PHYLOGENOMIC INSIGHTS INTO THE RADIATION OF AN ANDEAN GROUP OF PLANTS

A Dissertation

Presented in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

with a

Major in Biology

in the

College of Graduate Studies

University of Idaho

by

Simon Uribe-Convers

August 2014

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

Authorization to Submit Dissertation

This dissertation of Simon Uribe-Convers, submitted for the degree of Doctor of Philosophy with a Major in Biology and titled "Phylogenomic insights into the radiation of an Andean group of plants", 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 M. Sullivan, Ph.D.

_____ Date: _____
Luke J. Harmon, Ph.D.

_____ Date: _____
Eric H. Roalson, Ph.D.

Chair of the Department
of Biological Science: _____ Date: _____
James J. Nagler, Ph.D.

Dean of the College
of Science _____ Date: _____
Paul Joyce, Ph.D.

Final Approval and Acceptance

Dean of the College
of Graduate Studies: _____ Date: _____
Jie Chen, Ph.D.

Abstract

This dissertation focuses on the evolutionary history of the *Neobartsia* clade of the plant genus *Bartsia* (Orobanchaceae) distributed throughout the high elevation páramo ecosystem in the Andes. This group of plants was the subject of a detailed taxonomic study in the 1990's, as well as various phylogenetic studies at higher taxonomic levels. However, the work presented in this dissertation constitutes the most comprehensive study of the group so far. I conducted extensive fieldwork throughout the Andes and was able to sample every species described to date, including multiple individuals per species for many, and I have used these collections to elucidate phylogenetic relationships among closely related genera, and between the ~45 species in the clade. My molecular systematic work was accompanied by comparative phylogenetic studies focused on the time of divergence, historical biogeography, and the rate of diversification of the group, which showed that the clade is the result of an ongoing recent and rapid radiation that began ~3 million years ago – a time period when the northern Andes had the necessary elevation to generate alpine conditions.

It was clear from these studies, that the genetic diversity within *Neobartsia* was low, given the young age of the genus, and therefore, I began to focus on developing genomic approaches to generate large amounts of phylogenetic data, both efficiently and cost effectively. This last part of my dissertation resulted in two genomic methods being developed, one for sequencing complete chloroplast genomes, and one for generating targeted subgenomic datasets, which were used to generate the data necessary to elucidate interspecific relationships. Finally, I revised the taxonomy of the clade and its closely allied genera to reflect these phylogenetic results, alleviating a history of taxonomic instability by

elevating the South American *Bartsia* clade to the newly formed genus *Neobartsia*, thereby recognizing its unique evolutionary history.

Acknowledgements

First and foremost, I would like to thank my advisor David Tank, who has been instrumental during all these years of work. Not only has Dave been a role model for me in what a scientist, professor, and mentor should be; he has been supportive, understanding, and encouraging, which have made me be the type of scientist and person that I am today. Over the years we have developed a strong foundation for future collaboration based on mutual respect, but most importantly, we have developed a great friendship. Gracias Dave.

I would also like to thank my committee members Eric Roalson, Jack Sullivan, Luke Harmon, and George Newcombe, who have helped directed, and encouranged me throughout my graduate work. To my great friends Diego Morales-Briones and Matt Pennell, I owe gratitude for not only keeping me sane and entertained over the years, but also for stimulating discussions about science and life. I would also like to acknowledge Hannah Marx, Sarah Jacobs, and Maribeth Latvis, who supported me both in and outside of my academic life. Matt Settles, Sam Hunter, Dan New, and Tamara Max were instrumental for my development as a scientist in the genomic era, and have planted the 'computer coding' seed in me. My fiends Nicolai Nürk, John Chau, Laura Frost, Segund Leyva, and Mario Zapata, contributed inmensely during the fieldwork in the Andes. I would also like to thank the herbaria ID, PUCE, HAO, HUT, LPB, and ANDES for hosting my collections and providing support in the field. Finally, the opportunities would have not happened without the funding support of The National Science Foundation (NSF), the University of Idaho Student Grant Program, the Society of Systematic Biologists (SSB), the Botanical Society of America (BSA), the American Society of Plant Taxonomists (ASPT), and the University of Idaho Stillinger Herbarium Expedition Funds.

Dedication

Para Marilas y Tuchi, por su apoyo incondicional

To Tracy, for loving me

Table of Contents

List of Figures

List of Tables

Chapter 1: Shifts in diversification rates linked to biogeographic movement into new areas: an example of a recent radiation in the Andes

Abstract

To properly understand the history of a group of organisms, it is imperative that we infer their evolutionary relationships among them. New phylogenetic comparative methods allow us to investigate the time of divergence between species, as well as to calculate the rates at which they are diversifying. Here, we are interested in studying the history of the plant clade Rhinantheae in the parasitic family Orobanchaceae, with a special focus on the Andean clade of the genus *Bartsia* L.. Previous studies have implied that the genus, as formerly circumscribed, is polyphyletic and that the South American species form a distinct clade. However, none of the previous studies have sampled more than a few of the ~45 species that are distributed throughout the high elevation shrublands and grasslands from Colombia to Chile and Argentina. In this study, we sampled representative species from the South American diversity and analyzed them in the phylogenetic context of the mostly European clade Rhinantheae. We present a robust and well-resolved phylogenetic hypothesis for relationships in this clade, which are then used to infer divergence times among the major lineages. We then focus on elucidating the biogeographic history and diversification dynamics of this clade. Finally, we examine the elevated rates of diversification rates found in the South American clade and discuss these in relation to biogeographic movements into new geographic ranges.

Introduction

The investigation of global patterns of biodiversity has a long history (e.g., Mittelbach et al. 2007). With the increase in our knowledge of phylogenetic relationships as well as methods for using phylogenies to understand diversification rates and biogeographic patterns (e.g., Ree et al. 2005; Alfaro et al. 2009), these global patterns can now be placed in an explicitly historical context (sensu Moore and Donoghue 2007). Along these lines, differences in species richness between geographic areas have often been explained by climatic stability, age of the region, and/or niche conservatism that contributes to the slow, but steady, accumulation of species over time (Wiens and Donoghue 2004). Likewise, clade specific bursts in net diversification (speciation minus extinction) are often associated with the evolution of novel morphologies, referred to as key innovations, such as nectar spurs in angiosperms (e.g., Hodges 1997), molar characters in mammals (Woodburne et al. 2003), and feathers in birds (Ostrom 1979). More recently, Moore and Donoghue (2007) demonstrated that in the plant families Adoxaceae and Valerianaceae, shifts in diversification rates were not correlated with the evolution of novel floral characters, but rather, with the movement into new geographic areas, and hypothesized that "dispersification" (dispersal and diversification) may play a larger role in shaping global biodiversity patterns than previously recognized. This is concordant with the hypothesis that in newly emerging environments, as long as the corridors for biogeographic movements are in place, these new areas will most likely be filled with lineages from environmentally similar areas where the relevant morphological and/or physiological adaptations are already in place (Donoghue 2008). Empirical tests of these hypotheses not only require a robust estimate of phylogenetic

relationships, but also the estimation of divergence times, diversification rates, and biogeographic patterns for the group of interest.

Various approaches have been taken to assess phylogenetic relationships, divergence times, and rates of diversification - each increasing our understanding of biodiversity and the way in which it has been produced. Bayesian analyses are now regularly used to estimate divergence times (e.g., Bacon et al. 2012; Drummond et al. 2012), most often performed in the program BEAST (Drummond and Rambaut 2007), because with the use of probabilistic priors they accommodate for both phylogenetic uncertainty (i.e., topology and branch lengths), as well as the timing of calibration points. Diversification rate analyses have been instrumental to our understanding of disparities in clade richnesses across the tree of life. For example, Alfaro et al. (2009) suggested that several pulses of diversification instead of single events have shaped the current diversity of jawed vertebrates. Additionally, in the plant genus *Asclepias* L. (milkweeds) it has been shown that increases in the rate of diversification are tightly associated with the evolution of defense traits that prevent or minimize herbivory, and that this resulted in an adaptive radiation in the group (Agrawal et al. 2009). Finally, studies of Andean plants, e.g., the family Valerianaceae and the genus *Lupinus* L. (Bell and Donoghue 2005; Hughes and Eastwood 2006, respectively), have shown that groups with North American temperate ancestors have elevated diversification rates in the Andes, given that they were "pre-adapted" to the conditions of the newly and unoccupied niche at the time that they colonized the Andes.

To investigate the influence of biogeographic movements on rates of diversification, we have chosen to study the mostly European clade Rhinantheae of the parasitic plant family Orobanchaceae (Wolfe et al. 2005; Bennett and Mathews 2006; McNeal et al. 2013), with a

particular focus on the genus *Bellardia* All., a clade of 48 species that are disproportionally distributed across two disjunct geographic regions. The majority of the species in *Bellardia* were formerly part of the genus *Bartsia* L., but it has been recently recircumscribed (Scheunert et al. 2012) to better reflect the evolutionary history of its species. Prior to this taxonomic rearrangement, *Bartsia* (49 spp.) had two species distributed in the mountains of northeastern Africa (*B. decurva* Benth. and *B. longiflora* Benth.), one in the Mediterranean region (*B. trixago* L.), one in Scandinavia, the Alps, Greenland and the Hudson Bay region of northeastern North America (*B. alpina* L.), and the remaining 45 species distributed throughout the páramos of Andean South America (Molau 1990). Broad-scale phylogenetic studies of Orobanchaceae (Wolfe et al. 2005; Bennett and Mathews 2006) and the Rhinantheae clade (Těšitel et al. 2010) had suggested that *Bartsia* was not monophyletic, but Scheunert et al. (2012) were the first to include species from the complete geographic distribution of the genus, as well as the two species of the related Mediterranean genus *Parentucellia*. However, because their sampling only included two species of the South American clade of *Bartsia*, they chose to only reclassify these two species, leaving ca. 43 species in a large polyphyletic group with the monotypic lineage of *B. alpina* in Europe. The South American *Bartsia* species, which we will refer to here as the *Neobartsia* clade, are quite distinct from their Mediterranean counterparts (i.e., *Bellardia trixago,* and the two species of *Parentucellia* that were also moved to the expanded genus—*Bellardia latifolia* and *B.* viscosa) in multiple aspects. Ecologically, *Neobartsia* species grow at high elevation (ca. 3,000–5000 m) in wet environments while the Mediterranean species grow at low elevation (ca. 0–500 m) in seasonally dry environments. Geographically, *Neobartsia* is restricted to the Andes while the Mediterranean taxa are native to the Mediterranean region and more

recently been introduced to Australia, coastal Chile, and coastal western North America. Finally, the Mediterranean species all have reflexed corolla lips, usually associated with bee pollination, whereas a large number of the species in *Neobartsia* have erect corolla lips that are thought to be associated with hummingbird pollination due to their tubular shape and the placement of reproductive parts.

Previous studies of the group have only included a minor fraction of the South American species richness, usually sampling only one or two species, making it difficult to assess the influence of biogeographic movements on rates of diversification across the clade. Here, we included representatives from all of the major lineages comprising the former genus *Bartsia*, as well as a representative sampling of all known allied genera of the Rhinantheae clade of Orobanchaceae, to establish a robust and well-supported phylogeny of the clade based on both chloroplast and nuclear ribosomal DNA sequence data. We then use this phylogeny to estimate divergence times across the clade and to investigate the biogeographic history of the clade, with a special focus on the origin of the *Neobartsia* clade in Andean South America. Finally, we use all these analyses to test if increases in rates of diversification are indeed associated with biogeographic movements into newly formed environments, i.e., "dispersification" sensu Moore and Donoghue (2007).

Methods

Sampling

A total of 49 taxa were included in this study (Table 1.1). Because our main focus is the diversification dynamics of the South American *Neobartsia* clade in the context of the disparate geographic distributions of the Old World species and the remainder of the mostly

European Rhinantheae clade of Orobanchaceae, our sampling effort included representatives of 10 of the 11 genera thought to comprise the clade (Wolfe et al. 2005; Bennett and Mathews 2006; Těšitel et al. 2010; Scheunert et al. 2012; McNeal et al. 2013). *Bellardia* and the *Neobartsia* clade are represented here by 15 South American species and two of the three Mediterranean taxa, *Bellardia trixago* (L.) All. and *Bellardia viscosa* (L.) Fisch. & C.A. Mey. Based on previous results (Olmstead et al. 2001; Wolfe et al. 2005; Bennett and Mathews 2006; Těšitel et al. 2010; Scheunert et al. 2012; McNeal et al. 2013) *Melampyrum* L. was used as a functional outgroup for the Rhinantheae clade.

Molecular methods

Total genomic DNA was extracted from silica gel-dried tissue or herbarium material using a modified 2X CTAB method (Doyle and Doyle 1987). Two chloroplast (cp) regions–*trnT-trnF* region and the *rps16* intron–were amplified via polymerase chain reaction (PCR) using the *trn*-a and *trn*-f (Taberlet et al. 1991) and the *rps16*_F and *rps16*_R primers (Oxelman et al. 1997), respectively. The nuclear ribosomal (nr) internal transcribed spacer (ITS) and external transcribed spacer (ETS) regions were amplified using the ITS4 and ITS5 primers (Baldwin 1992) and the ETS-B (Beardsley and Olmstead 2002) and 18S-IGS (Baldwin and Markos 1998), respectively. PCR profiles for all regions followed Tank and Olmstead (Tank and Olmstead 2008). When amplification of a region in one fragment was not possible, internal primers were used to amplify the region in multiple fragments. The primer pairs *trn-a/trnb, trnc/trn-d,* and *trn-e/trn-f* (Taberlet et al. 1991) were used to amplify the *trnT-trnF* region. Additionally, *Bellardia* specific internal primers were designed and used when these primer combinations failed (*trnT/trnL* intergenic spacer: *trnT-L_iF* 5-

CTTGGTTTTCATCCGTAAAGG-3 and *trnT-L_iR* 5-CCTTTACGGATGAAAACCAAG-3). Following Tank and Olmstead (2008), the *rps16*_F/*rps16*_iR and *rps16*_iF/*rps16*_2R primer combinations were used to amplify the *rps16* intron in two fragments. Similarly, the ITS5/ITS2 and ITS3/ITS4 primer combinations (Baldwin 1992) were used to amplify the ITS region in two fragments.

PCR products were purified by precipitation in a 20% polyethylene glycol 8000 (PEG)/2.5 M NaCl solution and washed in 70% ethanol prior to sequencing. To ensure accuracy, we sequenced both strands of the cleaned PCR products on an ABI 3130xl capillary DNA sequencer (Applied Biosystems, Foster City, California, USA) using ABI BigDye v.3.1 cycle sequencing chemistry. Sequence data were edited and assembled for each region using the program Sequencher v.4.7 (Gene Codes Corp., Ann Arbor, Michigan, USA), and consensus sequences were generated. When sequencing was not possible for any given species or gene region, GenBank sequences were used to reduce the amount of missing data in the final matrix (Table 1.1).

Table 1.1
Taxa and voucher information for plant material from which DNA was extracted. GenBank accession numbers are also provided. Herbarium abbreviations are as follow: FHO = University of Oxford Herbarium , K = Royal Botanic Gardens Kew , ID, University of Idaho = Stillinger Herbarium , ANDES = Museo de Historia Natural Universidad de los Andes , WTU = University of Washington Herbarium , GH = Harvard University Herbarium , LJU = University of Ljubljana Herbarium , USFS = United States Forest Service , CBFS = University of South Bohemia České Budějovice. Empty cells will be filled with GenBank accession numbers upon publication.

		GenBank Accession Number				
Species	DNA Voucher/ Herbarium	ITS	ETS	*trnT-L*	*trnL-F*	*rps16*
Bartsia alpina L.	Lampinen s/n /ID	FJ790046				
B. crenoloba Wedd.	Solomon 7152/K					
B. laniflora Benth.	SU-24/ANDES					
B. laticrenata Benth.	Ramsay & Merrow-Smith 771/K					
B. melampyroides (Kunth) Benth.	Tank 2005-07/WTU					
B. orthocarpiflora Benth.	Ollgaard 34129/K					
B. pedicularoides Benth.	Jorgenson 1729/K	FJ790047		FJ790077		
B. pyricarpa Molau	Tank 2005-36/WTU					
B. ramosa Molau	CG-016/ANDES					
B. santolinifolia (Kunth) Benth.	SU-18/ANDES					
B. sericea Molau	Tank 2005-06/WTU					
B. cf *sericea*	Tank 2005-25/WTU					
B. cf *inaequalis ssp. duripilis*	Tank 2005-29/WTU					
B. stricta (Kunth) Benth.	SU-1b/ANDES					
B. tenuis Molau	Tank 2005-02/WTU					
B. thiantha Diels	RGO 2009-23/WTU					
Bellardia trixago (L.) All.	Bennett s/n /FHO	FJ790063				
B. viscosa (L.) Fisch. & C.A. Mey	Halse 2249/ID	AY911244				
Euphrasia alsa F.Muell.	Zich 220/GH					
E. collina R.Br.	Zich 209/GH					
E. mollis (Ledeb.) Wettst.	Mancuso 107/ID					
E. regelii Wettst.	Ho 1741/GH					
E. stricta D. Wolff ex J.F. Lehm	Musselman 4872/ID					
Hedbergia abyssinica (Benth.) Molau var. *abyssinica*	Etuge 3488/K	FJ790061				
H. abyssinica (Benth.) Molau var. *nykiensis*	Carter et al 2386/K					
H. abyssinica (Benth.) Molau var. *petitiana*	Paton s/n /K					
H. decurva (Hochst. ex Benth.) A. Fleischm. & Heubl	Wesche 9/K					
H. longiflora ssp. longiflora (Hochst. ex Benth.) A. Fleischm. & Heubl	Kisalye van Heist 109/K					
Lathraea squamaria L.	Frajman s/n /LJU	FJ790044		EU264174		
Melampyrum carstiense Fritsch	Krajsek s/n /LJU	GU445314		EU264177		
M. lineare Lam.	Bjork 6465/ID					
M. sylvaticum L.	Krajsek s/n /LJU	EU624134				

Table 1.1 (continued)

		GenBank Accession Number				
Species	DNA Voucher	ITS	ETS	*trnT-L*	*trnL-F*	*rps16*
Odontites corsicus (Loisel.) G. Don	J. Stefani/ID					
O. aucheri (Boiss.) Rothm	Oganesian et al. 03-1575/K					
O. linkii Heldr. & Sart. ex Boiss. *ssp cyprius*	Ferguson 4537/K					
O. maroccanus Bolliger	Gattefose s/n /K					
O. vulcanicus Bolliger	Bolliger & Moser O-M3/K					
O. vulgaris Moench	Kharkevich s/n /K					
Rhinanthus crista-galli L.	Bjork 6656/ID					
R. freynii Fiori	Mathews 04-05 / n/a	GU445319				
R. kyrollae Chabert	Stickney 1236/USFS					
R. serotinus (Schönh.) Oborny	Musselman 4871/ID					
Rhynchocorys elephas Griseb.	Tesitel 5044/CBFS	FJ790055		FJ790085		
R. kurdica Nábělek	Tesitel 5042/CBFS	FJ790037		FJ790067		
R. maxima Richter	Tesitel 5040/CBFS	FJ790036		FJ790066		
R. odontophylla R.B.Burbidge & I.Richardson	Tesitel 5038/CBFS	FJ790034		FJ790064		
R. orientalis Benth.	Tesitel 5039/CBFS	FJ790035		FJ790065		
R. stricta Albov	Tesitel 5047/ CBFS	FJ790057		FJ790087		
Tozzia alpina L.	Mathews 04-04/ n/a	AY911258				

Phylogenetic analyses

Although two separate cpDNA regions were sequenced, the evolutionary histories of the *trnT-trnF* region and the *rps16* intron are tightly linked due to the nonrecombining nature of the chloroplast genome, and thus, were treated as a single locus. The nrDNA regions (ITS and ETS) were also treated as a single locus, given that they are linked because of their physical proximity in the nrDNA repeat. We created three primary datasets with our two independent loci: 1) cpDNA only, 2) nrDNA only, and 3) a combined cpDNA and nrDNA dataset. Global alignments across the Rhinantheae clade were created for each gene region using the group-to-group profile alignment method as implemented in Muscle v.3.6 (Edgar

2004). This method takes advantage of previous knowledge about monophyly of the major lineages (e.g., Těšitel et al. 2010; Scheunert et al. 2012) and consists of lineage-specific alignments that are then iteratively aligned to one another resulting in fewer alignment ambiguities (Smith et al. 2009). These alignments were visually inspected and minor adjustments were made manually using Se-Al v.2.0a11 (Rambaut 1996). Sites that could not be unambiguously aligned were excluded from the analyses. File format conversions and matrix concatenations were performed using the program Phyutility v.2.2 (Smith and Dunn 2008).

A statistical selection of the best-fit model of nucleotide substitution according to the Akaike information criterion (AIC) was conducted independently for each gene region using the program jModelTest (Guindon and Gascuel 2003; Posada 2008). Based on these results, partitioned (by gene region) maximum likelihood (ML) analyses were performed on our three primary datasets using RAxML v. 7.2.4 (Stamatakis 2006) with 1,000 replicates of nonparametric bootstrapping using the rapid bootstrap algorithm (Stamatakis et al. 2008). Every fifth bootstrap tree generated by the rapid bootstrap analyses was used as a starting tree for full ML searches and the trees with the highest ML scores were chosen. Likewise, partitioned Bayesian inference (BI) analyses were performed using the parallel version of MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003) with the individual parameters unlinked across the data partitions. Analyses consisted of two independent runs with four Markov chains using default priors and heating values. Each independent run consisted of 15 million generations and was started from a randomly generated tree and was sampled every 1,000 generations. Convergence of the chains was determined by analyzing the plots of all parameters and the –lnL using Tracer v.1.5 (Rambaut and Drummond 2004). Stationarity

was assumed when all parameters values and the –lnL had stabilized; the likelihoods of independent runs were considered indistinguishable when the average standard deviation of split frequencies was < 0.001. Consensus trees were obtained for each dataset using the sumt command in MrBayes. Finally, incongruencies between the cpDNA and the nrDNA topologies were investigated using the approximately unbiased (AU) test (Shimodaira 2002) and the Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa 1999), as implemented in the program CONSEL (Shimodaira and Hasegawa 2001).

Divergence time estimation

To maximize the number of taxa and minimize the amount of missing data in our dating analyses, we reduced our combined dataset to include sequences for only the cpDNA *trnT-trnL* intergenic spacer and the nrDNA ITS region. This resulted in a dataset that included all 49 taxa and only 2% missing data, compared to 15% missing data for the complete dataset. Each gene was treated as a separate partition. To ensure convergence in divergence times, five independent runs were conducted using BEAST v.1.5.4 (Drummond and Rambaut 2007). BEAST implements Markov Chain Monte Carlo (MCMC) methods that allow for uncertainty in both the topology and the calibration points, i.e., calibration points are treated as probabilistic priors, rather than point estimates (Ho and Phillips 2009). It also implements an uncorrelated lognormal relaxed clock (UCLN) (Drummond et al. 2006), allowing every branch to have an independent substitution rate.

Each run was started from the resulting ML tree obtained for the dataset containing all regions, after performing a semiparametric rate smoothing based on penalized likelihood

(Sanderson 2002) in R (R Development Core Team 2013) using the package Ape (Paradis et al. 2004). Each run consisted of 100,000,000 generations sampled every 1000 trees. The models of nucleotide substitution were kept unlinked for both partitions and the tree priors were kept as default under the birth-death process.

Because of the mostly herbaceous habit of the species in Orobanchaceae, there are no known fossils for the family. This lack of fossils made the dating of our analyses dependent on secondary calibrations obtained from a previous study. Based on age estimates of an ITS molecular clock (Wolfe et al. 2005), a calibration point at the node containing every genus except *Melampyrum* (i.e.*,* one node higher than the root) was used. This was done with a lognormal distribution prior with an offset of 25 million years (Ma), a mean of 0.9, and a standard deviation of 0.8, this way incorporating uncertainty in the calibration point. Because the use of this secondary calibration is far from ideal, to corroborate our calibration strategy, an additional analysis using the most recent uplift of the Andes as the calibration point (Simpson 1975; Burnham and Graham 1999; Gregory-Wodzicki 2000; Antonelli et al. 2009) was conducted with a lognormal distribution prior (offset of 1.7 Myr, a mean of 0.2 and an standard deviation of 0.6). This calibration prior was set at the node where the species in *Neobartsia* diverge from *B. viscosa*. This additional calibration scenario was conducted to assess the impact of alternative calibration points in the node ages.

Convergence of the parameters was monitored using Tracer v. 1.5 and the resulting trees were summarized using TreeAnnotator v.1.5.4 (Drummond and Rambaut 2007) after 25% of the trees had been discarded as burn-in. Each of the five topologies and their node heights were visualized using FigTree v. 1.3.1 (Rambaut 2006) and a final tree, representing the maximum clade credibility tree with information of the 95 percent highest posterior

density (HPD), was obtained by combining the five runs using LogCombiner v.1.5.4 (Drummond and Rambaut 2007) and by summarizing them with TreeAnnotator v.1.5.4.

Biogeographic analyses

The biogeographic history of *Bellardia* and allied genera was reconstructed using the program Lagrange v. C++ (Ree and Smith 2008). Lagrange implements the Dispersal-Extinction-Cladogenesis (DEC) model (Ree et al. 2005) to estimate the most likely ancestral geographic range based on current distributions of extant lineages. This model assumes extinction or dispersal by contraction or expansion of the ancestral geographic range, respectively. Additionally, the user is given the option to assign a dispersal probabilities matrix based on prior knowledge of connectivity between areas, incorporating valuable ancestral geographic information. However, most of this knowledge, at least for Northern Hemisphere temperate plants, is based on macrofossils of woody mesophytic taxa, e.g., *Quercus* (Tiffney and Manchester 2001). Because the herbaceous genus *Bellardia* is almost completely restricted to alpine-like conditions, which separates this lineage from the ecological conditions in which mesophytic forest species are found, and the vast majority of the Rhinantheae clade is also herbaceous, we consider that biological routes for these types of taxa are less well understood (Donoghue and Smith 2004). Therefore, we did not include a dispersal probability matrix in our analyses (see also Smith and Donoghue 2010).

We used Lagrange on a posterior distribution of 1,000 randomly chosen trees (post burn-in) from our dating analyses. By inferring ancestral ranges over a posterior distribution of trees we are incorporating uncertainty in both topology as well as times of divergence (Smith 2009; Smith and Donoghue 2010; Beaulieu et al. 2013). We conducted three

independent analyses with varying distributions of current taxa. The first analysis was performed with conservative geographic ranges following Mabberley's Plant-Book (Mabberley 2008), in which the genera have wider distributions, e.g,. the genus *Euphrasia* L. has a north temperate distribution (Eurasia, Europe and Eastern North America). The second analysis included prior expert knowledge about the distribution of the genera based on published work, e.g., we followed the proposed Eurasian origin for the genus *Euphrasia* (Gussarova et al. 2008). The final analysis was based on species-specific distributions based on the explicit species that we sampled, i.e., species within a genus can have different distributions to account for endemisms and/or disparate distributions within a genus.

We considered species to be distributed in five distinct geographic areas (Fig. 1.1): i) Eurasia (western Eurasia: the Balkan Peninsula and the Caucasus region), ii) Europe (including the Mediterranean climatic region in southern Europe and northern Africa), iii) Africa (montane northeastern Africa), iv) North America (Hudson Bay region of northeastern North America), and v) South America (including only the Andes). The results of the analyses were summarized in R. Following Beaulieu et al. (2013), we calculated Akaike weights for every biogeographic scenario reconstructed at every node in each tree separately. We then summed the Akaike weights for each node and averaged them across the distribution of trees, which resulted in composite Akaike weights (w_i) for our biogeographic reconstructions. Furthermore, we examined the evidence for the most supported scenario by calculating an evidence ratio of this model versus all models (Burnham and Anderson 2002). These were interpreted as relative evidence of one scenario being the most supported when comparing it with competing biogeographic hypotheses (Beaulieu et al. 2013).

Diversification rates

Diversification rates analyses were conducted on the same posterior distribution of 1,000 trees, as well as on the maximum clade credibility (MCC) tree using MEDUSA (Alfaro et al. 2009), which is an extension of the method described by Rabosky et al. (2007) and is available in the R package geiger 2.0 (Pennell et al. 2014). In Rabosky et al. (2007), two likelihoods are estimated for a dated tree: i) a phylogenetic likelihood that uses the timing of the splits on the backbone to estimate ML values for birth and death rates following the equations of Nee et al. (1994), and ii) a taxonomic likelihood that uses species richness along with the date of the splits, estimating diversification rates following Magallón and Sanderson (2001). MEDUSA (Alfaro et al. 2009) looks for shifts in diversification rates in a stepwise manner by comparing AIC scores of successively more complex models. This method requires complete sampling that is achieved by collapsing every clade to a single tip and then assigning clade richnesses to these tip lineages. We collapsed our trees into tips representing each of the major lineages of Rhinantheae, which in most cases corresponds to each of the genera. The following clade richnesses were used: *Neobartsia* clade (45 spp.), *Parentucellia* clade (2 spp.), *Bellardia* clade (1 sp.), *Odontites* clade (32 spp.), *Euphrasia* clade (350 spp.), *Rhinanthus* clade (45 spp.), *Melampyrum* clade (35 spp.), *Lathraea* clade (7 spp.), *Rhynchocorys orientalis* (1 spp.), *Rhynchocorys elephas* (1 spp.), *Rhynchocorys odontophylla* (1 spp.), *Rhynchocorys kurdica* (1 spp.), *Rhynchocorys stricta* (1 spp.), *Rhynchocorys maxima* (1 spp.), *Bartsia alpina* (1 spp.), *Tozzia alpina* (1 spp.), *Hedbergia abyssinica* (1 spp.), *Bartsia decurva* (1 spp.), *Bartsia longiflora* (1 spp.).

To compare our MEDUSA results we conducted an additional analysis using the program SymmeTREE v1.0 (Chan and Moore 2005). SymmeTREE is based on the

topological distribution of species on the whole tree, which is compared to a distribution simulated on a tree under the equal-rates Markov random branching model (EMR), where the probability of a branching event is constant throughout the tree (Yule 1924). If a clade shows an unbalanced distribution of species richness when compared to its sister clade, then a shift in the rate of diversification is identified. SymmeTREE also estimates several whole tree statistics that are evaluated against their own simulated null distribution, i.e., a constant pure-birth model (Chan and Moore 2005). To accommodate topological and temporal uncertainty, we assessed diversification rate shifts with SymmeTREE using default settings across a random set of 542 trees from the posterior distribution of trees from our divergence time analysis; the full set of 1,000 trees was not used due to computational limitations.

Results

Molecular methods

The cpDNA data set included the *trnT-trnF* region and the *rps16* intron and had a total length of 2,686 bp with 13% missing data (Table 1.1). Similarly, the nrDNA dataset included the ITS and ETS regions with a total of 1,134 bp and 17% missing data (Table 1.1). A combined data set was created from the cpDNA and the nrDNA matrices, with a total length of 3,820 bp and 15% missing data.

Phylogenetic analyses

Alignment of individual gene regions was straightforward requiring minor adjustments to the automated alignment strategy implemented in MUSCLE v. 3.6. Some regions that could not be unambiguously aligned in the *trnT-trnL* intergenic spacer and in the

ETS region were excluded from the analyses (*trnT-L*: alignment positions 519–529 and 587–620; ETS: alignment positions 63–65, 83–85 and 152–158). Model selection for the cpDNA regions yielded the General Time Reversible model + Γ (GTR) (Rodríguez et al. 1990) for the *trnT-trnF* intergenic spacer, and the Transversion model + Γ (TVM) for the *rps16* intron. The ITS and ETS regions resulted in the selection of the GTR+I+Γ and Hasegawa-Kishino-Yano+ Γ (HKY) models, respectively. To avoid the difficulties of estimating Γ and the invariable sites simultaneously (Ronquist and Huelsenbeck 2003; Yang 2006), the model of substitution GTR+Γ with an increase in the number of rate categories from four to six was preferred in the case of the ITS region.

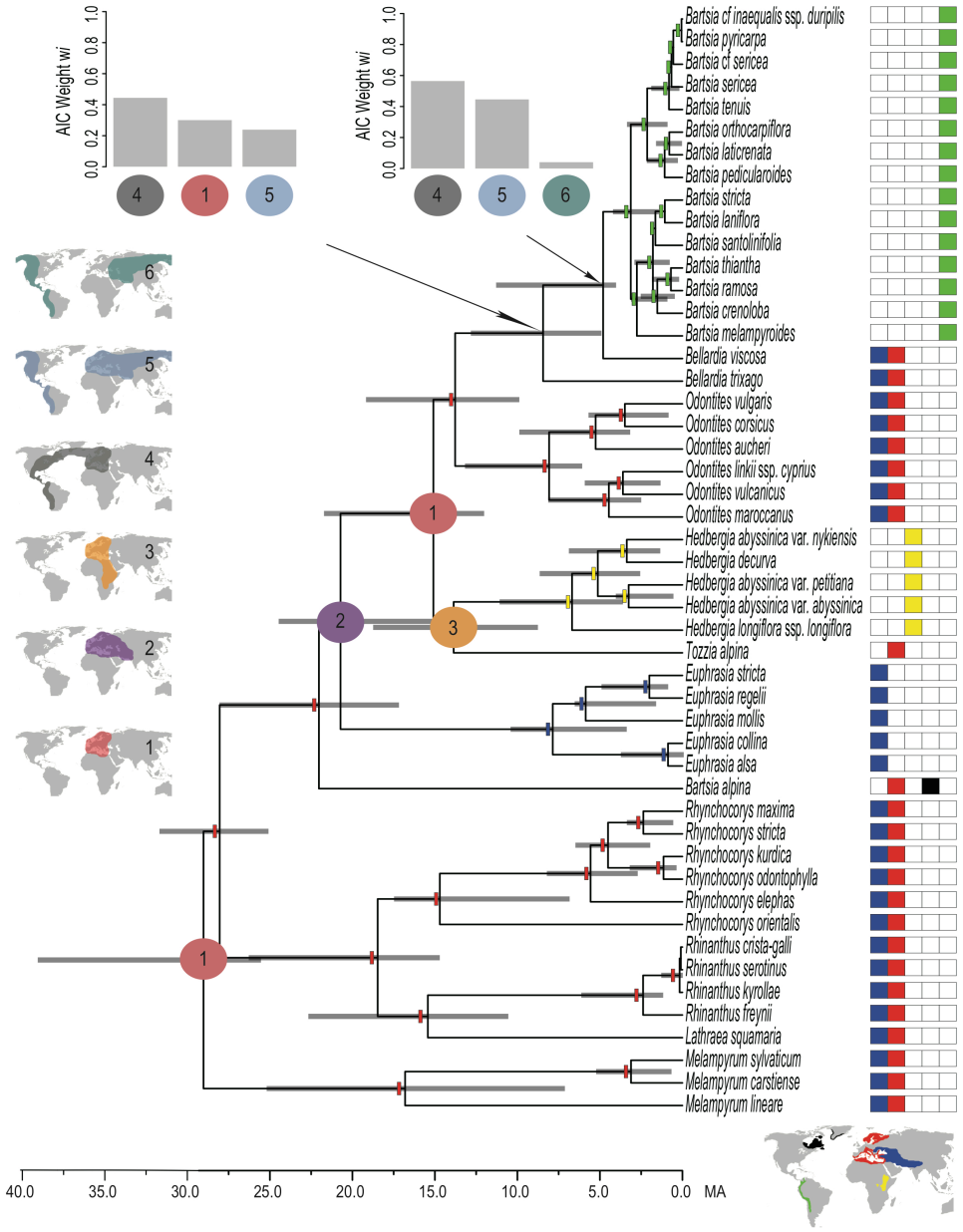
Our results from every dataset (Fig. 1.2 for the combined dataset and Fig. 1.3 for the cpDNA and nrDNA datasets) are in concordance with those presented in previous Rhinantheae studies (Těšitel et al. 2010; Scheunert et al. 2012), and assessment of incongruences between the cpDNA and nrDNA datasets showed that these were either not significant, or if they were, the alternative topology was only weakly supported. For example, the well-supported relationships in the cpDNA dataset between *Tozzia alpina* and *Hedbergia*—or between *Odontites* and *Bellardia*—are not statistically significant when constrained in the nrDNA dataset. Conversely, the relationship between *H. abyssinica* var. *nykiensis* and *H. decurva* found in the cpDNA dataset is significant in the AU Test, but it is only moderately supported on the tree (BS 72, PP 0.96) and it does not exist in the combined dataset (Table 1.2). An important new result from this study, which is based on the first comprehensive sampling of the group, is that the South American species indeed form a distinct clade, the *Neobartsia* clade, that is very well supported with a posterior probability (PP) of 1.0 and a bootstrap support (BS) of 100. *Bellardia*–including the *Neobartsia* clade–is

sister to *Odontites* (PP 1.0, BS 92) and together are sister to a clade comprised by *Hedbergia* and *Tozzia alpina* (PP 1.0, BS 93). The placement of the genus *Tozzia* was uncertain until now, although the support of our analyses is marginal (PP 0.94, BS 80). Finally, the genus *Euphrasia* is sister to the latter genera (PP 1, BS 100) and together form a clade sister to *Bartsia alpina* (PP 1, BS 100). This last clade is what Scheunert et al. (2012) referred to as the core Rhinantheae.

Table 1.2
Results for the Approximately Unbiased (AU) and the Shimodaira-Hasegawa (SH) tests at $p < 0.05$ for different constrained relationships. Log likelihood scores for the original analysis are given, as well as the difference in log likelihood between the original and the constraint topology (∂). Values in bold are significant with 95% confidence.

nrDNA analysis constraint compared to clades from the cpDNA analysis	ln likelihood	∂	AU	SH
Unconstrained nrDNA analysis	-8445.56			
Tozzia + *Hedbergia*	-8449.04	3.48	0.166	0.186
H. decurva + *H. abyssinica* var. *nykiensis*	-8464.51	18.95	**0.004**	**0.026**
Odontites + *Bellardia*	-8456.43	10.87	0.141	0.131
cpDNA analysis constraint compared to clades from the nrDNA analysis				
Unconstrained cpDNA analysis	-10044.47			
Tozzia + *Bellardia*	-10064.45	19.98	**0.001**	**0.017**
H. decurva + *H. longiflora* ssp. *longiflora*	-10053.20	8.73	0.07	0.101
Odontites + *Euphrasia*	-10071.61	27.14	**0.0004**	**0.008**

Figure 1.1
Topology obtained after combining and annotating five independent BEAST analyses. The calibration point was set at the node where all genera are included except for *Melampyrum*. The calibration had a prior with a lognormal distribution, offset 25 Ma, a mean of 0.9, and a standard deviation of 0.8 following dates by by Wolfe et al. (2005). Time in millions of years ago (Ma) is represented by the scale below the tree. Current distributions of the species are color-coded after the species names. The current distributions are plotted on a map below the species names and correspond to blue for Eurasia, red for Europe, yellow for Africa, black for northeastern North America, and green for South America. The most supported ancestral range reconstructions obtained from a Lagrange analysis, are plotted on the tree with color rectangles or circles with numbers that represent different biogeographic hypotheses. Ancestral range reconstruction scenarios are plotted on five different maps on the left of the figure, each with a number that distinguishes it. Composite Akaike weights (w_i) are plotted in the form of histograms for nodes where the reconstruction had competing hypotheses. Two possible routes of migration, one including the North Atlantic Land Bridge (NALB) and one including the Bering Strait, are shown on maps 5 and 6.



Divergence time estimations

When the calibration point was placed at the node where *Melampyrum* diverged from the remaining genera, the South American *Neobartsia* clade was inferred to have a mean age of 3.13 Ma (1.53–4.11, 95% HPD) (Table 1.3). The split between *Bellardia trixago* and the remaining species in the clade was estimated to have a mean age of 8.44 Ma (5.01–12.72 Ma). The African clade diverged from *Tozzia alpina* 13.86 Ma (8.86–18.64 Ma), while the split of the European *Bartsia alpina* occurred 22.03 Ma (17.28–27.95 Ma). The root of the tree was estimated to have a mean age of 29.02 Ma (25.65–38.96 Ma). Likewise, when the geological constraint was imposed, the *Neobartsia* clade had a mean age of 1.85 Ma (1.97–3.58 Ma), and the divergence of *Bellardia trixago* from the remaining *Bellardia-Neobartsia* species occurred 6.73 Ma (4.95–12.48 Ma). The African clade had a mean age of 12.31 Ma (8.69–18.75 Ma), *Bartsia alpina* of 17.09 Ma (16.23–28.36 Ma), and the root of the tree was estimated at 29.80 Ma (29.13–35.96 Ma). The age of the root is consistent with the date (35.5 Ma) inferred for this clade in an angiosperm wide analysis (Zanne et al. 2014). Because the results using different calibration strategies were within the 95 percent HPD of each other (see Table 1.3), we used the root calibration analysis for subsequent biogeographic and diversification rate analyses.

Table 1.3
Divergence time estimates for the main clades, with each node representing the most recent common ancestor (mrca) of the taxa mentioned. The first value was obtained by calibrating the node of divergence of *Melampyrum* from its sister clade. The calibration point had a prior with lognormal distribution, offset 25 Myr, mean of 0.9, and standard deviation of 0.8, using the results of Wolfe et al. (2005), but incorporating considerable temporal uncertainty. The second value corresponds to an additional analysis where the uplift of the Andes was used as the calibration point of the node of divergence for S. Am. *Bellardia*. This last calibration had a prior with a lognormal distribution, offset of 1.7 Myr, mean of 0.2, and standard deviation of 0.6. Mean age estimates as well as the 95% highest posterior density (HPD) are shown for both analyses. In addition, the composite Akaike weights (w_i) from our biogeographic analyses are shown for the 'expert based' coding of current geographic distributions with the following abbreviations: A (Africa), EU (Europe), EUR (Eurasia), ENA (Eastern North America), SAM (South America). Evidence ratio is presented for the most supported geographic reconstruction.

Node (mrca)	Mean Age (Ma)	95% HPD (Ma)	w_i	Evidence Ratio
Root	29.02 / 29.80	25.65-38.96 / 29.13-35.96	EU 0.31	1.82
Melampyrum	16.80 / 14.70	7.22-25.09 / 7.14-24.32	EU 0.18	1.8
Lathraea-Bellardia	28.04 / 21.73	25.19-31.59 / 20.87-33.57	EU 0.53	3.8
Rhynchocorys-Lathraea	18.47 / 17.72	14.81-26.17 / 13.97-26.88	EU 0.58	2.4
Rhinanthus-Lathraea	15.43 / 12.77	10.65-22.57 / 10.37-23.14	EU 0.53	2.2
Rhynchocorys	14.70 / 7.65	6.94-17.37 / 6.69-17.45	EU 0.49	2.13
Bartsia alpina-Bellardia	22.03 / 17.09	17.28-27.95 / 16.23-28.36	EU 0.44	1.69
Euphrasia-Bellardia	20.72 / 15.72	14.34-24.37 / 13.72-24.60	EU+EUR 0.43	1.95
Euphrasia	7.86 / 4.93	3.48-10.31 / 3.60-10.38	EUR 1.0	-
Tozzia-Bellardia	15.09 / 14.40	12.12-21.62 / 11.61-21.48	EU 0.53	1.76
Tozzia-Hedbergia	13.86 / 12.31	8.86-18.64 / 8.69-18.75	A+EU 0.94	23.5
H. longiflora-H. decurva	6.87 / 3.29	3.71-10.97 / 3.60-10.96	A 0.99	247.5
Odontites-Bellardia	13.77 / 12.21	9.99-19.07 / 9.93-19.19	EU 0.80	8.88
Odontites	8.08 / 7.32	6.18-13.07 / 5.94-13.29	EU 1.0	-
Bellardia trixago-Neobartsia clade	8.44 / 6.73	5.01-12.72 / 4.95-12.48	EU+SAM 0.26 EU 0.18 EU+EUR+SAM 0.16	1.44 1.13 -
Bellardia viscosa-Neobartsia clade	4.80 / 5.63	4.12-11.19 / 4.03-10.84	EU+SAM 0.53 EU+EUR+SAM 0.35	1.51 11.66
Neobartsia clade	3.13 / 1.85	1.53-4.11 / 1.97-3.58	SAM 1.0	-

Biogeographic analyses

Our three different codings of current geographic distribution resulted in very similar ancestral reconstructions (Table 1.4). Given that so much work has been done in recent years for several of these groups, e.g., *Bartsia/Bellardia* (Molau 1990)*, Euphrasia* (Gussarova et al. 2008), and *Odontites* (Bolliger 1996), we favored the second coding scenario where current distributions were based on expert knowledge, including recent phylogenetic and biogeographic studies (for a wide-scale example on campanulids see Beaulieu et al. 2013). The most recent common ancestor (mrca) of the Rhinantheae clade of Orobanchaceae was likely distributed in Europe with a composite Akaike weight (w_i) of 0.31 and an evidence ratio of 1.82 (Table 1.2). This ancestral range is maintained throughout the backbone of the tree until the node where *Euphrasia* diverges from the rest of the genera (w_i = 0.43, evidence ratio = 1.95). Nevertheless, a European ancestral range becomes the most supported reconstruction again at the node of divergence of *Odontites* (w_i = 0.80, evidence ratio = 8.88). A South American ancestral range is included for the first time at the crown node of *Bellardia*, where a w_i 0.26 supports a split between Europe and South America and a w_i of 0.18 supports an entirely European ancestral range; the evidence ratio between these two reconstructions is 1.44. The node where the *Neobartsia* clade diverges from the Mediterranean *Bellardia viscosa* is again supported by two competing models i) a split between Europe and South America (w_i = 0.53, evidence ratio = 1.51) and ii) one between South America and an area comprised of Europe and Eurasia (w_i = 0.35). Additional results for other genera can be seen on Figure 1.1 and summarized in Table 1.4.

