenetics and Evolution. https://doi.org/10.1016/s1055-7903(03)00210-0

Schneider, A. C. (2016). Resurrection of the genus Aphyllon for New World broomrapes (Orobanche s.l., Orobanchaceae). PhytoKeys, (75), 107–118.

Schneider, A. C., & Moore, A. J. (2017). Parallel Pleistocene amphitropical disjunctions of a parasitic plant and its host. American Journal of Botany. https://doi.org/10.3732/ajb.1700181

Schwery, O., & O'Meara, B. C. (2016). MonoPhy: a simple R package to find and visualize monophyly issues. PeerJ Computer Science. https://doi.org/10.7717/peerj-cs.56

Smith, S. A., Beaulieu, J. M., & Donoghue, M. J. (2009). Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. BMC Evolutionary Biology. https://doi.org/10.1186/1471-2148-9-37

Smith, S. A., & Brown, J. W. (2018). Constructing a broadly inclusive seed plant phylogeny. American Journal of Botany. https://doi.org/10.1002/ajb2.1019

Smith, S. A., & Dunn, C. W. (2008). Phyutility: a phyloinformatics tool for trees, alignments and molecular data. Bioinformatics , 24(5), 715–716.

Smith, S. A., & O'Meara, B. C. (2012). treePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics. https://doi.org/10.1093/bioinformatics/bts492

Smith, S. A., & Walker, J. F. (2019). Py PHLAWD : A python tool for phylogenetic dataset construction. Methods in Ecology and Evolution / British Ecological Society, 10(1), 104–108.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. https://doi.org/10.1093/bioinformatics/btu033

Tank D.C. (2006). Molecular phylogenetics of Castilleja and subtribe Castillejinae (Orobanchaceae)

Stevens, P. F. (2001). Angiosperm Phylogeny Website. Version 14, July 2017." http://www.mobot.org/MOBOT/research/APweb/.

University of Washington, Seattle, Washington, USA.

Tank, D. C., Eastman, J. M., Pennell, M. W., Soltis, P. S., Soltis, D. E., Hinchliff, C. E., … Harmon, L. J. (2015). Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. The New Phytologist, 207(2), 454–467.

Tank, D. C., Mark Egger, J., & Olmstead, R. G. (2009). Phylogenetic Classification of Subtribe Castillejinae (Orobanchaceae). Systematic Botany. https://doi.org/10.1600/036364409787602357

Tank, D. C., & Olmstead, R. G. (2008). From annuals to perennials: phylogeny of subtribe Castillejinae (Orobanchaceae). American Journal of Botany, 95(5), 608–625.

Tank, D. C., & Olmstead, R. G. (2009). The evolutionary origin of a second radiation of annual Castilleja (Orobanchaceae) species in South America: The role of long distance dispersal and allopolyploidy. American Journal of Botany, 96(10), 1907–1921.

Tkach, N., Ree, R. H., Kuss, P., Röser, M., & Hoffmann, M. H. (2014). High mountain origin, phylogenetics, evolution, and niche conservatism of arctic lineages in the hemiparasitic genus Pedicularis (Orobanchaceae). Molecular Phylogenetics and Evolution. https://doi.org/10.1016/j.ympev.2014.03.004

Uribe-Convers, S., Settles, M. L., & Tank, D. C. (2016). A Phylogenomic Approach Based on PCR Target Enrichment and High Throughput Sequencing: Resolving the Diversity within the South American Species of Bartsia L. (Orobanchaceae). PLOS ONE. https://doi.org/10.1371/journal.pone.0148203

Uribe-Convers, S., & Tank, D. C. (2015). Shifts in diversification rates linked to biogeographic movement into new areas: An example of a recent radiation in the Andes. American Journal of Botany. https://doi.org/10.3732/ajb.1500229

Uribe-Convers, S., & Tank, D. C. (2016). Phylogenetic Revision of the Genus Bartsia (Orobanchaceae): Disjunct Distributions Correlate to Independent Lineages. Systematic Botany, 41(3), 672–684.

Wolfe, A. D., Randle, C. P., Liu, L., & Steiner, K. E. (2005). Phylogeny and biogeography ofOrobanchaceae. Folia Geobotanica, 40(2), 115–134.

Xia, Z., Wang, Y.-Z., & Smith, J. F. (2009). Familial placement and relations of Rehmannia and Triaenophora (Scrophulariaceae s.l.) inferred from five gene regions. American Journal of Botany. https://doi.org/10.3732/ajb.0800195

Xi, Z., Liu, L., Rest, J. S., & Davis, C. C. (2014). Coalescent versus Concatenation Methods and the Placement of Amborella as Sister to Water Lilies. Systematic Biology. https://doi.org/10.1093/sysbio/syu055

Young, N. D., Steiner, K. E., & dePamphilis, C. W. (1999). The Evolution of Parasitism in Scrophulariaceae/Orobanchaceae: Plastid Gene Sequences Refute an Evolutionary Transition Series. Annals of the Missouri Botanical Garden. Missouri Botanical Garden, 86(4), 876–893.

Yu, W.-B., Liu, M.-L., Wang, H., Mill, R. R., Ree, R. H., Yang, J.-B., & Li, D.-Z. (2015). Towards a comprehensive phylogeny of the large temperate genus Pedicularis (Orobanchaceae), with an emphasis on species from the Himalaya-Hengduan Mountains. BMC Plant Biology. https://doi.org/10.1186/s12870-015-0547-9

Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., FitzJohn, R. G., … Beaulieu, J. M. (2015). Corrigendum: Three keys to the radiation of angiosperms into freezing environments. Nature, 521(7552), 380.

Table 2.1. The twelve gene regions mined from PyPHLAWD showing with the number of genera and species represented from each of the four largest clades of Orobanchaceae.

	Orobanchaceae		Orobancheae		Rhinantheae		Buchnereae		Pedicularideae	
Locus	**Genera**	**Species**	Genera	Species	Genera	Species	Genera	Species	Genera	Species
matK	**71**	**523**	10	22	18	99	19	40	14	344
psbA-trnH	**15**	**129**	0	0	5	8	0	0	7	115
rbcL	**47**	**436**	5	25	13	53	10	27	12	315
rpl32	**1**	**24**	0	0	1	24	0	0	0	0
rps16	**45**	**275**	3	3	17	71	6	8	11	171
rps2	**57**	**217**	9	68	13	42	12	28	13	60
trnL-trnF	**52**	**560**	4	46	12	89	13	52	15	352
ETS	**7**	**71**	0	0	0	0	0	0	7	71
ITS	**73**	**800**	11	102	17	147	17	66	17	455
GBSSI	**2**	**32**	0	0	0	0	0	0	2	32
PHYA	**52**	**96**	9	20	11	20	13	23	11	22
PHYB	**46**	**80**	7	16	12	21	12	17	8	17

Table 2.2. Individual sequences that were removed from the final concatenated data matrix.

Taxon (original tip name)	Gene regions removed	Reason for removal
Agalinis peduncularis	*rps16*	Data distribution
Aphyllon californicum ssp. *californicum*	*rps2,* ITS, *rbcL*	Sub-specific taxon overlap
Aphyllon californicum ssp. *condensum*	rps2, ITS, *trnL-trnF*	Sub-specific taxon overlap
Aphyllon californicum ssp. *feudgei*	rps2, ITS, *trnL-trnF*	Sub-specific taxon overlap
Aphyllon californicum ssp. *grande*	rps2, ITS, *trnL-trnF*	Sub-specific taxon overlap
Aphyllon californicum ssp. *grayanum*	r*ps2,* ITS, *trnL-trnF*	Sub-specific taxon overlap
Aphyllon californicum ssp. *jepsonii*	ITS, *trnL-trnF*	Sub-specific taxon overlap
Aphyllon cooperi ssp. *palmeri*	ITS	Sub-specific taxon overlap
Aphyllon corymbosum ssp. *mutabile*	ITS	Sub-specific taxon overlap
Aphyllon parishii	ITS	Sub-specific taxon overlap
Aphyllon parishii ssp. *brachylobum*	*rps2,* ITS, *trnL-trnF*	Sub-specific taxon overlap
Aphyllon validum ssp. *howellii*	*rps2,* ITS, *trnL-trnF*	Sub-specific taxon overlap
Castilleja coccinea	PHYB	Paralogy
Castilleja crista	PHYB	Paralogy
Castilleja lasiorhyncha	PHYB	Paralogy
Castilleja lasiorhyncha	PHYA, PHYB	Paralogy
Castilleja rubicundula	PHYB	Paralogy
Castilleja sessiliflora	PHYB	Paralogy
Castilleja sulphurea	PHYB	Paralogy
Castilleja tenuis	PHYB	Paralogy
Castilleja coccinea	PHYB	Paralogy
Castilleja densiflora ssp. *densiflora*	ITS	Data distribution
Castilleja elegans	ITS	Data distribution
Castilleja exserta ssp. *exserta*	ITS	Sub-specific taxon overlap
Castilleja rubicundula ssp. *rubicundula*	ITS	Sub-specific taxon overlap
Castilleja sulphurea	PHYA	Paralogy
Castilleja unalaschcensis	ITS	Data distribution
Chloropyron maritimum ssp. *canescens*	ITS	Sub-specific taxon overlap
Cistanche phelypaea ssp. *lutea*	*rps2,* ITS	Sub-specific taxon overlap
Cistanche phelypaea ssp. *phelypaea*	*rps2,* ITS	Sub-specific taxon overlap
Cordylanthus kingii	*trnL-trnF*	Data distribution
Cyclocheilon kelleri	ITS	Data distribution
Euphrasia regelii	*trnL-trnF*	Gappy sequence
Euphrasia sevanensis	*trnL-trnF*	Gappy sequence
Hedbergia longiflora ssp. *longiflora*	*rps16*	Sub-specific taxon overlap
Hedbergia longiflora ssp. *macrophylla*	ITS, *rps16*	Sub-specific taxon overlap
Lamourouxia rhinanthifolia	*PHYB*	Paralogy

Melampyrum nemorosum	*trnL-trnF*	Gappy sequence
Melampyrum pratense	*trnL-trnF*	Gappy sequence
Melampyrum saxosum	*trnL-trnF*	Gappy sequence
Melampyrum sylvaticum	*trnL-trnF*	Gappy sequence
Orobanche aconiti-lycoctoni	*rbcL*	Data overlap
Orobanche multicaulis	ITS	Data overlap
Orthocarpus bracteosus	PHYB	Paralogy
Orthocarpus bracteosus	PHYA	Paralogy
Parentucellia latifolia	*trnL-trnF*	Gappy sequence
Pedicularis anthemifolia	PHYB	Paralogy
Pedicularis canadensis	PHYB	Paralogy
Pedicularis confertiflora	PHYB	Paralogy
Pedicularis cranolopha	PHYB	Paralogy
Pedicularis procera	*trnL-trnF*	Gappy sequence
Pedicularis comosa	*trnL-trnF*	Gappy sequence
Pedicularis confertiflora ssp. *parvifolia*	ITS, *rbcl, matK, trnL-trnF, PHYA*	Sub-specific taxon overlap
Pedicularis exaltata	*trnL-trnF*	Gappy sequence
Pedicularis hacquetii	*trnL-trnF*	Gappy sequence
Pedicularis kansuensis ssp. *villosa*	*rbcl,, matK, trnl-trnF*	Sub-specific taxon overlap
Pedicularis langsdorffii ssp. *arctica*	ITS	Sub-specific taxon overlap
Pedicularis palustris ssp. *karoi*	ITS, *matK*	Sub-specific taxon overlap
Pedicularis palustris ssp. *opsiantha*	ITS, *matK*	Sub-specific taxon overlap
Pedicularis przewalskii ssp. *australis*	ITS	Sub-specific taxon overlap
Pedicularis pygmaea ssp. *deqinensis*	*rbcL, matK*	Sub-specific taxon overlap
Pedicularis racemosa ssp. *alba*	ITS	Sub-specific taxon overlap
Pedicularis rex ssp. *lipskyana*	ITS, *rbcL,, matK*	Sub-specific taxon overlap
Pedicularis rex ssp. *rex*	ITS, *matK*	Sub-specific taxon overlap
Pedicularis rex ssp. *zayuensis*	ITS, *matK*	Sub-specific taxon overlap
Pedicularis sudetica	ITS, *rbcL, trnL-trnF*	Sub-specific taxon overlap
Pedicularis sudetica ssp. *gymnostachya*	ITS, *matK*	Sub-specific taxon overlap
Pedicularis sudetica ssp. *interioroides*	ITS, *matK*	Sub-specific taxon overlap
Pedicularis sudetica ssp. *pacifica*	ITS, *matK*	Sub-specific taxon overlap
Phelipanche purpurea ssp. *bohemica*	*rps2,* ITS, *trnL-trnF*	Sub-specific taxon overlap
Rhinanthus glacialis	*trnL-trnF*	Gappy sequence
Rhinanthus major	*trnL-trnF*	Singular long branch
Rhinanthus rumelicus	*trnL-trnF*	Gappy sequence
Rhynchocorys elephas	*trnL-trnF*	Gappy sequence
Rhynchocorys kurdica	*trnL-trnF*	Gappy sequence
Rhynchocorys maxima	*trnL-trnF*	Gappy sequence
Rhynchocorys odontophylla	*trnL-trnF*	Gappy sequence