Table 1.4
The composite Akaike weights (w_i) are shown for our three different coding scenarios: conservative, expert based, and species specific. Abbreviations are as follow: A (Africa), EU (Europe), EUR (Eurasia), ENA (Eastern North America), SAM (South America). Evidence ratio is presented for the most supported geographic reconstruction.

Coding Scheme	*Conservative*		*Expert based*		*Species specific*	
Node (mrca)	w_i	Evidence Ratio	w_i	Evidence Ratio	w_i	Evidence Ratio
Root	EU 0.52	3.71	EU 0.31	1.82	EU+ENA 0.40	1.05
Melampyrum	EU 0.25	3.57	EU 0.18	1.8	EU+ENA 0.93	4.65
Lathraea-Bellardia	EU 0.81	11.6	EU 0.53	3.8	EU 0.76	8.4
Rhynchocorys-Lathraea	EU 0.73	8.11	EU 0.58	2.4	EU 0.92	13.14
Rhinanthus-Lathraea	EU 0.67	6.1	EU 0.53	2.2	EU 0.96	13.7
Rhynchocorys	EU 0.62	5.16	EU 0.49	2.13	EU 0.70	2.59
Bartsia alpina-Bellardia	EU 0.84	16.8	EU 0.44	1.69	EU 0.59	3.10
Euphrasia-Bellardia	EU 0.80	26.6	EU+EUR 0.43	1.95	EU+EUR 0.40	1.81
Euphrasia	EU 0.26	2.88	EUR 1.0	-	EUR 1.0	-
Tozzia-Bellardia	EU 0.74	5.28	EU 0.53	1.76	EU 0.53	1.96
Tozzia-Hedbergia	A+EU 0.90	11.25	A+EU 0.94	23.5	A+EU 0.91	15.16
H. longiflora-H. decurva	A 0.98	122.5	A 0.99	247.5	A 0.98	81.66
Odontites-Bellardia	EU 0.79	26.03	EU 0.80	8.88	EU 0.76	7.6
Odontites	EU 0.74	6.16	EU 1.0	-	EU 0.93	15.5
B. trixago-Neobartsia	EU+SAM 0.21	1.31	EU+SAM 0.26	1.44	EU+SAM 0.24	1.33
clade	EU 0.16	1.45	EU 0.18	1.13	EU 0.18	1.06
	EU+EUR+SAM 0.11	-	EU+EUR+SAM 0.16	-	EU+EUR+SAM 0.17	-
B. viscosa-Neobartsia	EU+SAM 0.48	1.77	EU+SAM 0.53	1.51	EU+SAM 0.49	1.29
clade	EU+EUR+SAM 0.27	6.75	EU+EUR+SAM 0.35	11.66	EU+EUR+SAM 0.38	12.66
Neobartsia clade	SAM 1.0	-	SAM 1.0	-	SAM 1.0	-

Diversification rates

Our analyses discovered six shifts in the rate of net diversification (r = speciation minus extinction) in the Rhinantheae clade when performed over the posterior distribution of trees; three of these were also identified on the MCC tree. Importantly, the three shifts identified on the MCC tree corresponded to the shifts that occurred at the highest frequency in the analyses across the posterior distribution of trees. Because most of these analyses were conducted on a posterior distribution of trees to incorporate phylogenetic uncertainty (both temporal and topological), we report the mean net diversification rate of each shift (r_{mean}) on the text and the ranges of these shifts in table 1.5. For the three shifts found in the MCC tree, we also report that value (r_{mcc}). The first two shifts found in our analyses correspond to shifts that were only present in less than 15 percent of the trees and show minimal deviation from the background rate of the tree. One of these shifts is on the node subtending the core Rhinantheae (r_{mean} = 0.11, frequency = 0.07) and the other one involves the hemiparasitic genus *Rhinanthus* L. and the holoparasite *Lathraea* L. (r_{mean} = 0.17, frequency = 0.12). The latter shift could correspond to a change in life history from hemiparasitism to holoparasitism in *Lathraea*, but given the limited sampling of these two groups and the low frequency at which the shift was found we dare not comment further. The next shift involves a slowdown in the rate of *Bartsia alpina* (r_{mean} = –0.4, r_{mcc} = 0) and was the most frequent shift in the analyses (frequency = 1.17). The frequency higher than 1.0 for this node is an artifact of the way MEDUSA adds the shifts. When two sister clades each have a shift at their crown nodes, MEDUSA adds the parameters from both shifts and places the result on the stem leading to the two clades. Thus these shifts do not occur with a frequency higher than 1.0, but are very common. The fourth shift corresponds to an increase in net diversification (r_{mean} = 0.09) in

the clade sister to *Bartsia alpina* and was found with a frequency of 0.32. An additional shift was found in the clade comprised of *Tozzia alpina* and the genus *Hedbergia*, the shift was found in 75% of the trees (r_{mean} = 0.06; r_{mcc} = 0.05). Finally, a shift showing an uptick in net diversification rate was present for the *Neobartsia* clade, with a frequency of 0.40 (r_{mean} = 0.40; r_{mcc} = 0.79).

In comparison, the results obtained with SymmeTREE evidenced fewer diversification shifts on the whole tree ($p < 0.05$). Like the MEDUSA results, an increase in diversification rate was also leading to the clade sister to *Bartsia alpina* and was consistently found in every tree we analyzed (table 1.5). A shift showing a slowdown in the *Tozzia+Hedbergia* clade was found to be marginally significant at $p < 0.05$ ($p = 0.067$) in every tree. If we were to choose a less stringent significance threshold (e.g., $p < 0.10$), this shift would be significant in 506 trees (93% of the distribution). The same is true for the shift involving the *Neobartsia* clade, where it was only found to be significant in 77 trees at $p < 0.05$ (14%), but increased to 258 trees (48%) when the less stringent p value was chosen.

Table 1.5
Results from our diversification rate analyses using MEDUSA and SymmeTREE over a posterior distribution of trees. The shifts were found at the nodes subtending the taxa specified in the first column, followed by the frequency of that shift in the posterior distribution of trees, the net diversification rate (*r*) for the maximum clade credibility tree (mcc), and the mean, median, minimum (min), maximum (max), and standard deviation (sd) summarized across 1,000 trees from the posterior distribution. In the results for SymmeTREE, two different significance values (α) were examined, $\alpha < 0.05$, and $\alpha < 0.10$.

Node	Freq. shift	*r* mcc	r mean	r median	r min.	r max.	r sd
MEDUSA							
Bartsia alpina	1.17	0	-0.04	-0.15	-0.32	0.33	0.18
Tozzia-Hedbergia	0.75	0.05	-0.06	0.00	-0.36	0.73	0.14
Neobartsia clade	0.40	0.79	0.40	0.38	0.15	1.03	0.13
Clade sister to *Bartsia alpina*	0.32	n/a	0.09	0.12	-0.18	0.26	0.07
Rhinanthus -Lathraea	0.12	n/a	0.17	0.16	0.12	0.28	0.03
Core *Rhinantheae*	0.07	n/a	0.11	0.12	-0.11	0.47	0.07

SymmeTREE	**Freq. shift at p < 0.05**	**Freq. shift p < 0.10**
Bartsia alpina	1.00	1.00
Tozzia-Hedbergia	0.00	0.93
Neobartsia clade	0.14	0.48

Discussion

Systematic implications

Molau (1990) published a comprehensive monograph on the genus "*Bartsia*," where he hypothesized that the species formed a monophyletic group that was sister to the African monotypic genus *Hedbergia.* Our phylogenetic results (Fig. 1.2 and Fig. 1.3), which are in agreement with those of Těšitel et al. (2010) and Scheunert et al. (2012), show clearly that *Bartsia* sensu Molau is polyphyletic, and that the new classification (sensu Scheunert et al. 2012) better reflects the disparate geographic distributions of these lineages, as well as their evolutionary histories. Previous studies have recovered the basal relationships within

Bellardia as a polytomy (Těšitel et al. 2010; Scheunert et al. 2012), where the position of *B. trixago, B. viscosa*, and *B. latifolia* is uncertain. Here, we recovered *Bellardia trixago* as the earliest divergent lineage but because we did not sample *B. latifolia*, we cannot be certain of the position of the other two taxa. The South American *Neobartsia* clade was highly supported in every analysis, and this is the first study to sample a geographically and morphologically representative diversity of the richness in this clade. These results provide strong evidence of the evolutionary distinctiveness of the *Neobartsia* clade with respect to the Mediterranean members of the expanded genus *Bellardia* (sensu Scheunert et al. 2012) – i.e., its unique geographic distribution and biogeographic history, the long divergence times from their Mediterranean relatives (~4.2 Ma), and the elevated diversification rates. Along with diagnostic morphological characters, we feel this justifies a reanalysis of the generic revision of Scheunert et al. (2012) with respect to the taxonomy in this clade, and this is the subject of ongoing taxonomic work in this clade.

Figure 1.2
Majority rule consensus tree (excluding burn-in trees) with mean branch lengths from the partitioned Bayesian analysis of the combined dataset. Branch lengths are proportional to the number of substitutions per site as measured by the scale bar. Values above the branches represent Bayesian posterior probabilities (PP) and maximum likelihood bootstrap support (BS). Major clades are summarized following species names with the current species diversity in parenthesis.

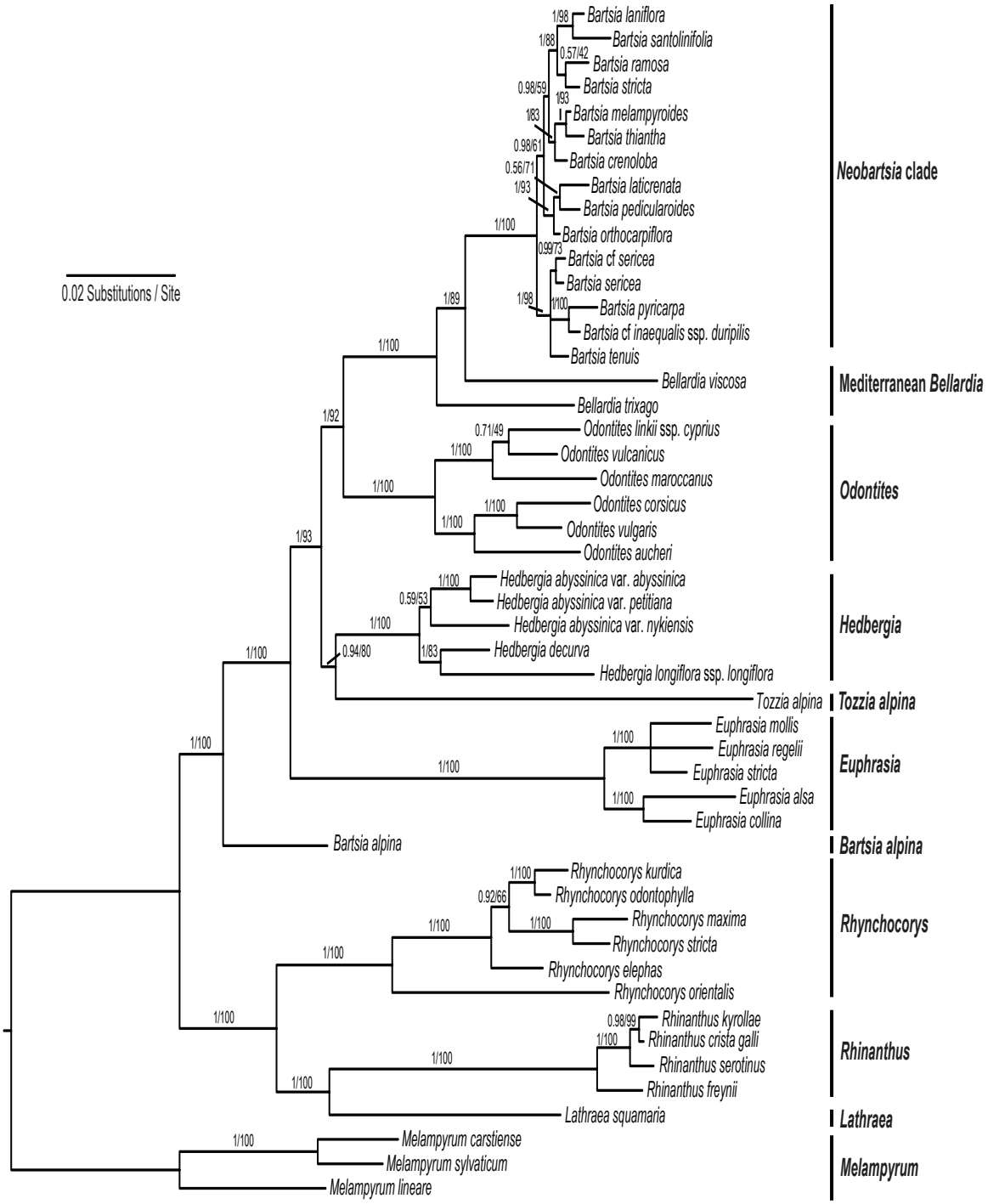
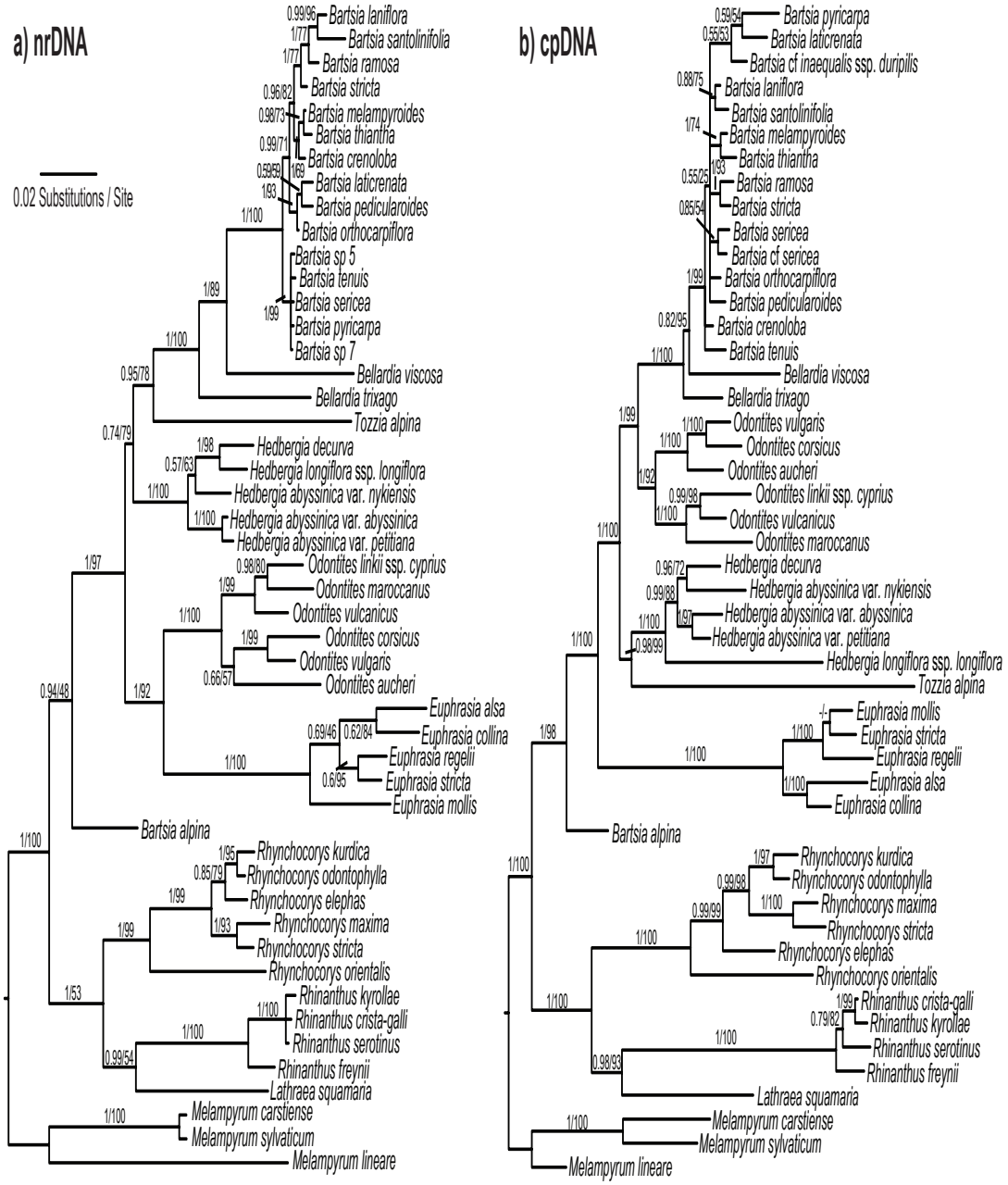


Figure 1.3
Majority rule consensus tree (excluding burn-in trees) with mean branch lengths from the partitioned Bayesian analysis of the a) nuclear ribosomal (nr) DNA and b) the chloroplast (cp) DNA datasets. Branch lengths are proportional to the number of substitutions per site as measured by the scale bar. Values above the branches represent Bayesian posterior probabilities (PP) and maximum likelihood bootstrap support (BS).



Our cpDNA and nrDNA analyses placed the genus *Tozzia* in different positions in the tree, although these differences were not statistically significant (Table 1.2). Our combined analysis placed the genus as sister to *Hedbergia*, albeit with marginal support (PP 0.94, BS 80). While this relationship is in agreement with previous studies (Těšitel et al. 2010; Scheunert et al. 2012), further work will be necessary to confidently place this genus in the Rhinantheae clade. Lastly, the African genus *Hedbergia* showed interesting and likely problematic species delimitations. The taxon *H. abyssinica* var. *nykiensis* was sister to *H. decurva* in our cpDNA analyses, and its relationship to the other *H. abyssinica* varieties in the nrDNA dataset was weakly supported. This is the first time that varietal taxa for this group have been included in a molecular study, and highlights the necessity for a more detailed study on the clade.

Biogeography and diversification rates

Our divergence time and biogeographic results depicted in Figure 1.1 illustrate interesting evolutionary hypotheses regarding the current distribution of the Rhinantheae clade. As a reminder, our analyses were calibrated using dates obtained with the molecular rate of the ITS (Wolfe et al. 2005), and thus, they should be taken as estimates where some uncertainty is expected. Nevertheless, they provide an evolutionary foundation and valid and useful information that help explain the current distribution and the diversity of the South American *Neobartsia* clade. With no doubt, Europe played a major role, almost at every node, in the reconstruction of ancestral ranges in the Rhinantheae clade of Orobanchaceae. Although this is the first formal biogeographic analysis in the clade, these results are in line with the verbal biogeographic scenarios described in previous studies (Wolfe et al. 2005;

Těšitel et al. 2010), but with slight differences in the description of the ancestral areas. Diversification of the majority of the genera was achieved in the European continent with subsequent migration events to Eurasia, northeastern North America, the Mediterranean region, Africa and South America. *Bartsia alpina* is a good example of a taxon with a purely European ancestral distribution but that is currently distributed in other parts of the world. This suggests that the current distribution was the result of a second and more recent migration into Greenland and northeastern North America sometime along its very long branch (Fig. 1.1). *Odontites* is another good example of a European radiation that has expanded its range to include Eurasia after the initial divergence. Moreover, the genus *Euphrasia*, which accounts for more than half of the members of the clade with ~400 species, was reconstructed as having a Europe/Eurasian ancestral range. The few species sampled in this study all have Eurasian distributions but the genus is currently considered to have a "bipolar" distribution (Gussarova et al. 2008), with species distributed in north temperate regions and extreme Austral areas. This pattern is extremely interesting since it suggests that extinction and/or long distance dispersal have played a large role in shaping the current distribution of this large clade..

The Mediterranean region was not included as a distinct area in our reconstruction analyses, and therefore, some of the genera with current Mediterranean distributions were treated as European, e.g., *Rhynchocorys, Odontites*, *Bellardia trixago*, and *B. viscosa*. The Mediterranean climate, as recognized today, is a young environment formed only 2.3–3.2 Ma and it is the result of two main events: i) the establishment of the Mediterranean rhythm of dry summers and mild-cold winters ~3.2 Ma, and ii) the oldest xeric period know for the region ~2.2 Ma (Zagwin 1960; 1974; Suc 1984). The crown clades for each of these genera

was reconstructed to have a European ancestral distribution, which implies that their current ranges are the results of independent evolutions into the European Mediterranean climatic region not earlier than ~3.2 Ma.

This study is mainly focused on studying the disproportionate diversity of the *Neobartsia* clade in the Andes, and to propose plausible hypotheses for its distribution. The Andes are thought to have begun uplifting in the late Miocene (~10 Ma) but only reaching the necessary elevation to host alpine conditions in the late Pliocene or early Pleistocene 2–4 Ma (Simpson 1975; Burnham and Graham 1999; Gregory-Wodzicki 2000; Antonelli et al. 2009). In our biogeographic analyses, South America is reconstructed for the first time at the crown node of *Bellardia*, which has a mean age of 8.44 Ma (5.01–12.72 Ma), making it too old for the páramos to be available for colonization. The clade containing *Bellardia viscosa* and the *Neobartsia* clade, on the other hand, was inferred to have a mean age of 4.8 Ma (4.12–11.19 Ma), which correlates better with the earliest estimates of the páramo.

South America is first included in the ancestral range in our analyses between 12.72 and 4.12 Ma, which defines a nine million year window for the ancestor to have reached South America. There are two main land routes that were present during this time period, the North Atlantic Land Bridge (NALB) uniting northeastern North America and western Europe, and the Bering Land Bridge between eastern Asia and western North America. Previous studies in the plant family Malpighiaceae (Davis et al. 2002; 2004), have suggested a migration route from South America to Africa starting in the early Oligocene (~30 Ma) via North America, the NALB, and Europe. The NALB was available from the early Eocene (~50 Ma) until the middle to late Miocene (~10–8 Ma) (Tiffney 1985; Tiffney and Manchester 2001; Denk et al. 2010; 2011), dates which overlap with our divergence time

estimates (Table 1.3) and with the appearance of South America as an ancestral range in our biogeographic analyses. This allows for the possibility of an early dispersal from Europe into North America over this land bridge. Colonization of North America would have followed a stepwise migration to South America over the forming Isthmus of Panama and/or island chains sometime in the last 4.5 Ma (Coates et al. 2004; Kirby and MacFadden 2005; Retallack and Kirby 2007). An alternative stepwise migration scenario for the South American clade's colonization of the Andes involves a migration route through Beringia. This land bridge, which was available on-and-off from ~58–3.5 Ma (Hopkins 1967; Tiffney and Manchester 2001; Tiffney 2008), has been proposed as a route for several groups found both in eastern Asia, western north America**,** and the Andes—e.g., Valerianaceae (Moore and Donoghue 2007)**.** This migration scenario is also plausible since Eurasia, Europe, and South America were reconstructed as the second most supported ancestral range at the node of divergence of the South American clade ($w_i = 0.35$).

Both of these stepwise migration scenarios rely completely on North America as an intermediary step where the South American ancestor possibly diversified, migrated, and finally went extinct. Unfortunately, there is no fossil record in the Rhinantheae clade (or in Orobanchaceae), and thus, no physical evidence is available to support either of these hypotheses. An interesting, albeit currently inconclusive, part of the biogeographic history of the *Neobartsia* clade, concerns a small, isolated, and disjunct population of *Bartsia glandulifera* in the southern mountains of Oaxaca (Mexico)*,* a species otherwise restricted to northern Colombia. This is the only known species in the *Neobartsia* clade that is distributed outside of the Andes. This taxon was unfortunately not sampled here but will be crucial to elucidate more in depth biogeographic patterns in the clade.

Molau (1990) hypothesized that the *Neobartsia* clade had colonized the Andes via a long-distance dispersal from Africa, sometime in the early Pliocene (~5 Ma). This hypothesis seemed plausible at the time when no phylogenetic evidence was available for the clade, but now that it is clear that the former genus *Bartsia* is polyphyletic and the two African species (*Hebergia decurva, H. longifolia*) are not sister to the South American species, there is no longer support for this hypothesis. Nevertheless, there is a third hypothesis that does rely on long-distance dispersal, but rather from Mediterranean Europe/north Africa to Andean South America (or, alternatively, from somewhere in North America following a land bridge migration from the Old World). Many plants are dispersed over long distances by water (e.g., *Cocos* L.), birds (e.g., *Pisonia* L.), or wind (e.g., *Taraxacum* F.H. Wigg.), and physiological and morphological adaptations to float, adhere, or fly are common (reviewed in Howe and Smallwood 1982). The seeds of *Bellarida-Neobartsia* are enclosed in a dry dehiscent capsule that contains between 20–200 small seeds (0.3–2 mm) per fruit, each equipped with 6–13 short wings or ridges (Molau 1990). Although these seeds are light and have wings making them at first glance suitable for long distance traveling, it has been estimated that their mean dispersal distance is 0.3 meters, at least in *Bartsia alpina* (Molau 1990). Nevertheless, there is a known constant storm track from western Africa (including the northwestern African Mediterranean climatic region) that crosses the Atlantic Ocean into the Caribbean and the Americas, and recent evidence has shown that there are major influxes of African dust in southern North America (Bozlaker et al. 2013), northeastern South America (Prospero et al. 2014), and the Caribbean basin (Prospero and Mayol-Bracero 2013). This opens the possibility for seeds of a Mediterranean ancestor—even if just one seed in a period of nine million years—to have been picked up and carried over to the New World.

At this point we cannot accept or reject any of the biogeographic hypotheses described above—the two stepwise migrations through North America or the long-distance dispersal from the Mediterranean climatic region—and it highlights the difficulty of inferring ancestral colonization routes even when using modern ancestral range reconstruction methods (see Tripp and McDade 2014), especially with a non-existent fossil record. Nevertheless it is interesting to study and discuss some of the caveats that these hypotheses have. The route over the NALB requires that the migration from the Old World occurred sometime between 12–8 Ma, which based on palynological evidence (Denk et al. 2010; 2011) is the latest time that this land bridge was available. This timeframe overlaps with the oldest estimates of our dating analyses (5.01–12.71 Ma), but leaves a narrower window of time for a stepwise migration to have occurred. Furthermore, the warmer temperatures in eastern North America during the late Miocene would possibly have affected the migration of the ancestor through the NALB. Conversely, the Bering Land Bridge was available until ~3.5 Ma, which overlaps completely with both the divergence of the South American clade from *Bellardia viscosa* (4.12–11.19 Ma), as well as with the split between *Bellardia trixago* and the other members of the *Bellardia-Neobartsia* clade 5.01–12.71 Ma ago. Moreover, this more recent route allows for the world to cool down during the Pliocene (Tiffney and Manchester 2001), which would facilitate the migration. Importantly, with either route, a stepwise migration hypothesis implies that the ancestral lineage (and any of its descendants) would have then gone extinct thereafter in North America and in eastern Asia (if the Bering Land Bridge route is considered). Both of these routes have been used to describe possible migrations and current distributions of large groups, e.g., Malpighiales (Davis et al. 2002) and Valerianaceae (Moore and Donoghue 2007). However, in cases like Valerianaceae and

Malpighiales there is either current North American diversity (Valerianaceae) or a fossil record in North America (Malpighiales) that corroborates the biogeographic scenario; we have neither with respect to the *Neobartsia* clade. Finally, the long-distance dispersal hypothesis requires that seeds (or at least one seed) be carried over the Atlantic Ocean into the New World, a distance of ~7,000 km (~4,000 mi). However, and as mentioned earlier, the species in the *Neobartsia* clade have very small seeds (0.3–2.0 mm) with wings or ridges that adhere to surfaces when wet. Given that there is a constant occurrence of storms that are generated in western and northern Africa that cross the Atlantic Ocean and deposit large amounts of dust in the Americas, one could envision one of these seeds easily being transported to the New World. Although at first this may seem unlikely, it is important to point out that a single seed may be sufficient for the colonization of a new habitat, and that the nine million year time window coupled with the large amounts (~200) of seeds that are produced in each capsule, increase the probability for this event to have happened.

To investigate if these biogeographic movements have affected the rate at which clades are diversifying (i.e., "dispersification"), we need to assess if the shifts found in our analyses correlate with a movement into a new area or if there is something else, e.g., a morphological change, that has triggered them. Regardless of the reason, investigating shifts of diversification and the location of these on a phylogenetic tree is extremely helpful when trying to understand disparities in species richnesses across related clades. The comparison of two methods that are based on different tenets, a stepwise model testing approach vs. a topological imbalance approach (MEDUSA and SymmeTREE, respectively), allowed us to i) better evaluate the performance of different approaches used to identify shifts in diversification, while ii) making results shared by both methods robust and reliable. This

comparison also showed the advantages of using a stepwise model testing approach and a method that incorporates extinction. Our MEDUSA analysis found six shifts across the posterior, and three when using the MCC tree; two of these six identified shifts represent a slowdown in net diversification. One of these slowdowns, which is the only shift consistently found by SymmeTREE at $p < 0.01$, across the posterior, and in the MCC tree, corresponds to the node where *Bartsia alpina* diverges from the rest of the core Rhinantheae 22.03 Ma. The extremely low diversification rate and its very long branch indicate that this species is likely the only extant member of a lineage that has had historically very low speciation rates or high extinction rates, or both. The first significant increase in net diversification rates was found at the node where the genus *Euphrasia* diverges from other genera 20.72 Ma. This genus includes ~400 species that encompass more than 80% of the species richness of Rhinantheae, estimated to be ~528 spp. (Mabberley 2008). Based on our limited sampling of this group, we cannot identify an apparent change in morphology or geography in the genus, and thus, no evident cause for this shift can be assessed with these data. Nevertheless, given the age and very high diversity of the clade, this shift is not surprising. However, it is important to point out that because we collapsed clades at the generic level to incorporate unsampled diversities, the present shift might not be the only one in *Euphrasia* and that clades within the genus may also have shifts of their own, where there might be an apparent change in either morphology or geography.

We also identified an increased rate of net diversification in the South American *Neobartsia* clade. We hypothesize that the clade underwent a similar pattern as seen in other Andean radiations, e.g. the family Valerianaceae and the genus *Lupinus* (Bell and Donoghue 2005; Hughes and Eastwood 2006, respectively), where their North American ancestor was

"pre-adapted" to cold environments making the colonization of the high Andes, and further radiation, easier (Donoghue 2008). The *Neobartsia* clade has a mean divergence time of 3.13 Ma and a mean diversification rate of 0.40, however, when the analysis is performed on the MCC tree, the net diversification almost doubles (r_{mcc} = 0.79). The large difference in values implies that although the shift was only identified in 40% of the posterior distribution of trees, when detected, the rate can be nearly four times higher than the background rate of the tree (background r_{mean} = 0.22). Based on the very short branches within the clade, its young age, and the genetic similarity between the species included in this study, this shift likely resulted in a rapid radiation event where the movement to and colonization of the high Andes acted as a trigger to increased diversification. As the Andes were uplifting, the creation of new vacant niches and the simulation of alpine conditions promoted the radiation into the diversity that we see today. Accordingly, this is a another example of how phylogenetic niche conservatism and the movement into a new geographic area, can lead to a high number of species in a relatively short period of time without the appearance of morphological key innovation, which is what Moore and Donoghue (2007) referred to as "dispersification".

Conclusion

This study places the *Neobartsia* clade in the context of a robust and well-supported phylogeny within the Rhinantheae clade of Orobanchaceae. This is the first study to study this clade in an explicitly temporal framework, with detailed divergence time estimates for the clade. Here, we focused primarily on the colonization and diversification of Andean South America ~3.13 Ma. This date correlates well with the necessary age for the Andes to have acquired the adequate elevation to simulate alpine conditions for the establishment of

this temperate, largely alpine clade in South America. Given that the South American clade is sister to a Mediterranean taxon, we hypothesized three biogeographic scenarios for the colonization of the Andes. The first route involves the NALB and North America as a stepwise migration route from Europe ~12–8 Ma, whereas the second hypothesis involves a westerly route from Europe through Asia, the Bering Land Bridge, and North America ~12–4 Ma. Both of these scenarios share a second migration from North America to South America over the forming Isthmus of Panama and/or island chains in the mid to late Pliocene ~ 4.5–3.13 Ma, which gave rise to the *Neobartsia* clade, and high levels of extinction throughout Asia and/or North American. Finally, the third hypothesis involves a long-distance dispersal from the Mediterranean climatic region (Europe and northern Africa) to South America. At this point however, we cannot accept or reject any of the previously described hypotheses. Regardless of the biogeographic route taken, once the South American ancestor reached the Andes, it was able to diversify rapidly in the vacant niches in the páramos. The greater diversification rates in the *Neobartsia* clade help explain the species richness found in the Andes today and support the idea that, movement into a new geographical area may trigger high diversification without the necessity of the evolution of morphological key innovations. This is especially true when the colonizing ancestral lineage is “pre-adapted” to the new conditions it encounters.

Literature Cited

Agrawal A.A., Fishbein M., Halitschke R., Hastings A.P., Rabosky D.L., Rasmann S. 2009. Evidence for adaptive radiation from a phylogenetic study of plant defenses. Proceedings of the National Academy of Sciences. 106:18067–18072.

Alfaro M.E., Santini F., Brock C., Alamillo H., Dornburg A., Rabosky D., Carnevale G., Harmon L. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proceedings of the National Academy of Sciences. 106:13410.

Antonelli A., Nylander J.A.A., Persson C., Sanmartín I. 2009. Tracing the impact of the Andean uplift on Neotropical plant evolution. Proceedings of the National Academy of Sciences. 106:9749–9754.

Bacon C.D., Baker W.J., Simmons M.P. 2012. Miocene Dispersal Drives Island Radiations in the Palm Tribe Trachycarpeae (Arecaceae). Systematic Biology. 61:426–442.

Baldwin B.G. 1992. Phylogenetic Utility of the Internal Transcribed Spacers of Nuclear Ribosomal DNA in Plants: An Example from the Compositae. Molecular Phylogenetics and Evolution. 1:3–16.

Baldwin B.G., Markos S. 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.

Beardsley P.M., Olmstead R.G. 2002. Redefining Phrymaceae: the placement of Mimulus, tribe Mimuleae, and Phryma. American Journal of Botany. 89:1093–1102.

Beaulieu J.M., Tank D.C., Donoghue M.J. 2013. A Southern Hemisphere origin for campanulid angiosperms, with traces of the break-up of Gondwana. BMC Evol Biol. 13:80.

Bell C., Donoghue M. 2005. Dating the Dipsacales: comparing models, genes, and evolutionary implications. American Journal of Botany. 92:284.

Bennett J.R., Mathews S. 2006. Phylogeny of the parasitic plant family Orobanchaceae inferred from phytochrome A. American Journal of Botany. 93:1039–1051.

Bolliger M. 1996. Monographie der Gattung Odontites (Scrophulariaceae) sowie der verwandten Gattungen Macrosyringion, Odontitella, Bornmuellerantha und Bartsiella. Willdenowia. 26:37–168.

Bozlaker A., Prospero J.M., Fraser M.P., Chellam S. 2013. Quantifying the contribution of long-range Saharan dust transport on particulate matter concentrations in Houston, Texas, using detailed elemental analysis. Environ. Sci. Technol. 47:10179–10187.

Burnham K.P., Anderson D.R. 2002. Model selection and multi-model inference: a practical information-theoretic approach. Springer New York.

Burnham R.J., Graham A. 1999. The history of neotropical vegetation: new developments and status. Annals of the Missouri Botanical Garden.:546–589.

Chan K., Moore B. 2005. SymmeTREE: whole-tree analysis of differential diversification rates. Bioinformatics. 21:1709.

Coates A., Collins L., Aubry M., Berggren W. 2004. The geology of the Darien, Panama, and the late Miocene-Pliocene collision of the Panama arc with northwestern South America. Geological Society of America Bulletin. 116:1327.

Davis C.C., Bell C.D., Mathews S., Donoghue M.J. 2002. Laurasian migration explains Gondwanan disjunctions: Evidence from Malpighiaceae. Proc. Natl. Acad. Sci. U.S.A. 99:6833–6837.

Davis C.C., Fritsch P.W., Bell C.D., Mathews S. 2004. High-latitude Tertiary migrations of an exclusively tropical clade: evidence from Malpighiaceae. International Journal of Plant Sciences. 165:S107–S121.

Denk T., Grimsson F., Zetter R. 2010. Episodic migration of oaks to Iceland: Evidence for a North Atlantic "land bridge" in the latest Miocene. American Journal of Botany. 97:276–287.

Denk T., Grímsson F., Zetter R., Símonarson L.A. 2011. The Biogeographic History of Iceland – The North Atlantic Land Bridge Revisited. Topics in Geobiology. Dordrecht: Springer Netherlands. p. 647–668.

Donoghue M., Smith S. 2004. Patterns in the assembly of temperate forests around the Northern Hemisphere. Philosophical Transactions B. 359:1633.

Donoghue M.J. 2008. A phylogenetic perspective on the distribution of plant diversity. Proc. Natl. Acad. Sci. U.S.A. 105:11549–11555.

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

Drummond A., Ho S., Phillips M., rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology. 4:699.

Drummond A., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 7:214.

Drummond C.S., Eastwood R.J., Miotto S.T.S., Hughes C.E. 2012. Multiple Continental Radiations and Correlates of Diversification in Lupinus (Leguminosae): Testing for Key Innovation with Incomplete Taxon Sampling. Systematic Biology. 61:443–460.

Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 32:1792–1797.

Gregory-Wodzicki K. 2000. Uplift history of the Central and Northern Andes: a review. Geological Society of America Bulletin. 112:1091.

Guindon S., Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology. 52:696–704.

Gussarova G., Popp M., Vitek E., Brochmann C. 2008. Molecular phylogeny and biogeography of the bipolar Euphrasia (Orobanchaceae): recent …. Molecular Phylogenetics and Evolution.

Ho S., Phillips M. 2009. Accounting for Calibration Uncertainty in Phylogenetic Estimation of Evolutionary Divergence Times. Systematic Biology. 58:367–380.

Hodges S.A. 1997. Floral Nectar Spurs and Diversification. International Journal of Plant Sciences. 158:S81–S88.

Hopkins D.M. 1967. The Bering land bridge. Stanford University Press, Stanfor, CA.

Howe H.F., Smallwood J. 1982. Ecology of Seed Dispersal. Annu. Rev. Ecol. Syst. 13:201–228.

Hughes C., Eastwood R. 2006. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. Proceedings of the National Academy of Sciences. 103:10334.

Kirby M.X., MacFadden B. 2005. Was southern Central America an archipelago or a peninsula in the middle miocene? A test using land-mammal body size. Palaeogeography, Palaeoclimatology, Palaeoecology. 228:193–202.

Mabberley D.J. 2008. Mabberley's plant-book: a portable dictionary of plants, their classification and uses. Cambridge, UK. University Press.

Magallón S., Sanderson M.J. 2001. Absolute diversification rates in angiosperm clades. Evolution. 55:1762–1780.

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

Mittelbach G.G., Schemske D.W., Cornell H.V., Allen A.P., Brown J.M., Bush M.B., Harrison S.P., Hurlbert A.H., Knowlton N., Lessios H.A., McCain C.M., McCune A.R., McDade L.A., McPeek M.A., Near T.J., Price T.D., Ricklefs R.E., Roy K., Sax D.F., Schluter D., Sobel J.M., Turelli M. 2007. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. Ecology Letters. 10:315–331.

Molau U. 1990. The genus Bartsia (Scrophulariaceae - Rhinanthoideae). Opera Botanica. 102:1–100.

Moore B.R., Donoghue M.J. 2007. Correlates of diversification in the plant clade Dipsacales:

geographic movement and evolutionary innovations. Am Nat. 170 Suppl 2:S28–55.

Nee S., May R.M., Harvey P.H. 1994. The reconstructed evolutionary process. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 344:305–311.

Olmstead R.G., Depamphilis C.W., Wolfe A.D., Young N.D., Elisons W.J., Reeves P.A. 2001. Disintegration of the Scrophulariaceae. American Journal of Botany. 88:348–361.

Ostrom J.H. 1979. Bird flight: how did it begin? Am. Sci. 67:46–56.

Oxelman B., Lidén M., Berglund D. 1997. Chloroplast rps16 intron phylogeny of the tribe Sileneae (Caryophyllaceae). Plant Systematics and Evolution. 206:393–410.

Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 20:289–290.

Pennell M.W., Eastman J.M., Slater G.J., Brown J.W., Uyeda J.C., FitzJohn R.G., Alfaro M.E., Harmon A.L.J. 2014. geiger v2. 0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics.

Posada D. 2008. jModelTest: phylogenetic model averaging. Mol Biol Evol. 25:1253–1256.

Prospero J.M., Collard F.X., Molinié J., Jeannot A. 2014. Characterizing the annual cycle of African dust transport to the Caribbean Basin and South America and its impact on the environment and air quality. Global Biogeochemical Cycles. 29.

Prospero J.M., Mayol-Bracero O.L. 2013. Understanding the Transport and Impact of African Dust on the Caribbean Basin. Bull. Amer. Meteor. Soc. 94:1329–1337.

R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rabosky D.L., Donnellan S.C., Talaba A.L., Lovette I.J. 2007. Exceptional among-lineage variation in diversification rates during the radiation of Australia's most diverse vertebrate clade. Proceedings of the Royal Society B: Biological Sciences. 274:2915–2923.

Rambaut A. 1996. Se-Al: Sequence alignment editor. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK. Available at http://tree.bio.ed.ac.uk/software/seal/.

Rambaut A. 2006. FigTree. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK. Available at http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut A., Drummond A.J. 2004. Tracer. University of Edinburgh, Edinburgh, UK. Available at http://tree.bio.ed.ac.uk/software/tracer/.

Ree R., Moore B., Webb C., Donoghue M. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution. 59:2299–2311.

Ree R.H., Smith S.A. 2008. Maximum Likelihood Inference of Geographic Range Evolution by Dispersal, Local Extinction, and Cladogenesis. Systematic Biology. 57:4–14.

Retallack G., Kirby M. 2007. Middle Miocene global change and paleogeography of Panama. Palaios. 22:667–679.

Rodríguez F., Oliver J.L., Marín A., Medina J.R. 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. 142:485–501.

Ronquist F., Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.

Sanderson M.J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19:101–109.

Scheunert A., Fleischmann A., Olano-Marin C., Bräuchler C., Heubl G. 2012. Phylogeny of tribe Rhinantheae (Orobanchaceae) with a focus on biogeography, cytology and re-examination of generic concepts. Taxon. 61:1269–1285.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Systematic Biology.

Shimodaira H., Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114.

Shimodaira H., Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics. 17:1246–12461247.

Simpson B. 1975. Pleistocene changes in the flora of the high tropical Andes. Paleobiology.:273–294.

Smith S. 2009. Taking into account phylogenetic and divergence-time uncertainty in a parametric biogeographical analysis of the Northern Hemisphere plant clade Caprifolieae. Journal of Biogeography. 36:2324–2337.

Smith S., Donoghue M. 2010. Combining Historical Biogeography with Niche Modeling in the Caprifolium Clade of Lonicera (Caprifoliaceae, Dipsacales). Systematic Biology.

Smith S., Dunn C. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. Bioinformatics. 24:715.

Smith S.A., Beaulieu J.M., Donoghue M.J. 2009. Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. BMC Evol Biol. 9:37.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22:2688–2690.

Stamatakis A., Hoover P., Rougemont J. 2008. A Rapid Bootstrap Algorithm for the RAxML Web Servers. Systematic Biology. 57:758–771.

Suc J.P. 1984. Origin and evolution of the Mediterranean vegetation and climate in Europe. Nature. 307:429–432.

Taberlet P., Gielly L., Pautou G., Bouvet J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology. 17:1105–1109.

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

Těšitel J., Říha P., Svobodová Š., Malinová T., Štech M. 2010. Phylogeny, Life History Evolution and Biogeography of the Rhinanthoid Orobanchaceae. Folia Geobotanica. 45:347–367.

Tiffney B., Manchester S. 2001. The use of geological and paleontological evidence in evaluating plant phylogeographic hypotheses in the Northern Hemisphere Tertiary. International Journal of Plant Sciences. 162:3–17.

Tiffney B.H. 1985. Perspectives on the Origin of the Floristic Similarity Between Eastern Asia and Eastern North-America. Journal of the Arnold Arboretum. 66:73–94.

Tiffney B.H. 2008. Phylogeography, Fossils, and Northern Hemisphere Biogeography: The Role of Physiological Uniformitarianism 1. Annals of the Missouri Botanical Garden. 95:135–143.

Tripp E.A., McDade L.A. 2014. A Rich Fossil Record Yields Calibrated Phylogeny for Acanthaceae (Lamiales) and Evidence for Marked Biases in Timing and Directionality of Intercontinental Disjunctions. Systematic Biology. In press. DOI:10.1093/sysbio/syu029

Wiens J.J., Donoghue M.J. 2004. Historical biogeography, ecology and species richness. Trends in Ecology & Evolution. 19:639–644.

Wolfe A., Randle C., Liu L., Steiner K. 2005. Phylogeny and biogeography of Orobanchaceae. Folia Geobotanica. 40:115–134.

Woodburne M.O., Rich T.H., Springer M.S. 2003. The evolution of tribospheny and the antiquity of mammalian clades. Molecular Phylogenetics and Evolution. 28:360–385.

Yang Z. 2006. Computational Molecular Evolution. oxford University Press.

Yule G.U. 1924. A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S. Philosophical Transactions of the Royal Society B: Biological Sciences. 213:21–87.

Zagwin W.H. 1960. Aspects of the Pliocene and early Pleistocene vegetation and climate in the Netherlands. Mededelingen Geologische Strichting. 3:1–78.

Zagwin W.H. 1974. The Pliocene-Pleistocene boundary in western and southern Europe. Boreas. 3:75–97.

Zanne A.E., Tank D.C., Cornwell W.K., Eastman J.M., Smith S.A., FitzJohn R.G., McGlinn D.J., O'Meara B.C., Moles A.T., Reich P.B., Royer D.L., Soltis D.E., Stevens P.F., Westoby M., Wright I.J., Aarssen L., Bertin R.I., Calaminus A., Govaerts R., Hemmings F., Leishman M.R., Oleksyn J., Soltis P.S., Swenson N.G., Warman L., Beaulieu J.M. 2014. Three keys to the radiation of angiosperms into freezing environments. Nature. 506:89–92.

Chapter 2: Phylogenetic revision of the genus *Bartsia* (Orobanchaceae), disjunct distributions correlate to independent lineages

Abstract

We propose a new classification for the South American species of the genus Bartsia L. and relatives recently included in an expanded treatment of the genus Bellardia (L.) All. This new classification reflects their evolutionary history and is based on morphological and molecular evidence, biogeographic hypotheses, and rates of diversification for these species. Additionally, we rearranged the current taxonomic classification of close relatives so that the current circumscriptions encompass only monophyletic groups. Some of these changes include the creation of a new genus, Neobartsia Uribe-Convers and Tank (47 spp.), as well as the reclassification of Bellardia latifolia (L.) Cuatrec. back to Parentucellia latifolia (L.) Caruel. These taxonomic changes are important for proper communication within the large Rhinantheae clade of Orobanchaceae and for the interpretation of biogeographic patterns and diversification processes of these species.

Introduction

A great number of molecular phylogenetic studies have made it evident that some well-established morphological classifications are the result of taxa that are not each other closets relatives (e.g., Olmstead et al. 2001; Angiosperm Phylogeny Group 2009). Systematists are now able to approach classifications from an evolutionary and phylogenetic perspective, allowing them to erect or update classifications so that they reflect the evolutionary history of the groups they work on, as well as generating hypotheses about

times of divergence among taxa (e.g., Swenson et al. 2012), historical biogeography (e.g., Olmstead et al. 2001; Angiosperm Phylogeny Group 2009; Beaulieu et al. 2013), and hybridization (e.g., Swenson et al. 2012; Clay et al. 2012). This is the case with the genus *Bartsia* L., a genus historically of ca. 49 species in the family Orobanchaceae that has recently been included in several family-wide phylogenetic studies (Wolfe et al. 2005; Bennett and Mathews 2006; Těšitel et al. 2010; Scheunert et al. 2012). *Bartsia* was the subject of a detailed taxonomic treatment in which the major sections and taxa were revised and a few new species described (Molau 1990). In this treatment, Molau (1990) divided *Bartsia* into seven sections based on morphological characters and geographic distributions. Section *Bartsia* included the type species *B. alpina* L. that is distributed in Scandinavia, the Alps, Greenland, and the Hudson Bay region of northeastern North America. Section *Longiflorae* contained two species (*B. decurva* Benth. and *B. longiflora* Benth.) restricted to the mountains of northeastern Africa, while section *Bellardia* comprised a single species (*B. trixago* L.) with a Mediterranean origin but with an introduced, and somewhat weedy, distribution in subtropical dry areas, e.g., coastal western North America, especially in the Californian Chaparral, Australia, and Chilean lowlands. Finally, four different sections (*Strictae*, *Orthocarpiflorae*, *Laxae*, and *Diffusae*) formed the largest group in the genus, with ca. 45 species distributed throughout the high elevation páramo ecosystem in the Andes.

An Orobanchaceae-wide study (Bennett and Mathews 2006) had hinted that *Bartsia*, as circumscribed at the time, may not be monophyletic but their sampling of only three species was too limited to confidently conclude anything. Těšitel et al. (2010) expanded the sampling of the genus including five species in a study of the Rhinantheae clade (sensu McNeal et al. 2013), but Scheunert et al. (2012) were the first to include species from all

morphological sections and from the complete geographic range of the genus. Their molecular results showed that the genus was indeed polyphyletic and that it was comprised of four distinct lineages, each corresponding to its geographic distribution, i.e., Europe, the Mediterranean, northeastern Africa, and Andean South America. Based on their phylogenetic results, various taxonomic rearrangements were made to better reflect the evolutionary history of the taxa. The European species *B. alpina* was conserved as the generic type, while the two African species were included in the African genus *Hedbergia* Molau, which was shown to be their closest relative. The Mediterranean species *B. trixago* was found to form a well-supported clade with the two South American species included in their study, as well as with the two species of the genus *Parentucellia* Viv. (*P. latifolia* (L.) Caruel and *P. viscosa* (L.) Caruel). However, the relationship between *P. viscosa, B. trixago,* and a well-supported clade containing *P. latifolia* as the sister-group to the two South American species was unresolved. Both *Parentucellia* species have a Mediterranean origin and the same current introduced distribution of *B. trixago. Bartsia trixago* has been classified in its own genus in various European floras as *Bellardia trixago* (L.) All. (e.g., Allioni 1785; Tutin et al. 1973), and thus, Scheunert et al. (2012) decided to expand the circumscription of *Bellardia* All. to include the two species of *Parentucellia* and the two South American *Bartsia* species they had sampled; they purposefully avoided reclassifying the remaining 43 South American species, due to their poor sampling of the Andean taxa. However, we believe that this taxonomic rearrangement makes little sense, as it leaves ca. 45 South American species classified in a large paraphyletic genus along with the European type species *B. alpina*.

Here, we update the taxonomic classification of these taxa making use of the results of a recent study on the Rhinantheae clade that focuses on the diversification of the South

American *Bartsia* clade in the context of its mostly European relatives (Chapter 1). In addition, we present a phylogenetic analysis of ~25,000 bp of chloroplast DNA sequence data (Chapter 4) that includes 43 of the 47 described species of South American *Bartsia* (sensu Molau 1990; Cabrera and Botta 1992; Sylvester 2014). These two scales of phylogeny estimation clarify the relationships between lineages of the Rhinantheae clade, and confirm the monophyly of the South American clade. Finally, based on biogeographic analyses, analyses of diversification rates across the clade, morphological differences, and phylogenetic results, we propose a new generic classification for the South American species of *Bartsia* and closely related taxa.

Materials and Methods

Sampling.

A total of 77 taxa were included in this study based on the sampling in two previous studies (Chapter 1; Chapter 4). Because we wanted to both place the South American diversity in the context of the Rhinantheae clade and confirm the monophyly of the South American taxa, this sampling included representatives from all of the major lineages that comprise the clade, as well as 43 of the 47 South American *Bartsia* species (sensu Molau 1990; Cabrera and Botta 1992; Sylvester 2014).

Molecular datasets

This study is based on two different datasets: i) a representative sampling of the Rhinantheae clade using Sanger sequences of the Internal and External Transcribed Spacer regions of the nuclear ribosomal repeat, ITS and ETS, respectively (Chapter 1), the Sanger

dataset henceforth, and ii) a comprehensive sampling of the South American species diversity and close Mediterranean relatives based on ~25,000 base pairs of chloroplast data from a single individual of each species generated via high-throughput sequencing (HTS) (Chapter 4), the HTS dataset henceforth. Because the Sanger dataset from Uribe-Convers and Tank (in review) did not include sequences of *P. latifolia*, we complemented it with data for these two regions obtained via HTS (Chapter 4). We used these two different datasets to answer phylogenetic questions in this group at different scales and with different foci. The Sanger dataset was compiled to place the South American diversity within the large Rhinantheae clade, whereas the HTS dataset was used to confirm the monophyly of the South American taxa, and provide an initial estimate of interspecific relationships in this clade.

Phylogenetic analyses

Both the sequences from the Sanger dataset, including the added sequences for *P. latifolia*, and the HTS dataset were aligned independently using MUSCLE v3.8.31 in its default settings (Edgar 2004), and the alignments were visually inspected in Geneious R7 (Biomatters, Auckland, New Zealand). Model selection of sequence evolution and partition schemes followed Uribe-Convers and Tank (in review) and Uribe-Convers et al. (in review) for each the Sanger and HTS datasets, respectively. Each dataset was analyzed in a maximum likelihood (ML) framework in the program RAxML v.8.0.3 (Stamatakis 2014). ML support was assessed with 1,000 replicates of nonparametric bootstrapping using the rapid bootstrap algorithm (Stamatakis et al. 2008). Additionally, we conducted Bayesian analyses on the Sanger dataset in the program MrBayes v.3.2.2 (Ronquist et al. 2012), and consisted of two

independent runs with four Markov chains using default priors and heating values and were performed with the individual parameters unlinked across the data partitions. Each independent run was started from a randomly generated tree, was sampled every 1,000 generations, and consisted of 15 million generations. Convergence of the chains was determined by analyzing the plots of all parameters and the –lnL using Tracer v.1.5 (Rambaut and Drummond 2004). Stationarity was assumed when all parameters values and the –lnL had stabilized; the likelihoods of independent runs were considered indistinguishable when the average standard deviation of split frequencies was < 0.001. Consensus trees were obtained for each dataset using the sumt command in MrBayes.

Results

Convergence of the Bayesian analysis of the Sanger dataset was assessed using Tracer v.1.5, and a consensus tree was generated after 25% of the trees had been discarded as burn-in. The topologies recovered from both the ML and Bayesian analyses of the Sanger dataset were identical with respect to the relationships among the major lineages of Rhinantheae. A summary of the current evolutionary hypothesis of the Rhinantheae clade based on the Sanger dataset is given in Figure 2.1. In short, our results are concordant with those of previous studies (Těšitel et al. 2010; Scheunert et al. 2012; McNeal et al. 2013), except for the relationship between *Odontites* and *Euphrasia*, which in these analyses were recovered as sister groups. Uribe-Convers and Tank (in review) obtained this same result when only using data from the nuclear ribosomal DNA (as was used here) but not when they used data from the chloroplast genome. They investigated the incongruences between these two topologies using the approximately unbiased test (AU test) (Shimodaira 2002), as implemented in the

program Consel (Shimodaira and Hasegawa 2001), and showed that these discordant relationships were not statistically significant. Since their combined analysis of chloroplast and nuclear ribosomal DNA is congruent with previous results (Těšitel et al. 2010; Scheunert et al. 2012; McNeal et al. 2013), and we are using the data from that study, we will adopt those relationships for the major lineages of the Rhinantheae clade with the addition of *Bellardia latifolia.* The major result is that the former genus *Bartsia* is recovered as a polyphyletic group. *Bartsia alpina* is sister to the rest of the core Rhinantheae (sensu Scheunert et al. 2012) with high posterior probability (PP) and bootstrap support (BS) (0.99 PP, 95 BS). The two African species (*B. decurva* and *B. longiflora*) form a clade with the African genus *Hedbergia* (1.0 PP, 100 BS), and this clade is sister to the European species *Tozzia alpina* L. (0.66 PP, 41 BS). The genus *Odontites* Ludw. forms a clade (1.0 PP, 99 BS) that is sister to a clade comprised by the current species of *Bellardia* (sensu Scheunert et al. 2012) and the South American *Bartsia* species. Within the latter clade, *Bellardia trixago* is sister to *Bellardia viscosa* (L.) Fisch. & C.A. Mey (0.99 PP, 84 BS), and this clade is sister to a clade comprised of *Bellardia latifolia* (L.) Cuatrec. and the South American species (1.0 PP, 100 BS). Finally, *Bellarida latifolia* is sister to the South American clade (1.0 PP, 100 BS), which is a highly supported monophyletic group (1.0 PP, 99 BS).

The HTS dataset represents the first comprehensive sampling for the group and includes 43 species of South American *Bartsia,* the three closely related Mediterranean species of *Bellardia,* and *Bartsia alpina* as the outgroup. The resulting ML phylogeny, which is congruent with the Sanger dataset with respect to relationships between *Bellardia* and the South American species, is shown in Figure 2.2. Briefly, *Bellardia trixago* and *Bellardia viscosa* are sister taxa (95 BS) and form a clade that is sister to a clade comprised by

Bellardia latifolia and the South American species (97 BS). Finally, the South American species form a clade that is highly supported (100 BS). It is noteworthy to mention that the position of *Bellardia latifolia* has not been stable in previous studies (e.g., Těšitel et al. 2010) depending on the data used, being either sister to the South American clade or to a clade formed by *Bellardia viscosa* and the new world taxa. Here, the position of this taxon is robust using both datasets.

Within the South American clade, the initial split is between a clade that is mostly comprised of species in section *Diffusae* (Fig. 2.2; clade A), and a clade mostly comprised of taxa from the other three sections, i.e., *Orthocarpiflorae*, *Strictae*, and *Laxae* (Fig. 2.2., clade B). Although not highly supported (clade A, 20 BS; clade B, 29 BS), this split is interesting as it can be associated with the considerably different corolla morphologies among the sections. Species in section *Diffusae* are all characterized by having a deflected lower corolla lip, as opposed to the straight corolla lip found in the other three sections. The shape of the lip has been hypothesized to be of evolutionary importance as it influences the type of pollinator that visits the flowers. A deflected lip would provide a landing site for insect pollinators, in particular bees, whereas a straight corolla lip would suggests pollination by hummingbirds (Molau 1990). Support for other groups within the South American clade is sparse, and can likely be attributed to the relatively slow mutation rate of the chloroplast genome, the young age of the group, and the rapid radiation of this clade (Chapter 4). Nevertheless, there are some interesting results worth mentioning. For example, one of the few well-supported clades (78 BS) within clade A (clade A'; Fig. 2.2) is comprised of species collected in dry environments in Peru, while clade B' (100 BS) is comprised of species that are distributed in the northern and wetter parts of the Andes in Colombia and

Ecuador. However, to fully elucidate the relationships of the South American species, it will be necessary to include multiple independent nuclear loci that will aid in teasing apart evolutionary processes such as coalescent stochasticity, hybridization, and introgression, which are likely playing a large role in the diversification of this group.

Figure 2.1
Summary of the phylogenetic hypotheses for the Rhinantheae clade of Orobanchaceae based on taxonomic (Molau 1990) and molecular phylogenetic data (Scheunert et al. 2012 and Chapter 1). The size of each triangle corresponds to the sampling proportion in the Sanger dataset. The grey horizontal boxes highlight the recent taxonomic rearrangements in the clade with Molau's (1990) classification in the left column, Scheunert et al. (2012) in the middle column, and this paper's in the right column. Values above the branches represent Bayesian posterior probabilities (PP) and maximum likelihood bootstrap support (BS). The asterisk above the branches represents places where our Sanger dataset is incongruent with previous studies. These incongruences have been shown to not be statistically significant using the approximately unbiased test (Shimodaira 2002), and do not affect the taxonomic conclusions made here. See text and Uribe-Convers and Tank (in review) for more details.

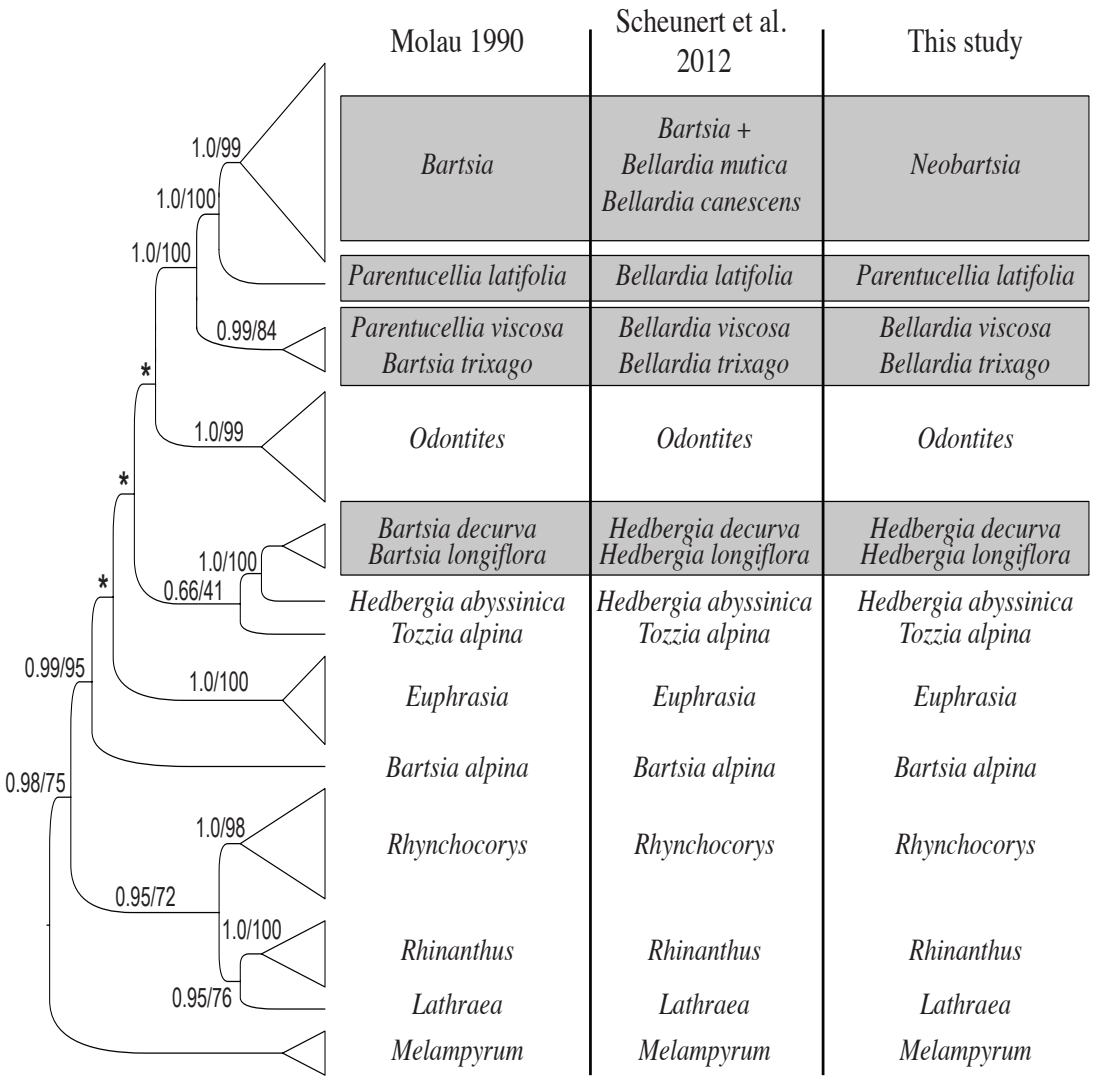
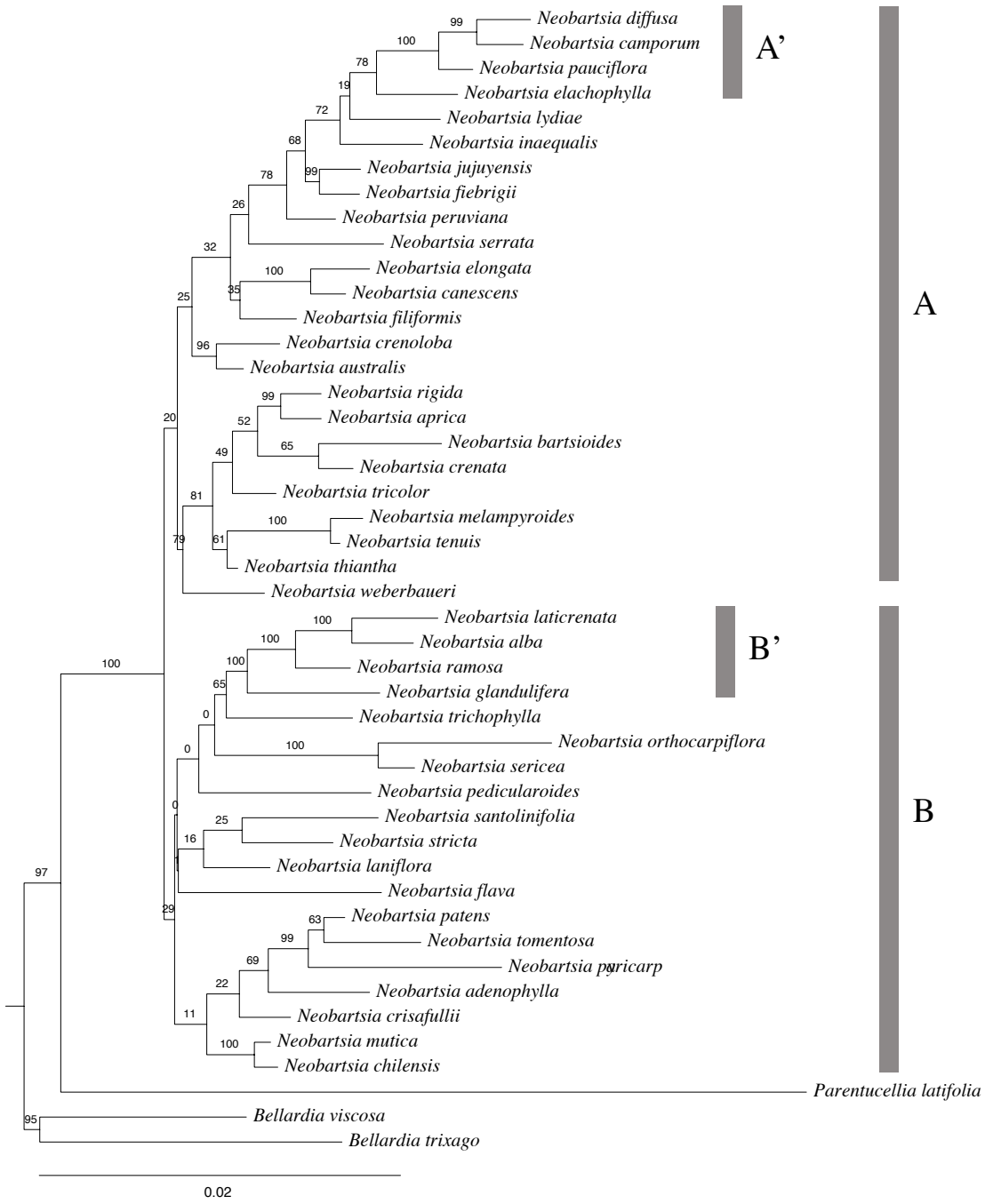


Figure 2.2
Current phylogenetic hypothesis based on a maximum likelihood analysis of the South American clade and its close Mediterranean relatives using the high-throughput sequence dataset. Clade A is comprised of species mostly in section *Diffusae*, while clade B is comprised mostly of species in the other three sections. Clade A' is comprised of species collected in dry environments in Peru, while clade B' is comprised of species that are distributed in the northern and wetter parts of the Andes in Colombia and Ecuador.
Branch lengths are proportional to the number of substitutions per site as measured by the scale bar. Values above the branches represent maximum likelihood bootstrap support (BS). The outgroup, *Bartsia alpina* L., was removed for better visualization.



Taxonomic Treatment

South American Species

One of the main goals in systematics is to create robust and well-supported classifications in order to better understand the world's diversity. These classifications should be stable and long lasting, and should reflect the evolutionary histories of the species they encompass. In the past, taxa were primarily classified by morphological similarities described by an authority in the group but without taking evolutionary processes such as convergent evolution or homoplasy into consideration, which sometimes resulted in groups of unrelated taxa (e.g., Olmstead et al. 2001).

The major phylogenetic relationships of the Rhinantheae clade shown here are consistent with previous Orobanchaceae-wide (Wolfe et al. 2005; Bennett and Mathews 2006; McNeal et al. 2013) and Rhinantheae clade studies (Těšitel et al. 2010; Scheunert et al. 2012). Although previously implied (Těšitel et al. 2010; Scheunert et al. 2012), this is the first time that the monophyly of the South American taxa has been recovered in a study that incorporates a nearly complete sampling of the species richness of the clade and not just a few species. The members of this clade are the only species with a South American origin in the Rhinantheae clade, as shown in a recent biogeographic study of the group (Chapter 1). In that study, the authors proposed two hypotheses for how the ancestor of the South American species possibly colonized the Andes from Mediterranean Europe/north Africa—a stepwise migration through North America or long-distance dispersal. The first hypothesis relies on a stepwise migration via either the North American Land Bridge (NALB), a connection that was available between northeastern North America and western Europe between ca. 60–12 million years ago (Ma) (Tiffney 1985), with recent palynological evidence extending the

connection to up to ca. 8 Ma (Denk et al. 2010; 2011), or from Eurasia to North America via the Bering Land Bridge which was available between the late Paleocene and the early to mid-Pliocene, ca. 58–3.5 Ma (Hopkins 1967; Tiffney and Manchester 2001; Tiffney 2008). Via either of these stepwise migration routes, after colonization of North America, the ancestral lineage would follow a migration into South America over the forming Isthmus of Panama and/or island chains in the mid to late Pliocene ca. 4.5 Ma (Coates et al. 2004; Kirby and MacFadden 2005; Retallack and Kirby 2007). Importantly, with either route, a stepwise migration hypothesis implies that the ancestral lineage (and any of its descendants) would have then gone extinct thereafter in North America (and in eastern Asia via the Bering Land Bridge route). Both of these routes have been used to describe possible migrations and current distributions of large groups, e.g., Malpighiales (Davis et al. 2002) and Valerianaceae (Moore and Donoghue 2007), and both hypothesized routes fit well with the divergence times reconstructed for the Rhinantheae clade by Uribe-Convers and Tank (chapter 1), the crown node for South America has a mean age of 3.13 Ma (4.11–1.53 Ma 95% highest posterior density [HPD]) and the divergence between the New World species and the Old World *Bellardia viscosa* was estimated to have a mean date of 4.8 Ma (11.19–4.12 Ma 95% HPD). However, in cases like Valerianaceae and Malpighiales there is either current North American diversity (Valerianaceae) or a fossil record in North America (Malpighiales) that corroborates the biogeographic scenario; we have neither with respect to the South American *Bartsia* clade. Lastly, the divergence of *Bellardia trixago* from *Bellardia viscosa* and the South American clade was estimated to have a mean age of 8.4 Ma (12.72–5.0 Ma 95% HPD). The reconstructed dates for these splits allow for a nine million year window for the South American ancestor to reach the then uplifting Andes, and fit well with the proposed

age for both land bridges. The second hypothesis for the movement of this lineage from Mediterranean Europe/north Africa to the Andes is via long-distance dispersal. Although the seeds of the South American *Bartsia* species are not particularly adapted for flying, they do have short wings and ridges that adhere to surfaces when wet. Furthermore, each capsule produces ~200 seeds that are each between 0.3-2.0 mm (Molau 1990). It has been shown recently, that there is a constant storm track from western Africa (including the northwestern African Mediterranean climatic region) that crosses the Atlantic Ocean into the Caribbean and the Americas, and large influxes of African dust have been found in southern North America, northeastern South America, and the Caribbean basin (Prospero and Mayol-Bracero 2013; Bozlaker et al. 2013; Prospero et al. 2014). This opens the possibility for seeds of a Mediterranean ancestor—even if just one seed in a period of nine million years—to have been picked up and carried over to the New World. However, Uribe-Convers and Tank (chapter 1) where not able to accept or reject any of these competing hypotheses.

Moore and Donoghue (2007) demonstrated that upticks in the rate of net diversification (speciation minus extinction) in the plant families Adoxaceae and Valerianaceae were associated not with the appearance of a new key morphological character but rather to the movement to a new geographic area, a process they referred to as "dispersification" (dispersion and diversification). Uribe-Convers and Tank (chapter 1) demonstrated that the same might be true for the Andean species of Rhinantheae, as they found elevated rates of net diversification in this clade when compared to the background rate of the tree associated with the biogeographic movement into the Andes. Other young groups of Andean plants have been shown to have elevated rates of diversification in the páramos, e.g., *Lupinus* L. (Hughes and Eastwood 2006), and processes like "dispersification"

may be associated with the great diversity that we see today in these rapid radiations. This could especially be true when the lineage filling the newly emerging environment already has the relevant morphological and/or physiological adaptations in place, which is concordant with the hypothesis that it is easier to move than it is to evolve (Donoghue 2008).

All of the evidence gathered for the Andean *Bartsia* clade, i.e., its unique geographic distribution and biogeographic history, the long divergence times from their Mediterranean relatives (~4.2 Ma), and the elevated diversification rates, point to the distinctiveness of this group of New World taxa. Not only are they distinct in terms of their evolutionary history, they are different in their reproductive and vegetative morphologies, and possibly their pollination syndromes. The Mediterranean *Bellardia* have a large deflected lower corolla lip thought to aid in pollination by bees (Molau 1990). While the Andean species in the section *Diffusae* share this character, which may be the ancestral character of the whole Rhinantheae clade, the ca. 26 species in the other three sections (i.e., *Strictae, Orthocarpiflorae,* and *Laxae*) have an erect lower lip that forms a tubular corolla, usually associated with hummingbird pollination (Molau 1990). Additionally, the Mediterranean *Bellardia* have coarsely dentate leaves whereas the South American *Bartsia* species have leaves with tightly crenate or serrate margins. Based on the large evidence of the distinctiveness of the Andean species, we propose that the South American species of *Bartsia* be classified in the new genus *Neobartsia* Uribe-Convers and Tank, and offer the following taxonomic rearrangements based on the species delimitations designated by Molau (1990).

Neobartsia Uribe-Convers & Tank, gen. nov.—TYPE: *Neobartsia santolinifolia* (Kunth) Uribe-Convers & Tank. Basionym: *Euphrasia santolinaefolia* Kunth in Humboldt, Bonpland & Kunth, Nov. gen. sp. 2: 333. 1818.

Annual or perennial hemiparasitic herbs or subshrubs, usually pubescent and often glandular; stems prostrate, scandent, or erect, branched mainly in the proximal parts. Leaves decussate, sessile, the blades herbaceous or subcoriaceous with dentate, serrate, crenate, or entire margins, deflexed or revolute. Inflorescences loose to rather dense, often spicate to subspicate. Bracts not or slightly modified, resembling foliage leaves and gradually smaller upwards. Calyx unequally or equally cleft, if the latter, the median clefts usually much deeper than the lateral ones, the dorsal cleft deepest, usually sinuate and widened at base. Calyx lobes straight to reflexed (section *Laxae*). Corolla pubescent, yellow, red, purple, or lavender with the lobes equal to unequal (section *Orthocarpiflorae*). The lower lip straight and flattened to deflect and with reduced gibbae (section *Diffusae*). Anthers included, glabrous to pilose, blunt or mucronate. Style included to exposed and in some cases exserted (section *Laxae*). Stigma bilobate, fusiform, or almost capitate. Capsule pilose or setose, rarely glabrous. Seeds 0.5–2.0 mm long, longitudinally winged with the wings finely cross-striate. Distribution: Andean South America.

Neobartsia new clade name. Definition (node-based, branch-modified): the least inclusive clade containing *Neobartsia santolinifolia* (Kunth) Uribe-Convers & Tank 2014 and *Neobartsia canescens* (Wedd.) Uribe-Convers and Tank 2014, but not *Bellardia trixago* (L.) All. 1785, *Parentucellia latifolia* (L.) Caruel 1885, *Bellardia viscosa* (L.) Fisch. & C.A. Mey

1836. There is no preexisting scientific name for this clade. To alleviate confusion with the new name, we have chosen to conserve part of the previous name of the group (i.e., *Bartsia*) but at the same time incorporating information about its New World distribution (i.e., *Neo*). We believe that the name *Neobartsia* will facilitate communication between botanists who work in the high Andes, where these taxa are very common. Our analyses show that every species found in South America is part of a well-supported clade using both the Sanger and the HTS datasets. The monophyly of these species has been suggested previously but this is the first time that it has been recovered with more than three species (e.g., Těšitel et al. 2010). The recognition of this clade is important for the interpretation of biogeographic patterns and diversification within *Neobartsia* and the Rhinantheae clade.

Neobartsia adenophylla (Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia adenophylla* Molau, Opera Bot. 102: 76. 1990. TYPE: PERU. Cajamarca: Prov. Hualgayoc, 8 km from Bambamarca on road to Hualgayoc, 2850 m, 26 Mar 1985, *Molau, Öhman & Sánchez Vega 7725* (Holotype: GB; Isotypes: CPUN, HUT).

Neobartsia alba (Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia alba* Molau, Opera Bot. 102: 76. 1990. TYPE: ECUADOR. Chimborazo: Cordillera Occidental, 34 km from Riobamba on road to El Triunfo, summit area, 3850 m, 18 Feb 1985, *Molau, Ohman, Arvidsson, Lindqvist & Lindstrom 1327* (Holotype: GB; Isotype: QCA).

Neobartsia aprica (Diels) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia aprica* Diels, Bot. Jahrb. Syst. 37: 430. 1906. TYPE: PERU. Junín: Prov. Tarma, above Tarma,

3300–3700 m, Feb 1903, *Weberbauer 2399* (Holotype: B, destroyed; no isotypes found). NEOTYPE: PERU. Junin: Prov. Tarma, 20 km W of (above) Tarma, 4000 m, 9 Apr 1952, *Hutchison 640* (Neotype: UC [designated by Molau in Opera Bot. 102: 76. 1990]; Isotype: US).

Euphrasia bicolor Ruiz & Pavón ex A. López, Anal. Inst. bot. Cavanilles 17: 454. 1959 (not validly published, nomenclatural type not designated).

Euphrasia dentata Ruiz & Pavón ex A. López, Anal. Inst. bot. Cavanilles 17: 456. 1959 (not validly published, nomenclatural type not designated).

Neobartsia australis (Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia australis* Molau, Opera Bot. 102: 76. 1990. TYPE: ARGENTINA. Jujuy: Dept. Yavi, Yavi, 3400 m, 1 Jan 1902, *Fries 1713a* (Holotype: S; Isotypes: CORD, LD).

Neobartsia bartsioides (Hook.) Uribe-Convers & Tank, comb. nov. Basionym: *Lamourouxia bartsioides* Hook., Bot. misc. 2: 234. 1831. *Bartsia bartsioides* Edwin, Field Mus. Nat. Hist., Bot. Ser. 13: 492. 1971. *Bartsia densiflora* Benth., nom. superfl., in Candolle, Prodr. 10: 548. 1846. TYPE: PERU. Lima: Prov. Canta, valley of Canta, *Cruckshanks s.n.* (Holotype: K).

Bartsia calycina Diels, Bot. Jahrb. Syst. 37: 432. 1906. TYPE: PERU. Ancash: Prov. Huaylas, below Hda. Cajabamba, W slopes of the Cordillera Negra, SW of Caraz, 3000–3500 m, May 1903, *Weberbauer 3166* (Holotype: B destroyed; Lectotype: G [designated by Molau in Opera Bot. 102: 76. 1990].

Euphrasia subulata Ruiz & Pavón ex A. López, An. Inst. bot. Cavanilles 17: 455. 1959 (not validly published, nomenclatural type not designated).

Neobartsia camporum (Diels) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia camporum* Diels, Bot. Jahrb. Syst. 37: 433. 1906. TYPE: PERU. Cuzco: Prov. Cuzco, ruins of Sacsayhuaman just above Cuzco, 3500–3600 m, May 1905, *Weberbauer 4864* (Holotype: B destroyed; Lectotype: G [designated by Molau in Opera Bot. 102: 76. 1990]).

Neobartsia canescens (Wedd.) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia canescens* Wedd., Chlor. Andina 2: 123. 1860. TYPE: PERU. Lima, Sine loco (probably Baños de Churín, prov. Cajatambo), *Dombey s.n.* (Isotype: PH, Lecototype: P, designated by Molau in Opera Bot. 102: 76. 1990). *Bellardia canescens* (Wedd.) A. Fleischm. & Heubl, Taxon 61(6): 1282. 2012. *Bartsia cinerea* Diels, Bot. Jahrb. Syst. 31: 432. 1906. TYPE: PERU. Ancash: Prov. Bolognesi, between Chiquián and Tallenga (N of Aquia), 3300–3600 m, Apr 1903, *Weberbauer 2858* (Holotype: B destroyed, not represented in photograph, Isotypes: not found).

Neobartsia chilensis (Benth.) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia chilensis* Benth. in Candolle, Prodr. 10: 548. 1846. TYPE: CHILE. Valparaiso: Dept. Quillota, La Palma, Quillota, Oct 1829, *Bertero 1072* (Lectotype: G [designated by Molau in Opera Bot. 102: 76. 1990]; Isolectotypes: F, G, GH, P).

Neobartsia crenata (Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia crenata* Molau, Opera Bot. 102: 76. 1990. TYPE: BOLIVIA. Cochabamba: Prov. Quillacollo, SE slopes of Cerro Tunari, between Quillacollo and Morochata. 40–45 km by road from Cochabamba, ca 3950 m, l0 Mar 1983, *Molau 690* (Holotype: S; Isotypes: GB, LPB).

Neobartsia crenoloba (Wedd.) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia crenoloba* Wedd., Chlor. And. 2: 124. 1860. TYPE: BOLIVIA. Potosí: Prov. Modesto Omiste, Quebrada Honda, *d'Orbigny 1315* (Holotype: P; Fragment: F).

Neobartsia crisafullii (N. Holmgren) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia crisafullii* N. Holmgren in Harling & Sparre (eds.), Flora of Ecuador 21: 165. 1984. TYPE: ECUADOR. Azuay: Páramo de Tinajillas, ca 42 km S of Cuenca on road to Loja, 1.6 km N of summit, 3560 m, 27 Jul 1982, *Holmgren, Crisafulli, Boeke & Clemants 10164* (Holotype: NY; Isotypes: AAU, GB, QCA, S).

Neobartsia crisafullii (N. Holmgren) Uribe-Convers & Tank **subsp. crisafullii**, comb. nov. Basionym: *Bartsia crisafullii* N. Holmgren subsp. *crisafullii* in Harling & Sparre (eds.), Flora of Ecuador 21: 165. 1984. TYPE: Ecuador. Azuay: Páramo de Tinajillas, ca 42 km S of Cuenca on road to Loja, 1.6 km N of summit, 3560 m, 27 Jul 1982, *Holmgren, Crisafulli, Boeke & Clemants 10164* (Holotype: NY; Isotypes: AAU, GB, QCA, S).

Neobartsia crisafullii (N. Holmgren) Uribe-Convers & Tank **subsp. acutiloba** (Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia crisafullii* N. Holmgren subsp.

acutiloba Molau, Opera Bot. 102: 76. 1990. TYPE: PERU. Huánuco: Prov. Huánuco, Mito, ca 25 km (air dist.) W of Huánuco, ca 2750 m, 8–22 Jul 1922, *Macbride & Featherstone 1495* (Holotype: F; Isotype: PH).

Neobartsia diffusa (Benth.) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia diffusa* Benth. in Candolle, Prodr. 10: 548. 1846. TYPE: PERU. Dept. Lima/Junín: Casapi, *Mathews s.n.* (Holotype: K). *Bartsia frigida* Diels, Bot. Jahrb. Syst. 37: 431.1906. TYPE: PERU. Junín: Prov. Yauli, near Arapa, above Yauli along the Lima-Oroya railroad, 4400 m, Jan 1902, *Weberbauer 271* (Holotype: B, destroyed).

Neobartsia elachophylla (Diels) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia elachophylla* Diels, Bot. Jahrb. Syst. 37: 431. 1906. TYPE: PERU. Junín: Prov. Tarma, mountains W of Huacapistana, 3000–3100 m, Jan 1903, *Weberbauer 2087* (Holotype: B destroyed; Lectotype G [designated by Molau in Opera Bot. 102: 76. 1990]; Isotype: MOL). *Euphrasia incana* Ruiz & Pavón ex A. López, Anal. Inst. bot. Cavanilles 17: 456. 1959 (not validly published, nomenclatural type not designated). *Euphrasia sagittata* Ruiz & Pavón ex A. López, Anal. Inst. bot. Cavanilles 17: 454. 1959 (not validly published, nomenclatural type not designated).

Bartsia glabra Edwin, Phytologia 19: 365. 1970. TYPE: PERU. Amazonas: Prov. Chachapoyas, Cerro de Fraijaco (Huaui-huni), NE of Tambo de Ventilla, 3500 m, 7 Jul 1948, *Pennell 15878* (Holotype: PH; Isotypes: BM, GH, NY, US, USM).

Neobartsia elongata (Wedd.) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia elongata* Wedd., Chlor. And. 2: 121. 1860. TYPE: PERU. Cuzco: "Cordillères de Cuzco", *Gay* 348 (Holotype: P). *Bartsia simulans* Edwin, Phytologia 19: 367. 1970. TYPE: PERU. Puno: Prov. Puno, San Antonio de Esquilache, ca 4700 m, 16 May 1937, *Stafford 743* (Holotype: BM; Isotype: K).