Rhynchocorys orientalis	*trnL-trnF*	Gappy sequence
Rhynchocorys stricta	*trnL-trnF*	Gappy sequence
Seymeria laciniata	PHYB	Paralogy
Seymeria pectinata	PHYB	Paralogy
Sopubia trifida	*rbcL*	Data overlap
Striga bilabiata ssp. *linearifolia*	ITS, PHYA, PHYB	Sub-specific taxon overlap
Striga hermonthica	ITS	Data overlap
Tozzia alpina ssp. *carpathica*	*trnL-trnF*	Sub-specific taxon overlap
Tozzia alpina ssp. *carpathica*	ITS, *matK*	Sub-specific taxon overlap
Triphysaria pusilla	PHYB	Paralogy
Triphysaria eriantha ssp. *rosea*	ITS, *trnL-trnF, rps16*	Sub-specific taxon overlap
Triphysaria versicolor ssp. *faucibarbata*	ITS, *trnL-trnF, rps16*	Sub-specific taxon overlap

Table 2.3. Subspecific taxa combined across gene regions. Tip names were changed from the original PyPHLAWD name so that each species of the phylogeny was only represented by a single sequence in the alignment, and a single tip in the tree.

Tree tip label	Taxon 1	Gene region	Taxon 2	Gene region	Taxon 3	Gene Region 3
Aphyllon californicum	*Aphyllon californicum*	ITS, PHYA, PHYB, *rbcL, rps2*	*Aphyllon californicum* ssp. *californicum*	*trnL*		
Castilleja ambigua	*Castilleja ambigua* ssp. *ambigua*	ETS, ITS, *rps2, trnL*	*Castilleja ambigua*	GBSSI, *rbcl*		
Castilleja campestris	*Castilleja campestris* ssp. *campestris*	ETS, ITS, *rps16, trnL*	*Castilleja campestris*	GBSSI,		
Castilleja densiflora	*Castilleja densiflora* ssp. *densiflora*	ETS, *rps16*, t*rnL*	*Castilleja densiflora*	GBSSI, ITS		
Castilleja exserta	*Castilleja exserta*	GBSSI, ITS, *matK* PHYA, PHYB, rps2	*Castilleja exserta* ssp. *exserta*	ETS, *rps16 trnL*		
Castilleja rubicundula	*Castilleja rubicundula*	GBSSI, ITS, *matk,* PHYA, rps2	*Castilleja rubicundula* ssp. *rubicundula*	ETS rps16 *trnL*		
Chloropyron maritimum	*Chloropyron maritimum* ssp. *maritimum*	ITS, *matK, rbcL*	*Chloropyron maritimum* ssp. *canescens*	ETS, *rps16 trnL*		
Cordylanthus rigidus	*Cordylanthus rigidus*	ITS	*Cordylanthus rigidus* ssp. *rigidus*	ETS, *rps16*, *trnL*		
Cycnium tubulosum	*Cycnium tubulosum*	ITS, PHYB, *rps2, trnL*	*Cycnium tubulosum* ssp. *montanum*	*matK*		
Hedbergia longiflora	*Hedbergia longiflora*	ITS, *rps16*	*Hedbergia longiflora* ssp. l*ongiflora*	*trnL*	*Hedbergia longiflora* ssp. *macrophylla*	*matK*
Pedicularis debilis	*Pedicularis debilis*	ITS, *matK, psbA, trnL, rbcL*	*Pedicularis debilis* ssp. *debilis*	*rps16*		
Pedicularis przewalskii	*Pedicularis przewalskii*	ITS, *matK, psbA, trnL, rbcL*	*Pedicularis przewalskii* ssp. *australis*	*rps16*		
Pedicularis racemosa	*Pedicularis racemosa*	ITS, *rbcL*	*Pedicularis racemosa* ssp. *alba*	*matK*		
Pedicularis rupicola	*Pedicularis rupicola*	ITS, *matK, psbA, trnL, rbcL*	*Pedicularis rupicola* ssp. *rupicola*	*rps16*		
Pedicularis stenocorys	*Pedicularis stenocorys*	ITS, *trnL*	*Pedicularis stenocorys* ssp. *melanotricha*	*matK, rbcL*		
Striga bilabiata	*Striga bilabiata*	ITS, *matK*, PHYA, PHYB, *trnL*	*Striga bilabiata* ssp. *linearifolia*	*rps2*		
Triphysaria eriantha	*Triphysaria eriantha*	GBSSI, ITS, *rps16, trnL*	*Triphysaria eriantha* ssp. *eriantha*	*matK, rbcL*	*Triphysaria eriantha* ssp. *rosea*	ETS
Triphysaria versicolor	*Triphysaria versicolor*	GBSSI, *rps2*, *rbcL*	*Triphysaria versicolor* ssp. *versicolor*	ITS, *rps16*, *trnL*		

Table 2.4. Genera of Orobanchaceae, major clade, number of species sampled, total species described, and general geography of the genus.