Neobartsia fiebrigii (Diels) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia fiebrigii* Diels, Bot. Jahrb. Syst. 31: 430. 1906. TYPE: ARGENTINA. Salta: Dept. Santa Victoria, Toldos, S of Río Bermejo, 2200 m, 11 Dec 1903, *Fiebrig 2390* (Holotype: B destroyed; Lectotype: K [designated by Molau in Opera Bot. 102: 76. 1990]; Isotypes: BM, G, L). *Bartsia curtiflora* Edwin, Phytologia 19: 363. 1970. TYPE: PERU. Cuzco: Prov. Urubamba, Ollantaytambo, 3000–3100 m, 26 Apr 1925, *Pennell 13643* (Holotype: K; Isotypes: BM, G, GH, NY, PH, S, US).

Neobartsia filiformis (Wedd.) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia filiformis* Wedd., Chl. And. 2: 126. 1860. TYPE: BOLIVIA. La Paz: Prov. Larecaja, between Tipuani and Apolobamba, 2500 m, May 1847, *Weddell 4596* (Lectotype P [designated by Molau in Opera Bot. 102: 76. 1990]; Isolectotypes: F, PH).
Bartsia viridis Edwin, Phytologia 19: 367. 1970. TYPE: PERU. Cuzco: Prov. Paucartambo, Paso de Tres Cruces, Cerro de Cusilluyocc, 3700–3900 m, 3 May 1925, *Pennell 13826* (Holotype: F; Isotypes: BM, G, GH, K, M, NY, PH, S, US).

Neobartsia flava (Molau) Uribe-Convers & Tank, comb.nov. Basionym: *Bartsia flava* Molau, Opera Bot. 102: 76. 1990. TYPE: PERU. Cajamarca: Prov. Celendín, Rio Sendamal Valley, NE of Cruz Conga, 68–72 km from Cajamarca on road to Celendín, 3250–3300 m, 27 Mar 1985, *Molau, Sánchez Vega & Öhman 1760* (Holotype: GB; Isotypes: CPUN, GB, HUT).

Neobartsia flava (Molau) Uribe-Convers & Tank **subsp. minor** (Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia flava* Molau subsp. *minor* Molau, Opera Bot. 102: 76. 1990. TYPE: PERU. Cajamarca: Prov. Cajamarca, Cerro Negro, 1–3 km SE, of (above) Abra Gavilán, ca 13 km (air dist.) SSE of Cajamarca, rocky jalca, 3400–3560 m, 18 Mar 1988, *Molau & Eriksen 3337* (Holotype: GB; Isotypes: AAU, CPUN).

Neobartsia glandulifera (Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia glandulifera* Molau, Opera Bot. 102: 76. 1990. TYPE: COLOMBIA. Santander: Edge of Páramo de Las Vetas, 3300–3700 m, 20 – 21 Dec 1926, *Killip & Smith 15605* (Holotype: NY; Isotypes: BR, G, GH, K, P, PH, S, UC, US).

Neobartsia inaequalis (Benth.) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia inaequalis* Benth. in Candolle, Prodr. 10: 548. 1846. TYPE: COLOMBIA. Cundinamarca: Bogotá, *Goudot s.n.* (Lectotype: K [designated by Molau in Opera Bot. 102: 76. 1990]; Isotypes: G, OXF, P, US).

Bartsia laxiflora Benth. in Candolle, Prodr. 10: 548. 1846. TYPE: ECUADOR. Pichincha: Cerro Pichincha, *Jameson s.n.* (Holotype: K).

Neobartsia inaequalis (Benth.) Uribe-Convers & Tank **subsp. brachyantha** (Diels) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia brachyantha* Diels, Bot. Jahrb. Syst. 37: 431. 1906. TYPE: PERU. Puno: Prov. Sandia, above Cuyocuyo, 3600–3800 m, May 1902, *Weberbauer 919* (Holotype: B, destroyed, no isotypes found). NEOTYPE: PERU. Puno: Prov. Sandia, Sandia, 13 Mar 1902, *Weberbauer 510* (Neotype: PH [designated by Molau in Opera Bot. 102: 76. 1990]). *Bartsia inaequalis* Benth. subsp. *brachyantha* Molau, Opera Bot. 102: 76. 1990.

Bartsia altissima Rusby, Phytologia 1: 75. 1934. TYPE: BOLIVIA. La Paz: Prov. Nor Yungas, Pongo, Unduavi Valley, 3650 m, Feb 1926, *Tare 222* (Holotype: NY; Isotype: NY).

Neobartsia inaequalis (Beth.) Uribe-Convers & Tank **subsp. duripilis** (Edwin) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia duripilis* Edwin, Phytologia 19: 364. 1970. TYPE: PERU. Ayacucho: Prov. Huanta, Putis, upper Río Chuimacota Valley, 3200–3300 m, 27 Feb–12 Mar 1926, *Weberbauer 7533* (Holotype: F; Isotypes: NY, PH, U, WIS). *Bartsia inaequalis* Benth. subsp. *duripilis* Molau, Opera Bot. 102: 76. 1990.

Euphrasia rubescens Ruiz & Pavón ex A. López, An. Inst. bot. Cavanilles l7: 457. 1959 (not validly publiihed, nomen- clatural type not designated).

Neobartsia integrifolia (Wedd.) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia integrifolia* Wedd., Chlor. And. 2: 131. 1860. TYPE: PERU. Dept. unknown (dept. Lima according to the protologue, uncertain): Sine loco, Jun 1847, *Castelnau s.n.* (Lectotype: P

[designated by Molau in Opera Bot. 102: 76. 1990]; Isolectotype: PH).

Neobartsia jujuyensis (Cabrera & Botta) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia jujuyensis* Cabrera & Botta, Hickenia 2(9): 46. 1992. TYPE: ARGENITNA, Provincia de Jujuy, Departamento Valle Grande, camino a Alto Calilegua, Tolditos, 30-Dec-1977. *R. Kiesling, E. Ulibarri y A. López 1553* (Holotype: SI).

Neobartsia laniflora (Benth.) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia laniflora* Benth. in Candolle, Prodr. 10: 548. 1846. TYPE: COLOMBIA. Magdalena: Sierra Nevada de Santa Marta, July 1844, *Purdie s.n.* (Holotype: K).
Bartsia spissifolia Pennell in Steyermark, Fieldiana Bot. 28 (3): 516. 1953. TYPE: VENEZUELA. Mérida: Páramo de Pozo Negro, between San José and Beguilla, 2590–3220 m, 3 May 1944, *Steyermark 56286* (Holotype: F).
Bartsia tachirensis Pennell in Steyermark, Fieldiana Bot. 28 (3): 516. 1953.
TYPE: VENEZUELA. Táchira: limestone outcrops of Páramo de Tamá. 3045–3475 m, l5 Jul 1944. *Steyermark 57409* (Holotype: PH; Isotypes: F, NY, US).

Neobartsia laticrenata (Benth.) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia laticrenata* Benth. in Candolle, Prodr. 10: 548. 1846. TYPE: ECUADOR. Napo: On the boggy outlet of the Laguna Mica near Antisana, ca 4000 m, 1843, *Hartweg 1289* (Lectotype: K [designated by Molau in Opera Bot. 102: 76. 1990]; Isotypes: BM, BREM, CCE, E, G, LD, OXF, P, W).

Neobartsia lydiae (Sylvester) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia lydiae* Sylvester, Phytotaxa 164(1): 41. 2014. TYPE: PERU. Cusco, Prov. Urubamba: Distr. Urubamba, Área de Conservación Privada (ACP) Mantanay, 10 km up the valley from Yanahuara in the small valley 3 km E of laguna Ipsaycocha, forest on the SW side of laguna Manalloqsa, 4614m, S13° 11' 59.2" W72° 08' 39.9", 23 June 2012, *S.P. Sylvester 1754* (Holotype USM!, Isotypes CUZ!, GB!, ID!, MO!, Z!).

Neobartsia melampyroides (Kunth) Uribe-Convers & Tank, comb. nov. Basionym: *Euphrasia melampyroides* Kunth in Humboldt, Bonpland and Kunth, Nov. gen. sp. 2: 334. 1818. TYPE: ECUADOR. Pichincha: Cachabamba plain and near Píntag, foothills of Antisana, 3300–3900 m, May 1802, *Bonpland 1406* (Lectotype: B-WILLD [designated by Molau in Opera Bot. 102: 76. 1990]; Isotype: H). *Bartsia melampyroides* Benth. in Candolle, Prodr. 10: 548. 1846. *Bartsia gracilis* Benth. in Candolle, Prodr. 10: 548. 1846. TYPE: ECUADOR. Chimborazo: Hda. Chuquipucyu ("Hda. Chuquipoyo"), ca 6 km SW of summit of the Ambato-Riobamba road, ca 3300 m, *Hartweg* 1291 (Lectotype: K [designated by Molau in Opera Bot. 102: 76. 1990]; Isotypes: BM, BR, BREM, CGE, E, F, G, LD, LE, OXF, P, W). *Bartsia euphrasioides* Wedd., Chlor. And. 2: 130. 1860. TYPE: ECUADOR. Pichincha(?): Andes of Quito, ca. 3950 m, *Jameson 477* (Holotype: P; Isotypes: E, G, K). *Bartsia elongata* Wedd. β *Pusilla* Wedd., Chlor. And. 2: 127. 1860. TYPE: PERU. Puno: Prov. Carabaya, Jun – Jul 1847, *Weddell 4658* (Holotype: P).

Neobartsia mutica (Kunth) Uribe-Convers & Tank, comb. nov. Basionym: *Euphrasia mutica* Kunth in Humboldt, Bonpland and Kunth, Nov. gen. sp. 2: 334. 1818. TYPE: PERU.

Piura: Prov. Ayabaca, between Lucarque and Ayabaca, ca 2400 m, Aug 1802, *Bonpland 3466* (Lectotype: P, designated by Molau in Opera Bot. 102: 58. 1990, Isotype: H, F fragment).

Bellardia mutica (Kunth) A. Fleischm. & Heubl, Taxon 61(6): 1282. 2012.

Bartsia hispida Benth. in Humboldt, Bonpland and Kunth, Nov. gen. sp. 2: 547. 1818. TYPE: PERU. Amazonas: Province of Chachapoyas, *Mathews 788* (Holotype: K, Isotypes: G, OXF).

Bartsia campii N. Holmgren in Harling & Sparre (eds.), Flora of Ecuador 21: 159. 1984. TYPE: ECUADOR. Chimborazo: Canyon of the Río Chanchán, above Huigra. 1500–2150 m, 29–37 May 1945, *Camp E-3501* (Holotype: NY; Isotype PH). *Bartsia patriciae* N. Holmgren, in Harling & Sparre (eds.), Flora of Ecuador 21: 158. 1984. TYPE: ECUADOR. Pichincha: 23 km E of (above) Tandapi on road from Santo Domingo to Alóag, 2510 m, 8 Jul 1982, *Holmgren* & *Holmgren 10072* (Holotype: NY; Isotypes: AAU, cB, QCA, S).

Neobartsia orthocarpiflora (Benth.) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia orthocarpiflora* Benth. in Candolle, Prodr. 10: 548. 1846. TYPE: ECUADOR. Pichincha: Andes of Quito, *Jameson s.n.* (Holotype: K). *Bartsia breviflora* Benth., in Candolle, Prodr. 10: 548. 1846. TYPE: ECUADOR. Pichincha: Cerro Pichincha, *Hall 49* (Lectotype: K [designated by Molau in Opera Bot. 102: 76. 1990]).

Bartsia heterophylla Wedd., Chl. And. 2: 129. 1860. TYPE: COLOMBIA. Nariño: Province of Tuquerres, 3000–3500 m, *Triana s.n.* (Lectotype: P [designated by Molau in Opera Bot. 102: 76. 1990]; Isotypes: BM, G, W; Fragments: F, PH).

Bartsia laxissima Danguy & Chermezon, Bull. Mus. Hist. Nat. (Paris) 28: 436. 1922. TYPE: ECUADOR, Prov. unknown: Naes, Feb 1903, *Rivet 354* (Holotype: P).

Neobartsia orthocarpiflora (Benth.) Uribe-Convers & Tank **subsp. villosa (**Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia orthocarpiflora* Benth. subsp. *villosa* Molau, Opera Bot. 102: 76. 1990. TYPE: COLOMBIA. Risaralda: Cerro Tatamá, Cordillera Occidental, moist grassy páramo, 3400-3700 m, 8-10 Sep 1922, *Pennell 10577* (Holotype: GH; Isotypes: F, NY, PH, US).

Neobartsia patens (Benth.) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia patens* Benth. in Candolle, Prodr. 10: 548. 1846. TYPE: PERU. Sine loco, *Mathews 898* (Lectotype: CGE [herb. Lindley] [designated by Molau in Opera Bot. 102: 76. 1990; Isotype: CGE [herb. Lemann]).

Neobartsia pauciflora (Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia pauciflora* Molau, Opera Bot. 102: 76. 1990. TYPE: BOLIVIA. La Paz: Prov. Murillo, La Rinconada, NE of La Paz, just below summit of road to Unduavi, ca 4200 m, 10 Feb 1921, *Asplund 2392* (Holotype: UPS; Isotypes: S, US).

Neobartsia pedicularoides (Benth.) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia pedicularoides* Benth. in Candolle, Prodr. 10: 548. 1846. TYPE: ECUADOR. Napo: On the boggy outlet of Laguna Micacocha (Mica), plains SW of Nevado de Antisana, ca 4100 m, *Hartweg 1290* (Holotype: K; Isotypes: BM, CGE, E; G, LD, OXF, P, W).

Bartsia parvifolia Benth., in Candolle, Prodr. 10: 545. 1846. TYPE: VENEZUELA. Mérida: Sierra Nevada de Mérida, ca 3300 m, Aug 1842, *Linden 423* (Holotype: K; Isotypes: BM, G, OXF, P, W; Fragment: F).

Bartsia biloba Wedd., Chl. And. 2: 123. 1860. TYPE: PERU. Cuzco: Sine loco, 1839–40, *Gay 909* (Lectotype: P [designated by Molau in Opera Bot. 102: 76. 1990]).

Neobartsia peruviana (Walp.) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia peruviana* Walp., Nov. Act. Acad. Caes. Leopold. Carol. 19 Suppl. 1: 400. 1843. TYPE: PERU. Puno: Prov. Chuquito, Pizacoma, ca 4600 m, 31 Apr 1831, *Meyen s.n.* (Lectotype: K [designated by Molau in Opera Bot. 102: 76. 1990]; Isotype: BR).

Bartsia meyeniana Benth. in Candolle, Prodr. 10: 545. 1846. TYPE: CHILE. Parinacota: Tacora, close to the Peruvian border, ca 70 km (air dist.) NE of Tacna, ca 4600 m, 1831, *Meyen s.n.* (Holotype: K).

Bartsia subinclusa Benth. in Candolle, Prodr. 10: 545. 1846. TYPE: PERU. Puno: Prov. Chuquito, Pizacoma, ca 4600 m, Apr 1831, *Meyen s.n.* (Holotype: K).

Neobartsia pumila (Benth.) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia pumila* Benth. in Candolle, Prodr. 10: 548. 1846. TYPE: ECUADOR. Pichincha/Napo: Andes of Quito, 1845, *Jameson s.n.* (Holotype: K).

Neobartsia pyricarpa (Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia pyricarpa* Molau, Opera Bot. 102: 76. 1990. TYPE: PERU. Cajamarca: Prov. Cajamarca, Cumbemayo ca. 11 km (air dist.) WSW of Cajamarca, 3400–3450 m, 24 Mar 1985, *Molau,*

Sánchez Vegas & Öhman 1681 (Holotype: GB; Isotypes: HUT, CPUN).

Neobartsia ramosa (Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia ramosa* Molau, Opera Bot. 102: 76. 1990. TYPE: ECUADOR. Napo: Páramo de Guamaní, 1.5–2 km N of the summit of the Quito-Baeza road, 4200–4300 m, 21 Dec 1987, *Molau & Eriksen 2114* (Holotype: GB; Isotypes: AAU QCA, QCNE).

Neobartsia remota (Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia remota* Molau, Opera Bot. 102: 76. 1990. TYPE: PERU. Lima: Prov. Canta, along Río Chillón, near Obrajillo (ca 2 km NE of Canta), 2200–2500 m, 10–15 Jun 1925, *Pennell 14326* (Holotype: PH; Isotypes: BM, F, GH, NY, US).

Neobartsia rigida (Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia rigida* Molau, Opera Bot. 102: 76. 1990. TYPE: PERU. Lima: Prov. Huarochirí, Río Blanco, upper Rímac Va1ley, ca 3650 m, 8–19 May 1922, *Macbride & Featherstone 735* (Holotype: F; Isotype: PH).

Neobartsia santolinifolia (Kunth) Uribe-Convers & Tank, comb. nov. Basionym: *Euphrasia santolinaefolia* Kunth in Humboldt, Bonpland & Kunth, Nov. gen. sp. 2: 333. 1818. TYPE: COLOMBIA. Cundinamarca: Between Soacha and Santa Fé de Bogotá, just SW of Bogotá, ca 2500 m, Jul 1802, *Bonpland s.n.* (Lectotype: P [designated by Molau in Opera Bot. 102: 76. 1990]; Isotypes: H, P).

Bartsia santolinifolia (Kunth) Benth in Candolle, Prodr. 10: 548. 1846.

Neobartsia sericea (Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia sericea* Molau, Opera Bot. 102: 76. 1990. TYPE: PERU. Cajamarca: Cumbemayo, ca 11 km (air dist.) WSW of Cajamarca, 3400–3500 m, 24 Mar 1985, *Molau, Öhman & Sánchez Vega 1682* (Holotype: GB; Isotypes: CPUN, HUT).

Neobartsia serrata (Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia serrata* Molau, Opera Bot. 102: 76. 1990. TYPE: PERU. Arequipa: Prov. Arequipa, canyon above Arequipa, 3100–3300 m, 26–28 May 1925, *Pennell 14276* (Holotype: NY; Isotypes: B, BM, F, G, GH, K, M, PH, S, US, WIS).

Neobartsia stricta (Kunth) Uribe-Convers & Tank, comb. nov. Basionym: *Euphrasia stricta* Kunth in Humboldt, Bonpland & Kunth, Nov. gen. sp. 2: 333. 1818. TYPE: ECUADOR. Prov. Pichincha/Napo: Above the village of Chillo and on the slopes of Volcán Antisana, 2500–3100 m, May 1802, *Bonpland s.n.* (Lectotype: B-WILLD no. 11166 [designated by Molau in Opera Bot. 102: 76. 1990]). *Bartsia stricta* (Kunth) Benth. in Candolle, Prodr. 10: 548. 1846.

Neobartsia strigosa (Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia strigosa* Molau, Opera Bot. 102: 76. 1990. TYPE: PERU. Lima: Prov. Huarochirí, Río Blanco, upper Rímac Valley, 3000–3500 m, 15–17 Apr 1929, *Killip & Smith 21608* (Holotype: US; Isotypes: F, NY, PH).

Neobartsia tenuis (Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia tenuis* Molau, Opera Bot. 102: 76. 1990. TYPE: PERU. Cajamarca: Prov. Cajamarca, Abra El Gavilán, S of Cajamarca, summit area along road to Chilete, 3100–3150 m, 23 Mar 1985, *Molau, Öhman & Sánchez Vega 1658* (Holotype: GB; Isotypes: CPUN, HUT).

Neobartsia thiantha (Diels) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia thiantha* Diels, Bot. Jahrb. Syst. 37: 432. 1906. TYPE: PERU. Cuzco: Prov. Cuzco, ruins of Sacsayhuaman, 3500 m, 17 Mar 1985, *Molau & Öhman 1606* (Neotype: GB [designated by Molau in Opera Bot. 102: 76. 1990]; Isoneotypes: CPUN, HUT); both syntypes cited in the protologue (i.e., *Weberbauer* 3834 and 4858) were destroyed at B and no duplicates have been found. *Bartsia anomala* Edwin, Phytologia 19: 362. 1970. TYPE: PERU. Cuzco: Prov. Paucartambo, on trail from Paucartambo to Vilcanota, Cerro de Colquepata, 3400–3700 m, 7 May 1925, *Pennell* 14176 (Holotype: BM; Isotypes: GH, NY, PH, US). *Bartsia aurea* Edwin, Phytologia 19: 363. 1970, no holotype was given in the protologue, not validly published.

Neobartsia tomentosa (Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia tomentosa* Molau, Opera Bot. 102: 76. 1990. TYPE: PERU. Cajamarca: Prov. Cajamarca, Cumbemayo, ca 11 km (air dist.) WSW of Cajamarca, 3450–3500 m, 29 Mar 1985, *Molau, Sánchez Vega & Öhman 1680* (Holotype: GB; Isotypes: CPUN, HUT).

Neobartsia trichophylla (Wedd.) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia trichophylla* Wedd., Chl. And. 2: 122. 1860. TYPE: PERU. Puno: between Puno and Arequipa, 4000 m, 1847, *Weddell s.n.* (Holotype: P).

Bartsia ciliolata Wedd., Chl. And. 2: 123. 1860. TYPE: BOLIVIA. La Paz: Prov. Larecaja, Cordillera de Sorata, 1850, *Mandon 114* (Holotype: P). *Bartsia sanguinea* Diels, Bot. Jahrb. Syst. 37: 433. 1906. TYPE: BOLIVIA. LaPaz: Prov. Larecaja, above Sorata, ca. 4000 m (13,000 ft, says incorrectly 8000 ft on some labels to the collection), Feb 1886, *Rusby 1092* (Lectotype: NY [designated by Molau in Opera Bot. 102: 76. 1990]; Isotypes: BM, E, F, G, GH, K, MICH, MO, NY, P, PH, US). *Bartsia guggenheimiana* Rusby, Phytologia 1: 76. 1934. TYPE: BOLIVIA. La Paz: Prov. Nor Yungas, Alaska Mine, above El Pongo, ca. 4200 m, 1–4 Mar 1926, *Tate 64* (Holotype: NY; Isotype: PH).

Bartsia pedicellata Edwin, Phytologia 19: 366. 1970. TYPE: PERU. Cuzco: Prov. Paucartambo, between Paucartambo and Tres Cruces, Cerro de Cusilluyocc, 3500–3800 m, 2-6 May 1925, *Pennell 13815* (Holotype: F; Isotypes: BM, G, K, M, NY, PH, S, US, USM, WIS).

Neobartsia tricolor (Molau) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia tricolor* Molau, Opera Bot. 102: 76. 1990. TYPE: PERU. Ancash: Prov. Huaraz, upper W slopes of the Cordillera Negra, 35–37 km from Huaraz on road to Casma, 2–3 km W of the pass (Punta Callán), ca. 4000 m, 29 Jan 1983, *Molau, Dillon & Matekaitis 538* (Holotype: GB; Isotypes: S, USM).

Neobartsia werberbaueri (Diels) Uribe-Convers & Tank, comb. nov. Basionym: *Bartsia weberbaueri* Diels, Bot. Jahrb. Syst. 37: 431. 1906. TYPE: PERU. Ancash: Prov. Bolognesi, above Ocros, 3500–3700 m, 28 Mar 1893, *Weberbauer 2692* (Holotype: B destroyed; Lectotype: G [designated by Molau in Opera Bot. 102: 76. 1990]; Isotype: PH).

Mediterranean Species

In order to reflect the monophyly and distinctiveness of the South American species in the classification proposed here, the taxon *Bellardia latifolia* needs to revert to its previous taxonomic classification, i.e., *Parentucellia latifolia*. Scheunert et al. (2012) expanded the genus *Bellardia* to include both species of the previous genus *Parentucellia* because these taxa did not form a clade in their analyses. However, given that our results show that *Parentucellia latifolia* is sister to the *Neobartsia* clade and that *Bellardia viscosa* forms a clade with *Bellardia trixago*, the name *Parentucellia latifolia* becomes useful again.

Interestingly, Scheunert et al. (2012) rearranged the two previous species of *Parentucellia* but disregarded the generic type *P. floribunda* Viv. and the implications concerning this third species. Although we were not able to include *P. floribunda* in our study, we have investigated the history of this species and how the genus was described. In the original publication, Domenico Viviani (1824) described the genus providing a description and a drawing of *P. floribunda,* but he did not specify a type specimen. We have examined a high-resolution photo of the single specimen of *P. floribunda* that we have been able to find in any major herbaria (Voucher No. P03934257, Muséum National d'Histoire Naturelle [P], Paris, France), and based on the photo and the original description for this

species (Viviani 1824), as well as the original description of *P. latifolia* and *B. viscosa* (Caruel 1885), it is clear that *P. floribunda* is the same taxon as *P. latifolia*. This means that *P. floribunda* has been collected and misidentified for a long time. Moreover, the Plant List (http://www.theplantlist.org, last accessed on April 28th 2014) lists *P. floribunda* as an unresolved name and suggests it is a synonym of *Bartsia latifolia* (L.) Sm., a synonym itself of *P. latifolia*. Flora Europaea (Tutin et al. 1973), the Families and Genera of Vascular Plants (Fischer 2004), and Mabberley's Plant-Book (Mabberley 2008) all recognize the genus *Parentucellia* to have only two species (*P. latifolia* and *B. viscosa*), supporting that the name *P. floribunda* has been neglected and that all the collections have been made under *P. latifolia. P. floribunda* and *P. latifolia* share diagnostic characters that separate them from *B. viscosa*, mainly i) the smaller size: 5–30 cm (10–70 cm for *B. viscosa*), ii) the purple corolla (white in *B. viscosa*), iii) the short corolla tube, almost as long as the calyx (long tube in *B. viscosa*), iv) the persistent corolla (caducous in *B. viscosa*), and v) the glabrous capsule (pubescent in *B. viscosa*). Based on all this, it is evident that the names *P. latifolia* and *P. floribunda* are synonyms and that they should be consolidated into one. Although the first assumption would be to synonymize *P. latifolia* under the generic type *P. floribunda*, after careful review of their taxonomic history and alternative synonymous names, one can discover that the name *P. latifolia* should be used instead, because it was described by Linnaeus in 1753 as *Euphrasia latifolia* L., and thus, has priority over *P. floribunda* (published in 1824).

Parentucellia latifolia (L.) Caruel. Basionym: *Euphrasia latifolia* L., Sp. Pl. 2: 604. 1753. LECTOTYPE (designated by Sutton in Jarvis, Order Out Of Chaos: 514. 2007): [icon.]

Euphrasia pratensis Italica latifolia Morison, Pl. Hist. Univ. 3: 431, s. 11, t. 24, f. 8. 1699.

Bartsia latifolia (L.) Sm., Flora Graeca 6: 69. 1827.

Bellardia latifolia (L.) Cuatrec., Trab. Mus. Ci. Nat. Barcelona 12: 428. 1929.

Lastly, we would like to propose a clade name for all the species mentioned above, as to facilitate communication within the Rhinantheae clade.

Molaua Uribe-Convers & Tank, new clade name. Definition (node-based, branch-modified): most inclusive clade that includes *Bellardia trixago* (L.) All. 1785, *Parentucellia latifolia* (L.) Caruel 1885, and *Neobartsia santolinifolia* (Kunth) Uribe-Convers & Tank 2014. There is no preexisting scientific name for this clade. The name *Molaua* is chosen to honor Prof. Ulf Molau, University of Gothenburg, for his dedication and excellent work on the former *Bartsia* species.

Key to the genera in the clade *Molau*

1. Annual plants; lower corolla lip deflected with conspicuous dorsal gibbae (humps), the lateral lobes wider than the central one. Mediterranean distribution or subtropical dry climate (Mediterranean-like climate in western USA, Australia, lowland Chile).....2

1. Annual or perennial plants; lower corolla lip erect or deflected, dorsal gibbae rare. If lower corolla lip is deflected, lateral lobes as wide as the central one. High elevation Andean (Páramo ecosystem) distribution.....*Neobartsia*

2. Corolla red to purple, plants 5-30 cm .....*Parentucellia latifolia*

2. Corolla white to pale yellow, plants 10-70 cm 3

3. Corolla white, the galea dorsally suffused with purple, diploid plants .....*Bellardia trixago*

3. Corolla yelow, the galea yellow, tetraploid plants.....*Bellardia viscosa*

Literature Cited

Allioni C. 1785. Flora pedemontana. Torino.

Angiosperm Phylogeny Group. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Botanical Journal of the Linnean Society. 161:105–121.

Beaulieu J.M., Tank D.C., Donoghue M.J. 2013. A Southern Hemisphere origin for campanulid angiosperms, with traces of the break-up of Gondwana. BMC Evol Biol. 13:80.

Bennett J.R., Mathews S. 2006. Phylogeny of the parasitic plant family Orobanchaceae inferred from phytochrome A. American Journal of Botany. 93:1039–1051.

Bozlaker A., Prospero J.M., Fraser M.P., Chellam S. 2013. Quantifying the contribution of long-range Saharan dust transport on particulate matter concentrations in Houston, Texas, using detailed elemental analysis. Environ. Sci. Technol. 47:10179–10187.

Cabrera A.L., Botta S.M. 1992. Novedades para la Flora de Jujuy (Argentina). Hickenia. 2:1–4.

Caruel T. 1885. Flora italiana; ossia, Descrizione delle piante che crescono spontanee o vegetano come tali in Italia e nelle isole ad essa aggiacenti; disposta secondo il metodo naturale. Tipografia Le Monnier, Firenze.

Clay D.L., Novak S.J., Serpe M.D., Tank D.C., Smith J.F. 2012. Homoploid hybrid speciation in a rare endemic Castilleja from Idaho (Castilleja christii, Orobanchaceae). American Journal of Botany. 99:1976–1990.

Coates A., Collins L., Aubry M., Berggren W. 2004. The geology of the Darien, Panama, and the late Miocene-Pliocene collision of the Panama arc with northwestern South America. Geological Society of America Bulletin. 116:1327.

Davis C.C., Bell C.D., Mathews S., Donoghue M.J. 2002. Laurasian migration explains Gondwanan disjunctions: evidence from Malpighiaceae. Proc. Natl. Acad. Sci. U.S.A. 99:6833–6837.

Denk T., Grimsson F., Zetter R. 2010. Episodic migration of oaks to Iceland: Evidence for a North Atlantic “land bridge” in the latest Miocene. American Journal of Botany. 97:276–287.

Denk T., Grímsson F., Zetter R., Símonarson L.A. 2011. The Biogeographic History of Iceland – The North Atlantic Land Bridge Revisited. Topics in Geobiology. Dordrecht: Springer Netherlands. p. 647–668.

Donoghue M.J. 2008. A phylogenetic perspective on the distribution of plant diversity. Proc. Natl. Acad. Sci. U.S.A. 105:11549–11555.

Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 32:1792–1797.

Fischer E. 2004. Scrophulariaceae. In: Kubitzki K., Kadereit J., editors. Flowering Plants — Dicotyledons: Lamiales. The Families and Genera of Vascular Plants. Springer. p. 333–432.

Hopkins D.M. 1967. The Bering land bridge. Stanford University Press, Stanfor, CA.

Hughes C., Eastwood R. 2006. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. Proceedings of the National Academy of Sciences. 103:10334.

Kirby M.X., MacFadden B. 2005. Was southern Central America an archipelago or a peninsula in the middle miocene? A test using land-mammal body size. Palaeogeography, Palaeoclimatology, Palaeoecology. 228:193–202.

Mabberley D.J. 2008. Mabberley's plant-book: a portable dictionary of plants, their classification and uses. Cambridge, UK. University Press.

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

Molau U. 1990. The genus Bartsia (Scrophulariaceae - Rhinanthoideae). Opera Botanica. 102:1–100.

Moore B.R., Donoghue M.J. 2007. Correlates of diversification in the plant clade Dipsacales: geographic movement and evolutionary innovations. Am Nat. 170 Suppl 2:S28–55.

Olmstead R.G., Depamphilis C.W., Wolfe A.D., Young N.D., Elisons W.J., Reeves P.A. 2001. Disintegration of the Scrophulariaceae. American Journal of Botany. 88:348–361.

Prospero J.M., Collard F.X., Molinié J., Jeannot A. 2014. Characterizing the annual cycle of African dust transport to the Caribbean Basin and South America and its impact on the environment and air quality. Global Biogeochemical Cycles. 29.

Prospero J.M., Mayol-Bracero O.L. 2013. Understanding the Transport and Impact of African Dust on the Caribbean Basin. Bull. Amer. Meteor. Soc. 94:1329–1337.

Rambaut A., Drummond A.J. 2004. Tracer. University of Edinburgh, Edinburgh, UK. Available at http://tree.bio.ed.ac.uk/software/tracer/.

Retallack G., Kirby M. 2007. Middle Miocene global change and paleogeography of Panama. Palaios. 22:667–679.

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

Scheunert A., Fleischmann A., Olano-Marin C., Bräuchler C., Heubl G. 2012. Phylogeny of tribe Rhinantheae (Orobanchaceae) with a focus on biogeography, cytology and re-examination of generic concepts. Taxon. 61:1269–1285.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Systematic Biology.

Shimodaira H., Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics. 17:1246–12461247.

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

Stamatakis A., Hoover P., Rougemont J. 2008. A Rapid Bootstrap Algorithm for the RAxML Web Servers. Systematic Biology. 57:758–771.

Swenson U., Nylinder S., Wagstaff S.J. 2012. Are Asteraceae 1.5 Billion Years Old? A Reply to Heads. Systematic Biology. 61:522–532.

Sylvester S.P. 2014. Bartsia lydiae, a new species of Bartsia sect. Laxae (Orobanchaceae) from the southern Peruvian Andes with a revised key to Bartsia sect. Laxae. Phytotaxa. 164:41–46.

Těšitel J., Říha P., Svobodová Š., Malinová T., Štech M. 2010. Phylogeny, Life History Evolution and Biogeography of the Rhinanthoid Orobanchaceae. Folia Geobotanica. 45:347–367.

Tiffney B., Manchester S. 2001. The use of geological and paleontological evidence in evaluating plant phylogeographic hypotheses in the Northern Hemisphere Tertiary. International Journal of Plant Sciences. 162:3–17.

Tiffney B.H. 1985. Perspectives on the Origin of the Floristic Similarity Between Eastern Asia and Eastern North-America. Journal of the Arnold Arboretum. 66:73–94.

Tiffney B.H. 2008. Phylogeography, Fossils, and Northern Hemisphere Biogeography: The Role of Physiological Uniformitarianism 1. Annals of the Missouri Botanical Garden. 95:135–143.

Tutin T.G., Heywood V.H., Burges N.A., Moore D.M., Valentine D.H., Walters S.M., Webb D.A. 1973. Flora Europaea: Vol. 3: Diapensiaceae to Myoporaceae. Cambridge, UK. University Press.

Viviani D. 1824. Scrophulariaceae *Parentucellia* Viv. Florae libycae specimen.:31–32.

Wolfe A., Randle C., Liu L., Steiner K. 2005. Phylogeny and biogeography of Orobanchaceae. Folia Geobotanica. 40:115–134.

Chapter 3: A long PCR based approach for DNA enrichment prior to next-generation sequencing for systematic studies

Coauthors: Justin R. Duke, Michael J. Moore

Abstract

We present an alternative approach for molecular systematic studies that combines long PCR and next-generation sequencing (NGS). Our approach can be used to generate templates from any DNA source for NGS. Here we test our approach by amplifying complete chloroplast genomes and we present a set of 58 potentially universal primers for angiosperms to do so. Additionally, this approach is likely to be particularly useful for nuclear and mitochondrial regions. Chloroplast genomes of 30 species across angiosperms were amplified to test our approach. Amplification success varied depending on whether PCR conditions were optimized for a given taxon. To further test our approach, some amplicons were sequenced on an Illumina HiSeq 2000. Although here we tested this approach by sequencing plastomes, long PCR amplicons could be generated using DNA from any genome, expanding the possibilities of this approach for molecular systematic studies.

Introduction

Advancements in next-generation sequencing (NGS) technologies have permitted the assembly of large, genome-scale datasets that have shed light on the evolutionary history of many taxa (e.g., Parks et al. 2009; Moore et al. 2010; Xi et al. 2012; Tennessen et al. 2013; Eaton and Ree 2013). For plant phylogenetics, there has been a major focus on methods for chloroplast phylogenomics (e.g., Parks et al. 2009; Moore et al. 2010), although methods for

collecting phylogenomic datasets from the nuclear and mitochondrial genomes have also been developed (e.g., Straub et al. 2012; Eaton and Ree 2013). Stull et al. (Stull et al. 2013) developed a custom RNA probe set designed to capture angiosperm plastomes via solution-based hybridization. While their capture system was broadly successful, Stull et al. (2013) found that the most variable spacer regions were often captured at much-reduced coverage compared to more conserved regions, and were sometimes missed entirely if the target taxon was phylogenetically divergent from one of the 22 plastomes used in the bait design. Moreover, the current cost of the capture probes makes this method most efficient for projects dealing with hundreds of species. Another commonly employed method for plant phylogenomic studies is genome skimming (Straub et al. 2012), which takes advantage of the fact that organellar DNA and nuclear ribosomal DNA are present at high copy numbers in genomic DNA. However, a significant limitation of this method for systematic studies is that only high-copy number regions are recovered consistently across all samples, whereas regions with lower representation are only recovered in some samples and missed completely in others (Straub et al. 2011). This can be problematic for molecular systematic studies where missing data may result in misleading phylogenetic results (Lemmon et al. 2009). Moreover, being limited to high-copy regions in the genome becomes restrictive for experimental design as it excludes putatively highly informative regions in the genome such as single copy nuclear genes—e.g., the COSII genes (Wu et al. 2006; Bombarely et al. 2011) and the PPR gene family (Yuan et al. 2009).

As an alternative, we present a NGS approach that combines long PCR and Illumina sequencing to strategically compile phylogenomic datasets for molecular systematic studies. Long PCR, or long-range PCR, uses a combination of two polymerases, a nonproofreading

polymerase at high concentration and a proofreading polymerase at a lower concentration, to amplify DNA fragments that range between 3 and 15 kilobases (kb), although cases of extremely large fragments (22 kb to 42 kb) have been reported (e.g., Cheng et al. 1994). Long PCR has been used extensively in human genome projects (e.g., Craig et al. 2008) and to sequence complete mitochondrial genomes (Alexander et al. 2013), using both Sanger sequencing and NGS technologies. Here, we use long PCR to generate chloroplast DNA templates for systematic studies using NGS. While we focus on whole chloroplast amplification, this approach is directly translatable to targeted studies where only particular regions of the plastome are of interest (e.g., the inverted repeat or the small single copy region). In addition, long PCR could also be very useful for the enrichment of mitochondrial and/or nuclear regions where intron sizes are large or unknown, as well as for regions that are difficult to assemble bioinformatically such as repetitive regions.

Our focus on chloroplast genomes is driven by its phylogenetic informativeness at essentially all taxonomic scales and its relative ease of amplification (e.g., Downie and Palmer 1992; Graham and Olmstead 2000; Moore et al. 2007; Parks et al. 2009; Moore et al. 2010), which have made the chloroplast the workhorse of molecular plant systematics since the beginning of the field. Moreover, the availability of a large number of angiosperm plastome sequences had facilitated the design of potentially universal PCR primers. To test this approach we amplified the chloroplast genomes of 30 species (17 genera) across angiosperms using a set of 58 chloroplast PCR primers that were designed to potentially be universal in angiosperms and that may work in some gymnosperm lineages.

Methods And Results

Representatives of 17 different genera (30 spp.) spanning 12 orders of angiosperms *sensu* APG III (Angiosperm Phylogeny Group 2009) were chosen to test this approach (Table 3.1). Special focus was given to three genera in Orobanchaceae: *Lamourouxia* Kunth (one species); *Bartsia* L. (two species), and *Castilleja* Mutis ex L.f. (12 species). High quality genomic DNA was extracted from ca. 0.02 g of silica gel-dried or herbarium tissue using a modified 2X CTAB method (Doyle and Doyle 1987), yielding 30 to 70 ng/µL of DNA per sample. Using the 83 plastid gene angiosperm alignments of Moore et al. (Moore et al. 2010), we developed 58 primers with a goal of maximizing universality across angiosperms (Table 3.2). Conserved regions for primer design were identified by eye and the primers were tested with IDT OligoAnalyzer tools (Integrated DNA Technologies, Inc http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/) to ensure that melting temperatures (Tm) were over 50°C, and that there were no significant hairpins or self-dimerization problems. From these, 16 overlapping primer combinations were chosen to amplify the entire chloroplast genome in appropriately sized, overlapping fragments, making sure to allow at least 100 bp of overlap between regions (Fig. 3.1, Table 3.2) to minimize the drop in sequencing depth usually associated with the ~30 bp immediately adjacent to the primer sites (Cronn et al. 2008; Harismendy and Frazer 2009; Cronn et al. 2012).

Table 3.1

List of species included in this study, with voucher information, tissue sources, and NGS assembly statistics when available. The number of the regions is the same as the order in Figure 3.1. Base pairs (bp) sequenced is the sum of all contigs when including only one copy of the inverted repeat, CAL = Contig Average Length, Ave. Assem. Depth = Average Assembly Depth, No. of masked bp = Number of bases masked because the minimum sequencing depth of 5x was not achieved , % of masked bp = Percentage of masked bases, % called bases = Percentage of unambiguously called bases. All data from the 16 chosen primer combinations.