Genus	Major_Clade	Sampled	Total	Geography
Brandisia	Brandisia	3	13	Myanmar & China
Aeginetia	Buchereae	1	3	Indomalaya & E Asia
Alectra	Buchereae	10	40	Tropical Africa & Asia
Asepalum	Buchereae	1	1	NE & E Africa
Bardotia	Buchereae	1	1	N Madagascar
Baumia	Buchereae	0	1	Angola
Buchnera	Buchereae	9	100	Africa & Australia
Buttonia	Buchereae	0	2	Tropical & S Africa
Centranthera	Buchereae	1	6	China to Australia
Christisonia	Buchereae	1	17	SW China, SE Asia, Indomalaya
Cyclocheilon	Buchereae	2	3	NE Africa
Cycniopsis	Buchereae	0	2	Tropical Africa
Cycnium	Buchereae	5	16	Warm Africa
Escobedia	Buchereae	3	15	Tropical Americas
Gerardiina	Buchereae	0	2	Tropical & S Africa
Ghikaea	Buchereae	0	1	NE Africa
Graderia	Buchereae	2	5	Africa & Socotra
Harveya	Buchereae	18	40	Tropical & S Africa, Masc.
Hiernia	Buchereae	0	1	Angola & Namibia
Hyobanche	Buchereae	4	8	S Africa
Leucosalpa	Buchereae	1	4	Madagascar
Magdalenaea	Buchereae	0	1	SE Brazil
Melasma	Buchereae	4	20	S. Africa & Am.
Micrargeria	Buchereae	1	5	Trop. Africa & India
Micrargeriella	Buchereae	0	1	Congo & Zambia
Nesogenes	Buchereae	2	8	Tanzania, Madagascar, India, & Pacific Oceans.
Nothochilus	Buchereae	0	1	Brazil
Paraharveya	Buchereae	0	1	C & E Africa
Parasopubia	Buchereae	0	2	SE Asia
Petitmenginia	Buchereae	0	2	S China & SE Asia
Physocalyx	Buchereae	0	2	Brazil
Platypholis	Buchereae	0	1	Bonin Island, Japan
Pseudomelasma	Buchereae	0	1	Madagascar
Pseudosopubia	Buchereae	0	5	Tropical Africa
Pseudostriga	Buchereae	0	1	SE Asia
Radamaea	Buchereae	1	5	Madagascar
Rhamphicarpa	Buchereae	0	6	Russia, Turkey, trop. & S Africa, India, & Australia
Rhaphispermum	Buchereae	0	1	Madagascar
Sieversandreas	Buchereae	1	1	S. Madagascar
Sopubia	Buchereae	8	40	Trop. Africa, Indomalaya, Madagascar
Striga	Buchereae	7	33	OW tropics
Tetraspidium	Buchereae	0	1	Madagascar
Thunbergianthus	Buchereae	0	2	Sao Tome, trop. E. Africa
Tomanthera	Buchereae	0	2	C N America
Vellosiella	Buchereae	0	2	Brazil
Xylocalyx	Buchereae	3	5	Somalia, Socotra
Bungea	Cymbarieae	1	2	SW Asia, C Asia, China
Cymbaria	Cymbarieae	1	4	European, C & E Asia
Monochasma	Cymbarieae	2	4	E Asia
Schwalbea	Cymbarieae	1	1	E USA
Siphonostegia	Cymbarieae	3	3	E Mediterranean & E Asia
Lindenbergia	Lindenbergia	6	12	Old world tropics
Gleadovia	Orobacheae	1	2	W Himalaya & W China
Aphyllon	Orobacheae	19	22	N America to S America
Boschniakia	Orobacheae	2	2	N Arctic Russia, Asia to Japan
Boulardia	Orobacheae	1	1	Algeria, Morocco, Spain, & Portugal
Cistanche	Orobacheae	7	10	Mediterranean Europe & Ethiopia to W. India
Conopholis	Orobacheae	2	3	SE USA to Panama
Diphelypaea	Orobacheae	2	4	Mediterranean
Epifagus	Orobacheae	1	1	N America
Kopsiopsis	Orobacheae	2	2	W. N America
Mannagettaea	Orobacheae	1	3	E Siberia to W China
Orobanche	Orobacheae	50	100	Old world
Phacellanthus	Orobacheae	1	1	E Asia
Phelipanche	Orobacheae	20	50	Old world
Agalinis	Pedicularideae	32	40	Tropical & Warm Americas

Anisantherina	Pedicularideae	0	1	Tropical Americas
Aureolaria	Pedicularideae	5	11	E USA & Mexico
Brachystigma	Pedicularideae	1	1	SW N America
Castilleja	Pedicularideae	78	190	N America, Eurasia, C American, Andes
Chloropyron	Pedicularideae	4	4	W N America
Cordylanthus	Pedicularideae	9	13	W N America
Dasistoma	Pedicularideae	1	1	SE USA
Dicranostegia	Pedicularideae	1	1	Baja California
Esterhazya	Pedicularideae	1	5	Bolivia & Brazil
Lamourouxia	Pedicularideae	3	30	N Mexico To Peru
Leptorhabdos	Pedicularideae	1	1	Caucuses & Iran to C Asia & Himalayas
Macranthera	Pedicularideae	1	1	SE USA
Orthocarpus	Pedicularideae	9	9	W N America
Pedicularis	Pedicularideae	364	750	N hemisphere, esp. mtns C & E Asia, Europe
Phtheirospermum	Pedicularideae	1	1	E Asia
Seymeria	Pedicularideae	3	25	S N America
Seymeriopsis	Pedicularideae	0	1	Cuba
Silviella	Pedicularideae	0	2	Mexico
Triphysaria	Pedicularideae	5	5	W N America incl. Mexico
Rehmannia	Rehmannieae	9	9	E Asia
Triaenophora	Rehmannieae	2	3	NE Asia
Bartsia	Rhinantheae	1	1	Alpine European & NE N America
Bartsiella	Rhinantheae	1	1	N African Atlas Mtns
Bellardia	Rhinantheae	2	2	Mediterranean, Coastal Australia, Chile and SW USA
Bornmuellerantha	Rhinantheae	0	1	Turkey to Iran
Euphrasia	Rhinantheae	80	350	N Temperate Europe, NG, Ausralia, NZ.
Hedbergia	Rhinantheae	3	3	Tropical African Mountains
Lathraea	Rhinantheae	3	7	Temperate Eurasia
Macrosyringion	Rhinantheae	0	2	Mediterranean
Melampyrum	Rhinantheae	13	35	N Temperate Europe
Neobartsia	Rhinantheae	15	47	High elevation Andes
Nothobartsia	Rhinantheae	3	3	W Mediterranean (Mabberly 4th = 2)
Odontitella	Rhinantheae	1	1	Iberia
Odontites	Rhinantheae	29	34	W & S Europe
Omphalotrix	Rhinantheae	1	1	NE Asia
Parentucellia	Rhinantheae	1	1	Mediterranean
Psuedobartsia	Rhinantheae	1	1	W Himalaya & China
Pterygiella	Rhinantheae	6	6	S China
Rhinanthus	Rhinantheae	9	45	N Hemisphere
Rhynchocorys	Rhinantheae	6	6	S Europe to Iran
Tozzia	Rhinantheae	1	1	Alps, Carpathians, Pyrenees
Xizangia	Rhinantheae	1	1	Tibet & China

Table 2.5. Remaining non-monophyletic genera identified by MonoPhy after tip names were changed to reflect an updated taxonomy.