Species	Order/Family	Collection no.	Herbarium	Type of tissue	Collection date	No. of amplified regions	Region no. not amplified[b]	Base pairs sequenced[c]	No. of contigs	CAL bp (min–max)	Ave. assembly depth	No. of masked bp[d]	% of masked bp	N50	% called bases[e]	No. of ambiguous bases	% of ambiguous bases
Bartsia inaequalis Benth.	Lamiales/Orobanchaceae	Uribe-Convers 2010-22	ID	Silica gel–dried	5 July 2010	16	n/a	125,283	25	5011 (204–28,257)	656	2126	1.7	19,294	99.9729	34	0.02714
Castilleja covilleana L. F. Hend.	Lamiales/Orobanchaceae	Tank 1046	ID	Silica gel–dried	13 July 2009	16	n/a	133,595	10	13,360 (1222–48,767)	641	101	0.08	37,107	99.9948	7	0.00524
Castilleja elmeri Fernald	Lamiales/Orobanchaceae	Olmstead 2001-78	WTU	Silica gel–dried	4 July 2001	16	n/a	122,614	11	11,147 (464–34,602)	664	440	0.36	33,049	99.9976	3	0.00245
Castilleja linariifolia Benth.	Lamiales/Orobanchaceae	Tank 2001-49	WTU	Silica gel–dried	21 July 2001	16	n/a	122,046	8	15,256 (819–50,680)	642	260	0.21	28,529	99.9984	2	0.00164
Castilleja miniata Douglas ex Hook.	Lamiales/Orobanchaceae	Tank 1048-b	ID	Silica gel–dried	13 July 2009	16	n/a	134,704	4	33,676 (6157–75,123)	844	35	0.03	75,123	99.9970	4	0.00297
Castilleja pallescens (A. Gray) Greenm.	Lamiales/Orobanchaceae	Tank 2009-8	ID	Silica gel–dried	6 June 2009	16	n/a	125,490	4	31,372 (3039–73,629)	764	29	0.02	73,629	99.9984	2	0.00159
Bartsia stricta (Kunth) Benth.	Lamiales/Orobanchaceae	Uribe-Convers 2010-24	ID	Silica gel–dried	7 July 2010	15	13	119,828	14	8559 (425–67,195)	707	1045	0.87	67,195	99.9967	4	0.00334
Castilleja applegatei Fernald	Lamiales/Orobanchaceae	Tank 2001-35	WTU	Silica gel–dried	24 June 2001	15	10	119,647	14	8546 (204–28,559)	642	394	0.33	18,856	99.9983	2	0.00167
Castilleja virgata (Domb. ex Wedd.) Edwin	Lamiales/Orobanchaceae	Olmstead 2009-22	WTU	Silica gel–dried	5 Mar. 2009	15	7	113,650	21	5412 (178–39,914)	698	1525	1.34	14,541	99.9938	7	0.00616
Castilleja ortegae Standl.	Lamiales/Orobanchaceae	Egger 1213	WTU	Silica gel–dried	22 Feb. 2002	15	13	108,071	3	36,024 (269–97,615)	925	198	0.18	97,615	99.9991	1	0.00093
Castilleja lineariloba (Benth.) T. I. Chuang & Heckard	Lamiales/Orobanchaceae	Tank 2002-04	WTU	Silica gel–dried	27 Apr. 2004	14	9, 10	122,182	23	5312 (179–36,972)	540	810	0.66	11,656	99.9844	19	0.01555
Castilleja victoriae Fairbarns & J. M. Egger	Lamiales/Orobanchaceae	Fairbarns s.n.	WTU	Silica gel–dried	21 July 2005	14	10, 14	111,371	10	11,137 (616–44,011)	688	547	0.49	18,398	99.9982	2	0.00180
Lamourouxia virgata Kunth	Lamiales/Orobanchaceae	Zak & Jaramillo, 3387	F	Herbarium	16 Jan. 1988	14	9, 10	108,767	30	3626 (214–36,850)	652	2255	2.07	11,012	99.9669	36	0.03310
Castilleja oresbia Greenm.	Lamiales/Orobanchaceae	Tank 2001-27	WTU	Silica gel–dried	19 June 2001	10	6, 9, 10, 13, 14, 16	83,384	20	4169 (222–36,830)	717	1544	1.85	9986	99.9676	27	0.03238
Castilleja arvensis Cham. & Schltdl.	Lamiales/Orobanchaceae	Tank 2005-27	WTU	Silica gel–dried	16 Apr. 2005	6	4, 6, 7, 8, 9, 10, 13, 14, 15, 16	73,378	15	4892 (186–36,621)	701	1187	1.62	9803	99.9877	9	0.01227
Penstemon montanus Greene var. *idahoensis* (D. D. Keck) Cronq.	Lamiales/Plantaginaceae	Brunsfeld 4159	ID	Herbarium	14 June 2001	16	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Balsamorhiza sagittata (Pursh) Nutt.	Asterales/Asteraceae	Willard 2013-42	ID	Silica gel–dried	3 July 2013	15	5	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Table 3.1 (continued)

Species	Order/Family	Collection no.	Herbarium	Type of tissue	Collection date	No. of amplified regions	Region no. not amplified[b]	Base pairs sequenced[c]	No. of contigs	CAL bp (min–max)	Ave. assembly depth	No. of masked bp[d]	% of masked bp	N50	% called bases[c]	No. of ambiguous bases	% of ambiguous bases
Lomatium dissectum (Nutt.) Mathias & Constance	Apiales/Apiaceae	Poor 21	ID	Herbarium	27 May 2004	15	14	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Nuphar polysepala Engelm.	Nymphaeales/Nymphaeaceae	Morales-Briones 412	ID	Silica gel–dried	8 July 2013	15	5	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Salix scouleriana Barratt ex Hook.	Malpighiales/Salicaeae	Brunsfeld 7213	ID	Herbarium	11 June 2008	15	9	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Crataegus columbiana Howell	Rosales/Rosaceae	Hetrick 1005	ID	Herbarium	10 Apr. 1996	13	9, 14, 17	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Polygonum douglasii Greene	Caryophyllales/Polygonaceae	Smith 8040	ID	Herbarium	23 June 2005	12	5, 6, 9, 15	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Umbellularia californica (Hook. & Arn.) Nutt.	Laurales/Lauraceae	Halse 6901	ID	Herbarium	28 Mar. 2002	12	6, 8, 9, 10	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Bromus tectorum L.	Poales/Poaceae	Clippinger 2	ID	Herbarium	1 May 2004	11	5, 6, 9, 11, 17	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Alnus rhombifolia Nutt.	Fagales/Betulaceae	Gray 52	ID	Herbarium	7Aug. 1989	10	5, 6, 8, 9, 10, 14	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Poa bulbosa L.	Poales/Poaceae	Willard 2013-26	ID	Silica gel–dried	3 July 2013	10	5, 6, 9, 12, 13, 14	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Senecio integerrimus Nutt. var. *exaltatus* (Nutt.) Cronq.	Asterales/Asteraceae	Willard 2013-21	ID	Silica gel–dried	3 July 2013	10	3, 5, 6, 8, 9, 11	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Abies amabilis Douglas ex J. Forbes	Pinales/Pinaceae	1419-46	WA Park Arb.	Silica gel–dried	24 May 2009	9	4, 6, 7, 9, 10, 11, 12	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Capsella bursa-pastoris (L.) Medik.	Brassicales/Brassicaceae	Brunsfeld 6313	ID	Herbarium	1 June 2005	8	4, 6, 8, 9, 10, 13, 14, 17	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Lupinus leucophyllus Douglas ex Lindl.	Fabales/Fabaceae	Willard 2013-03	ID	Silica gel–dried	3 July 2013	8	1, 6, 8, 9, 10, 12, 13, 14	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Abies fraseri (Pursh) Poir.	Pinales/Pinaceae	1005-47	WA Park Arb.	Silica gel–dried	24 May 2009	7	4, 5, 6, 7, 8, 9, 10, 11, 12	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Balsamorhiza hookeri Nutt.	Asterales/Asteraceae	Smith 9421	ID	Herbarium	4 June 2007	7	4, 5, 6, 7, 8, 9, 10, 11, 13	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Abies grandis (Douglas ex D. Don) Lindl.	Pinales/Pinaceae	1084-49	WA Park Arb.	Silica gel–dried	24 May 2009	6	1, 3, 4, 6, 7, 8, 9, 10, 11, 12	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Average								11,493	14.13	13,166.60	698.73	833.07	0.79	35,052.87	99.99	10.6	0.01

PCRs were performed using a combination of two high quality *Taq* polymerases—QIAGEN *Taq* DNA Polymerase (5 units/μL) and QIAGEN HotStar HiFidelity DNA Polymerase (2.5 units/μL) (QIAGEN, Valencia, California, USA)—to obtain amplification of fragments between 5 kb and 12 kb. The QIAGEN HotStar HiFidelity DNA Polymerase was diluted to 0.2 units/μL by combining 0.1 μL of 5x QIAGEN HotStar HiFidelity PCR buffer, 0.36 μL of double-deionized water (ddH_2O), and 0.04 μL of QIAGEN HotStar HiFidelity DNA Polymerase (2.5 units/μL). Each PCR had a total volume of 25 μL, was prepared on ice, and contained the following reagents: 2.5 μL of 10x PCR buffer (QIAGEN CoralLoad or colorless, with 15mM $MgCl_2$), 1.0 μL $MgCl_2$ (QIAGEN 25mM), 0.75 μL of deoxyribonucleotide triphosphates (dNTPs, each at 10 mM), 5.0 μL of 5x QIAGEN Q solution, 2.5 μL of both forward and reverse primers (each at 5μM), 0.25 μL (1.25 units) of QIAGEN *Taq* DNA Polymerase, 0.5 μL of the diluted QIAGEN HotStar HiFidelity DNA Polymerase solution, 9 μL of ddH_2O, and 1.0 μL of DNA template. Long PCR profiles were as follows: preheat at 93°C, initial denaturation at 93°C for 3 min followed by 35 cycles of denaturation at 93°C for 15 sec., annealing at 48-68°C (depending on the primer pair) for 30 sec., and extension at 68°C for 5-12 min (1 min/kb of target). To assess amplification, 2 μL of the final reactions were examined on a 1% agarose gel with appropriate size standards and the final products were kept at 4°C. The complete, step-by-step long PCR protocol can be found in Appendix 1.

For the three genera of Orobanchaceae in which PCR optimization was performed, amplification of the fragments was straightforward and had an average success rate of 89.7% (range = 73% – 100%). The most difficult regions to amplify were regions 2 (*trnQ*$^{(UUG)}$-*rpoC2*), 9 (*petA-psbB*), 10 (*psbB-rps3*), and 14 (*trnN*$^{(GUU)}$-*ndhA*), which are among the

largest fragments (10.3 kb, 9.8 kb, 10.9 kb, and 11.2 kb respectively; Table 3.2). It was possible to split region 2 into two smaller fragments, 2a (*trnQ*$^{(UUG)}$-*atpH*: 6.3 kb) and 2b (*atpF*-*rpoC2*: 4 kb), which facilitated its amplification in several taxa. This was not the case for regions 9, 10 and 14 for which multiple long PCR experiments using varying amounts of DNA template were necessary to obtain successful amplifications. Amplification outside of Orobanchaceae was highly variable with an average success rate of 70.8% (range = 22% – 100%) with regions 5, 6, 9, 10, and 11 showing the lowest success. Importantly, the results for these taxa were obtained after just two rounds of PCR where the annealing temperatures were changed to either 48°C or 55°C. Although we did not optimize the long PCRs for each group, we are confident that optimization on a per group basis (e.g., increasing template volume, altering annealing temperatures, and/or long PCR profiles) and/or the use of fresh tissue for DNA extractions would improve success rates. Furthermore, if genomic rearrangements and/or primer mismatches are present in certain groups, primer combinations other than the 16 that were used here could be tested (Table 3.2). Nevertheless, we successfully amplified all 16 regions in seven species, whereas in the remaining 23 species it was only possible to amplify between 6 (1 sp.) and 15 (8 spp.) regions (Table 3.1). These results translate to 21 species having at least 12 regions amplified (114.7 kb based on potential amplicon size), representing ca. 74% of the chloroplast genome when considering only one copy of the inverted repeat. Even the species with the smallest number of amplified fragments (*Castilleja arvensis* Cham. & Schltdl.) was represented by ~73 kb of data, exemplifying the effectiveness of this approach.

Table 3.2
Universal angiosperm primers used for chloroplast genome amplifications. The 16 primer combinations chosen for this study are in bold with approximate amplicon sizes in kilobases (kb) indicated.a All primers are shown in the 5′ to 3′ direction; the name of each primer consists of three parts: the gene in which the primer is anchored, the approximate position of the primer within that gene, and either an "F" or an "R." It is important to note that the F and R designations do not indicate that the primer should be used as a forward or reverse primer; rather, they indicate the 5′ to 3′ orientation of the primer with respect to the gene—i.e., a primer that is designated as an "F" primer has its 5′ to 3′ orientation in the same orientation as the gene (i.e., on the forward strand), whereas an "R" primer is oriented in the direction opposite to the 5′ to 3′ orientation of the gene (i.e., on the reverse strand).b Overlap between regions is given in number of base pairs (bp), without taking the length of the primers into consideration.

Region No.	Approx. size in kb	Primer Name	Sequence (5'-3')	Overlap between regions in bp
1	8	**trnH.GUG.6R**	CCTTRATCCACTTGGCTACAT	Regions 1 & 2= 542
1		**psbK.195R**	ACTTACAGCAGCTTGCCAAAC	Regions 1 & 2a= 542
2/2a	10.3/6.3	**trnQ.UUG.50R**	GGACGGAAGGATTCGAACC	Regions 2a & 2b= 627
2a		**atpH.17F**	CTGCYGCTTCYGTTATTGCT	Regions 2b & 3= 2059
2b	4	**atpF.65R**	CGGTATTAAACCCGAAACTCC	Regions 2 & 3= 2059
2/2b		**rpoC2.4805F**	GYCGTATYGATTGGTTRAAAGG	Regions 3 & 4= 1274
3	7	**atpI.705R**	CRGCTAAAGTTGCAAAAATAAGAGCT	Regions 4 & 5= 860
3		**rpoC1.1670F**	GRGATCAAATGGCTGTTCAT	Regions 5 & 6= 618
4	9	**rpoC2.520R**	GTTCGTACAGCAGTATCYACAAC	Regions 6 & 7= 764
4		**petN.3R**	GCCCAAGCRAGACTTACTATATCC	Regions 7 & 8= 153
5	10.5	**trnC.GCA.47F**	CCCAGTTCAAATCCGGGT	Regions 8 & 9= 1216
5		**psaB.2170F**	GCRGCTTTCTTGATTGCYTC	Regions 9 & 10= 135
6	10	**trnfM.CAU.21R**	GGTTATGAGCCTTGCGAGCTA	Regions 10 & 11= 771
6		**trnT.UGU.17F**	GGTTAGAGCATCGCATTTGTAATG	Regions 11 & 12= 2781
7	10.3	**rps4.380R**	GGTTTGCARCGATAACTTGGKATATC	Regions 12 & 13= 142
7		**rbcL.178R**	GTCCATGTACCAGTAGARGATTC	Regions 13 & 14= 392
8	9.2	**rbcL.2F**	TGTCACCACAAACAGARACTAAAG	Regions 14 & 15= 1911
8		**psbJ.3F**	GGCYGATACTACTGGAAGRAT	Regions 16 & 1= 840
9	9.8	**petA.920F**	CTTCAAGAYCCATTACGTGTHCAAG	
9		**psbB.160R**	TRCCYTGTCTCCACATTGGAT	
10	10.9	**psbB.3F**	GGGTTTRCCTTGGTATCGTGT	
10		**rps3.17F.new**	ATCCACTTGGTTTYMGACTTGG	
11	8.7	**rpl16.3R**	AACCAACGAGTCACACACTAAGC	
11/16		**ycf2.5100R**	CAGATCATGAATGTTTGGAATCCAT	
12	10	**ycf2.2300F**	TCGGGATCCTRATGCATATAGATAC	
12		**rps12.190F**	GTTGCCAGAGTACGMTTAACCT	
13	11	**rps12.360R**	CCCTTGTTGACGATCCTTTACTC	
13		**ycf1.59R**	CCGACCACAACGACCGAAT	
14/15	11.2	**trnN.GUU.7R**	CCGCTCTACCACTGAGCTAC	
14		**ndhA.535F**	GCTGCTCAATCDATTAGTTATGAA	
15	10.5	**ndhI.194R**	CGAACRCATACTTCACAAGCAA	
16	8.2	**psbA.640F**	GCTATGCATGGTTCYTTGGTAAC	

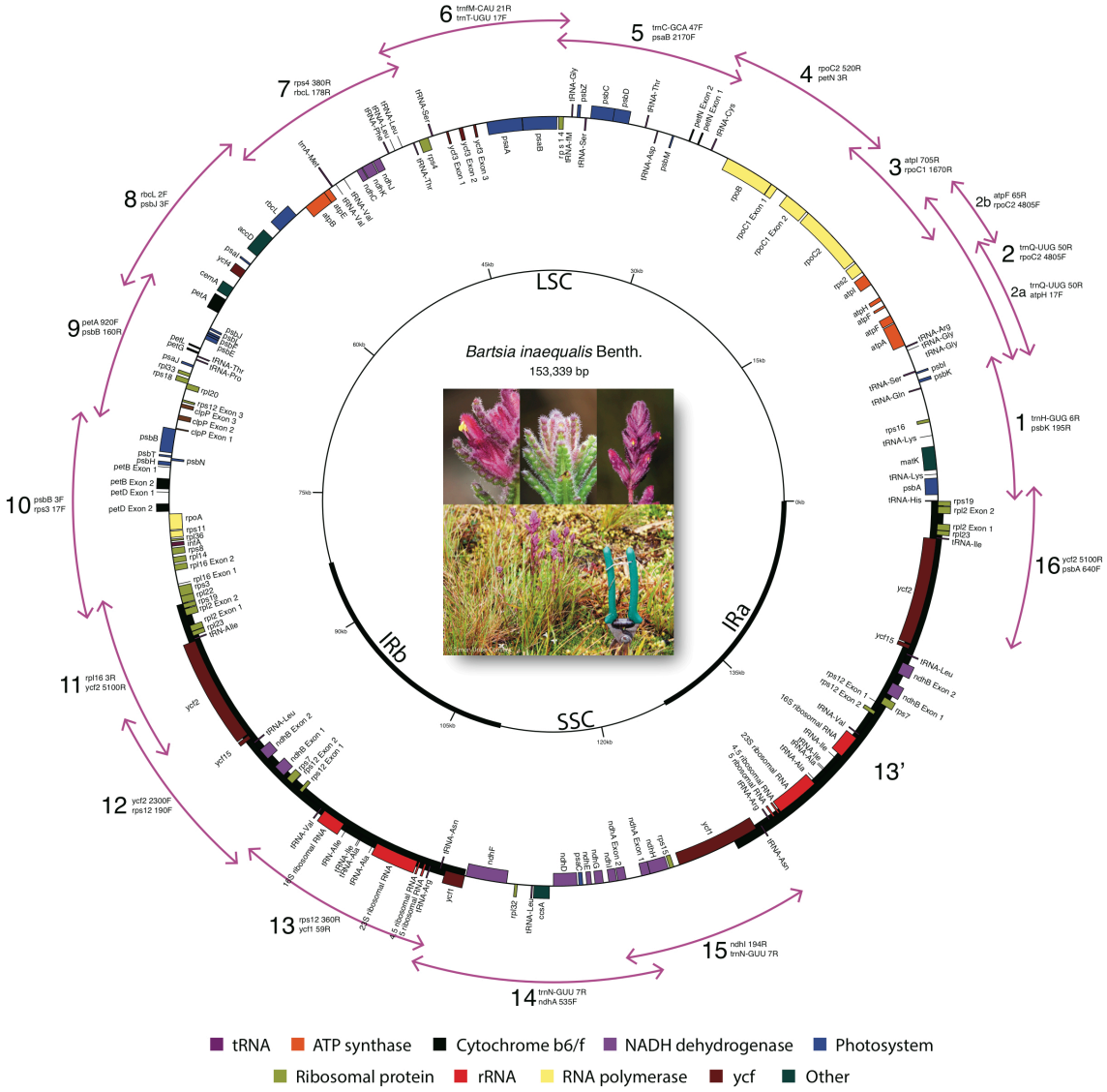
Table 3.2 (continued)

Region No.	Approx. size in kb	Primer Name	Sequence (5'-3')	Overlap between regions in bp
		rps16.50R	CGAACATCAATTGCAACGATTCGATA	
		rps16.50F	TATCGAATCGTTGCAATTGATGTTCG	
		psbK.200F	GGCAAGCTGCTGTAAGTTTTCGA	
		atpF.70F	GGGTTTAATACCGATATTTTAGCAAC	
		trnR.UCU.45F	GGTATAGGTTCAAATCCTATTGGAC	
		trnQ.UUG.47F	CGGAGGTTCGAATCCTTCC	
		trnK.UUU.3R	GAGATGGCAACTCAATCGTTG	
		trnK.UUU.3F	CAACGATTGAGTTGCCATCTC	
		atpA.430F	CGTTCYGTATATGARCCTCTTCAAAC	
		atpA.820F	ATCGMCAAATGTCTCTTCTATTAMG	
		ccsA.890R	TCCAAGTAATAAANGCCCAAGTTTC	
		trnR.ACG.15F	GAGGATTAGAGCACGTGG	
		ycf1.70F	GTGGTCGGACTCTATTATGGAT	
		trnL.UAG.18F	GGTAGACACGCTGCTCTTAGG	
		trnL.UAG.19F	GTAGACACGCTGCTCTTAGGAAG	
		rps12.320R	GGGTTCCTCGAACAATGTGATATC	
		rpl2.550F	GTGCTGTAGCGAAACTGATTG	
		rpl2.640F	TCAGCAACAGTCGGACARGT	
		psbT.3F	TGGAAGCATTGGTTTATACATTYCT	
		atpB.1290R	ARGGTTGTGATAAGAAACGYTCAA	
		trnT.UGU.42F	GATGGTCATCGGTTCGATTC	
		psbC.3R	AGTTCCATTAAAGAGCGTTTCC	
		psbD.860F	CYGGTTTATGGATGAGYGCT	
		rpoB.900R	CGTCGACCAATCYTTCCTAATTC	
		rpoB.470R	CCRGGRCTTTGCAATATTTGATTG	
		rpoC2.430R	ATRGGTAAATCAATCATTTGYCCTTG	

It is notable that many of the DNAs that were tested were extracted from herbarium tissues that ranged from 5-25 years old when isolated. In addition, we tested these primers in several species of *Abies* (Pinaceae; Table 3.1) with surprising success, amplifying between 6 and 9 regions without any PCR optimization. We caution that our long PCR protocol works best using recent DNA extractions that have not been through multiple freeze-thaw cycles. Ideally, long PCR should be conducted using new DNA extractions that are stored at 4°C

while performing experiments. Additionally, discrete PCR bands were only obtained using high quality *Taq* polymerases. When conventional polymerases were used (e.g., GoTaq, [Promega, Madison, Wisconsin, USA] or TopTaq [QIAGEN, Valencia, California, USA]), the resulting PCR products were smears rather than discrete bands and were not used for sequencing.

Figure 3.1
The final annotated chloroplast genome assembly of *Bartsia inaequalis* with the 16 overlapping primer combinations indicated. Note that the primer combinations for regions 11, 12, 13, and 16 amplify both inverted repeat A and B in a single reaction. Photos by Simon Uribe-Convers.



To confirm that our long PCR approach was compatible with NGS and that our primers would yield complete chloroplast genomes, the amplicons from each of the 15 Orobanchaceae taxa were purified by precipitation in a 20% polyethylene glycol 8000 (PEG)/2.5 M NaCl solution and washed in 70% ethanol. The amplicons were sheared by nebulization at 30 psi for 70 sec, yielding an average shear size of 500 bp as measured by a Bioanalyzer High-Sensitivity Chip (Agilent Technologies, Inc., Santa Clara, California, USA). DNA normalization is a critical step when pooling samples for multiplexing in NGS, however, due to the large number of plastomes per cell and the very few samples that were being sequenced in such a high-throughput sequencing platform, no DNA quantification was made and the sheared amplicons were pooled by species at equal volume ratios. Sequencing libraries were constructed using the Illumina TruSeq library preparation kit and protocol (Illumina Inc., San Diego, California, USA) and were standardized at 2nM prior to sequencing. Library concentrations were determined using the KAPA qPCR kit (KK4835) (Kapa Biosystems, Woburn, Massachusetts, USA) on an ABI StepOnePlus Real-Time PCR System (Life Technologies, Grand Island, New York, USA). The resulting libraries were multiplexed in one Illumina HiSeq 2000 lane—ca. 187.5 million reads per lane (Glenn 2011)—at the Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley, yielding ~12.5 million 100-bp single end reads for each taxon (GenBank Sequence Read Archive accessions: SRR1023085, SRR1023089, SRR1023095, SRR1023112, SRR1023113, SRR1023126, SRR1023128 – SRR1023136). Average depth of coverage of our sequencing experiment was ~8333x (taking 150 kb as the average plastome size). The results obtained here clearly do not maximize the potential of the Illumina HiSeq 2000 for plastome sequencing. To take full advantage of the large amount of data produced

by a HiSeq 2000 for plastome sequencing, it would be theoretically possible to sequence ~4170 samples per lane and still reach the 30x minimum threshold generally regarded as ideal for plastome sequencing (Straub et al. 2012). However, high-level multiplexing in NGS with this or any other high-throughput method requires careful normalization of DNA concentrations across samples and sufficient adapter barcodes; commonly used commercial kits currently offer either 96 (NEXTflex DNA Barcode kit, Bioo Scientific, Austin, Texas, USA) or 386 (Fluidigm, San Francisco, California, USA). Alternatively, one could choose to perform this type of experiment on a NGS platform that yielded a lesser amount of data, e.g. 1 million 250 bp paired-end reads on an Illumina MiSeq Reagent Nano Kit v2, which would allow a 30x sequencing depth for 96 samples (or 50x sequencing depth for 64 samples).

Because of the high depth of coverage of our sequencing experiment, reads were cleaned at high stringency (minimum quality = 30/40, maximum number of low-quality bases per read = 5, maximum number of duplicate reads = 10, minimum number of duplicate reads = 2), and assembled against a reference genome (*Sesamum indicum* L. GenBank accession JN637766) using the Alignreads pipeline v. 2.25 (Straub et al. 2011) with the following options: percent identity = medium, minimum coverage depth = 5, and Single Nucleotide Polymorphism (SNP) minimum coverage depth = 25 with 80% of those reads supporting the SNP. The resulting assemblies had an average depth of ~700x, an average of 0.79% bases that were masked for not reaching the minimum sequencing depth of 5x, and an average N50 of 35,053 bp (Table 3.1; Contigs and ACE files deposited in the Dryad Digital Repository: http://doi.org/10.5061/dryad.kc75n). We noticed a small drop in sequencing depth in regions immediately adjacent to some primer sites, which is a phenomenon that has been reported in the past (Njuguna et al. 2010; Whittall et al. 2010; Knaus et al. 2011; Cronn

et al. 2012). Given that our shortest overlap between amplicons is 135 bp (between regions 9 and 10; Table 3.2), with the rest spanning hundreds of bp (Table 3.2) and that our experiment yielded a high sequencing depth, we had no problems calling bases unambiguously (99.99% on average, Table 3.1). The *Bartsia inaequalis* Benth. assembly (Fig. 3.1; Genbank Accession KF922718) was annotated using DOGMA (Wyman et al. 2004) and visualized in GenomeVx (Conant and Wolfe 2008).

Conclusions

We present an alternative approach for systematic studies that combines long PCR and NGS to strategically compile phylogenomic datasets for molecular systematic studies. This approach is on par with genome skimming in terms of costs, but it has the advantage of being a targeted approach and has the potential to produce data more uniformly across samples, i.e. minimizing missing data across taxa. Although this approach was only tested with chloroplast data, we emphasize that the long PCR amplicons can be generated using DNA from any genome, expanding the possibilities of long PCR and NGS for molecular systematic studies. This last point is important for studies targeting the mitochondrion or low copy regions of the genome that otherwise might be missed or not shared across all samples using genome skimming approaches. For example, this approach may be particularly useful for the enrichment of nuclear regions, where intron sizes are large or unknown.

Literature Cited

Alexander A., Steel D., Slikas B., Hoekzema K., Carraher C., Parks M., Cronn R., Baker C.S. 2013. Low diversity in the mitogenome of sperm whales revealed by next-generation sequencing. Genome Biology and Evolution. 5:113–129.

Angiosperm Phylogeny Group. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Botanical Journal of the Linnean Society. 161:105–121.

Bombarely A., Menda N., Tecle I.Y., Buels R.M., Strickler S., Fischer-York T., Pujar A., Leto J., Gosselin J., Mueller L.A. 2011. The Sol Genomics Network (solgenomics.net): growing tomatoes using Perl. Nucleic Acids Research. 39:D1149–55.

Cheng S., Fockler C., Barnes W.M., Higuchi R. 1994. Effective amplification of long targets from cloned inserts and human genomic DNA. Proc. Natl. Acad. Sci. U.S.A. 91:5695–5699.

Conant G.C., Wolfe K.H. 2008. GenomeVx: simple web-based creation of editable circular chromosome maps. Bioinformatics. 24:861–862.

Craig D.W., Pearson J.V., Szelinger S., Sekar A., Redman M., Corneveaux J.J., Pawlowski T.L., Laub T., Nunn G., Stephan D.A., Homer N., Huentelman M.J. 2008. Identification of genetic variants using bar-coded multiplexed sequencing. Nat Methods. 5:887–893.

Cronn R., Knaus B.J., Liston A., Maughan P.J., Parks M., Syring J.V., Udall J. 2012. Targeted enrichment strategies for next-generation plant biology. American Journal of Botany. 99:291–311.

Cronn R., Liston A., Parks M., Gernandt D.S., Shen R., Mockler T. 2008. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. Nucleic Acids Research. 36:e122–e122.

Downie S.R., Palmer J.D. 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In: Soltis P.S., Soltis D.E., Doyle J.J., editors. Molecular systematics of plants. Chapman and Hall, New York, NY. p. 14–35.

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

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

Glenn T.C. 2011. Field guide to next-generation DNA sequencers. Molecular Ecology Resources. 11:759–769.

Graham S.W., Olmstead R.G. 2000. Utility of 17 chloroplast genes for inferring the

phylogeny of the basal angiosperms. American Journal of Botany. 87:1712–1730.

Harismendy O., Frazer K. 2009. Method for improving sequence coverage uniformity of targeted genomic intervals amplified by LR-PCR using Illumina GA sequencing-by-synthesis technology. Biotech. 46:229–231.

Knaus B.J., Cronn R., Liston A., Pilgrim K., Schwartz M.K. 2011. Mitochondrial genome sequences illuminate maternal lineages of conservation concern in a rare carnivore. BMC Ecol. 11:10.

Lemmon A.R., Brown J.M., Stanger-Hall K., Lemmon E.M. 2009. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. Systematic Biology. 58:130–145.

Moore M.J., Bell C.D., Soltis P.S., Soltis D.E. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proceedings of the National Academy of Sciences. 104:19363–19368.

Moore M.J., Soltis P.S., Bell C.D., Burleigh J.G., Soltis D.E. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proceedings of the National Academy of Sciences. 107:4623–4628.

Njuguna W., Liston A., Cronn R., Bassil N.V. 2010. Multiplexed Fragaria chloroplast genome sequencing. Acta Horticulturae. 859:315–321.

Parks M., Cronn R., Liston A. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol. 7:84.

Straub S.C.K., Fishbein M., Livshultz T., al E. 2011. Building a model: developing genomic resources for common milkweed (Asclepias syriaca) with low coverage genome sequencing. BMC Genomics. 12.

Straub S.C.K., Parks M., Weitemier K., Fishbein M., Cronn R.C., Liston A. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. American Journal of Botany. 99:349–364.

Stull G.W., Moore M.J., Mandala V.S., Douglas N.A., Kates H.-R., Qi X., Brockington S.F., Soltis P.S., Soltis D.E., Gitzendanner M.A. 2013. A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes. Applications in Plant Sciences. 1:1–7.

Tennessen J.A., Govindarajulu R., Liston A., Ashman T.-L. 2013. Targeted sequence capture provides insight into genome structure and genetics of male sterility in a gynodioecious diploid strawberry, Fragaria vesca ssp. bracteata (Rosaceae). G3; Genes|Genomes|Genetics. 3:1341–1351.

Whittall J.B., Sying J., Parks M., Buenrostro J., Dick C., Liston A., Cronn R. 2010. Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and

widespread pines. Molecular Ecology. 19 Suppl 1:100–114.

Wu F., Mueller L.A., Crouzillat D., Pétiard V., Tanksley S.D. 2006. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. Genetics. 174:1407–1420.

Wyman S.K., Jansen R.K., Boore J.L. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics. 20:3252–3255.

Xi Z., Ruhfel B.R., Schaefer H., Amorim A.M., Sugumaran M., Wurdack K.J., Endress P.K., Matthews M.L., Stevens P.F., Mathews S., Davis C.C. 2012. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. Proceedings of the National Academy of Sciences. 109:17519–17524.

Yuan Y., Liu C., Marx H., Olmstead R. 2009. The pentatricopeptide repeat (PPR) gene family, a tremendous resource for plant phylogenetic studies. New Phytologist. 182:272–283.

Chapter 4: A targeted subgenomic approach for phylogenomics based on microfluidic PCR and high throughput sequencing

Coauthor: Matt L. Settles

Abstract

Advances in high-throughput sequencing (HTS) have allowed researchers to obtain large amounts of genomic information at speeds and costs unimaginable a decade ago, and it is clear that regardless of the HTS method used to enter the realm of big data, the field of phylogenetics, and the study of evolution in general, is quickly migrating towards larger and larger molecular datasets. We present a method for generating large amounts of genomic data that is based on microfluidic PCR and HTS. This approach uses a microfluidic PCR array and two sets of PCR primers to amplify 48 targeted gene regions and incorporate sample-specific barcodes and HTS adapters to 48 samples simultaneously (2,304 amplicons per microfluidic array). The amplification is done in a way that the amplicons are ready to be sequenced, and thus, there is no need to construct genomic libraries for each sample. Moreover, we present a bioinformatic pipeline to process the raw reads in a a way that the user can decide to either generate consensus sequences (with or without ambiguities) for every locus in every sample or – more importantly – recover the alleles from heterozygous gene regions in each sample. This is important not only because it adds allelic information that is well suited for coalescent-based phylogenetic analyses, and the detection of allopolyploids, but also because it allows for the estimation of minimum ploidy levels. To test our genomic method and our processing pipeline, we sequenced 576 samples belonging to the South American clade of the genus *Bartsia* in the plant family Orobanchaceae.

Introduction

Advances in high-throughput sequencing (HTS) have allowed researchers to obtain large amounts of genomic information at speeds and costs unimaginable a decade ago. The fields of phylogenetics and population genetics have benefitted greatly from these advancements, and large phylogenomic and population genomic datasets are becoming more and more common (sensu Molau 1990; Lemmon et al. 2012; Faircloth et al. 2012; Stull et al. 2013). Driven by the need to generate homogenous, informative, and affordable multilocus datasets, we present a new approach for obtaining large, multilocus datasets for phylogenetic and population genetic studies, based on microfluidic PCR amplification and HTS. Microfluidic PCR technology has been used extensively in the fields of cancer research (e.g., Walter et al. 2012; Gaedcke et al. 2012), genotyping of single nucleotide polymorphism (SNP) (e.g., Byers et al. 2012; Bhat et al. 2012; Lu et al. 2012), gene expression (e.g., Moignard et al. 2013; Shalek et al. 2013; Dominguez et al. 2013), and targeted resequencing (e.g., Lohr et al. 2012; Moonsamy et al. 2013) but, to our knowledge, it has not yet been used to assemble molecular phylogenetic datasets for systematic studies (but see Godden et al. 2012 for a discussion of its potential use for phylogenetics). This approach uses the Fluidigm Access Array System (Fluidigm, San Francisco, CA, USA) and two sets of PCR primers to amplify 48 targeted gene regions, and to incorporate sample-specific barcodes and HTS adapters to 48 samples simultaneously (2,304 amplicons per microfluidic array). By conducting a four-primer reaction, this approach circumvents the necessity to construct genomic libraries for every sample, avoiding the high costs and time requirements involved during this step. Using a dual barcoding strategy, we multiplexed 24 microfluidic arrays, representing two distinct sets of 48 gene regions for 576 samples from the South American

clade of the plant genus *Bartsia* L. (Fig. 4.1) and its close relatives in Orobanchaceae, and demonstrate the power of this approach for species-level phylogenetics.

Figure 4.1
Floral diversity in the South American *Bartsia* clade. The sections from top left to bottom right are: *Strictae*, *Diffusae*, *Laxae*, and *Orthocarpiflorae*. Photos by Simon Uribe-Convers.

In plant phylogenomics, there has been a special focus on the chloroplast genome, also know as the plastome, given its phylogenetic informativeness at all taxonomic scales (e.g., Graham and Olmstead 2000; Moore et al. 2007; Parks et al. 2009; Moore et al. 2010), the straightforwardly interpreted results due to its non-recombining nature, conserved gene order, and gene content (Downie and Palmer 1992), and its historical importance since the beginning of the field (e.g., Chase et al. 1993). Large datasets have been produced with approaches that have involved massively parallel sequencing (e.g., Parks et al. 2009), a compilation of coding regions from both whole plastome sequences (e.g., Moore et al. 2010) and targeted approaches (e.g., Moore et al. 2011), transcriptomics (e.g., Xi et al. 2012), RNA hybridization or capture probes (e.g., Stull et al. 2013), and long range PCR coupled with HTS (e.g., Uribe-Convers et al. 2014). Because the chloroplast genome evolves relatively slowly, ~3–5 times slower than the nuclear genome in plants (Wolfe et al. 1987; 1989; Drouin et al. 2008), the power of these datasets for phylogenetic studies lies in their size; at ~150 kilobases (kb), plastome datasets can provide phylogenetic resolution from the interspecific level (e.g., Parks et al. 2009; Njuguna et al. 2013) to the level of major clades (e.g., Moore et al. 2010; Ruhfel et al. 2014). However, because it is inherited as a single unit, plastome sequences only provide information from a single locus, and although often well-supported, phylogenies based solely on plastome-scale datasets may be misleading because of the well known problem of gene tree-species tree discordance (Maddison 1997). This may be especially problematic at low-taxonomic scales where processes such as coalescent stochasticity and gene flow may be more prevalent. Thus, data from multiple independently evolving loci is necessary to fully understand the evolutionary history of a group of organisms, and to take full advantage of the emerging species tree paradigm made possible

by the integration of population genetic processes into phylogenetic reconstruction via the multispecies coalescent (Rannala and Yang 2003; Edwards 2009; Knowles and Kubatko 2010).

For the nuclear genome, phylogenomic datasets have been obtained in plant systems using genome skimming (e.g., Straub et al. 2012), sequence capture (e.g., Tennessen et al. 2013; Mandel et al. 2014), and restriction-site associated DNA—RadSeq, or generically know as genotyping by sequencing [GBS]) (e.g., Eaton and Ree 2013). Likewise, the field of phylogenomics in animals has advanced with datasets obtained with targeted amplicon sequencing (TAS) in Pancrustacea (Bybee et al. 2011) and North American tiger salamanders (O'Neill et al. 2013), GBS in butterflies (Nadeau et al. 2012), fish (Jones et al. 2013), and beetles (Cruaud et al. 2014). However, genome-scale datasets for animal phylogenetics has been most heavily influenced by sequence capture approaches focused on ultraconserved genomic elements (UCEs) at various taxonomic scales, e.g., vertebrates (Lemmon et al. 2012), amniotes (McCormack et al. 2012; Faircloth et al. 2012), turtles (Crawford et al. 2012), birds (McCormack et al. 2013), ray-finned fishes (Faircloth et al. 2013), and chipmunks (Bi et al. 2012). Furthermore, UCEs have been shown to be an important resource for gathering information from museum specimens (Bi et al. 2013), and to be useful at shallow evolutionary time scales in birds (Smith et al. 2014).

In plant systems, genome skimming (Straub et al. 2012) has perhaps had the most impact for assembling phylogenetic datasets from HTS data. In contrast to sequence capture approaches that require preliminary genomic data for bait design, genome skimming requires no existing genomic information. Genome skimming is a reference-guided approach that takes advantage of the high copy numbers of some regions in the genome, e.g., nuclear

rDNA, the plastome, and the mitochondrial genome. By using reference sequences, this method 'skims' out targeted regions from low-coverage genomic data. This approach has been used to recover the mitochondrial and chloroplast genomes in the genus *Asclepias* L. (Straub et al. 2012), study introgression in *Fragaria* L. species (Salamone et al. 2013), identify horizontal transfer of DNA from the mitochondrion to the chloroplast in *Asclepias syriaca* L. (Straub et al. 2013), resolve phylogenetic relationships in the Chrysobalanaceae family (Malé et al. 2014), recover plastomes across multiple genera (Uribe-Convers et al. 2014), and to assemble the plastomes used for microfluidic PCR primer design in this study.

Both the UCE sequence capture and genome skimming approaches share similar technical and fundamental constraints that make their utility for phylogenetics at low-taxonomic levels with large sampling strategies limited. First, both of these methods are limited by the necessity to construct genomic libraries for each sample in the study, a step that greatly increases the time and costs of the experiment. Even when libraries are constructed in-house, the time and costs to construct hundreds of them increases rapidly. Second, variable regions flanking the UCEs are often captured at much reduced depth as one moves away from the UCE, or they are lost completely if the target taxon is phylogenetically divergent from the one used in the bait design (Stull et al. 2013; Smith et al. 2014). Smith et al. (2014) found that UCEs containing variable flanking regions were usually not recovered across all samples if the variable regions extended more than 300 bp from the UCE probe. This is unfortunate, given that the more variable regions are of potentially greater use for interspecific phylogenetic and population genetic studies. Likewise, genome skimming from low-coverage genomic data is most useful for recovering high-copy number regions in the genome; however, regions with lower representation numbers, such as single copy nuclear

genes, are likely to be recovered in some samples and missed completely in others (Straub et al. 2011), depending on the depth of the low-coverage genomic data and the phylogenetic distance of the references used for mapping. Both of these cases result in the introduction of missing data, which could potentially lead to incorrect or misleading phylogenetic inferences (Lemmon et al. 2009). In contrast, the large scale targeting of chloroplast, nuclear rDNA, and multiple independent single-copy nuclear genes using microfluidic PCR arrays and HTS circumvents many of these problems.

Our approach is similar in theory to targeted amplicon sequencing (TAS) methods (e.g., Bybee et al. 2011; O'Neill et al. 2013), but contains major improvements in efficiency. For example, Bybee et al. (2011) implemented a first round of PCRs to amply each target region, which was then reamplified in a second round of PCRs to incorporate barcodes and HTS adapters. While using this two-reaction approach allows for more flexibility in the annealing temperature of the target specific primers, this approach is labor intensive and thus difficult to scale to hundreds of samples and/or a large number of targets to take full advantage of the current yield of most HTS platforms. In their study, Bybee et al. (2011) amplified six genes for 44 taxa from Pancrustacea, which translates to performing 12 PCRs for each of the 44 taxa to amplify and tag each amplicon. At this scale, both in terms of the number of samples and the number of loci, this method may be more favorable than the approach proposed here, however, once 48 or 96 different primer pairs are used to amplify hundreds of samples, this method becomes inefficient. We believe that experiments with high numbers of samples and loci are quickly becoming more common, and that the fields of systematics, phylogenetics, and population biology need the tools to deal with this type of sampling.

Here we test the performance and utility of our targeted, subgenomic approach on the Neotropical clade of the plant genus *Bartsia* L. (Orobanchaceae) (Fig. 4.1). This clade is comprised of ~45 closely related species that are part of an ongoing rapid and recent radiation in the páramo ecosystem above tree line (~2900 m in elevation) throughout the Andes (Chapter 1). Using minimal genomic resources collected via plastome sequencing (Uribe-Convers et al. 2014) and low-coverage genome sequencing in representative species of *Bartsia*, we present an approach for designing microfluidic PCR primer combinations for amplifying i) the most variable regions of the plastome (referred to as the chloroplast set henceforth), ii) the commonly sequenced ITS and ETS regions of the nuclear rDNA repeat, and iii) a suite of putatively single-copy nuclear loci (ii and iii are referred to as the nuclear set henceforth). Our targeted subgenomic approach generated a large multilocus dataset, which allowed us to investigate evolutionary relationships at the species level. While whole genome approaches (e.g., whole plastome sequencing) yield more data, the great majority of these data are highly conserved across samples and thus phylogenetically uninformative. By focusing on targeted loci and not whole genomes, we were able to maximize the yield of shared and phylogenetically informative data, which is ideal for phylogenetic studies at low taxonomic levels.

Methods

Preliminary data acquisition

Initial data used for primer design was generated with an approach that uses long PCR to generate DNA templates for HTS (Uribe-Convers et al. 2014) and via genome skimming (Straub et al. 2012). Two taxa were used from the plastome data generated by

Uribe-Convers et al. (2014) as well as newly generated low-coverage genome data for three species (4 samples, Table 4.1). For the newly generated data, DNA was extracted from ~0.02 g of silica gel-dried tissue using a modified 2X CTAB method (Doyle and Doyle 1987), yielding 30 to 70 ng/μL of DNA per sample. Genomic DNAs were sheared by nebulization at 30 psi for 70 sec, yielding an average shear size of 500 bp as measured by a Bioanalyzer High-Sensitivity Chip (Agilent Technologies, Inc., Santa Clara, California, USA). Sequencing libraries were constructed using the Illumina TruSeq library preparation kit and protocol (Illumina Inc., San Diego, California, USA) and were standardized at 2nM prior to sequencing. Library concentrations were determined using the KAPA qPCR kit (KK4835) (Kapa Biosystems, Woburn, Massachusetts, USA) on an ABI StepOnePlus Real-Time PCR System (Life Technologies, Grand Island, New York, USA). One of the resulting libraries was sequenced on an Ilumina HiSeq 2000 at the Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley, whereas the other three libraries were sequenced on an Illumina HiSeq 2000 at the Center for Genome Research and Biocomputing at Oregon State University (Table 4.1). The raw reads were cleaned using Seqyclean v 1.8.10 (Zhbannikov et al. in prep.) using defaults settings.