Remaining Non-monophyletic genera	# tips in genus	Intruding tips	Outlying tips
Aureolaria	5	*Dasistoma macrophylla*	None
Melasma	4	None	*M. physalodes, M. rhinanthoides*
Cycnium	5	None	*C. volkensii, C. adonense*
Graderia	2	None	*G. fruticosa*
Euphrasia	80	*Omphalotrix longipes*	None
Odontites	29	*Bartsiella rameauana*	None
Bellardia	2	None	*B. viscosa*
Orobanche	50	*Phacellanthus tubiflorus*	None

Table 2.6. Estimated divergence times of Orobanchaceae based on secondary calibrations from (**A**) Fu et al. (2017) and (**B**) Schneider and Moore (2017)

A

				Crown							Stem			
Clade	mean	q0.025	q0.975	median	min	max	Point Est.	Mean	q0.025	q0.975	Median	Min.	Max	Point Est.
Orobanchaceae	36.6516	30.4171	47.0387	36.0161	25.9824	60.3390	36.5731	40.2709	32.0934	49.5133	39.3885	29.6923	63.2453	39.4089
Rehmannieae	12.4685	9.9856	14.9597	12.4623	6.6073	18.6738	12.5473	36.6516	30.4171	47.0387	36.0161	25.9824	60.3390	36.5731
Lindenbergia	12.9656	5.6278	20.7939	12.3311	4.9166	27.7800	13.2171	33.1638	27.2001	43.1421	32.5654	23.2877	59.7236	32.6234
Cymbarieae	19.1609	11.6221	28.4022	19.1378	4.3283	34.1941	19.3566	31.1248	24.2502	39.2645	30.4266	22.0972	59.2201	30.5507
Brandisia	4.3876	3.5454	5.3404	4.3792	2.1572	6.6295	4.3982	28.2913	22.1023	37.1766	27.6604	20.3456	58.7881	27.7418
Orobancheae	23.7193	16.2079	31.3977	23.4454	7.4621	44.8198	23.6997	29.5411	23.0409	37.9631	28.8947	21.0839	59.0087	29.0063
Rhinantheae	21.6944	14.9307	31.0010	21.6011	7.8004	38.7060	21.9191	26.3773	19.9126	33.9409	25.6046	18.9127	58.4191	25.9156
Pterygiella Alliance	11.0057	8.2128	13.9498	11.0603	5.2123	16.9442	11.1459	-	-	-	-	-	-	-
Core Rhinantheae	15.6630	9.5389	21.5724	15.7211	5.6195	24.8997	15.9474	-	-	-	-	-	-	-
Bartsia alpina	-	-	-	-	-	-	-	12.2889	7.9847	17.3904	12.3686	4.3728	19.5499	12.4833
Buchnereae	18.3661	14.3216	23.4861	17.8456	11.5544	40.9561	18.2109	23.7887	18.3738	32.1693	23.0463	17.2854	57.9533	23.5285
Pedicularideae	15.3279	10.8787	20.3135	15.2553	6.6244	25.1731	15.4009	23.7887	18.3738	32.1693	23.0463	17.2854	57.9533	23.5285
Core Pedicularideae	14.1126	10.7903	18.9230	14.0143	6.4152	23.2040	14.1878	-	-	-	-	-	-	-
Castillejinae	8.2543	5.6658	10.9593	8.1572	3.5126	14.1637	8.2313	10.8437	8.0547	14.0893	10.6926	5.6887	19.5635	10.7611

B

Clade	Mean	q0.025	q0.975	Median	Min.	Max	Point Est.	Mean	q0.025	q0.975	Median	Min.	Max	Point Est.
Orobanchaceae	30.2038	25.5633	36.4131	29.5004	23.2411	39.8775	30.2303	31.4411	28.0881	37.4981	30.7023	28.0881	42.9571	31.3920
Rehmannieae	9.4229	7.9645	11.4445	9.2656	7.3228	12.2848	9.4981	30.2038	25.5633	36.4131	29.5004	23.2411	39.8775	30.2303
Lindenbergia	15.6450	11.5187	19.7667	15.4670	11.0297	23.1054	15.5855	29.1495	24.7077	34.8179	28.6081	22.9411	37.7062	29.1784
Cymbareae	15.6713	11.3823	20.2247	15.4812	9.7290	23.9512	15.5620	27.7152	23.4739	33.3110	27.2212	22.3048	36.0180	27.7235
Brandisia	3.1782	2.6468	3.8692	3.1169	2.4520	4.2051	3.2064	25.3509	21.6193	30.7714	24.8506	20.3093	33.3543	25.3668
Orobancheae	16.5187	13.5612	19.5898	16.2838	13.3457	22.5974	16.4802	26.5742	22.5404	31.9130	26.0914	21.1212	34.7040	26.5766
Rhinantheae	20.4858	17.2308	24.7887	20.1618	15.7193	27.6719	20.5426	24.1355	20.4937	29.3430	23.6779	19.4264	32.0588	24.1434
Pterygiella Alliance	8.3293	7.0065	10.1964	8.1803	6.4002	11.0766	8.4077	-	-	-	-	-	-	-
Core Rhinantheae	12.2678	9.5621	15.3439	12.1239	8.8369	17.0509	12.2009	-	-	-	-	-	-	-
Bartsia alpina	-	-	-	-	-	-	-	8.5095	6.7921	10.7208	8.3345	5.9536	12.2979	8.5185
Buchnereae	12.9902	10.7040	15.5127	12.7747	10.4004	17.2902	12.9718	21.9165	18.4237	26.4260	21.4473	17.7823	29.6403	21.9049
Pedicularideae	13.5000	10.9497	16.3256	13.3082	10.3655	18.5946	13.4613	21.9165	18.4237	26.4260	21.4473	17.7823	29.6403	21.9049
Core Pedicularideae	12.5208	10.0825	15.2564	12.3427	9.5204	17.5900	12.4727	-	-	-	-	-	-	-
Castillejinae	7.6756	6.3085	9.3402	7.5393	5.7874	10.6687	7.6617	10.9513	8.9913	13.6710	10.8301	8.3801	14.5673	10.9456

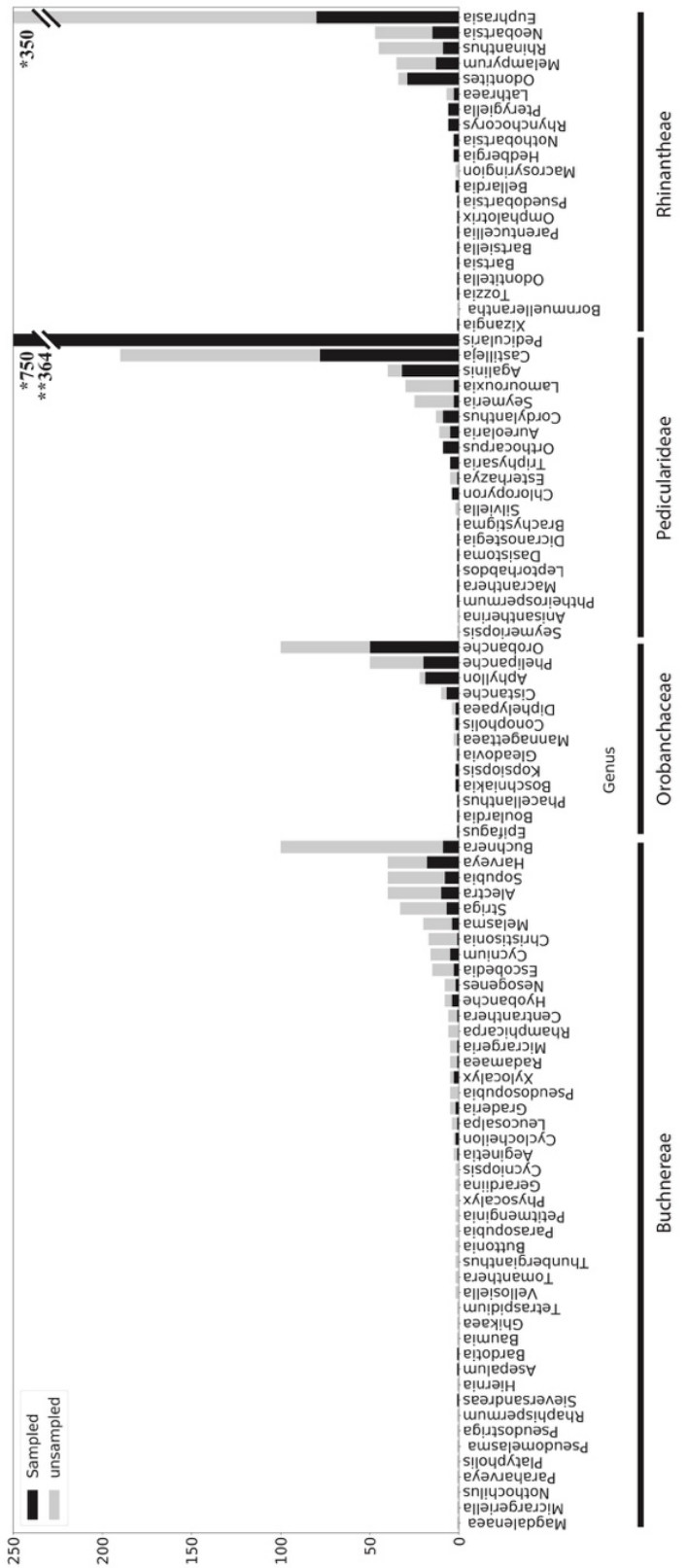

Figure 2.1. Proportion of sampled (black bars) and unsampled (gray bars) species of each genus in the four largest major clades of Orobanchaceae: Buchnereae, Orobancheae, Pedicularideae, and Rhinantheae.

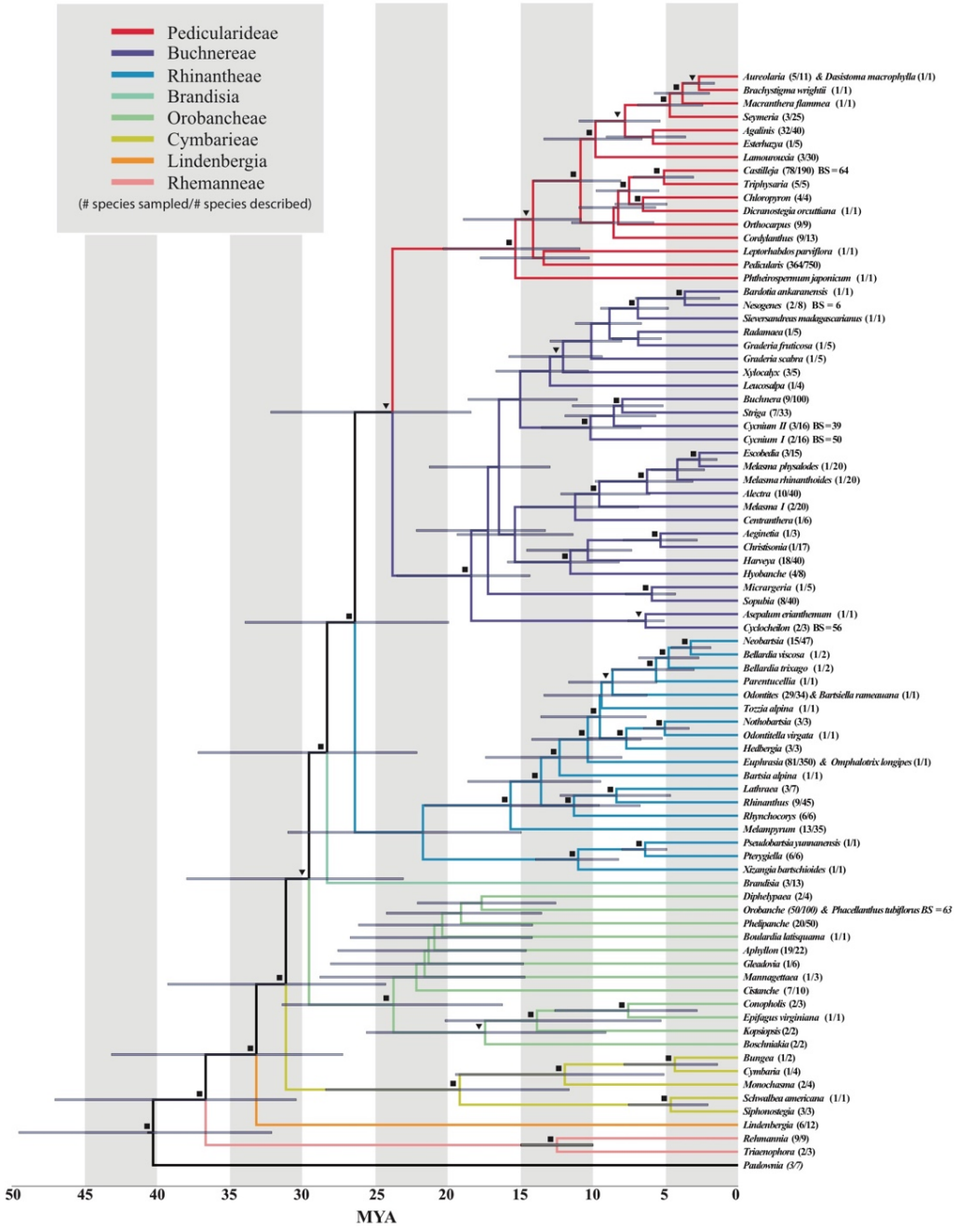

Figure 2.2. Maximum likelihood phylogeny of Orobanchaceae representing 922 species. Each tip represents a single monophyletic genus unless otherwise noted. Divergence times were estimated based on 'congruification' with data from Fu et al. (2017), and bars denote the 95% confidence intervals summarized from our 500 independent analyses. To the right of each tip name is the number of sampled species represented by the tip out of the number of species described in the genus. Nodes labeled with squares represent strong bootstrap support (85 – 100%), triangles represent moderately supported nodes (70 – 85%).