Table 4.1
Sequencing information of the samples used during the preliminary data acquisition step. Type of read refers to the length of the read in base pairs (bp), and if it was single or paired end. Number of reads denotes the number of raw reads in millions. Berkeley = Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley; OSU = Center for Genome Research and Biocomputing at Oregon State University.

Species	Collector	Platform	Type of read (bp)	No. clean reads	Sequencing Facility	Source
Bartsia inaequalis Benth.	Uribe-Convers 2010-022	Illumina HiSeq 2000	100 single end	0.93	Berkeley	Uribe-Convers et al. 2014
Bartsia stricta (Kunth) Benth.	Uribe-Convers 2010-024	Illumina HiSeq 2000	100 single end	0.9	Berkeley	Uribe-Convers et al. 2014
Bartsia pedicularoides Benth.	Antonelli 574	Illumina HiSeq 2000	100 single end	46.8	Berkeley	This study
Bartsia santolinifolia (Kunth) Benth.	Uribe-Convers 2010-41	Illumina HiSeq 2000	100 paired end	46.4	OSU	This study
Bartsia pedicularoides Benth.	Uribe-Convers 2011-64	Illumina HiSeq 2000	100 paired end	52.9	OSU	This study
Bartsia serrata Molau	Uribe-Convers 2012-15	Illumina HiSeq 2000	100 paired end	65.1	OSU	This study

Primer design and validation

Chloroplast set

For the chloroplast set, the cleaned reads were assembled against a reference genome (*Sesamum indicum* L., GenBank accession JN637766) using the Alignreads pipeline v. 2.25 (Straub et al. 2011), and visually inspected using Geneious R6 v6.1.5 (Biomatters, Auckland, New Zealand). Six complete plastomes—two from the long PCR approach and four from the low-coverage genome data—were aligned using MAFFT v.7.017b in its default settings (Katoh and Standley 2013). To target the putatively most phylogenetically informative regions of the plastome, using custom R-scripts (R Development Core Team 2013) we identified the most variable regions in the alignment that were flanked by conserved regions, and that spanned between 400 bp and 1000 bp. This allowed us to rank and prioritize regions in the plastome for primer design.

Nuclear Set

For the nuclear set, the cleaned reads were compared to two publicly available genomic databases, i.e. a list of the pentatricopeptide repeat genes (PPR) and the conserved orthologous set II (COSII), using the BLAST-Like Alignment Tool (BLAT) (Kent 2002) with the options 'tileSize' and 'minIdentity' set to 7 and 80 respectively. We chose these two reference databases because a list of 127 PPR loci was shown to have a single ortholog in both rice (*Oryza sativa* L.) and *Arabidopsis thaliana* (L.) Heynh. (Yuan et al. 2009), and was used successfully to infer the phylogenetic relationships of the plant family Verbenaceae and the *Verbena* L. complex (Yuan et al. 2010). Similarly, the COSII have been identified to be putatively single-copy and orthologous across the Euasterid plant clade (Wu et al. 2006), and several loci have been used for phylogenetic reconstructions of closely related species in the plant family Orobanchaceae (Li et al. 2008) and Solanaceae (Tepe et al. 2011).

Using a custom R-script, the reads that matched any of the references genes from these two databases were kept, binned with their respective reference locus, and aligned using MAFFT in its default settings. Furthermore, we used the online tool IntronFinder (http://solgenomics.net/, last accessed in January 2014) from the Sol Genomics Network (Bombarely et al. 2011) to predict intron position in the COSII genes. The PPR genes do not contain introns (Yuan et al. 2009) and thus this step was not necessary for this database. Reference loci that had at least two groups of reads aligned to them forming conserved 'islands' separated by 400-800 bp, including estimated introns with an assumed average length of 100 bp, were selected for primer design (Fig. 4.2a). Additionally, an alignment of nuclear rDNA internal and external transcribed spacers sequences—ITS and ETS,

respectively (Chapter 1)—as well as an alignment of sequences of the *PHOT1* gene and one of the *PHOT2* gene (Yuan and Olmstead 2008) were made in MAFFT with default settings.

Primer design

Forward and reverse primers for the selected chloroplast regions and nuclear loci were designed using Primer3 (Rozen and Skaletsky 2000; Koressaar and Remm 2007; Untergasser et al. 2012) following the recommended criteria specified in the Access Array System protocol (Fluidigm, San Francisco, CA, USA), e.g., annealing temperature was set to 60°C (+/- 1°C) for all primers, and no more than three continuous nucleotides of the same base were allowed (Max Poly-X=3). Furthermore, regions identified as appropriate for primer design that were not present in every sample in the alignment or that contained ambiguous bases (due to missing data and/or low coverage in our assemblies) were discarded. A complete list of the chosen primers can be found in Appendix 2. Once the initial primer design was completed, a conserved sequence (CS) tail was added to the 5' end of both the forward and reverse primers, CS1 and CS2 respectively (Fluidigm), resulting in the final target specific primers (TS) with universal tails (CS1-TS-F and CS2-TS-R, respectively). The purpose of the added tails (CS1 and CS2) is to provide an annealing site for the second pair of primers, which, starting from the 5' end, are composed of the HTS adapters (e.g., PE1 or PE2 for Illumina sequencing), a sample specific forward and reverse barcode combination (e.g., BC1 and BC2), and the complementary CS sequence (CS1' or CS2'; Fig. 4.2b and 4.2c). To avoid confusion, the first pair of primers with universal tails (CS1-TS-F and CS2-TS-R) will be referred to as the 'target specific primers', whereas the second pair of primers—with complementary universal tails, barcodes, and HTS adapters; PE1-BC-CS1'

and PE2-BC-CS2'—will be referred to as the 'barcoded primers'. The CS1 and CS2 sequences were obtained from the Fluidigm Access Array System protocol, whereas the barcoded primers were custom designed to allow for dual barcoding, which dramatically increases the number of samples that can be multiplexed in one sequencing experiment (Appendix 3).

Primer Validation

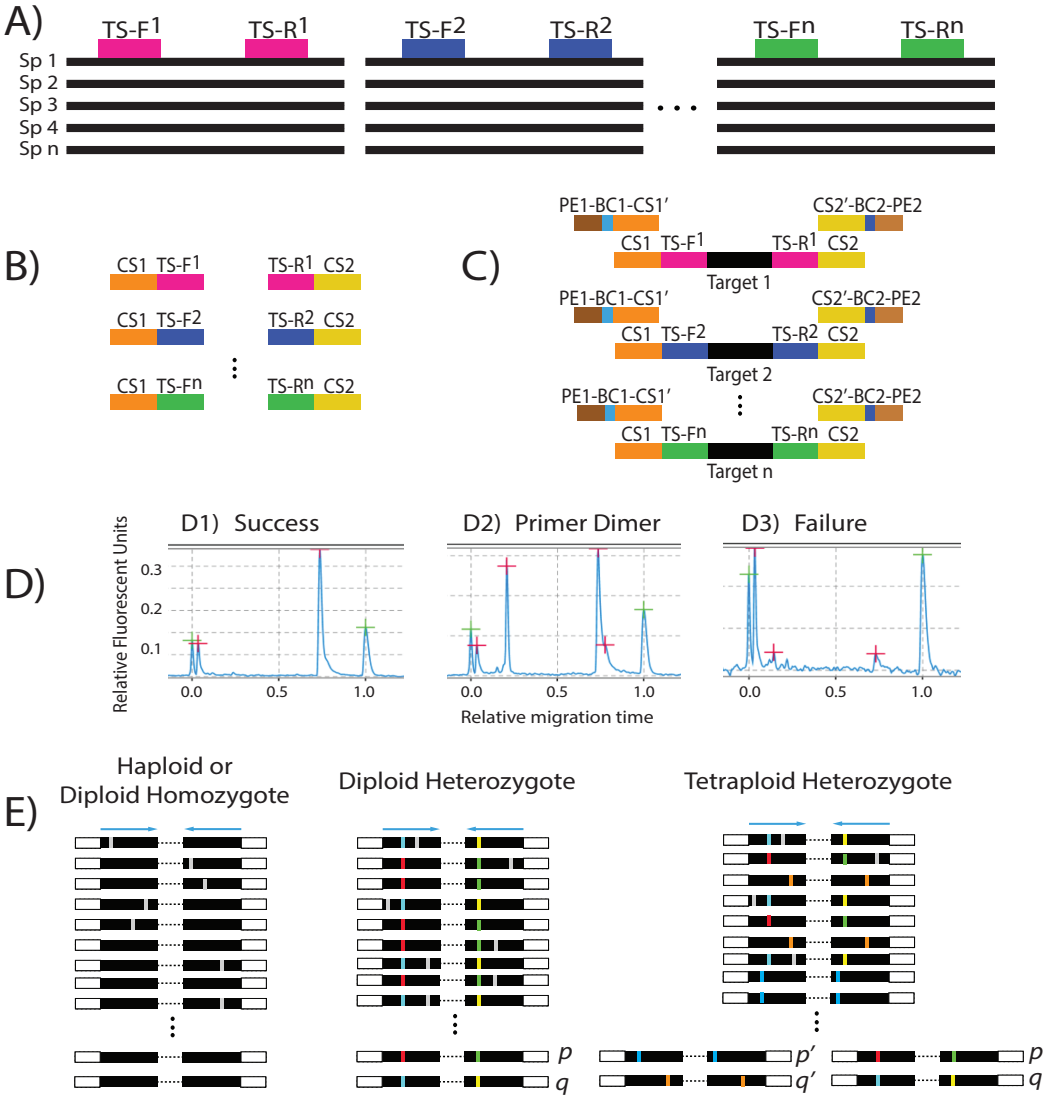
Due to the complexity of simultaneously using two sets of primers in one PCR, it is necessary to validate each set of primers prior to the actual microfluidic PCR amplification. Primer validation is a crucial step to ensure that no primer dimers are formed and that no interaction and/or competition between the barcoded and target specific primer pairs are negatively affecting the amplification. Primer validation was performed for each primer combination in 10 µL reactions in an Eppendorf Mastercycler ep thermocycler, following the Fluidigm Access Array System protocol. Validation reactions were performed on three species of *Bartsia* (*B. mutica* (Kunth) Benth., *B. crisafullii* N. H. Holmgren, and *B. melampyroides* (Kunth) Benth.), which represent the morphological and geographical diversity in the genus, and a negative control (using water instead of DNA), and included the following: 1µL of 10X FastStart High Fidelity Reaction Buffer without $MgCl_2$ (Roche Diagnostic Corp., Indianapolis, Indiana, USA), 1.8 µL of 25 mM $MgCl_2$ (Roche), 0.5 µL DMSO (Roche), 0.2 µL 10mM PCR Grade Nucleotide Mix (Roche), 0.1 µL of 5 U/µL FastStart High Fidelity Enzyme Blend (Roche), 0.5 µL of 20X Access Array Loading Reagent (Fluidigm), 2µL of 2 µM barcoded primers, 2µL of 50nM target specific primers, 0.5 µL of 30-70 ng/µL genomic DNA, 1.4 µL of PCR Certified Water (Teknova, Hollister,

California, USA). Amplicons from these reactions were analyzed in a QIAxcel Advance System (Qiagen, Valencia, California, USA), and primer pairs that produced a single amplicon and had no (or minimal) primer dimers were selected (Fig. 4.2d, Appendix 2).

Sampling, microfluidic PCR and sequencing

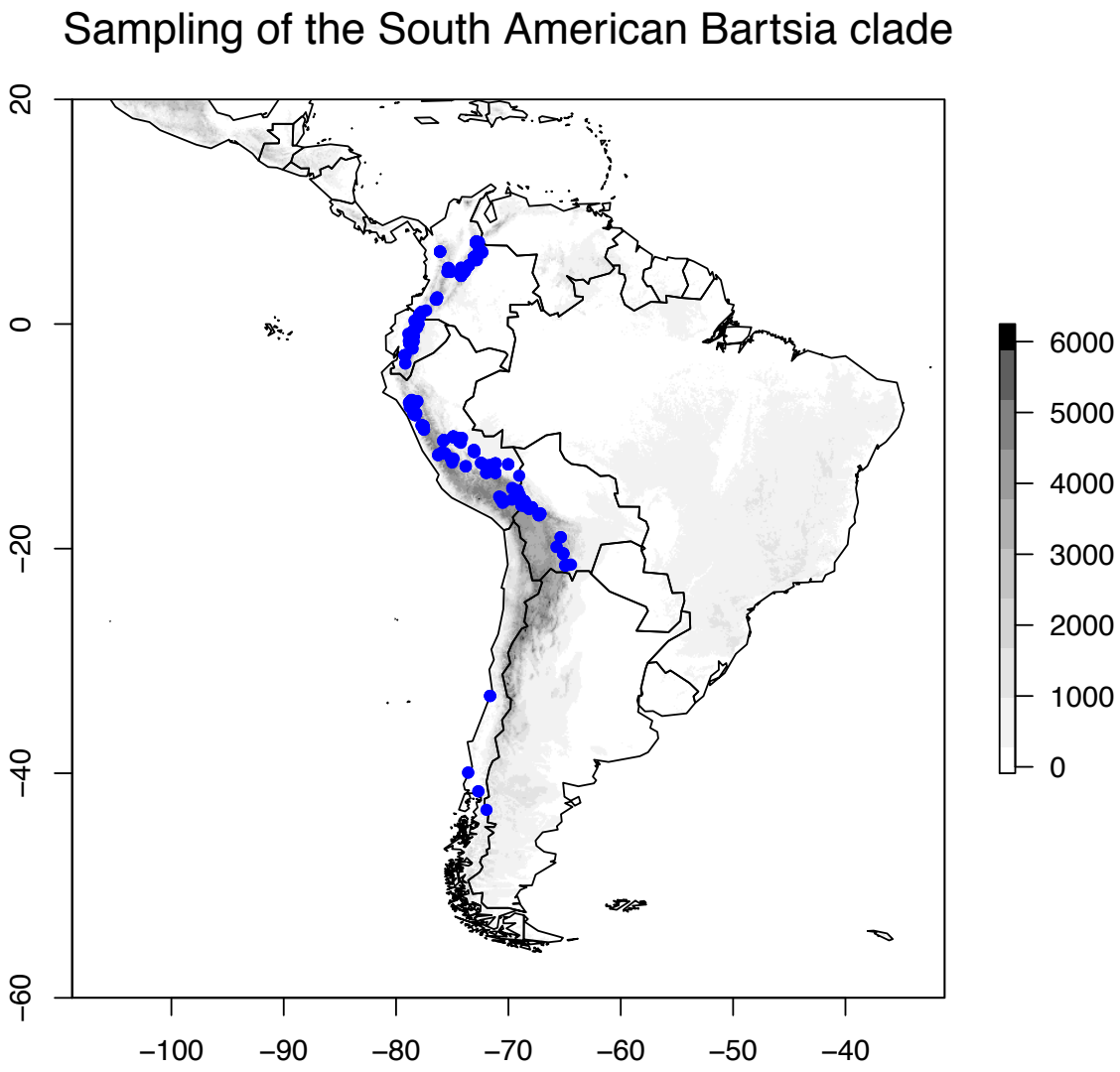
We were interested in generating data to investigate the evolutionary history of the Neotropical *Bartsia* clade (Chapter 1), and thus, we sampled the complete species richness of the group, including multiple individuals per species, and some of its close relatives. A total of 74 species were represented in 576 samples. These samples encompassed the entire geographic breadth of the South American clade, with samples ranging from northern Colombian to southern Chile (Appendix 4, Fig. 4.3). The majority of samples were collected in the field, dried in silica-gel desiccant, and stored in airtight bags. When field-collected tissue was not available, leaf tissue was sampled from herbarium specimens (Appendix 4). For all samples, DNA was extracted from ~0.02 g of silica gel-dried tissue using a modified 2X CTAB method (Doyle and Doyle 1987), yielding ~30 to 70 ng/µl of DNA per sample.

Figure 4.2
Flowchart describing the method used in this study. A) Forward and reverse target specific primer combinations (TS-F and TS-R, respectively) designed in Primer3 from a multiple sequence alignment of existing genomic resources obtained in the preliminary data acquisition step. B) The conserved sequences (CS1 and CS2) are added to the target specific primers at the time of synthesis. C) Each target specific combination needs to be validated to ensure amplification. This is done by simulating the microfluidic amplification step in a standard thermocycler, with both the first and the second pair of primers added (4-primer reaction). The second pair of primers is comprised of the sequencing adapters (e.g., PE1 and PE2, for Illumina), a barcode combination (BC) specific for each sample, and the reverse complement of the conserved sequences (CS1' and CS2'). D) Each reaction is analyzed for successful amplifications (D1), primer dimers (D2), or failed amplifications (D3). Only primer combinations with successful amplification and no primer dimers are chosen. E) After sequencing, the reads are demultiplexed, sample-specific pools of amplicon sequences are generated, and groups of identical reads are identified in each pool. Pools of identical sequence reads represented by at least 5 reads and representing at least 5% of the total reads for that amplicon/sample are kept as alleles. Three examples are shown: a haploid or diploid homozygote sample with just one sequence, a diploid heterozygote sample with two different sequences (p & q), and a tetraploid heterozygote sample with two sets of homeologs (p & q and p' & q').



Microfluidic PCRs were performed in an Access Array System (Fluidigm) using 24 (12 for the chloroplast and 12 for the nuclear set) of the 48.48 Access Array Integrated Fluidic Circuits (Fluidigm) following the manufacturer's protocols. This particular array allows for 48 samples to be simultaneously amplified with 48 distinct primer pairs, resulting in 2,304 PCRs per array. While we chose here to separate our 48 chloroplast and 48 nuclear targets in separate microfluidic arrays, we have also had success with multiplexing genomically divergent regions such as these and performing amplifications of all 96 primer pairs in a single array (i.e., primers for one chloroplast region and one nuclear locus pooled prior to amplification; D. F. Morales-Briones and D. C. Tank, unpubl. data). The amplicons were harvested from each array as per the Fluidigm Access Array System protocol and pooled per sample in equal volumes. To remove unused reagents and/or undetected primer dimers smaller than 350 bp, each pool was purified with 0.6X AMPure XP beads (AgencourtT, Beverly, Massachusetts, USA). The purified pools were analyzed in a Bioanalyzer High-Sensitivity Chip (Agilent Technologies, Santa Clara, California, USA) and standardized at 13 pM using the KAPA qPCR kit (KK4835; Kapa Biosystems, Woburn, Massachusetts, USA) on an ABI StepOnePlus Real-Time PCR System (Life Technologies, Grand Island, New York, USA). The resulting pools were multiplexed in an Illumina MiSeq using the Reagent Kit version 3 with a yield of 21.4 million 300 bp paired-end reads. Microfluidic PCR, downstream quality control and assurance, and Illumina sequencing was performed in the University of Idaho Institute for Bioinformatics and Evolutionary Studies Genomics Resources Core facility.

Figure 4.3
Sampling effort in the South American *Bartsia* clade. A total of 576 samples, represented by dots, were collected for this study. Samples of *Bellardia trixago* L., *Bellardia viscosa* (L.) Fisch. & C.A. Mey, and *Parentucellia latifolia* (L.) Caruel not shown. The Y and X axis represent latitude and longitude, respectively, and the gray-scale to the right denotes elevation in meters.



Data processing

The reads from the Illumina MiSeq run were bioinformatically separated (demultimplexed) by sample using the sample-specific dual barcode combinations, and then further by chloroplast region or nuclear locus using the target specific primers using a custom preprocessing pipeline, available at https://github.com/msettles. This pipeline allows for an additional and optional trimming step, where the amount of trimming can be set independently for both the forward and reverse reads. This step is especially useful for the more error-prone reverse reads or in instances of high sequencing errors rates, which most often occurs at the end of reads. Our sequencing run falls into the latter category, and thus, we trimmed 75 bp and 150 bp of our forward and reverse reads, respectively.

To maximize the number of amplicons recovered for each sample and each DNA region, the preprocessing pipeline allows for target specific primer matching errors less than or equal to four, as long as the last four bases of the 3' end of the target specific primers were successfully identified, thus yielding firm ends. Additionally, sister reads that overlap by at least 10 bp are joined into a continuous sequence. Finally, for every sample in every locus, our pipeline has the ability to either produce consensus sequences consensus with IUPAC ambiguity codes for individual sites represented by more than one base in more than 5% of the reads, or – more importantly – to recover the alleles present in each sample. Allele recovery is accomplished by creating groups of sister reads (or joined reads) that are identical in each sample for every locus. The number of reads conforming each group are then counted and if they represent at least five percent of the total number of reads for that sample in that amplicon, and there is a minimum sequencing depth of 5x (Fig. 4.2e) the group of reads is retained. Conversely, if both the sequencing depth and minimum total percentage criteria are

not met the group is discarded. Our allele-recovering method is based on the assumption that sequencing errors are mostly random and that reads containing errors will be represented at much lower frequencies. An additional benefit of recovering the alleles and not just the consensus sequence, is that it is possible to assess the minimum ploidy level of each sample, which is especially of interest in plant systems where polyploidy is very common.

Because the chloroplast genome is haploid, recovering alleles made no difference and consensus sequences instead were generated for each of the 48 regions. In cases where the forward and reversed consensus sequences did not overlap, the reads were concatenated into a single continuous sequence. A multiple sequence alignment was performed for each of these regions using MUSCLE v3.8.31 (Edgar 2004) in its default settings. The alignments were cleaned with Phyutility v2.2.4 (Smith and Dunn 2008) at a 50 percent threshold to minimize missing data due to ambiguous alignment sites, visually inspected in Geneious R6 v6.1.5 (Biomatters), and any misaligned or ambiguous sequences discarded. Finally, the 48 chloroplast alignments were concatenated with Phyutility into a single locus.

For the nuclear set, we generate consensus sequences for every sample in every locus as well as recovered the alleles for each sample. Similarly to the chloroplast set, sister reads that did not overlap were joined into a continuous sequence, and each region was independently aligned using MUSCLE and cleaned with Phyutility at a 50 percent threshold. For ITS and ETS regions, because these two regions are physically linked on the nuclear ribosomal repeat, the alignments were concatenated into a single locus using Phyutility. Our data processing resulted in a dataset containing 47 independent nuclear loci with allelic information and one large chloroplast locus that were used in downstream phylogenetic analyses. Finally, to compare alternative ways to analyze our data, the consensus sequences

(with IUPAC ambiguity codes) of the 47 nuclear loci were concatenated into a single locus (~13,500 bp). Furthermore, the consensus sequences of the nuclear loci and the chloroplast locus were concatenated into a large dataset with over 40,500 bp after cleaning.

Phylogenetic analyses

The concatenated chloroplast dataset was analyzed with PartitionFinder (Lanfear et al. 2012) to find the best partition scheme while also selecting for the best-fit model of sequence evolution for each possible partition. Using these partition schemes and models of sequence evolution, we conducted maximum likelihood (ML) analyses as implemented in GARLI v2.0.1019 (Zwickl 2006) with five independent runs, each with 200 nonparametric bootstrap replicates. Bootstrap support was assessed with the program SumTrees v3.3.1 of the DendroPy v3.12.0 package (Sukumaran and Holder 2010). Likewise, we analyzed the dataset in a Bayesian framework as implemented in MrBayes v3.2.1 (Ronquist et al. 2012) with the individual parameters unlinked across the data partitions. We ran two independent runs with four Markov chains each using default priors and heating values. Independent runs were started from a randomly generated tree and were sampled every 1000 generations. Convergence of the chains was determined by analyzing the plots of all parameters and the –lnL using Tracer v.1.5 (Rambaut and Drummond 2004). Stationarity was assumed when all parameters values and the –lnL had stabilized; the likelihoods of independent runs were considered indistinguishable when the average standard deviation of split frequencies was < 0.001. A consensus trees was obtained using the sumt command in MrBayes.

The nuclear dataset was analyzed in multiple ways. First, we inferred individual gene trees for each locus using RAxML v.8.0.3 (Stamatakis 2014) to ensure that the each locus

was indeed single copy. Second, we analyzed the nuclear concatenated dataset with RAxML with no topological restrictions. Third, a second analysis of the nuclear concatenated dataset with RAxML, but this time constraining the topology to make every species monophyletic (Linnen and Farrell 2008). Although not a formal coalescent-based species tree method, comparisons of this approach to coalescent-based species tree approaches have found them comparable and potentially the least sensitive to taxonomic sampling (Linnen and Farrell 2008). Furthermore, the concatenation with monophyly constraints (CMC) approach is a much more computationally tractable approach than currently available coalescent-based species tree approaches on datasets of the size that we are analyzing here (but see Chifman and Kubatko 2014 for a potentially scalable approach). Finally, the combined dataset (chloroplast and nuclear loci) - both constrained and unconstrained - was analyzed in RAxML. Although we understand the importance of analyzing this type of dataset in a coalescent framework, the scope of this study is not to infer the species tree for the species in question, which will be the focus of a detailed publication on the evolution of this clade, but rather to demonstrate the capability, effectiveness, and utility of the microfluidic method for systematics.

Results

Preliminary data acquisition

The newly generated low-coverage genomic data consisted of four samples representing three species of *Bartsia*: *B. pedicularoides* Benth. (two samples), *B. santolinifolia* (Kunth) Benth., and *B. serrata* Molau. One sample of *B. pedicularoides* was sequenced on an Illumina HiSeq 2000 at the Vincent J. Coates Genomics Sequencing

Laboratory at the University of California, Berkeley, yielding ~51.6 million 100 bp single-end reads. The other three samples were sequenced on an Illumina HiSeq 2000 at the Center for Genome Research and Biocomputing at Oregon State University, yielding in average ~51.4 million 100 bp paired-end reads per library (Table 4.1). The raw reads from these two sequencing experiments can be found on the GenBank Sequence Read Archive (SRR XXX). The raw reads were cleaned using Seqyclean v 1.8.10 (Zhbannikov et al in prep.) using defaults settings, resulting in ~46.8 million reads for the *B. pedicularoides* sample sequenced in California, and ~52.9 million for *B. pedicularoides*, ~46.4 million for *B. santolinifolia*, and ~65.1 million for *B. serrata* sequenced in Oregon. The plastomes assembled with the Alignreads pipeline from these samples had an average sequencing depth of 995x.

Primer design and validation

Chloroplast set

The *Bartsia* chloroplast alignment of six plastomes had a length of ~125 kilobases (kb) and included only one copy of the inverted repeat. Using this alignment, we were able to design a total of 74 primer pairs that spanned across the entire genome. Following validation, 53 primer pairs (72% success rate) passed the validation criteria. From these, a final set of the most variable 48 primer combinations was chosen, with an average variability of 2.7% (0.8% – 7.5%) (Appendix 2).

Nuclear set

For the nuclear set, we identified 51 PPR and 762 COSII loci that matched our criteria for further primer design (i.e., enough reads matching from low-coverage genomic

data to attempt primer design). The nuclear rDNA, *PHOT1*, and *PHOT2* alignments (6,711 bp, 578 bp, 1,272 bp long) all contained multiple places to design primers based on our criteria. A total of 188 primer pairs were designed from all datasets (Appendix 2). From those, 44 belonged to the PPR gene family, 130 to COSII, 8 to the nrDNA repeat, 3 to *PHOT1*, and 3 to *PHOT2*. After validation, 26 primer pairs were chosen for PPR (59.1% success rate), 25 for COSII (19.2 % success rate), 7 for the nuclear rDNA (87.5 % success rate), 0 for *PHOT1* (0 % success rate), and 3 for *PHOT2* (100 % success rate). Finally, the primer pair amplifying the longest target sequence was chosen among the various possibilities for the nuclear rDNA and *PHOT2* loci.

Sampling, microfluidic PCR and sequencing

To fully capture the morphological, genetic, and geographical diversity of the South American *Bartsia* clade, and to demonstrate the efficiency of this approach for molecular phylogenetic studies at low-taxonomic levels, we included 576 samples (Appendix 4) that represented 46 species of the clade and 28 related taxa as outgroups (included mostly to evaluate how far outside of the target group primers would work). Microfluidic amplification of the samples using 24 of the 48.48 Access Array Integrated Fluidic Circuits (Fluidigm) on the Access Array System (Fluidigm) resulted in 96 amplicons per sample (a total of 55,296 microfluidic reactions). After pooling and normalizing the amplicons for each sample, the pools were sequenced on an Illumina MiSeq platform with the Reagent Kit version 3, yielding ~20.3 million 300 bp paired-end reads (~35,200 reads per sample). From the total yield, ~16.9 million reads (77.7%) were successfully matched to both barcodes, and primers and were kept. The remaining discarded ~4.5 million reads were a combination of PhiX

Control v3 (Illumina) (~3.2 million reads or 15%) and ~1.3 million reads (7.3%) that did not match both the barcodes and the primers. Raw reads will be deposited in the GenBank Sequence Read Archive upon publication.

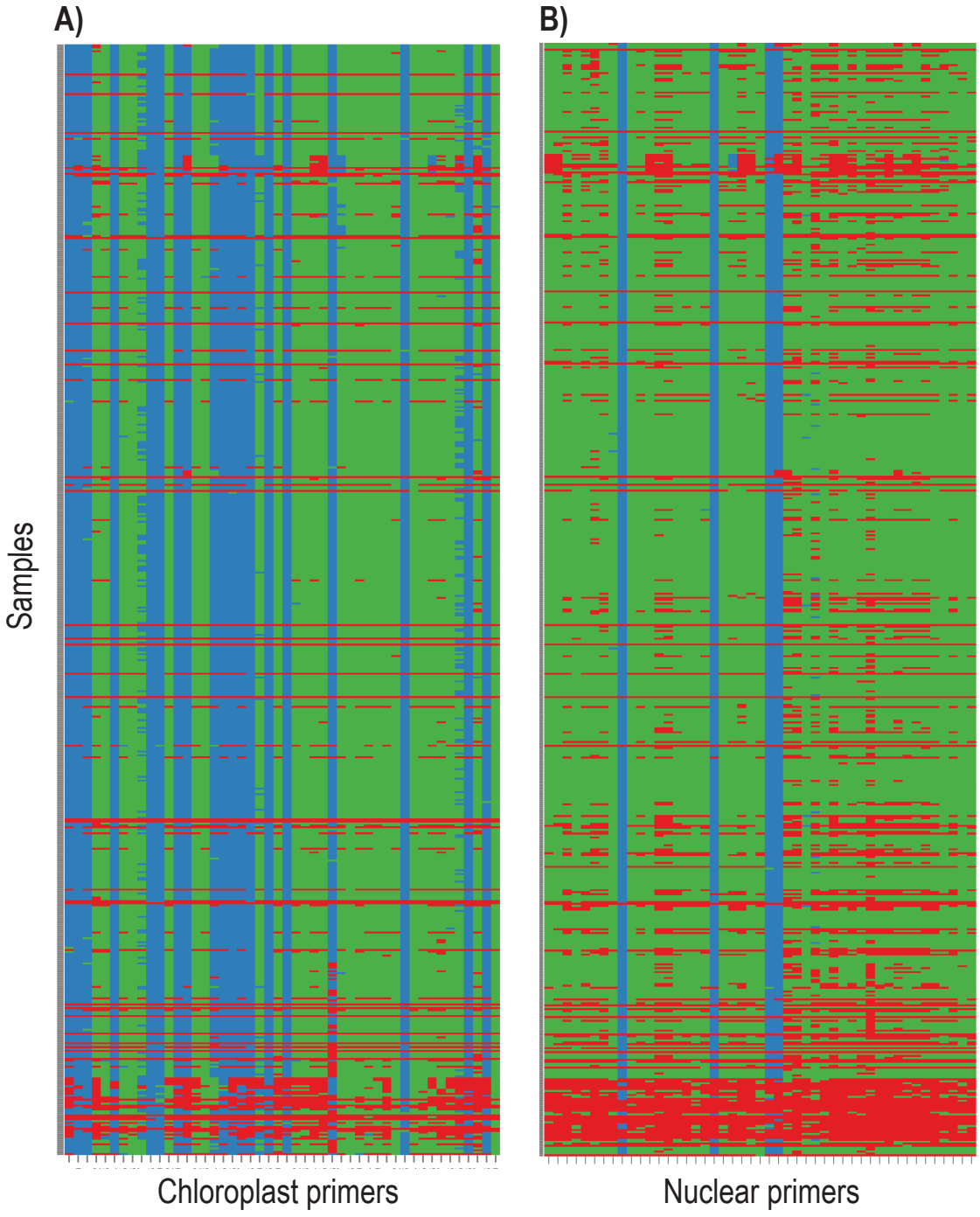
Data processing

We processed our data in two different ways by i) generating consensus sequences (with ambiguities) for every sample in every locus and ii) by recovering the alleles for each sample. Given its haploid nature, only consensus sequences were generated for the chloroplast set. From the 576 samples used in this study, 528 (91.7%) amplified at least one amplicon and 484 (84.0%) produced more than 40 amplicons (>21,300 bp) (Fig. 4.4a). The majority of the samples that did not amplify efficiently belonged to distantly related genera of the South American clade of *Bartsia*, suggesting that the designed primers are too specific to work efficiently outside this clade. This highlights the importance of a careful primer design and validation process that is in line with the taxonomic breath of the study in mind. Since our main focus was on the South American *Bartsia* clade and the primers were designed with plastomes from this clade, these results are not surprising. The final chloroplast dataset included 484 samples and was ~25,300 bp long after alignment and cleanup. Most of the samples belonged to the South American *Bartsia* clade (472), with the remaining 12 samples representing the three most closely related species (*Bellardia trixago* (L.) All., *Bellardia viscosa* (L.) Fisch. & C.A. Mey, and *Parentucellia latifolia* (L.) Caruel).

The nuclear set was processed either by generating consensus sequences (with ambiguities) for every sample in every locus or by recovering the alleles for each sample. For this set, 47 out the 48 regions were amplified for most of the samples (Fig. 4.4b) and only

one amplicon did not meet our demultiplexing criteria, i.e., did not match both barcodes and primers. From these 47 regions, we were able to recover an average of 442.7 samples (76.0%), ranging from 520 (90.3%) to 318 (55.2%). Each locus was aligned independently with an average length of 332 bp (from 267 to 459 bp) and a ML analysis in RAxML was performed independently on each locus. Based on these analyses we discovered that three loci had paralogous copies, and that another three were too variable (either due to true variability or sequencing errors) to be unambiguously aligned. These six loci were removed for downstream phylogenetic analyses. A nuclear concatenated dataset including 41 loci was constructed with a length of ~13,500 bp (after aligned and cleaned) and 363 samples. Finally, we constructed a combined concatenated matrix (nuclear and chloroplast) that had ~40,500 bp (after aligned and cleaned) for 349 samples. Each locus will be deposited in GenBank upon publication.

Figure 4.4
Bird's-eye view heat map showing the amplicon coverage (horizontal) for each sample (vertical) in A) the chloroplast set and in B) the nuclear set. These regions were processed to generate consensus sequences with ambiguities. Red indicates no amplicon was recovered either due to lack of successful amplification or mismatching of the barcodes and primers (see text for more details). Blue indicates that the forward and reverse sister reads overlapped by at least 10 bp and were joined. Green indicates that the forward and reverse reads did not overlap. The group of failed samples along the bottom of each panel are distantly-related taxa that were not included in primer design (see text).



Phylogenetic analyses

The concatenated chloroplast dataset was first analyzed with PartitionFinder to find the best partitioning scheme while also selecting for the best-fit model of sequence evolution for each possible partition. This analysis resulted in 11 partitions with the following models of sequence evolution: K81uf+I+G, K81uf+I, TrN+I+G, TVM+I+G, F81, K80+I, TVMef+I+G, TVMef+I+G, F81, TVM+I+G, TVM+I+G. Analyses in ML and Bayesian frameworks, in GARLI and MrBayes, respectively, resulted in the same overall phylogenetic relationships among the samples. The same is true for the nuclear concatenated, the combined (nuclear and chloroplast), and the concatenation with monophyly constraints (CMC) analyses, which resulted in the same overall relationships among species. Because every analysis resulted in very similar tree, the results and discussion will be based on the combined concatenated dataset, with and without constraints.

Four major clades were recovered with 100 support bootstrap support (BS) and 1.0 posterior probability (PP). The first clade is comprised of the two *Bellardia trixago* individuals included in this study. The second clade, which is sister to the latter one with 100 BS and 1.0 PP, encompasses all *Parentucellia viscosa* individuals. A third clade formed by all individuals of *Parentucellia latifolia* in this study is sister to the fourth clade, which is comprised by every individual of the South American *Bartsia* clade, and together, they are sister to the *B. trixago*-*P. viscosa* clade (Fig 4.5). The support of the backbone relationships within the South American *Bartsia* clade is low, and thus, very few systematic conclusions can be made at this point. First, we did not recover four monophyletic groups for the four morphological section (sensu Molau 1990). However, there are several clades that do contain multiple species from the same section with moderate support (Fig 4.5). Furthermore,

individuals of most of the species were recovered in multiple different clades, and in fact only three species (*B. filiformis* Wedd., *B. serrata* Molau, and *B. adenophylla* Molau) were monophyletic. This is not surprising, given the fact that the South American clade has been shown to be a recent and rapid radiation (Chapter 1), and processes like coalescent stochasticity, hybridization and introgression may be playing a large role in the evolution of these taxa. It will be necessary to conduct coalescent-based species trees estimation analyses to confidently elucidate the relationships in this clade, and this is the subject of ongoing work in this clade.

The CMC 'pseudo-species tree' analysis recovered most of the same clades containing species within the same taxonomic sections. Interestingly, enforcing monophyly reduced the BS support of the backbone relationships even further (Fig. 4.6), indicating that some of the individuals that were constrained clearly do not belong in those clades. There are several possibilities for this result, including violations of our *a priori* species designations (i.e., incorrect species delimitations, cryptic species, etc.), severe coalescent stochasticity, ancient and/or contemporary introgression, and/or hybrid speciation. Given the recent and rapid nature of this Andean diversification, any and all of these are potential problems that will be investigated in detail in future studies using these data.

Figure 4.5
Cladogram of the phylogenetic relationships of the South American *Bartsia* species and closely related taxa based on the maximum likelihood analysis in RAxML on the combined (chloroplast and nuclear) unconstrained dataset. Values above the branches represent maximum likelihood bootstrap support (BS). Clades containing individuals from species from only one of the four morphological sections of the South American clade (sensu Molau 1990) have been colored, as well as the three closely related taxa (*Bellardia trixago trixago* (L.) All., *Bellardia viscosa* Fisch. & C.A. Mey, and *Parentucellia latifolia* (L.) Caruel. The only three that were recovered as monophyletic are indicated on the tree and their clade collapsed in a triangle.

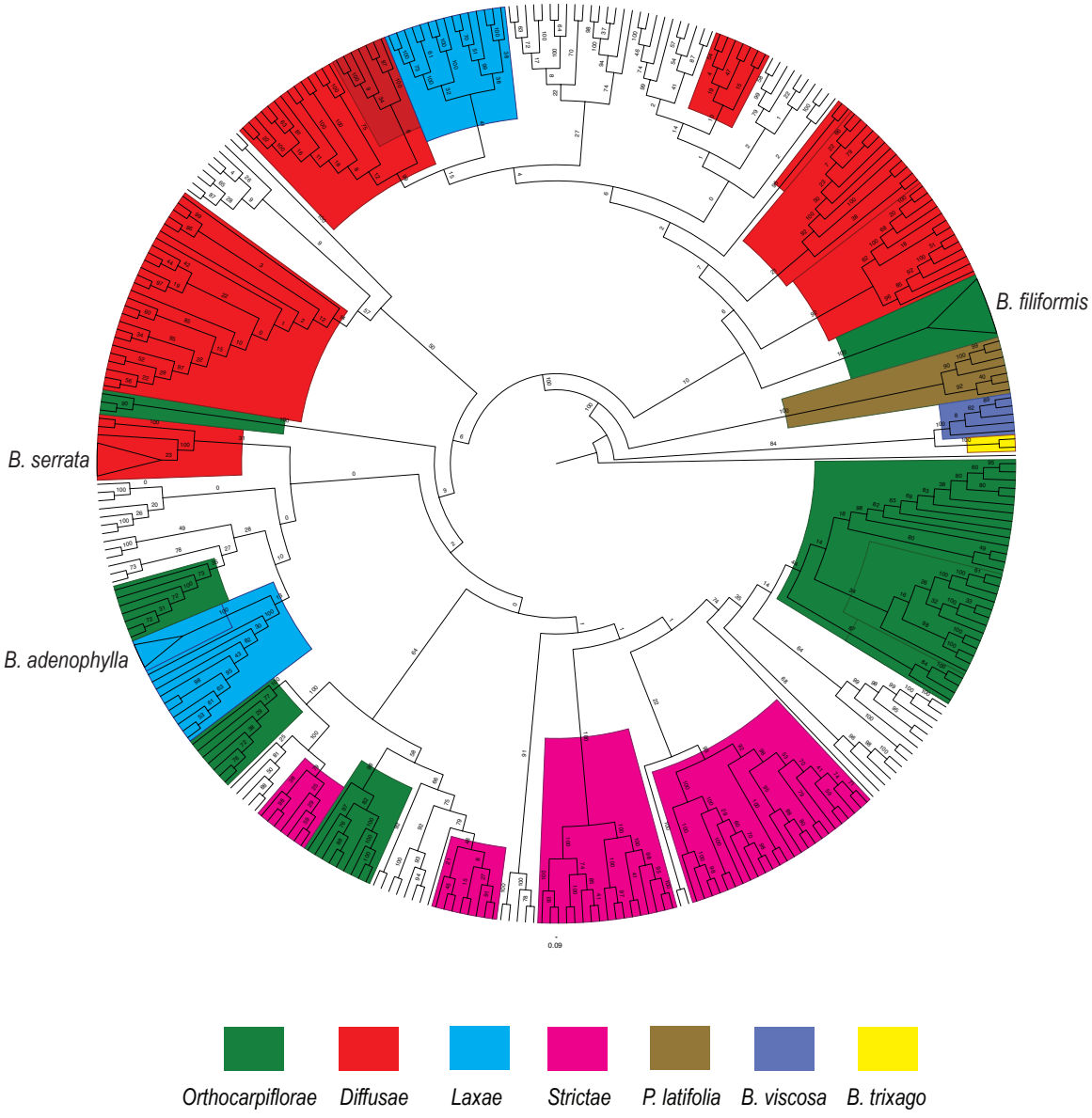
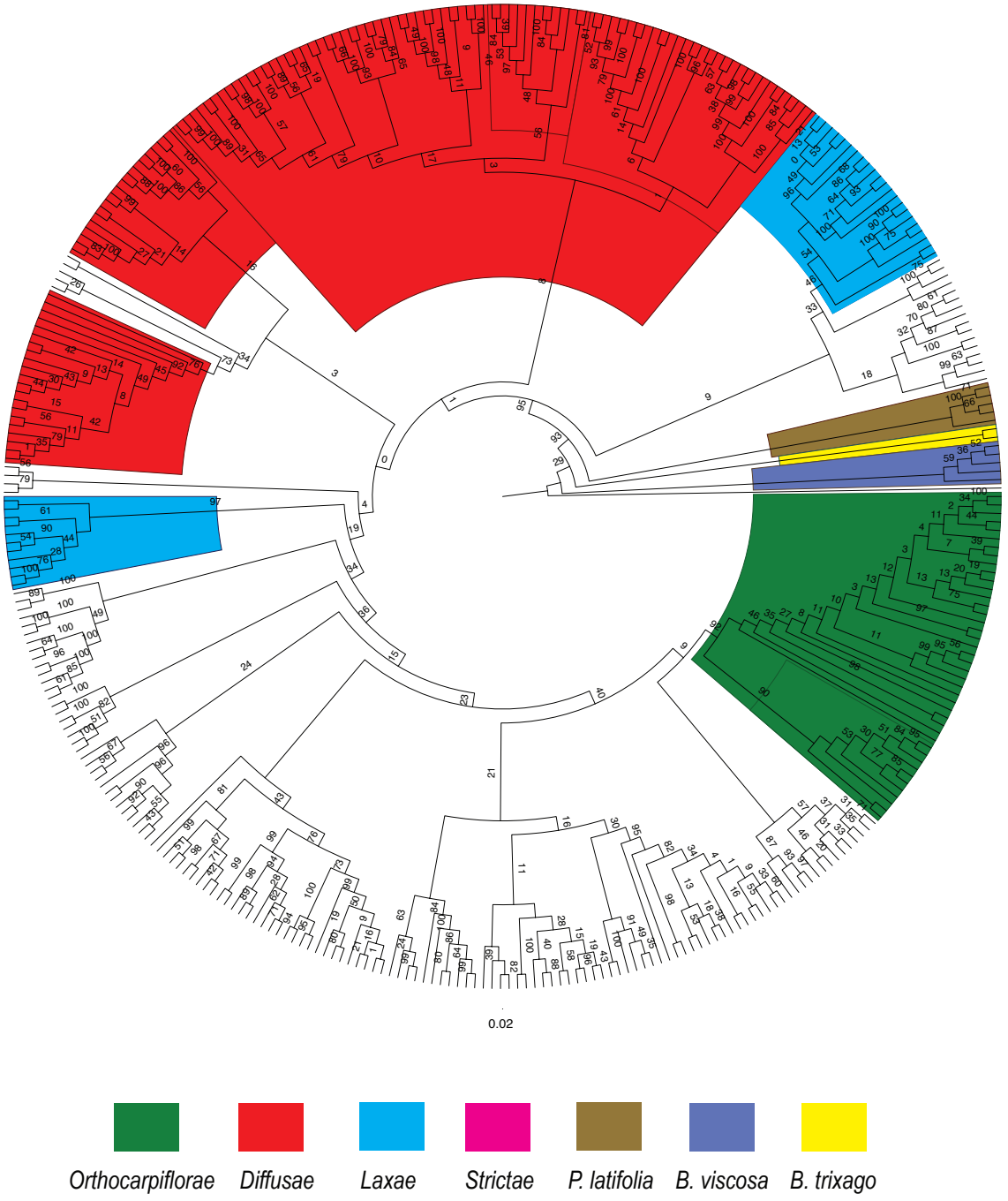


Figure 4.6
Cladogram of the phylogenetic relationships of the South American *Bartsia* species and closely related taxa based on the maximum likelihood analysis in RAxML on the combined (chloroplast and nuclear) dataset using the concatenation with monophyly constraints (CMC) approach on the combined dataset. Values above the branches represent maximum likelihood bootstrap support (BS). Clades containing at least two species from one of the four morphological section of the South American clade (sensu Molau 1990) have been colored, as well as the three closely related taxa (*Bellardia trixago trixago* (L.) All., *Bellardia viscosa* Fisch. & C.A. Mey, and *Parentucellia latifolia* (L.) Caruel.