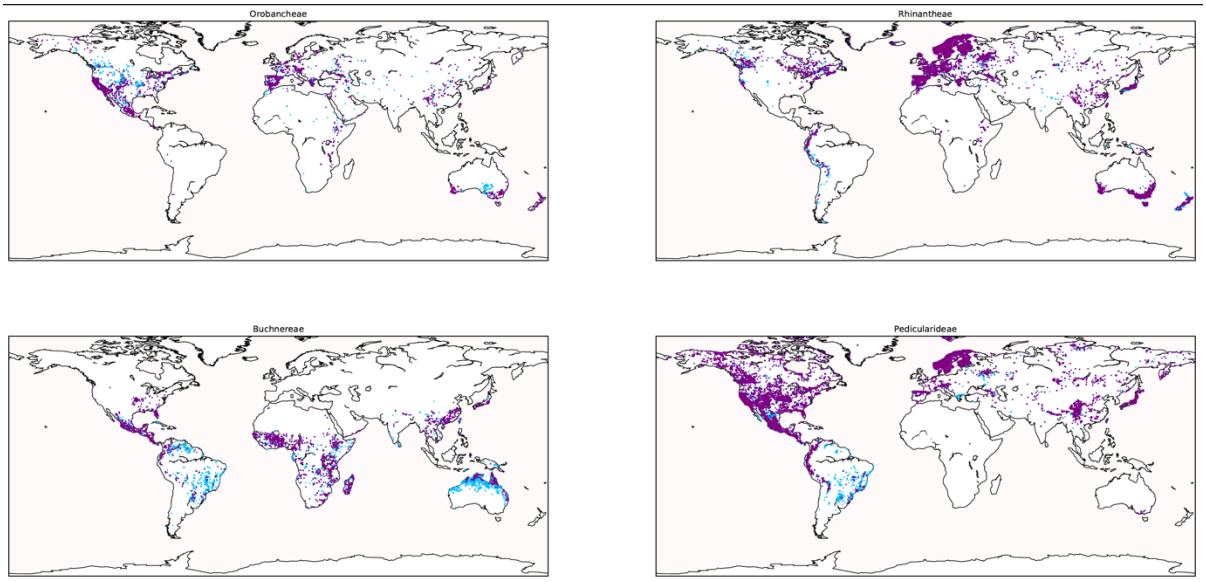

Figure 2.3. Geographic extent of unsampled (purple) and sampled (blue) taxa estimated from GBIF for four major clades of Orobanchaceae.

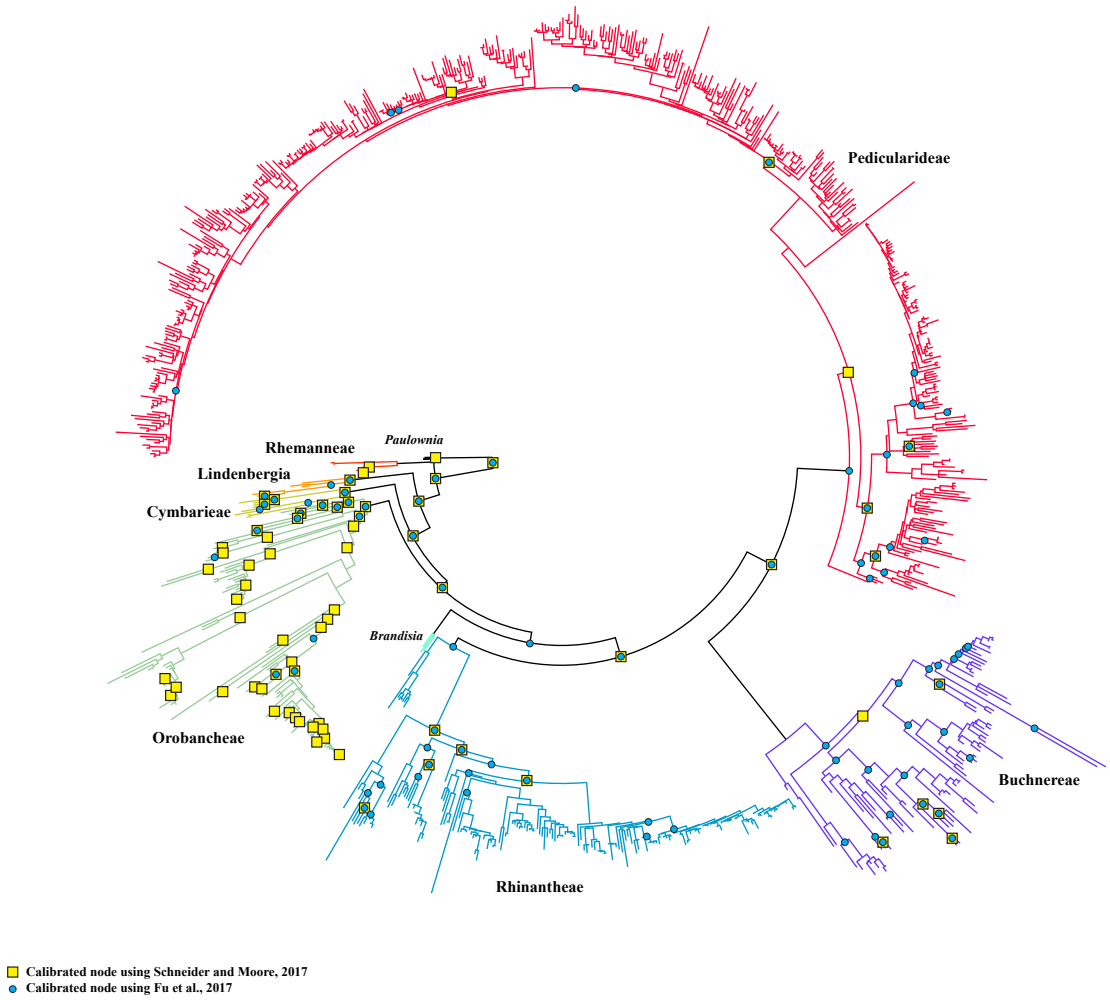

Figure 2.4. Maximum likelihood phylogeny of Orobanchaceae representing 922 species. Colored branches correspond to major named clades of the family. Node labels correspond to congruent nodes with previously published phylogenies used to estimate the divergence times. Yellow squares denote a node that is congruent with Schneider and Moore (2017), and blue circles denote a node that is congruent with Fu et al. (2017).