Discussion

Regardless of the HTS method used to enter the realm of big data, it is clear that the field of phylogenetics, and the study of evolution in general, is quickly migrating towards larger and larger molecular datasets. The ability to produce more and longer reads, as well as reduced sequencing costs and increased computing power are making this transition easier and faster. Here, we presented a novel approach to generate large multilocus, homogeneously distributed, and targeted subgenomic datasets using microfluidic PCR and HTS.

One of the main advantages of this approach is circumventing the necessity to construct genomic libraries for each sample in the experiment. This step greatly increases the time and cost of any HTS approach, particularly as new sequencing platforms have greater yield and we are able to multiplex more samples in a single sequencing experiment. The approach presented here takes advantage of a four-primer reaction to efficiently tag each amplicon with sample specific barcodes and HTS adapters. By doing so, the resulting amplicons are ready to be sequenced following standard pooling and quality control. Furthermore, the use of a sample-specific dual barcoding strategy allows for a high level of multiplexing with far fewer barcodes. Commonly used commercial barcoding kits currently offer either 96 (NEXTflex DNA Barcode kit; Bioo Scientific, Austin, Texas, USA) or 386 barcodes (Fluidigm), but we are able to multiplex up to 1,152 samples with only 72 indexes (48 forward and 24 reverse). This expands the possibilities during experimental design and takes full advantage of the yield of current HTS platforms, while maintaining low upfront costs.

An additional technical advantage of our microfluidic approach is the high throughput achieved with minimal amounts of DNA, reagents, and labor. A commercially available

platform – the Fluidigm Access Array System – facilitates simultaneous amplification of 48 samples with 48 distinct primer pairs (2,304 reactions) using only 15 U of *Taq* polymerase and 1μL of 50ng/μL genomic DNA per sample. By conducting a simultaneous four-primer reaction, one avoids the necessity of performing multiple rounds of PCR to incorporate barcodes and adapters – a limitation of the Targeted Amplicon Sequencing (TAS) strategy of Bybee et al. (2011) when the number of samples and/or gene regions is large. For example, following the TAS approach, to produce tagged amplicons for the 96 gene regions targeted in this study, it would be necessary to perform >1,100 96-well plates (or >275 384-well plates) of PCR to produce barcoded amplicons for the 576 accessions used here. While the TAS approach does allow for more flexibility in terms of primer design, i.e., primer annealing temperatures do not need to all be the same, and it may be possible to incorporate ambiguities into primer design, to take advantage of performing PCR in plates, significant PCR optimization would need to be performed. Nevertheless, with high levels of taxon-by-gene region samplings, TAS becomes unpractical. Using the microfluidic amplification approach to amplicon generation and tagging, only 24 microfluidic chips were necessary to amplify and tag the 55,296 amplicons. While studies with smaller sampling strategies (e.g., Bybee et al. 2011) would likely benefit from the two-reaction TAS approach, with the ever increasing sequencing read length and throughput of HTS platforms, the microfluidic PCR approach presented here allows researchers performing large phylogenetic or population genetic studies to maximize data collection using HTS techniques.

Perhaps the most significant advantage of generating phylogenetic datasets using a targeted amplicon approach is that using our 'read frequency' approach (Fig. 4.2e), it was possible to distinguish individual alleles at heterozygous nuclear loci without requiring

additional assembly or mapping steps. Because targeted loci were specifically enriched one locus per reaction per sample, when paired-end sequences from each specific barcode and primer combination were bioinformatically matched, at heterozygous loci in diploid species we identified two sets of reads at high frequency in the amplicon pools for those loci, allowing us to straightforwardly determine both alleles. Using this approach in systems with higher levels of ploidy and known polyploid complexes, we have been able to recover more than four alleles in known octoploids (D. F. Morales-Briones and D. C. Tank, unpubl. data). This contrasts with methods in which genomic DNA is sheared and selected for a specific length, e.g., sequence capture and GBS/RadSeq, where an assembly and/or reference-based mapping strategy is necessary to compile consensus sequences. These additional steps introduce known problems associated with the large number of de novo assemblers and mappers, e.g., varying numbers of contigs and N50 using different algorithms, performance of the algorithm based on the error model of the sequencing platform used, and computational power and time (see Bao et al. 2011 for further details). More importantly, most phylogenomic studies that have included nuclear data generated using HTS techniques like these, have ignored the challenge that heterozygosity presents by using ambiguity coding (e.g., (Lemmon et al. 2012) or by selecting only one allele and discarding others (e.g., (Faircloth et al. 2012; Smith et al. 2014). For phylogenetic and population genetic studies using modern coalescent-based approaches, allelic information is important when reconstructing the evolutionary histories of the genes sampled, and in cases where there is a large amount of coalescent stochasticity and/or gene flow, discarding or masking allelic information may be misleading. In a population genetic study of the North American tiger salamander (*Ambystoma tigrinum* Green) species complex, O'Neill et al. (2013) used a

haplotype phasing strategy to computationally determine individual alleles. For statistical phasing approaches, the number of individuals present in a sample is a critical factor in determining how well phase can be estimated (Browning and Browning 2011), and therefore may only be appropriate for the deep population-level sampling in population genetic studies such as this, but will likely not be useful for most phylogenetic studies.

Furthermore, polyploidy is common in many plant groups, as well as in select groups of insects (Lokki and Saura 1980), fish (Leggatt and Iwama 2003), amphibians (Cannatella and De Sá 1993), and reptiles (Bogart 1980), and therefore, this is an important consideration that complicates the issue of heterozygosity even more. For example, a tetraploid species may be heterozygous at both homeologous loci, and in this scenario, one would expect to identify four sets of reads with high frequencies dominating the amplicon pool. Likewise, a tetraploid may be homozygous at one homeolog and heterozygous at the other; in this case, we would expect to identify three-sets of high frequency reads dominating the amplicon pool. Finally, for many species of plants, ploidy levels are often unknown, or variable within a species (e.g., Judd et al. 2007), and material appropriate for determining ploidy via chromosome counts and/or flow cytometry is not available. While at any one nuclear locus, a polyploid species may or may not be heterozygous at one or more of the homeologs, by having multiple nuclear loci in one experiment, it is possible to calculate the frequencies of alleles across all loci and not only recover individual alleles, but also estimate ploidy level – or at least a minimum ploidy level depending on levels of heterozygosity. In plants, this is especially useful for evaluating hypothesized allopolyploid events, as well as the evolutionary and ecological consequences of polyploidy when these data are analyzed in a comparative phylogenetic context. Our study included seven known tetraploid species for

which chromosome numbers had been established (Molau 1990), allowing us to test the validity of this approach.

A notable limitation of the microfluidic approach that we present here is the necessity to design a relatively large number of target-specific primers to fill a microfluidic array. To do this in an efficient manner, it is necessary to first have at least some genomic resources available for your clade of interest. In our case, we had both whole plastome sequences, as well as low-coverage genomic data for a small, but representative, set of species. With these preliminary data in mind, we developed an effective approach for primer design that allowed us to target 1) the most variable regions of the plastome in the South American *Bartsia* clade, 2) the ITS (Baldwin 1992) and ETS (Baldwin and Markos 1998) regions of the nrDNA repeat that have been used extensively at the interspecific level in plants, 3) multiple, independent nuclear genes from the intronless PPR gene set developed by Yuan et al. (2009) and shown to be phylogenetically informative at the family level in Verbenaceae (Marx et al. 2010) and at the subfamily level in Campanuloideae (Crowl et al. 2014), and 4) intron-spanning regions from within the COSII gene set developed by Wu et al. (2006) and used within Orobanchaceae (Li et al. 2008), and the Phototropin 2 gene used at the interspecific level in *Glandularia, Junellia* and *Verbena* in the Verbenaceae (Yuan and Olmstead 2008). By specifically targeting the variable regions of the plastome, commonly sequenced regions of the nrDNA repeat (e.g., ITS, ETS), and multiple independent nuclear loci that range from more conserved (e.g., intronless PPR genes) to rapidly evolving nuclear introns (e.g., COSII), we were able to assemble a large, multilocus, homogeneously distributed dataset with high levels of intraspecific sampling for a complete clade of recently diverged Andean plants. Although we took, and advocate, a genome-skimming approach (sensu Straub et al. 2011;

2012) to develop the necessary genomic resources used here for primer development, there are a growing number of publically available databases that could also be used – e.g., for plants, Phytozome (http://www.phytozome.net), One Thousand Plants Project (1KP; http://onekp.com), IntrEST (Ilut and Doyle 2012), and Genome 10k for animals (Genome 10K Community of Scientists 2009).

The South American Bartsia *clade*

This is the first time that the interspecific relationships of the species in the South American *Bartsia* clade have been studied with such deep taxonomic sampling and with so much molecular data. From our results, it is clear that in order to fully understand the evolutionary history of the clade, coalescent-based methods that take incomplete lineage sorting into account are needed. However, this is not the scope of this study and the results presented here are preliminary. The monophyly of the group is highly supported by all analyses, and is in agreement with recent a study on the clade (Chapter 1). Interspecific relationships however, have very little support – a pattern that is expected to be seen in rapid radiations like this – and only three species formed clades with all of their individuals. These three species are all taxa with small and restricted geographic distributions and likely small effective population sizes. Given that the time to coalescence is directly linked to effective population size (Rannala and Yang 2003), it is not unexpected that individuals from these species were monophyletic in our unconstrained analyses. Conversely, when we look at a species with a large geographic distributions, and thus, larger effective population sizes (e.g., *B. pedicularoides* Benth.), we see that the individuals are recovered in multiple different groups across the tree (Fig 4.5).

Enforcing the monophyly of species has been used as a 'pseudo-species tree' method with good results (Linnen and Farrell 2008), and some relationships recovered here make evolutionary sense – *Bartsia sericea* Molau and *B. crisafullii* N. Holmgren were recovered as sister species (Fig. 4.5) with high support. Both species are extremely similar morphologically, only differing in their life history and ploidy level (perennial vs. annual, and diploid vs. tetraploid, respectively). However, in some instances, enforcing monophyly of the species reduced the BS support of deeper branches. This may signify that some of the individuals constrained are in real conflict and that they belong in other groups. Given the short time to coalescence within the South American clade, this is expected. There are several possibilities for this result, including violations of our *a priori* species designations (i.e., incorrect species delimitations, cryptic species, etc.), severe coalescent stochasticity, ancient and/or contemporary introgression, and/or hybrid speciation. Given the recent and rapid nature of this Andean diversification, any and all of these are potential problems that will be investigated in detail in future studies using these data. As mentioned earlier, this is the first attempt to recovering the interspecific relationships of the taxa in this clade, and it is clear that more powerful analyses, especially ones that explicitly incorporate coalescent stochasticity, will be necessary to confidently elucidate their evolutionary history.

Conclusion

We presented a novel approach to generate large multilocus phylogenomic datasets for a wide number of samples and species using microfluidic PCR and HTS. This approach allows for more control in selecting what parts of the genome are sequenced, resulting in datasets that are tailored to address the specific questions being asked, and that are

homogenous across samples. Additionally, this method is both cost effective and time efficient, as it does not require genomic libraries to be constructed for every sample, and takes full advantage of the large multiplexing capabilities of HTS platforms. As a case study, we focused on 576 samples of the South American *Bartsia* clade, amplifying and sequencing the 48 most variable regions of the chloroplast genome as well as 47 independently evolving loci from a range of coding and non-coding regions of the nuclear genome. This targeted, subgenomic strategy for the collection of multilocus data for phylogenetic studies provided us with a large, but modest set of loci appropriate for species tree estimation using coalescent-based inference methods, and provided us with the first species level phylogeny for the South American *Bartsia* clade. Furthermore, we were able to straightforwardly resolve alleles in heterozygous individuals and estimate ploidy levels, which is an important, and often overlooked, consideration at low taxonomic levels.

Literature Cited

Baldwin B.G. 1992. Phylogenetic Utility of the Internal Transcribed Spacers of Nuclear Ribosomal DNA in Plants: An Example from the Compositae. Molecular Phylogenetics and Evolution. 1:3–16.

Baldwin B.G., Markos S. 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.

Bao S., Jiang R., Kwan W., Wang B., Ma X., Song Y.-Q. 2011. Evaluation of next-generation sequencing software in mapping and assembly. J. Hum. Genet. 56:406–414.

Bhat S., Polanowski A.M., Double M.C., Jarman S.N., Emslie K.R. 2012. The Effect of Input DNA Copy Number on Genotype Call and Characterising SNP Markers in the Humpback Whale Genome Using a Nanofluidic Array. PLoS ONE. 7:e39181.

Bi K., Linderoth T., Vanderpool D., Good J.M., Nielsen R., Moritz C. 2013. Unlocking the vault: next-generation museum population genomics. Molecular Ecology. 22:6018–6032.

Bi K., Vanderpool D., Singhal S., Linderoth T., Moritz C., Good J.M. 2012. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC Genomics. 13:1–1.

Bogart J.P. 1980. Evolutionary implications of polyploidy in amphibians and reptiles. Basic Life Sci. 13:341–378.

Bombarely A., Menda N., Tecle I.Y., Buels R.M., Strickler S., Fischer-York T., Pujar A., Leto J., Gosselin J., Mueller L.A. 2011. The Sol Genomics Network (solgenomics.net): growing tomatoes using Perl. Nucleic Acids Research. 39:D1149–55.

Browning S.R., Browning B.L. 2011. Haplotype phasing: existing methods and new developments. Nature Reviews Genetics. 12:703–714.

Bybee S.M., Bracken-Grissom H., Haynes B.D., Hermansen R.A., Byers R.L., Clement M.J., Udall J.A., Wilcox E.R., Crandall K.A. 2011. Targeted Amplicon Sequencing (TAS): A Scalable Next-Gen Approach to Multilocus, Multitaxa Phylogenetics. Genome Biology and Evolution. 3:1312–1323.

Byers R.L., Harker D.B., Yourstone S.M., Maughan P.J., Udall J.A. 2012. Development and mapping of SNP assays in allotetraploid cotton. Theor Appl Genet. 124:1201–1214.

Cannatella D.C., De Sá R.O. 1993. Xenopus laevis as a model organism. Systematic Biology. 42:476–507.

Chase M.W., Soltis D.E., Olmstead R.G., Morgan D., Les D.H., Mishler B.D., Duvall M.R., Price R.A., Hills H.G., Qiu Y.-L. 1993. Phylogenetics of seed plants: an analysis of

nucleotide sequences from the plastid gene rbcL. Annals of the Missouri Botanical Garden.:528–580.

Chifman J., Kubatko L. 2014. Quartet Inference from SNP Data Under the Coalescent Model. Bioinformatics. In press. DOI: 10.1093/bioinformatics/btu530

Crawford N.G., Faircloth B.C., MCCORMACK J.E., BRUMFIELD R.T., Winker K., Glenn T.C. 2012. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. Biology Letters. 8:783–786.

Crowl A.A., Mavrodiev E., Mansion G., Haberle R., Pistarino A., Kamari G., Phitos D., Borsch T., Cellinese N. 2014. Phylogeny of Campanuloideae (Campanulaceae) with Emphasis on the Utility of Nuclear Pentatricopeptide Repeat (PPR) Genes. PLoS ONE. 9:e94199.

Cruaud A., Gautier M., Galan M., Foucaud J., Sauné L., Genson G., Dubois E., Nidelet S., Deuve T., Rasplus J.-Y. 2014. Empirical Assessment of RAD Sequencing for Interspecific Phylogeny. Mol. Biol. Evol. 31:1272–1274.

Dominguez M.H., Chattopadhyay P.K., Ma S., Lamoreaux L., McDavid A., Finak G., Gottardo R., Koup R.A., Roederer M. 2013. Journal of Immunological Methods. Journal of Immunological Methods. 391:133–145.

Downie S.R., Palmer J.D. 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In: Soltis P.S., Soltis D.E., Doyle J.J., editors. Molecular systematics of plants. Chapman and Hall, New York, NY. p. 14–35.

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

Drouin G., Daoud H., Xia J. 2008. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Molecular Phylogenetics and Evolution. 49:827–831.

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

Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 32:1792–1797.

Edwards S.V. 2009. Is A New And General Theory Of Molecular Systematics Emerging. Evolution. 63:1–19.

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

Faircloth B.C., Sorenson L., Santini F., Alfaro M.E. 2013. A Phylogenomic Perspective on the Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved Elements (UCEs). PLoS ONE. 8:e65923.

Gaedcke J., Grade M., Camps J., Sokilde R., Kaczkowski B., Schetter A.J., Difilippantonio M.J., Harris C.C., Ghadimi B.M., Moller S., Beissbarth T., Ried T., Litman T. 2012. The Rectal Cancer microRNAome - microRNA Expression in Rectal Cancer and Matched Normal Mucosa. Clinical Cancer Research. 18:4919–4930.

Genome 10K Community of Scientists. 2009. Genome 10K: A Proposal to Obtain Whole-Genome Sequence for 10 000 Vertebrate Species. Journal of Heredity. 100:659–674.

Godden G.T., Jordon-Thaden I.E., Chamala S., Crowl A.A., García N., Germain-Aubrey C.C., Heaney J.M., Latvis M., Qi X., Gitzendanner M.A. 2012. Making next-generation sequencing work for you: approaches and practical considerations for marker development and phylogenetics. Plant Ecology & Diversity. 5:427–450.

Graham S.W., Olmstead R.G. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. American Journal of Botany. 87:1712–1730.

Ilut D.C., Doyle J.J. 2012. Selecting Nuclear Sequences for Fine Detail Molecular Phylogenetic Studies in Plants: A Computational Approach and Sequence Repository. Syst Bot. 37:7–14.

Jones J.C., Fan S., Franchini P., Schartl M., Meyer A. 2013. The evolutionary history of Xiphophorusfish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. Molecular Ecology. 22:2986–3001.

Judd W.S., Soltis D.E., Soltis P.S., Ionta G. 2007. Tolmiea diplomenziesii: A new species from the Pacific Northwest and the diploid sister taxon of the autotetraploid T. menziesii (Saxifragaceae). Brittonia. 59:217–225.

Katoh K., Standley D.M. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. 30:772–780.

Kent W.J. 2002. BLAT––The BLAST-Like Alignment Tool. Genome Research. 12:656–664.

Knowles L.L., Kubatko L.S. 2010. Estimating species trees: practical and theoretical aspects. Wiley-Blackwell.

Koressaar T., Remm M. 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics. 23:1289–1291.

Lanfear R., Calcott B., Ho S.Y.W., Guindon S. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29:1695–1701.

Leggatt R.A., Iwama G.K. 2003. Occurrence of polyploidy in the fishes. Reviews in Fish Biology and Fisheries. 13:237–246.

Lemmon A.R., Brown J.M., Stanger-Hall K., Lemmon E.M. 2009. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. Systematic Biology. 58:130–145.

Lemmon A.R., Emme S.A., Lemmon E.M. 2012. Anchored Hybrid Enrichment for Massively High-Throughput Phylogenomics. Systematic Biology. 61:727–744.

Li M., Wunder J., Bissoli G., Scarponi E., Gazzani S., Barbaro E., Saedler H., Varotto C. 2008. Development of COS genes as universally amplifiable markers for phylogenetic reconstructions of closely related plant species. Cladistics. 24:727–745.

Linnen C., Farrell B. 2008. Comparison of Methods for Species-Tree Inference in the Sawfly Genus Neodiprion (Hymenoptera: Diprionidae). Systematic Biology. 57:876–890.

Lohr J.G., Stojanov P., Lawrence M.S., Auclair D., Chapuy B., Sougnez C., Cruz-Gordillo P., Knoechel B., Asmann Y.W., Slager S.L. 2012. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proc. Natl. Acad. Sci. U.S.A. 109:3879–3884.

Lokki J., Saura A. 1980. Polyploidy in insect evolution. Basic Life Sci. 13:277–312.

Lu X., Wang L., Chen S., He L., Yang X., Shi Y., Cheng J., Zhang L., Gu C.C., Huang J., Wu T., Ma Y., Li J., Cao J., Chen J., Ge D., Fan Z., Li Y., Zhao L., Li H., Zhou X., Chen L., Liu D., Chen J., Duan X., Hao Y., Wang L., Lu F., Liu Z., Yao C., Shen C., Pu X., Yu L., Fang X., Xu L., Mu J., Wu X., Zheng R., Wu N., Zhao Q., Li Y., Liu X., Wang M., Yu D., Hu D., Ji X., Guo D., Sun D., Wang Q., Yang Y., Liu F., Mao Q., Liang X., Ji J., Chen P., Mo X., Li D., Chai G., Tang Y., Li X., Du Z., Liu X., Dou C., Yang Z., Meng Q., Wang D., Wang R., Yang J., Schunkert H., Samani N.J., Kathiresan S., Reilly M.P., Erdmann J., Peng X., Wu X., Liu D., Yang Y., Chen R., Qiang B., Gu D. 2012. Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. Nature Publishing Group. 44:890–894.

Maddison W. 1997. Gene trees in species trees. Systematic Biology. 46:523–536.

Malé P.-J.G., Bardon L., Besnard G., Coissac E., Delsuc F., Engel J., Lhuillier E., Scotti-Saintagne C., Tinaut A., Chave J. 2014. Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. Molecular Ecology Resources.

Mandel J.R., Dikow R.B., Funk V.A., Masalia R.R., Staton S.E., Kozik A., Michelmore R.W., Rieseberg L.H., Burke J.M. 2014. A Target Enrichment Method for Gathering Phylogenetic Information from Hundreds of Loci: An Example from the Compositae. Applications in Plant Sciences. 2:1300085.

Marx H.E., O'Leary N., Yuan Y.-W., Lu-Irving P., Tank D.C., Múlgura M.E., Olmstead R.G. 2010. A molecular phylogeny and classification of Verbenaceae. American Journal of

Botany. 97:1647–1663.

McCormack J.E., Faircloth B.C., Crawford N.G., Gowaty P.A., Brumfield R.T., Glenn T.C. 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. Genome Research. 22:746–754.

McCormack J.E., Harvey M.G., Faircloth B.C., Crawford N.G., Glenn T.C., Brumfield R.T. 2013. A Phylogeny of Birds Based on Over 1,500 Loci Collected by Target Enrichment and High-Throughput Sequencing. PLoS ONE. 8:e54848.

Moignard V., Macaulay I.C., Swiers G., Buettner F., Schütte J., Calero-Nieto F.J., Kinston S., Joshi A., Hannah R., Theis F.J., Jacobsen S.E., de Bruijn M.F., Göttgens B. 2013. Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. Nature Cell Biology. 15:363–372.

Molau U. 1990. The genus Bartsia (Scrophulariaceae - Rhinanthoideae). Opera Botanica. 102:1–100.

Moonsamy P.V., Williams T., Bonella P., Holcomb C.L., Höglund B.N., Hillman G., Goodridge D., Turenchalk G.S., Blake L.A., Daigle D.A., Simen B.B., Hamilton A., May A.P., Erlich H.A. 2013. High throughput HLA genotyping using 454 sequencing and the Fluidigm Access Array™ system for simplified amplicon library preparation. Tissue Antigens. 81:141–149.

Moore M.J., Bell C.D., Soltis P.S., Soltis D.E. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proceedings of the National Academy of Sciences. 104:19363–19368.

Moore M.J., Hassan N., Gitzendanner M.A., Bruenn R.A., Croley M., Vandeventer A., Horn J.W., Dhingra A., Brockington S.F., Latvis M. 2011. Phylogenetic analysis of the plastid inverted repeat for 244 species: insights into deeper-level angiosperm relationships from a long, slowly evolving sequence region. International Journal of Plant Sciences. 172:541–558.

Moore M.J., Soltis P.S., Bell C.D., Burleigh J.G., Soltis D.E. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proceedings of the National Academy of Sciences. 107:4623–4628.

Nadeau N.J., Martin S.H., Kozak K.M., Salazar C., Dasmahapatra K.K., Davey J.W., Baxter S.W., Blaxter M.L., Mallet J., Jiggins C.D. 2012. Genome-wide patterns of divergence and gene flow across a butterfly radiation. Molecular Ecology. 22:814–826.

Njuguna W., Liston A., Cronn R., Ashman T.-L., Bassil N. 2013. Insights into phylogeny, sex function and age of Fragaria based on whole chloroplast genome sequencing. Molecular Phylogenetics and Evolution. 66:17–29.

O'Neill E.M., Schwartz R., Bullock C.T., Williams J.S., Shaffer H.B., Aguilar-Miguel X.,

Parra-Olea G., Weisrock D.W. 2013. Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (Ambystoma tigrinum) species complex. Molecular Ecology. 22:111–129.

Parks M., Cronn R., Liston A. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol. 7:84.

R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rambaut A., Drummond A.J. 2004. Tracer. University of Edinburgh, Edinburgh, UK. Available at http://tree.bio.ed.ac.uk/software/tracer/.

Rannala B., Yang Z. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics. 164:1645–1656.

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

Rozen S., Skaletsky H. 2000. Primer3 on the WWW for General Users and for Biologist Programmers. In: Misener S., Krawetz S.A., editors. Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ. p. 365–386.

Ruhfel B.R., Gitzendanner M.A., Soltis P.S., Soltis D.E., Burleigh J.G. 2014. From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol Biol. 14:23.

Salamone I., Govindarajulu R., Falk S., Parks M., Liston A., Ashman T.-L. 2013. Bioclimatic, ecological, and phenotypic intermediacy and high genetic admixture in a natural hybrid of octoploid strawberries. American Journal of Botany. 100:939–950.

Shalek A.K., Satija R., Adiconis X., Gertner R.S., Gaublomme J.T., Raychowdhury R., Schwartz S., Yosef N., Malboeuf C., Lu D., Trombetta J.T., Gennert D., Gnirke A., Goren A., Hacohen N., Levin J.Z., Park H., Regev A. 2013. Single-cell transcriptomics reveals bimodality inexpression and splicing in immune cells. Nature.:1–5.

Smith B.T., Harvey M.G., Faircloth B.C., Glenn T.C., Brumfield R.T. 2014. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Systematic Biology. 63:83–95.

Smith S., Dunn C. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. Bioinformatics. 24:715.

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

Straub S.C.K., Cronn R.C., Edwards C., Fishbein M., Liston A. 2013. Horizontal Transfer of DNA from the Mitochondrial to the Plastid Genome and Its Subsequent Evolution in Milkweeds (Apocynaceae). Genome Biology and Evolution. 5:1872–1885.

Straub S.C.K., Fishbein M., Livshultz T., al E. 2011. Building a model: developing genomic resources for common milkweed (Asclepias syriaca) with low coverage genome sequencing. BMC Genomics. 12.

Straub S.C.K., Parks M., Weitemier K., Fishbein M., Cronn R.C., Liston A. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. American Journal of Botany. 99:349–364.

Stull G.W., Moore M.J., Mandala V.S., Douglas N.A., Kates H.-R., Qi X., Brockington S.F., Soltis P.S., Soltis D.E., Gitzendanner M.A. 2013. A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes. Applications in Plant Sciences. 1:1–7.

Sukumaran J., Holder M.T. 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics. 26:1569–1571.

Tennessen J.A., Govindarajulu R., Liston A., Ashman T.-L. 2013. Targeted sequence capture provides insight into genome structure and genetics of male sterility in a gynodioecious diploid strawberry, Fragaria vesca ssp. bracteata (Rosaceae). G3; Genes|Genomes|Genetics. 3:1341–1351.

Tepe E.J., Farruggia F.T., Bohs L. 2011. A 10-gene phylogeny of Solanum section Herpystichum (Solanaceae) and a comparison of phylogenetic methods. American Journal of Botany. 98:1356–1365.

Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B.C., Remm M., Rozen S.G. 2012. Primer3 - new capabilities and interfaces. Nucleic Acids Research. 40:e115–e115.

Uribe-Convers S., Duke J.R., Moore M.J., Tank D.C. 2014. A Long PCR–Based Approach for DNA Enrichment Prior to Next-Generation Sequencing for Systematic Studies. Applications in Plant Sciences. 2:1300063.

Walter K., Holcomb T., Januario T., Du P., Evangelista M., Kartha N., Iniguez L., Soriano R., Huw L., Stern H., Modrusan Z., Seshagiri S., Hampton G.M., Amler L.C., Bourgon R., Yauch R.L., Shames D.S. 2012. DNA Methylation Profiling Defines Clinically Relevant Biological Subsets of Non-Small Cell Lung Cancer. Clinical Cancer Research. 18:2360–2373.

Wolfe K.H., Li W.H., Sharp P.M. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. U.S.A. 84:9054–9058.

Wolfe K.H., Sharp P.M., Li W.-H. 1989. Rates of synonymous substitution in plant nuclear genes. J Mol Evol. 29:208–211.

Wu F., Mueller L.A., Crouzillat D., Pétiard V., Tanksley S.D. 2006. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. Genetics. 174:1407–1420.

Xi Z., Ruhfel B.R., Schaefer H., Amorim A.M., Sugumaran M., Wurdack K.J., Endress P.K., Matthews M.L., Stevens P.F., Mathews S., Davis C.C. 2012. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. Proceedings of the National Academy of Sciences. 109:17519–17524.

Yuan Y., Liu C., Marx H., Olmstead R. 2009. The pentatricopeptide repeat (PPR) gene family, a tremendous resource for plant phylogenetic studies. New Phytologist. 182:272–283.

Yuan Y., Olmstead R. 2008. Evolution and phylogenetic utility of the PHOT gene duplicates in the Verbena complex (Verbenaceae): dramatic intron size variation and footprint of ancestral recombination. American Journal of Botany. 95:1166–1176.

Yuan Y.-W., Liu C., Marx H.E., Olmstead R.G. 2010. Molecular Phylogenetics and Evolution. Molecular Phylogenetics and Evolution. 54:23–35.

Zwickl D.J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin.

Appendices

Appendix 1

Protocol for long PCR for amplification of 4–20-kb targets

Product	Contents	Cat. No.
QIAGEN *Taq* DNA Polymerase[1]	250 units *Taq* DNA Polymerase, 10x PCR Buffer,† 5x Q-Solution, 25 mM MgCl2	201205
QIAGEN HotStar HiFidelity DNA Polymerase[2]	100 units HotStar HiFidelity DNA Polymerase[2], 10x HotStar PCR Buffer, 5x Q-Solution, 25 mM MgSO4	202602

[1]this should work using almost any high quality *Taq* polymerase – cheap taqs e.g. QIAGEN Top*Taq* or Promega Go*Taq* do not work and result in large smears, rather than discrete bands.
[2] QIAGEN HotStar HiFidelity DNA Polymerase was the only high-fidelity polymerase tested.
† Q-solution does seem to be an important additive, thus the use of QIAGEN *Taq*. However, this does work using Q-solution with other high-quality *Taq* polymerases such as Promega's or New England Biolab's standard *Taq* (i.e., if you have a stock of Q-solution, but no QIAGEN *Taq*).

Genomic DNA must be high quality. Run a 0.8% or 1% gel to check. Standard CTAB extractions from silica gel–dried or herbarium material work well if they (1) are recent (extraction and tissue), and (2) contain high-molecular-weight DNA. Most important, we have found that recent DNA extractions that have not been through numerous freeze-thaw cycles work best. **For best results, long PCR should be done using new DNA extractions stored at 4°C while performing long PCR experiments**.

All preparations should be done on ice.

1. Number tubes or prepare plate. Make sure to include appropriate negative controls.

2. Prepare QIAGEN HotStar HiFidelity DNA polymerase dilution:

Reagents to prepare the HotStar *Taq* dilution	Volumes for 25 reactions (total 12.5 µL)	Volumes for 50 reactions (total 25 µL)	Volumes for 100 reactions (total 50 µL)
5× HotStar HiFidelity PCR buffer	2.5 µL	5.0 µL	10 µL
H_2O	9.0 µL	18 µL	36 µL
QIAGEN HotStar *Taq*	1.0 µL	2.0 µL	4.0 µL

3. Prepare cocktail:

Cocktail	x 1 (25µL reaction)
10x PCR buffer (QIAGEN coralLoad or colorless, 15mM $MgCl_2$)	2.5 µL
$MgCl_2$ (25mM)	1.0 µL (3mM final conc.; adjustable)
dNTP (10mM each)	0.75 µL (3µL of 2.5mM each)
Q solution (5x)	5.0 µL
5' primer (5mM	2.5 µL (0.5mM final conc.)
3' primer (5mM)	2.5 µL (0.5mM final conc.)
Taq DNA polymerase (QIAGEN)	0.25 µL (1.25 units) [1]
*QIAGEN HotStar DNA polymerase (diluted)	0.50 µL
H_2O	to 25 µL (9µL if using 1.0µL DNA)

[1]The success rate was lower when a smaller quantity was used, but the best DNAs work with ≥0.125 µL.

4. Add 1-2 µL of template to each of the tubes.

5. While tubes/plate with template are on ice, add 24 µL of cocktail to each tube, being careful not to cross contaminate. Spin down to bring all liquid to the bottom of the tube.

6. Run appropriate long PCR profile. Generic temperatures and times are:

i. 93ºC infinity (important to go directly from ice to hot block)
ii. 93ºC for 3 min (initial denaturation)
iii. 93ºC for 15 s
iv. 48º—68ºC for 30 s (*T*a should be ~5°C below *T*m of primers)
v. 68º C for 5—20 min (1 min/kb of target)
vi. go to step 3, 34x
vii. 4º C infinity

7. Check reactions by running 2 µL on 1% agarose gel with appropriate size standards.

Primer combinations for long PCR amplification of the chloroplast genome [1,2]

Region	Approx. size	Primers (F|R)[3]	Sequence
1	8kb	trnH.GUG.6R	CCT TRA TCC ACT TGG CTA CAT
		psbK.195R	ACT TAC AGC AGC TTG CCA AAC
2	10.3kb	trnQ.UUG.50R	GGA CGG AAG GAT TCG AAC C
		rpoC2.4805F	GYC GTA TYG ATT GGT TRA AAG G
2a[4]	6.3kb	trnQ.UUG.50R	GGA CGG AAG GAT TCG AAC C
		atpH.17F	CTG CYG CTT CYG TTA TTG CT
2b[4]	4kb	atpF.65R	CGG TAT TAA ACC CGA AAC TCC
		rpoC2.4805F	GYC GTA TYG ATT GGT TRA AAG G
3	7kb	atpI.705R	CRG CTA AAG TTG CAA AAA TAA GAG CT
		rpoC1.1670F	GRG ATC AAA TGG CTG TTC AT
4	9kb	rpoC2.520R	GTT CGT ACA GCA GTA TCY ACA AC
		petN.3R	GCC CAA GCR AGA CTT ACT ATA TCC
5	10.5kb	trnC.GCA.47F	CCC AGT TCA AAT CCG GGT
		psaB.2170F	GCR GCT TTC TTG ATT GCY TC
6	10kb	trnfM.CAU.21R	GGT TAT GAG CCT TGC GAG CTA
		trnT.UGU.17F	GGT TAG AGC ATC GCA TTT GTA ATG
7	10.3kb	rps4.380R	GGT TTG CAR CGA TAA CTT GGK ATA TC
		rbcL.178R	GTC CAT GTA CCA GTA GAR GAT TC
8	9.2kb	rbcL.2F	TGT CAC CAC AAA CAG ARA CTA AAG
		psbJ.3F	GGC YGA TAC TAC TGG AAG RAT
9	9.8kb	petA.920F	CTT CAA GAY CCA TTA CGT GTH CAA G
		psbB.160R	TRC CYT GTC TCC ACA TTG GAT
10	10.9kb	psbB.3F	GGG TTT RCC TTG GTA TCG TGT
		rps3.17F.new	ATC CAC TTG GTT TYM GAC TTG G
11	8.7kb	rpl16.3R	AAC CAA CGA GTC ACA CAC TAA GC
		ycf2.5100R	CAG ATC ATG AAT GTT TGG AAT CCA T
12	10kb	ycf2.2300F	TCG GGA TCC TRA TGC ATA TAG ATA C
		rps12.190F	GTT GCC AGA GTA CGM TTA ACC T
13[5]	11kb	rps12.360R	CCC TTG TTG ACG ATC CTT TAC TC
		ycf1.59R	CCG ACC ACA ACG ACC GAA T
14	11.2kb	trnN.GUU.7R	CCG CTC TAC CAC TGA GCT AC
		ndhA.535F	GCT GCT CAA TCD ATT AGT TAT GAA
14'[6]	7kb	trnR.ACG.15F	GAG GAT TAG AGC ACG TGG
		ccsA.890R	TCC AAG TAA TAA ANG CCC AAG TTT C
15	10.5kb	ndhI.194R	CGA ACR CAT ACT TCA CAA GCA A
		trnN.GUU.7R	CCG CTC TAC CAC TGA GCT AC
16	8.2kb	psbA.640F	GCT ATG CAT GGT TCY TTG GTA AC
		ycf2.5100R	CAG ATC ATG AAT GTT TGG AAT CCA T

[1]Universal primers designed by Mike Moore (Oberlin College); compiled and tested by Simon Uribe-Convers and David Tank (University of Idaho).

[2]Ta should be ~5°C below Tm of primers; however, temperatures of 55°C have worked for all primer combinations.

[3]The name of each primer consists of three parts: (1) the gene in which the primer is anchored in, (2) the approximate position of the primer within that gene (based on all-angiosperm alignment per Moore et al., 2007), and (3) either an "F" or an "R". The F and R designations do not indicate that the primer should be used as a forward or reverse primer; rather, they indicate the 5' to 3' orientation of the primer with respect to the gene. In other words, a primer that is designated as an F primer has its 5' to 3' orientation in the same orientation as the gene (i.e., on the forward strand, or from start to stop), whereas an R primer is oriented in the direction opposite to the 5' to 3' orientation of the gene (i.e., on the reverse strand).