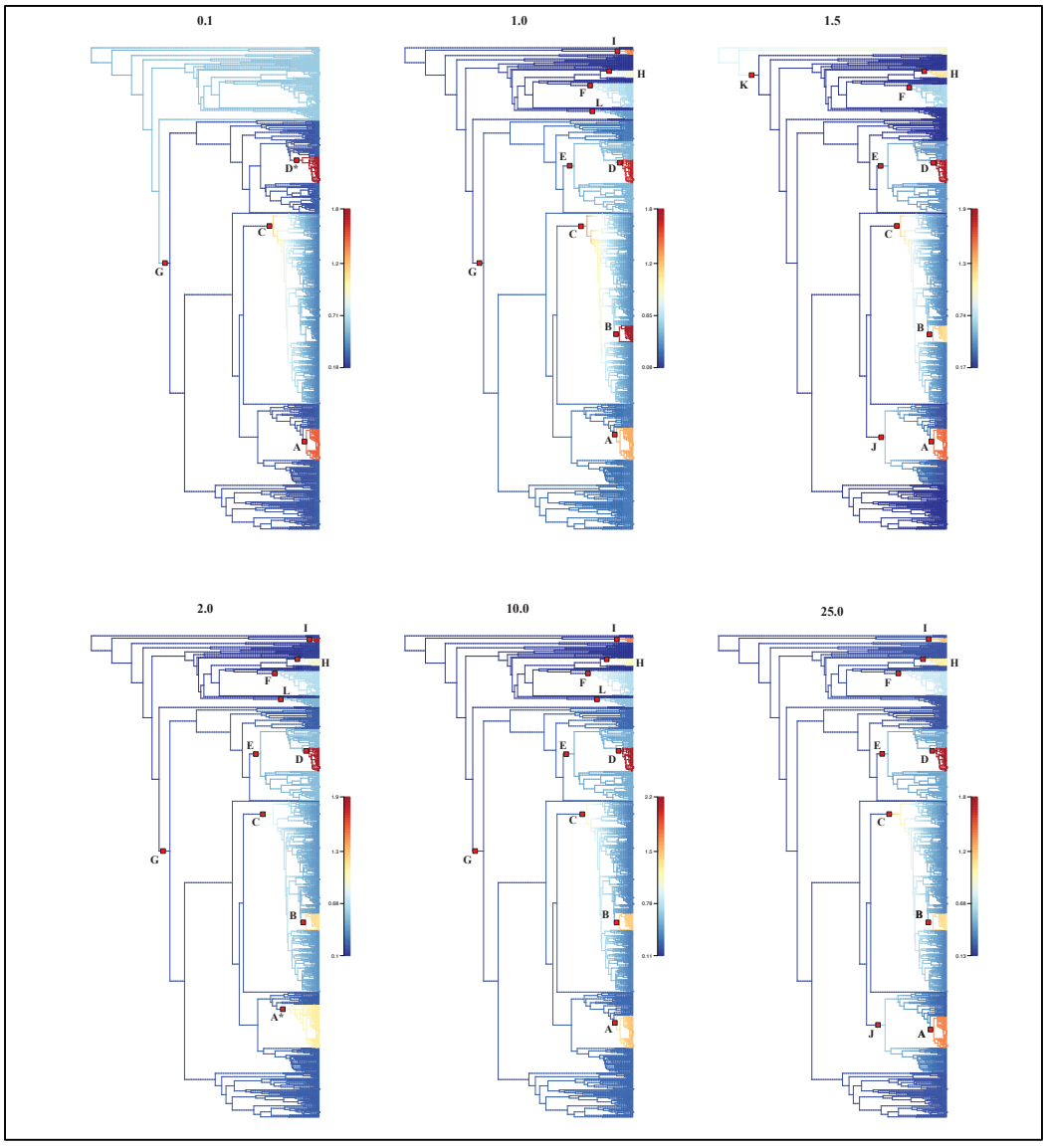

Figure 2.5. Orobanchaceae wide diversification rate plot for each of six independent analyses using a different prior distribution of expected number of speciation rate shifts in BAMM. MAP shifts are marked by red squares on branches of each phylogeny. Shifts are labeled according to clades in which they occur: **A**) denotes an uptick in speciation rate within the genus *Castilleja* (Pedicularideae), **A***) denotes a rate shift on stem branch of *Castilleja* + *Triphysaria* (Pedicularideae) found only in the prior=2.0 analysis, **B**) denotes an increased speciation rate within *Pedicularis* found in five analyses, **C**) denotes a slightly elevated speciation rate on the stem branch of *Pedicularis* that was found in every analysis, **D**) denotes a increased speciation rate within *Euphrasia* (Rhinantheae) and **D***) denotes a similar shift found on the branch adjacent to D, **E**) denotes a slightly elevated speciation rate in the clade sister to *Bartsia alpine* (Rhinantheae), **F**) denotes a rate shift found within *Orobanche* (Orobancheae), **G**) denotes a rate shift on the stem branch of the clade comprised of Rhinantheae, Buchnereae, and Pedicularideae, **H**) denotes a rate shift within *Phelipanche* (Orobancheae), **I**) denotes a rate shift found on the stem branch of (Rehmannieae), **J**) denotes a rate shift found on the stem branch of the clade sister to *Leptorhabdos* + *Pedicularis*, **K**) denotes a rate shift corresponding to all of parasitic Orobanchaceae + *Lindenbergia,* and **L**) denotes a rate shift found within *Aphyllon* (Orobancheae). Branch colors correspond to speciation rate.

Appendix A - Supplemental Files Ch. 1.

Supplementary data and files are housed at:

https://github.com/mortimersebastian/Lamourouxia_Supp

Appendix B - Supplemental Files Ch. 2.

Supplementary data and files are housed at:

https://github.com/mortimersebastian/Orobanchaceae_Supp