[4]Regions 2a and 2b can be used to amplify region 2 in two pieces

[5]Region 13 represents a large portion of the IR, thus, one amplification for both IRa and IRb

[6]Region 14' amplifies ~2/3 of region 14

Appendix 2

Primer pairs used in the microfluidic experiment
Available also as a supplementary file

Primer Pair	Forward Primer	Forward Primer Sequence	Reverse Primer	Reverse Primer Sequence	Region Variability	Locus (Forward - Reverse)
Chloroplast set						
1	psbZ_36342F	TGGATTGGGTTAGTCTTTCTGG	tRNA-fm(CAU)_36864R	TTGAGGTCACGGGTTCAAAT	0.124748491	psbZ - tRNA-fm(CAU)
2	psbC_35411F	AGAAGTTGGTTAGCTACCTCTCATTT	tRNA-Ser(UGA)_35783R	GGGTTCGAATCCCTCTCTCT	0.074285714	psbC - tRNA-Ser (UGA)
3	57279F	TGCCTGAATTAAGTGAAGTATCACA	accD_57871R	CATTACTGGAACTAGAATTGTCACG	0.071969697	57279F - accD_57871R
4	63714F	CCGTATTCCCAGTCATGTCA	64133R	TTTGTCCATTACTTTCTTGTACGC	0.067385445	Between petA and psbJ
5	psbM_30191F	TGCAGTAGCAATAAATGCAAGAA	tRNA-Asp_(GUC)_30787R	CTGTCAAGGCGGAAGCTG	0.0647	psbM - tRNA-Asp (GUC)
6	accD_57888F	GGGTCGTGACAATTCTAGTTCC	accD_58425R	TGCACATTGTTCACAAATATTCATT	0.052325581	accD
7	psbC_34841F	CGGAGGACATGTATGGTTAGG	psbC_35263R	TCTTGCCAAGGTTGTATGTCTTT	0.051546392	psbC
8	psbD_33131F	TCACTATGACTATAGCCCTTGGTAAA	psbD_33794R	CACGGAATGTATTTGCACCA	0.047285464	psbD
9	rrn16_100656F	CTCGTGCCGTAAGGTGTTG	101219R	CCGTCGAGAAACGAAAGAAG	0.040865385	rrn16_100656F - 101219R
10	psbD_33920F	TCCAATAAACGTTGGTTACATTTC	psbC_34553R	TTCCATACATAACCGAAGAACG	0.038461538	psbD_33920F - psbC_34553R
11	48343F	TCAATAACATTACTCCCACTGAAACT	48768R	CCAAATCTTGGTTATTTCAAGTTGT	0.035532995	tRNA-Leu (UAA) gene and tRNA-Phe (GAA) gene-tRNA-Phe (GAA) gene and ndhJ gene
12	rps2_16398F	CGCTTTGCAGAGATATAAGGTG	rpoc2_16985R	TCAAGAAACGGCTCGAGTTT	0.034700315	rps2_16398F - rpoc2_16985R
13	clpP_exon2_71340F	TTGATAAAGTCGGTTGAGTAGGG	clpP_intron1_71943R	GGAGCGTGAAGTGCAATTAGA	0.033519553	clpP_exon2_71340F - clpP_intron1_71943R
14	ycf1_122287F	ATAAAGAAAGAACCTTCCCATTTAGA	ycf1_122868R	TCGCTTTAATGACGGGAATC	0.027237354	ycf1
15	tRNA-Val(UAC)_52334F	TCCCTCAAAGTTATGGAGTAAGACA	atpE_52948R	AATAATTGAGGCAAATCTAACTCTCC	0.026132404	tRNA-Val (UAC) gene - atpE gene
16	ycf1_122216F	TGATTACCATCAGTTACGATTTGAG	ycf1_122856R	CGGGAATCAAATGGTCAGAA	0.025817556	ycf1_122216F - ycf1_122856R
17	72385F	CCCATTGCGTATTGGTACTTATC	psbB_72954R	GGGCAAATCCAAAGAAGGTT	0.025575448	72385F - psbB_72954R
18	psbK_7863F	AATTCTGCCCTTTCTTCGAGT	psbI_8380R	GCATTACACAATCTCCAAGATGA	0.024691358	psbK-psbI
19	rps15_117994F	GCAAACCTCTTTGGGATGAA	ycf1_118587R	TGAATCGTTATTGGTTTGATACGA	0.024439919	rps15_117994F - ycf1_118587R
20	rpoC2_17557F	TCCTGGAGTGGCCAAATAAG	rpoC2_18248R	AAAGGTTTGATTCGACATTTCC	0.020123839	rpoC2
21	atpF_13237F	GCCCAAAGAAACGAAAGAATC	atpH_13749R	AATTGGTCAAGGGACTGCTG	0.019512195	atpF - atpH
22	6587F	TGAATCCGCCAAATAACTCA	7152R	TCGATTATTGAAATATTACCCAACC	0.019493177	Between rps16 and tRNA-Gln (UUG)
23	ycf1_118709F	CAAGGTCCTTATTGTGAAGTGAAA	ycf1_119362R	TGGATCCATTCTCTCAGAACAC	0.019169329	ycf1
24	matK_2573F	CGCAATCAGAGGAATAATTGG	matK_3079R	AAGAAAGCCAGCTCCTCCTT	0.018099548	matK
25	4720F	GACCGGGTTCACTTATTACGTT	rps16_5273R	TGCTCAACCCACAACAACTG	0.017955801	Between matK and rps16 - rps16
26	tRNA-Lys(UUU)I_4051F	TCCTTCCCATATCAGGCACT	4676R	TTTAGTTCTGGCCCATAGGTTT	0.017699115	tRNA-Lys (UUU) intron - Between tRNA-Lys(UUU) and rps16
27	rps16_6142F	TTTCTACTCGAGCTCCATCGT	6568R	TGAGTTATTTGGCGGATTCA	0.016759777	rps16 - Between rps16 and tRNA-Gln (UUG)
28	atpB_54324F	CACGGCGATAAGGTGCTAA	54871R	GATCAACGTGCTATCGGACA	0.015936255	atpB gene - Between atpB and rbcL
29	tRNA-Lys(UUU)_2016F	TTTCATTGCACACGGCTTT	matK_2633R	CCCTTCAGTGGTACGGAGTC	0.015384615	tRNA-Lys (UUU) intron and matK
30	cemA_61963F	TCCACAAAGAAACGATCCAA	62483R	CTGGAATCATTTGACTGGAATATG	0.015250545	cemA gene - Between cemA and petA
31	301F	TCAAACTGATCTTGCTTACGATG	psbA_807R	CGTGGCCTGTAGTGGGTATC	0.014962594	Between tRNA-His (GUG) and psbA
32	ndhA_116264F	TTGATATGACTGCTATGGATGGTC	ndhH_116917R	CCAACAAGCTCTGGAAGGAA	0.014827018	ndhA exon - ndhH
33	ndhA_114844F	CGGTTTGATAACCTGCCACT	ndhA_115487R	CAATCCAGAGTATGTTCCTATCCA	0.013461538	ndhA exon - ndhA intron
34	107350F	AACTTGGGTGTGGGTCTTTG	tRNA-Asn(GUU)_107858R	GCTGTTAACCGATTGGTCGT	0.013043478	107350F - tRNA-Asn(GUU)_107858R
35	rpoC2_18414F	GGTTTACCGGTTCCATAAAGG	rpoC2_18902R	TTGATTGAGTATCGAGGAAGCA	0.012437811	rpoC2
36	ndhF_109439F	CCAATCCCATTCACAATTCC	ndhF_110059R	GGGTGTTCGAAACAAGTTAACAA	0.012280702	ndhF

Primer Pair	Forward Primer	Forward Primer Sequence	Reverse Primer	Reverse Primer Sequence	Region Variability	Locus (Forward - Reverse)
Chloroplast set						
37	69563F	TTTCACAACGAGACCCACCT	70067R	CATAATCATCCGGTTAGGATCAA	0.012254902	Between rpl20 and rps12 - Between rps12 and clpP
38	14350F	ATTCCAACCGACCGAATACA	atpI_14964R	CGGAAATATCTTAGCGGATGA	0.012068966	Between atpH and atpI - atpI
39	petD_intron_76880F	TGCATTCATTTCCTCTGCAT	petD_exon2_77491R	GGATCTGCTGGTTCACCAAT	0.011965812	petD_intron_76880F - petD_exon2_77491R
40	ndhF_110536F	CCATTCCCAGAGCTAACATCA	ndhF_111056R	TTGGGAATTGGTTGGAATGT	0.011547344	ndhF
41	atpF_12744F	AGGTCGTCGATTCCGTATTG	atpF_13219R	CCGATTCTTTCGTTTCTTTGG	0.011441648	atpF
42	61121F	GTTGGACGTGATACTTTAAGATGC	61595R	TCAAACATGAGTTCTAGACACGGTA	0.011061947	Between ycf4 and cemA gene
43	clpP_intron_70522F	ACCGTACGGGCATCATCTT	clpP_intron_70982R	ACTATGCCTTCGCCATATGAA	0.010666667	clpP_intron_70522F - clpP_intron_70982R
44	ndhJ_49618F	CCATGCTTAACTAACCAAGCAGA	ndhK_50285R	TTTATGGCCACTTCTCTACGG	0.010638298	ndhJ gene - ndhK gene
45	45315F	TCTTGGAACCTCGAAAGAAAGA	rps4_45806R	TCGTCAGACTTAAACCTAAATTCAAA	0.010416667	Between ycf3 and tRNA-Ser (GGA) - rps4
46	74994F	TTTGGTAGTTCGATCGTGGA	petB_75678R	TCATCTCGTACAGCTCAAGCA	0.010169492	74994F - petB_75678R
47	atpA_10694F	CAATTCACGCAATCGTTGAC	atpA_11299R	GGTGGCGCAGGTAGTAACTAAT	0.009560229	atpA
48	rps8_80497F	CAGCAATAGTGTCTCTACCCATGA	81105R	ACCAACAGAAACGAGTCTTCG	0.008264463	rps8_80497F - 81105R
Nuclear set						
1	AT1G02420_Bar_PPR_F	TCGACACCATGCTCTACGTC	AT1G02420_Bar_PPR_R	GTTGACCTGCCAATCCTAGC		AT1G02420 - AT1G02420
2	AT1G05600_Bar_PPR_F	GGCGAGTGACAGAGATGACA	AT1G05600_Bar_PPR_R	CCCTCCAGAACATTAAATACAACA		AT1G05600 - AT1G05600
3	AT1G10330_Bar_PPR_F	AAACGGCAGGGAAATTCAG	AT1G10330_Bar_PPR_R	CATGTGCCATTGCTTTCCT		AT1G10330 - AT1G10330
4	AT1G31430_Bar_PPR_F	CACGTACCCTTTCGTGTTGA	AT1G31430_Bar_PPR_R	CTTGTCCGGCTTAATTCTATTCA		AT1G31430 - AT1G31430
5	AT1G74600_Bar_PPR_F	GGGTTCGTTTCAGATGGGTA	AT1G74600_Bar_PPR_R	CCAGCAATGGATTTGTGATG		AT1G74600 - AT1G74600
6	AT1G80550-1_Bar_PPR_F	CTCGAAACGCTCAATTGCTA	AT1G80550-1_Bar_PPR_R	CTTTCTCCACGCCCTTCTTA		AT1G80550 - AT1G80550
7	AT2G15690_Bar_PPR_F	GCAGAGAGGGTAAGGTCAAGG	AT2G15690_Bar_PPR_R	TTCGAAACTCGGGTATCCTG		AT2G15690 - AT2G15690
8	AT2G18940_Bar_PPR_F	TACAAGCCCGACTTGGTTCT	AT2G18940_Bar_PPR_R	CTTGCAATACCCATCCACAA		AT2G18940 - AT2G18940
9	AT2G33680_Bar_PPR_F	TAACCATGGCGAGTGTCTTG	AT2G33680_Bar_PPR_R	CCTCACTCGTTCCACCTCAT		AT2G33680 - AT2G33680
10	AT3G14730-2_Bar_PPR_F	TATGAGAGAAGCCCGTTTGG	AT3G14730-2_Bar_PPR_R	CTCAGCTCTGCCACCTCTTC		AT3G14730 - AT3G14730
11	AT3G46790-1_Bar_PPR_F	CTATGCTCTCAAGGCGTGTG	AT3G46790-1_Bar_PPR_R	AAGAAGGTCGACCATGCAA		AT3G46790 - AT3G46790
12	AT4G01570_Bar_PPR_F	TTCGATTATGAGCTCGTTTGTC	AT4G01570_Bar_PPR_R	TCCAGCATCATCTTCAGCAC		AT4G01570 - AT4G01570
13	AT5G16420_Bar_PPR_F	TACCACTCCGTCATCCACAA	AT5G16420_Bar_PPR_R	TTCCTCCATCTCATCCATCAC		AT5G16420 - AT5G16420
14	AT5G39980_Bar_PPR_F	TTGGGTATCGAACCGAATGT	AT5G39980_Bar_PPR_R	CAAGGATACGGATGGCAGTT		AT5G39980 - AT5G39980
15	AT1G66345_Bar_PPR_F	GGTTAAGGCATGGGACTCAA	AT1G66345_Bar_PPR_R	ACCGAACGAGCATCTTCAAT		AT1G66345 - AT1G66345
16	At1g10500_BAR_COSII_F	GAAAATGCTAGACCCGACGA	At1g10500_BAR_COSII_R	ATTTACCACAACCGCACGTT		At1g10500 - At1g10500
17	At1g02140_BAR_COSII_F	GCAAGCTCCGATACGCTAAC	At1g02140_BAR_COSII_R	TTCGTAAACCCTCAGGATCTTT		At1g02140 - At1g02140
18	At1g14300_BAR_COSII_F	TGAAGTTGGTGAGGTCATTTTG	At1g14300_BAR_COSII_R	GATTTTCAGCTTCAAAACAGCA		At1g14300 - At1g14300
19	At1g71260_BAR_COSII_F	GTTGACATTCTCGCCTGCTA	At1g71260_BAR_COSII_R	GACAGGAACAACAAAACGATCA		At1g71260 - At1g71260
20	At1g63980_BAR_COSII_F	GGGCTTGGAAAAGAAAAACA	At1g63980_BAR_COSII_R	CATTTACAAGCTTCCCCCTCT		At1g63980 - At1g63980
21	At2g03510_BAR_COSII_F	AGAGGAGGTGCCCTTTTGA	At2g03510_BAR_COSII_R	TGTGCAGTCAGCCTGAAGAG		At2g03510 - At2g03510
22	At2g40980_BAR_COSII_F	AAGGTGTTCGATTCCAGCAG	At2g40980_BAR_COSII_R	TCCGAGCTTGTTCCAAATTC		At2g40980 - At2g40980
23	At2g27450_BAR_COSII_F	AGGATGCAGAATCTGGCAAA	At2g27450_BAR_COSII_R	GAAACCACTGATCCCAGCAT		At2g27450 - At2g27450
24	At3g08950_BAR_COSII_F	GCAAGGACCCTCTGTTGGTA	At3g08950_BAR_COSII_R	TGATCCACAAGATAGTCTGAACCT		At3g08950 - At3g08950
25	At3g09740_BAR_COSII_F	AGCCAATTTAGGCAAGAGTATGA	At3g09740_BAR_COSII_R	TCGATGCAGAAATTCCGACT		At3g09740 - At3g09740
26	At5g11480_BAR_COSII_F	ATGCGAGTATTCCTGCCAAG	At5g11480_BAR_COSII_R	CCTTGGTTCGTGACGCTACT		At5g11480 - At5g11480
27	At5g12200_BAR_COSII_F	AGCCCTATGGGATCCTGACT	At5g12200_BAR_COSII_R	AGTTGAATTGAAGGTGCAGTGA		At5g12200 - At5g12200

Primer Pair	Forward Primer	Forward Primer Sequence	Reverse Primer	Reverse Primer Sequence	Region Variability	Locus (Forward - Reverse)
Nuclear set						
28	At5g14520_BAR_C OSII_F	AATGAACCTGGTGCTTT GATG	At5g14520_BAR_C OSII_R	CAGGGATGATGAACACT AACGA		At5g14520 - At5g14520
29	At5g17990_BAR_C OSII_F	TGAGGGACTCGATGAGA TGA	At5g17990_BAR_C OSII_R	CGTTTAAGCACTTCAGCA TTGT		At5g17990 - At5g17990
30	At5g23120_BAR_C OSII_F	TGCTGTGCAGGAGACTG TTT	At5g23120_BAR_C OSII_R	CGGCTTGAGACAGCAAC ATA		At5g23120 - At5g23120
31	At5g23240_BAR_C OSII_F	TTGAAATGTGCCTTGTTT GC	At5g23240_BAR_C OSII_R	CCCGTGGTTGCTTAGAC ATT		At5g23240 - At5g23240
32	At5g41760_BAR_C OSII_F	TCACAAAGCAACGGTGG ATA	At5g41760_BAR_C OSII_R	TCGACGGGATCGGATAT AAT		At5g41760 - At5g41760
33	At5g42740_BAR_C OSII_F	AGGTGAAGTGGTCAGCA ACC	At5g42740_BAR_C OSII_R	CCTTCAAGCGGTTCTCCT TT		At5g42740 - At5g42740
34	At5g47390_BAR_C OSII_F	GGACGTGTCCGAATCGA G	At5g47390_BAR_C OSII_R	TCATCTGCCACAATATCA AACA		At5g47390 - At5g47390
35	ETS-2-F	GCACATGGTGTTGTTTG GTT	ETS-2-R	AATGAGCCATTCGCAGTT TC		External Transcribed Spacer (ETS) - External Transcribed Spacer (ETS)
36	ITS-2-F	AATGGTCCGGTGAAGTG TTC	ITS-2-R	GTTCGCTCGCCGTTACT AAG		Internal Transcribed Spacer (ITS) - Internal Transcribed Spacer (ITS)
37	Photo2-Exon14-F	CATGGCAAACAGATGAC CAC	Photo2-Exon12-R	TTCTCAACAGGTATTCCC TAAACC		Phototropin 2 - Phototropin 2
38	AT1G68930_Bar_P PR_F	GTGTGGGATGGTGGAA GACT	AT1G68930_Bar_P PR_R	ATCGAGAATTGGCTTGAT CC		AT1G68930 - AT1G68930
39	AT3G18020_Bar_P PR_F	TTCCACCGAGCTCCTAA ATG	AT3G18020_Bar_P PR_R	TTCTCCATGTCACAGCAT CC		AT3G18020 - AT3G18020
40	AT5G18475_Bar_P PR_F	CCCAATACGTGCATCTT CAA	AT5G18475_Bar_P PR_R	CTCGTTCGACGTCCCAA A		AT5G18475 - AT5G18475
41	AT1G20300_Bar_P PR_F	GGTGGTGTCGTGCGAAT AAT	AT1G20300_Bar_P PR_R	TTTCCAAAGCTTGATCAC CA		AT1G20300 - AT1G20300
42	AT1G53330-2_Bar_PPR_F	AATGCCATGAAGGTCTT GGA	AT1G53330-2_Bar_PPR_R	CTTCCTCAACCTTCCCTT CC		AT1G53330 - AT1G53330
43	AT2G22410_Bar_P PR_F	GTCGTGTGAGGGACTTG GTT	AT2G22410_Bar_P PR_R	GGCACTAGACCAGCTTC GAC		AT2G22410 - AT2G22410
44	At1g54390_BAR_C OSII_F	AAACGTCTTGATGAGGA TCTGAA	At1g54390_BAR_C OSII_R	ATTCACCTCCTTGGCACT GT		At1g54390 - At1g54390
45	At5g13420_BAR_C OSII_F	GGAAATGGAAGCATTGT CAAG	At5g13420_BAR_C OSII_R	AAGTTTGCATGCATCTTG GA		At3g56460 - At3g56460
46	At5g13710_BAR_C OSII_F	TTTGCTGCATACGAGTG GTG	At5g13710_BAR_C OSII_R	CAACAAGTCCTTCTGCA GCTT		At5g13420 - At5g13420
47	At5g38530_BAR_C OSII_F	CCTGTTATTAGAGCGGT TGAGC	At5g38530_BAR_C OSII_R	TCCACACACTGCCATTAG GA		At5g13710 - At5g13710
48	At3g56460_BAR_C OSII_F	GATTTGGTTGCAGCTGG TG	At3g56460_BAR_C OSII_R	GCAATGTTTGCTGGGATT ACA		At5g38530 - At5g38530

Appendix 3

Sequences for the sequencing adapters, barcodes, and conserved sequences used in the microfluidic experiment
Available also as a supplementary file

Sequencing Adapters		**Barcodes**		**Conserved Sequences**	
Adapter	Sequence	Adapter	Sequence	Name	Sequence
P5	AATGATACGGCGACCACCGAGATCTACAC	P5	TAGATCGC	CS1	ACACTGACGACATGGTTCTACA
P7	CAAGCAGAAGACGGCATACGAGAT	P5	CTCTCTAT	CS2	TACGGTAGCAGAGACTTGGTCT
		P5	TATCCTCT		
		P5	AGAGTAGA		
		P5	GTAAGGAG		
		P5	ACTGCATA		
		P5	AAGGAGTA		
		P5	CTAAGCCT		
		P5	TGAACCTT		
		P5	TGCTAAGT		
		P5	TGTTCTCT		
		P5	TAAGACAC		
		P5	CTAATCGA		
		P5	CTAGAACA		
		P5	TAAGTTCC		
		P5	TAGACCTA		
		P5	TATAGCCT		
		P5	ATAGAGGC		
		P5	CCTATCCT		
		P5	GGCTCTGA		
		P5	AGGCGAAG		
		P5	TAATCTTA		
		P5	CAGGACGT		
		P5	GTACTGAC		
		P7	TAAGGCGA		
		P7	CGTACTAG		
		P7	AGGCAGAA		
		P7	TCCTGAGC		
		P7	GGACTCCT		
		P7	TAGGCATG		
		P7	CTCTCTAC		
		P7	CAGAGAGG		
		P7	GCTACGCT		
		P7	CGAGGCTG		
		P7	AAGAGGCA		
		P7	GTAGAGGA		

Sequencing Adapters		**Barcodes**		**Conserved Sequences**	
Adapter	Sequence	Adapter	Sequence	Name	Sequence
		P7	ATCACGAC		
		P7	ACAGTGGT		
		P7	CAGATCCA		
		P7	ACAAACGG		
		P7	ACCCAGCA		
		P7	AACCCCTC		
		P7	CCCAACCT		
		P7	CACCACAC		
		P7	GAAACCCA		
		P7	TGTGACCA		
		P7	AGGGTCAA		
		P7	AGGAGTGG		
		P7	ATTACTCG		
		P7	TCCGGAGA		
		P7	CGCTCATT		
		P7	GAGATTCC		
		P7	ATTCAGAA		
		P7	GAATTCGT		
		P7	CTGAAGCT		
		P7	TAATGCGC		
		P7	CGGCTATG		
		P7	TCCGCGAA		
		P7	TCTCGCGC		
		P7	AGCGATAG		

Appendix 4

Information for the samples used in this study. Herbarium codes follow the Index Herbariorum
Available also as a supplementary file

Species	DNA Accession	Collector and Voucher No.	Herbarium	Country	State	Latitude, Longitude
Bartsia cf. integrifolia Wedd.	2013-193	Uribe-Convers 2012-117	ID	Peru	Junín	-10.1534, -74.2537
Bartsia cf. laniflora Benth.	2013-518	Uribe-Convers 2013-137 (Ind 1)	ID	Colombia	Caldas	4.9709, -75.3464
Bartsia cf. laniflora Benth.	2013-519	Uribe-Convers 2013-137 (Ind 2)	ID	Colombia	Caldas	4.9709, -75.3464
Bartsia cf. laniflora Benth.	2013-520	Uribe-Convers 2013-137 (Ind 3)	ID	Colombia	Caldas	4.9709, -75.3464
Bartsia cf. laticrenata Benth.	2011-219	Uribe-Convers 2011-078	ID	Ecuador	Tungurahua	-1.0899, -78.4387
Bartsia cf. melampyroides	2013-306	Uribe-Convers 2012-041	ID	Peru	Cusco	-12.7037, -71.9465
Bartsia cf. melampyroides (Kunth) Benth.	2011-261	Uribe-Convers 2011-121	ID	Peru	La Libertad	-7.9951, -78.4420
Bartsia cf. melampyroides (Kunth) Benth.	2012-126	Uribe-Convers 2011-154	ID	Peru	Cajamarca	-7.2463, -78.4697
Bartsia cf. melampyroides (Kunth) Benth.	2013-118	Uribe-Convers 2011-235	ID	Peru	Junín	-11.9790, -75.0942
Bartsia cf. melampyroides (Kunth) Benth.	2013-316	Uribe-Convers 2012-064	ID	Peru	Apurimac	-12.4105, -71.1526
Bartsia cf. mutica (Kunth) Benth.	2013-092	Uribe-Convers 2011-195	ID	Peru	Ancash	-9.0603, -77.6299
Bartsia cf. orthocarpiflora Benth.	2011-198	Uribe-Convers 2011-057	ID	Ecuador	Cotopaxi	-0.6419, -78.5084
Bartsia cf. orthocarpiflora Benth. ssp. villosa Molau	2013-505	Uribe-Convers 2013-126 (Ind 1)	ID	Colombia	Cundinamarca	4.5616, -74.0151
Bartsia cf. orthocarpiflora Benth. ssp. villosa Molau	2013-506	Uribe-Convers 2013-126 (Ind 2)	ID	Colombia	Cundinamarca	4.5616, -74.0151
Bartsia cf. patens Benth.	2011-272	Uribe-Convers 2011-132	ID	Peru	La Libertad	-8.1036, -78.2947
Bartsia cf. patens Benth.	2011-264	Uribe-Convers 2011-124	ID	Peru	La Libertad	-7.9966, -78.4230
Bartsia cf. pauciflora Molau	2013-188	Uribe-Convers 2012-107	ID	Peru	Junín	-10.3913, -74.3631
Bartsia cf. pauciflora Molau	2013-190	Uribe-Convers 2012-110	ID	Peru	Junín	-10.4462, -74.3734
Bartsia cf. pauciflora Molau	2013-115	Uribe-Convers 2011-229	ID	Peru	Junín	-11.9790, -75.0942
Bartsia cf. pauciflora Molau	2013-119	Uribe-Convers 2011-237	ID	Peru	Huancavelica	-12.3367, -75.0037
Bartsia cf. pauciflora Molau	2013-333	Uribe-Convers 2012-113	ID	Peru	Junín	-10.1746, -74.3008
Bartsia cf. pedicularoides	2013-304	Uribe-Convers 2012-039	ID	Peru	Cusco	-12.7041, -71.9464
Bartsia cf. pedicularoides	2013-508	Uribe-Convers 2013-128	ID	Colombia	Cundinamarca	4.5609, -74.0216
Bartsia cf. pedicularoides	2013-509	Uribe-Convers 2013-129 (Ind 1)	ID	Colombia	Antioquia	6.4453, -76.0849
Bartsia cf. pedicularoides	2013-510	Uribe-Convers 2013-129 (Ind 2)	ID	Colombia	Antioquia	6.4453, -76.0849
Bartsia cf. pedicularoides	2013-511	Uribe-Convers 2013-129 (Ind 3)	ID	Colombia	Antioquia	6.4453, -76.0849
Bartsia cf. pedicularoides Benth.	2011-189	Uribe-Convers 2011-047	ID	Ecuador	Pichincha	-0.1968, -78.1273
Bartsia cf. pedicularoides Benth.	2013-161	Uribe-Convers 2012-036	ID	Peru	Cusco	-12.7124, -71.9500
Bartsia cf. pedicularoides Benth.	2011-041	Uribe-Convers 2010-040	ID	Colombia	Boyacá	5.9124, -73.0769
Bartsia cf. pedicularoides Benth.	2011-058	Uribe-Convers 2010-059	ID	Colombia	Santander	6.9956, -72.6831
Bartsia cf. pedicularoides Benth.	2013-378	Uribe-Convers 2013-103	ID	Colombia	Cundinamarca	4.2905, -74.2070
Bartsia cf. peruviana	2011-269	Uribe-Convers 2011-129	ID	Peru	La Libertad	-8.0879, -78.2918
Bartsia cf. peruviana Walp.	2013-227	Uribe-Convers 2013-042	ID	Bolivia	Tarija	-21.5000, -64.9086
Bartsia cf. peruviana Walp.	2014-113	Uribe-Convers 2013-042 (Ind 2)	ID	Bolivia	Tarija	-21.5000, -64.9086
Bartsia cf. peruviana Walp.	2014-114	Uribe-Convers 2013-042 (Ind 3)	ID	Bolivia	Tarija	-21.5000, -64.9086
Bartsia cf. pyricarpa Molau	2013-206	Uribe-Convers 2013-016	ID	Bolivia	La Paz	-15.7334, -68.6376
Bartsia cf. pyricarpa Molau	2013-334	Uribe-Convers 2012-114	ID	Peru	Junín	-10.1746, -74.3008
Bartsia cf. pyricarpa Molau	2014-086	Uribe-Convers 2013-016 (Ind 2)	ID	Bolivia	La Paz	-15.7334, -68.6376
Bartsia cf. ramosa Molau	2013-503	Uribe-Convers 2013-125 (Ind 1)	ID	Colombia	Cundinamarca	4.5603, -74.0131
Bartsia cf. ramosa Molau	2013-504	Uribe-Convers 2013-125 (Ind 2)	ID	Colombia	Cundinamarca	4.5603, -74.0131
Bartsia cf. remota Molau	2012-145	Uribe-Convers 2011-173	ID	Peru	Cajamarca	-6.8692, -78.1130
Bartsia cf. rigida Molau	2013-159	Uribe-Convers 2012-033	ID	Peru	Cusco	-12.7044, -71.9460
Bartsia cf. rigida Molau	2013-172	Uribe-Convers 2012-068	ID	Peru	Apurimac	-12.3565, -72.4363
Bartsia cf. santolinifolia (Kunth) Benth.	2013-123	Uribe-Convers 2012-004	ID	Colombia	Boyacá	5.9270, -72.9135
Bartsia cf. sericea Molau	2011-249	Uribe-Convers 2011-109	ID	Ecuador	Azuay	-2.7813, -79.1853
Bartsia cf. sericea Molau	2012-124	Uribe-Convers 2011-152	ID	Peru	Cajamarca	-6.7760, -78.6428
Bartsia cf. serrata Molau	2013-124	Uribe-Convers 2012-005	ID	Peru	Arequipa	-15.5992, -70.6344

Species	DNA Accession	Collector and Voucher No.	Herbarium	Country	State	Latitude, Longitude
Bartsia cf. stricta	2011-056	Uribe-Convers 2010-056	ID	Colombia	Santander	7.2756, -72.8857
Bartsia cf. stricta (Kunth) Benth.	2011-200	Uribe-Convers 2011-059	ID	Ecuador	Cotopaxi	-0.6381, -78.4856
Bartsia cf. strigosa Molau	2013-168	Uribe-Convers 2012-057	ID	Peru	Cusco	-12.5606, -71.6402
Bartsia cf. tenuis Molau	2013-202	Uribe-Convers 2013-012	ID	Bolivia	La Paz	-15.7351, -68.6573
Bartsia cf. tenuis Molau	2013-226	Uribe-Convers 2013-041	ID	Bolivia	Tarija	-21.4595, -64.8645
Bartsia cf. tenuis Molau	2013-202	Uribe-Convers 2013-012	ID	Bolivia	La Paz	-15.7351, -68.6573
Bartsia cf. tenuis Molau	2013-226	Uribe-Convers 2013-041 (Ind 1)	ID	Bolivia	Tarija	-21.4595, -64.8645
Bartsia cf. tenuis Molau	2013-299	Uribe-Convers 2012-023	ID	Peru	Puno	-14.5924, -69.6603
Bartsia cf. tenuis Molau	2014-112	Uribe-Convers 2013-041 (Ind 2)	ID	Bolivia	Tarija	-21.4595, -64.8645
Bartsia cf. thiantha Diels	2011-210	Uribe-Convers 2011-069	ID	Ecuador	Cotopaxi	-0.9263, -78.8326
Bartsia cf. thiantha Diels	2011-211	Uribe-Convers 2011-070	ID	Ecuador	Cotopaxi	-0.9263, -78.8326
Bartsia cf. thiantha Diels	2011-214	Uribe-Convers 2011-073	ID	Ecuador	Cotopaxi	-0.8658, -78.9108
Bartsia cf. thiantha Diels	2012-118	Uribe-Convers 2011-145	ID	Peru	Cajamarca	-6.9937, -78.8142
Bartsia cf. thiantha Diels	2013-302	Uribe-Convers 2012-035	ID	Peru	Cusco	-12.7124, -71.9500
Bartsia cf. tricolor Molau	2013-095	Uribe-Convers 2011-200	ID	Peru	Ancash	-9.0478, -77.6100
Bartsia cf. tricolor Molau	2013-126	Uribe-Convers 2012-007	ID	Peru	Arequipa	-15.6098, -70.6823
Bartsia cf. tricolor Molau	2013-129	Uribe-Convers 2012-010	ID	Peru	Arequipa	-15.9251, -70.5138
Bartsia cf. tricolor Molau	2013-298	Uribe-Convers 2012-022	ID	Peru	Puno	-15.7964, -68.5346
Bartsia cf. weberbaueri Diels	2013-098	Uribe-Convers 2011-205	ID	Peru	Ancash	-9.0512, -77.6011
Bartsia chilensis Benth.	2013-287	Espíndola 12-001	ID	Chile		-33.1100, -71.6306
Bartsia chilensis Benth.	2013-288	Espíndola 12-002	ID	Chile		-33.1100, -71.6306
Bartsia chilensis Benth.	2013-289	Espíndola 12-003	ID	Chile		-33.1100, -71.6306
Bartsia crenata Molau	2013-178	Uribe-Convers 2012-086	ID	Peru	Huancavelica	-11.3555, -73.0832
Bartsia crenata Molau	2013-178	Uribe-Convers 2012-086	ID	Peru	Huancavelica	-11.3555, -73.0832
Bartsia crenata Molau	2013-284	Wood, JRI 18787	LPB	Bolivia		
Bartsia crenata Molau	2013-285	Isabell Hensen 492	LPB	Bolivia		
Bartsia crenata Molau	2013-286	Isabell Hensen 588	LPB	Bolivia		
Bartsia crenata Molau	2013-331	Uribe-Convers 2012-098	ID	Peru	Junín	-10.0161, -74.8994
Bartsia crenoloba Wedd.	2013-197	Uribe-Convers 2013-006	ID	Bolivia	La Paz	-16.4665, -68.1526
Bartsia crenoloba Wedd.	2013-200	Uribe-Convers 2013-010	ID	Bolivia	La Paz	-16.1651, -68.8393
Bartsia crenoloba Wedd.	2013-203	Uribe-Convers 2013-013	ID	Bolivia	La Paz	-15.7296, -68.6537
Bartsia crenoloba Wedd.	2013-208	Uribe-Convers 2013-018	ID	Bolivia	La Paz	-15.5952, -69.0703
Bartsia crenoloba Wedd.	2013-211	Uribe-Convers 2013-022	ID	Bolivia	La Paz	-15.1942, -69.0114
Bartsia crenoloba Wedd.	2009-108	K09-13	K	Bolivia	Murillo	-15.4667, -67.9167
Bartsia crenoloba Wedd.	2013-295	Uribe-Convers 2012-019	ID	Peru	Moquegua	-15.0101, -69.3047
Bartsia crenoloba Wedd.	2013-300	Uribe-Convers 2012-024	ID	Peru	Puno	-14.6721, -69.6513
Bartsia crenoloba Wedd.	2014-078	Uribe-Convers 2013-006 (Ind 2)	ID	Bolivia	La Paz	-16.4665, -68.1526
Bartsia crenoloba Wedd.	2014-083	Uribe-Convers 2013-010 (Ind 2)	ID	Bolivia	La Paz	-16.1651, -68.8393
Bartsia crenoloba Wedd.	2014-084	Uribe-Convers 2013-013 (Ind 2)	ID	Bolivia	La Paz	-15.7296, -68.6537
Bartsia crenoloba Wedd.	2014-092	Uribe-Convers 2013-022 (Ind 2)	ID	Bolivia	La Paz	-15.1942, -69.0114
Bartsia crisafullii N. H. Holmgren	2011-265	Uribe-Convers 2011-125	ID	Peru	La Libertad	-7.9966, -78.4230
Bartsia crisafullii N. H. Holmgren	2012-131	Uribe-Convers 2011-159	ID	Peru	Cajamarca	-7.1960, -78.5661
Bartsia crisafullii N. H. Holmgren	2009-107	K09-11	K	Peru	Cajamarca	-4.3333, -78.7333
Bartsia diffusa Benth.	2013-117	Uribe-Convers 2011-234	ID	Peru	Junín	-11.9790, -75.0942
Bartsia diffusa Benth.	2013-180	Uribe-Convers 2012-091	ID	Peru	Huancavelica	-11.4002, -73.0184
Bartsia diffusa Benth.	2013-329	Uribe-Convers 2012-096	ID	Peru	Junín	-10.0616, -74.9331
Bartsia diffusa Benth.	2013-337	Uribe-Convers 2012-121	ID	Peru	Junín	-10.1406, -74.1266
Bartsia elachophylla Diels	2013-171	Uribe-Convers 2012-066	ID	Peru	Apurimac	-12.4166, -71.1683
Bartsia elongata Wedd.	2013-127	Uribe-Convers 2012-008	ID	Peru	Arequipa	-15.6356, -69.6789

Species	DNA Accession	Collector and Voucher No.	Herbarium	Country	State	Latitude, Longitude
Bartsia elongata Wedd.	2013-151	Uribe-Convers 2012-014	ID	Peru	Moquegua	-15.3971, -70.7752
Bartsia elongata Wedd.	2013-212	Uribe-Convers 2013-023	ID	Bolivia	La Paz	-14.8044, -69.1822
Bartsia elongata Wedd.	2013-236	Uribe-Convers 2013-002	ID	Bolivia	La Paz	-16.3700, -68.1625
Bartsia elongata Wedd.	2014-074	Uribe-Convers 2013-002 (Ind 2)	ID	Bolivia	La Paz	-16.3700, -68.1625
Bartsia elongata Wedd.	2014-093	Uribe-Convers 2013-023 (Ind 2)	ID	Bolivia	La Paz	-14.8044, -69.1822
Bartsia elongata Wedd.	2014-094	Uribe-Convers 2013-023 (Ind 3)	ID	Bolivia	La Paz	-14.8044, -69.1822
Bartsia fiebrigii Diels	2013-220	Uribe-Convers 2013-032	ID	Bolivia	La Paz	-16.9267, -67.1665
Bartsia fiebrigii Diels	2013-221	Uribe-Convers 2013-034	ID	Bolivia	La Paz	-17.0022, -67.2690
Bartsia fiebrigii Diels	2013-223	Uribe-Convers 2013-038	ID	Bolivia	Potosí	-19.8458, -65.7097
Bartsia fiebrigii Diels	2013-228	Uribe-Convers 2013-043	ID	Bolivia	Tarija	-21.4217, -64.4295
Bartsia fiebrigii Diels	2013-230	Uribe-Convers 2013-045	ID	Bolivia	Chuquisaca	-18.9843, -65.3441
Bartsia fiebrigii Diels	2014-101	Uribe-Convers 2013-032 (Ind 2)	ID	Bolivia	La Paz	-16.9267, -67.1665
Bartsia fiebrigii Diels	2014-102	Uribe-Convers 2013-032 (Ind 3)	ID	Bolivia	La Paz	-16.9267, -67.1665
Bartsia fiebrigii Diels	2014-103	Uribe-Convers 2013-034 (Ind 2)	ID	Bolivia	La Paz	-17.0022, -67.2690
Bartsia fiebrigii Diels	2014-104	Uribe-Convers 2013-034 (Ind 3)	ID	Bolivia	La Paz	-17.0022, -67.2690
Bartsia fiebrigii Diels	2014-108	Uribe-Convers 2013-038 (Ind 2)	ID	Bolivia	Potosí	-19.8458, -65.7097
Bartsia fiebrigii Diels	2014-117	Uribe-Convers 2013-045 (Ind 2)	ID	Bolivia	Chuquisaca	-18.9843, -65.3441
Bartsia fiebrigii Diels	2014-118	Uribe-Convers 2013-045 (Ind 3)	ID	Bolivia	Chuquisaca	-18.9843, -65.3441
Bartsia filiformis Wedd.	2013-196	Uribe-Convers 2013-005	ID	Bolivia	La Paz	-16.3313, -67.9821
Bartsia filiformis Wedd.	2013-204	Uribe-Convers 2013-014	ID	Bolivia	La Paz	-15.7296, -68.6537
Bartsia filiformis Wedd.	2013-207	Uribe-Convers 2013-017	ID	Bolivia	La Paz	-15.7334, -68.6376
Bartsia filiformis Wedd.	2013-214	Uribe-Convers 2013-025	ID	Bolivia	La Paz	-16.3328, -67.9762
Bartsia filiformis Wedd.	2013-216	Uribe-Convers 2013-027	ID	Bolivia	La Paz	-16.3280, -67.9457
Bartsia filiformis Wedd.	2014-077	Uribe-Convers 2013-005 (Ind 2)	ID	Bolivia	La Paz	-16.3313, -67.9821
Bartsia filiformis Wedd.	2014-085	Uribe-Convers 2013-014 (Ind 2)	ID	Bolivia	La Paz	-15.7296, -68.6537
Bartsia filiformis Wedd.	2014-087	Uribe-Convers 2013-017 (Ind 2)	ID	Bolivia	La Paz	-15.7334, -68.6376
Bartsia filiformis Wedd.	2014-096	Uribe-Convers 2013-025 (Ind 2)	ID	Bolivia	La Paz	-16.3328, -67.9762
Bartsia flava Molau	2013-086	Uribe-Convers 2011-183	ID	Peru	Cajamarca	-6.9167, -78.6143
Bartsia flava Molau ssp. minor Molau	2014-066	Sylvester 2114	Z	Peru	Cusco	-13.2546, -71.1606
Bartsia flava Molau ssp. minor Molau	2014-067	Sylvester 1623	Z	Peru	Cusco	-13.2439, -71.1622
Bartsia flava Molau ssp. minor Molau	2014-068	Sylvester 2135	Z	Peru	Cusco	-13.2539, -71.1602
Bartsia glandulifera Molau	2011-051	Uribe-Convers 2010-051	ID	Colombia	Santander	7.3333, -72.8514
Bartsia glandulifera Molau	2011-054	Uribe-Convers 2010-054	ID	Colombia	Santander	7.2819, -72.8989
Bartsia glandulifera Molau	2013-361	Uribe-Convers 2013-065	ID	Colombia	Norte De Santander	7.2889, -72.6403
Bartsia glandulifera Molau	2013-363	Uribe-Convers 2013-069	ID	Colombia	Santander	6.9931, -72.6823
Bartsia glandulifera Molau	2013-368	Uribe-Convers 2013-076	ID	Colombia	Santander	6.9557, -72.6860
Bartsia glandulifera Molau	2013-372	Uribe-Convers 2013-087	ID	Colombia	Boyacá	6.3791, -72.3399
Bartsia glandulifera Molau	2013-375	Uribe-Convers 2013-094	ID	Colombia	Boyacá	6.1767, -72.7653
Bartsia glandulifera Molau	2013-376	Uribe-Convers 2013-100	ID	Colombia	Cundinamarca	4.2898, -74.2084
Bartsia glandulifera Molau	2013-496	Uribe-Convers 2013-114	ID	Colombia	Boyacá	5.9271, -73.0862
Bartsia glandulifera Molau	2013-499	Uribe-Convers 2013-117	ID	Colombia	Cundinamarca	5.2161, -73.5264
Bartsia glandulifera Molau	2013-501	Uribe-Convers 2013-121	ID	Colombia	Cundinamarca	4.5604, -74.0069
Bartsia inaequalis Benth.	2013-120	Uribe-Convers 2012-001	ID	Colombia	Boyacá	5.8988, -72.9398
Bartsia inaequalis Benth.	2013-342	Uribe-Convers 2012-127	ID	Colombia	Boyacá	5.8988, -72.9398
Bartsia inaequalis Benth. ssp. brachyantha (Diels) Molau	2013-169	Uribe-Convers 2012-061	ID	Peru	Apurimac	-12.4105, -71.1526
Bartsia inaequalis Benth. ssp. brachyantha (Diels) Molau	2013-218	Uribe-Convers 2013-029	ID	Bolivia	La Paz	-16.3141, -67.9065
Bartsia inaequalis Benth. ssp. brachyantha (Diels) Molau	2013-219	Uribe-Convers 2013-031	ID	Bolivia	La Paz	-16.8643, -67.1551
Bartsia inaequalis Benth. ssp. brachyantha (Diels) Molau	2014-098	Uribe-Convers 2013-029 (Ind 2)	ID	Bolivia	La Paz	-16.3141, -67.9065
Bartsia inaequalis Benth. ssp. brachyantha (Diels) Molau	2014-099	Uribe-Convers 2013-029 (Ind 3)	ID	Bolivia	La Paz	-16.3141, -67.9065

Species	DNA Accession	Collector and Voucher No.	Herbarium	Country	State	Latitude, Longitude
Bartsia inaequalis Benth. ssp. brachyantha (Diels) Molau	2014-100	Uribe-Convers 2013-031 (Ind 2)	ID	Bolivia	La Paz	-16.8643, -67.1551
Bartsia inaequalis Benth. ssp. inaequalis	2011-026	Uribe-Convers 2010-022	ID	Colombia	Cundinamarca	4.7418, -73.8663
Bartsia inaequalis Benth. ssp. inaequalis	2011-027	Uribe-Convers 2010-023	ID	Colombia	Cundinamarca	4.7419, -73.8665
Bartsia inaequalis Benth. ssp. inaequalis	2013-377	Uribe-Convers 2013-101	ID	Colombia	Cundinamarca	4.2898, -74.2084
Bartsia inaequalis Benth. ssp. inaequalis	2013-507	Uribe-Convers 2013-127	ID	Colombia	Cundinamarca	4.5617, -74.0202
Bartsia jujuyensis Cabrera & Botta	2013-225	Uribe-Convers 2013-040	ID	Bolivia	Tarija	-21.4833, -64.9410
Bartsia jujuyensis Cabrera & Botta	2014-111	Uribe-Convers 2013-040 (Ind 2)	ID	Bolivia	Tarija	-21.4833, -64.9410
Bartsia laniflora Benth.	2011-029	Uribe-Convers 2010-025	ID	Colombia	Cundinamarca	4.2889, -74.2108
Bartsia laniflora Benth.	2011-014	Uribe-Convers 2010-004	ID	Colombia	Cundinamarca	4.5595, -73.9996
Bartsia laniflora Benth.	2011-024	Uribe-Convers 2010-020	ID	Colombia	Cundinamarca	4.7417, -73.8661
Bartsia laniflora Benth.	2011-025	Uribe-Convers 2010-021	ID	Colombia	Cundinamarca	4.7417, -73.8661
Bartsia laticrenata Benth.	2011-145	Uribe-Convers 2011-005	ID	Ecuador	Carchi	0.6770, -77.8785
Bartsia laticrenata Benth.	2011-160	Uribe-Convers 2011-021	ID	Ecuador	Imbabura	0.1423, -78.2799
Bartsia laticrenata Benth.	2011-174	Uribe-Convers 2011-035	ID	Ecuador	Pichincha	0.0018, -78.0276
Bartsia laticrenata Benth.	2011-190	Uribe-Convers 2011-048	ID	Ecuador	Pichincha	-0.2225, -78.1302
Bartsia laticrenata Benth.	2011-208	Uribe-Convers 2011-067	ID	Ecuador	Cotopaxi	-0.8976, -78.7711
Bartsia laticrenata Benth.	2011-239	Uribe-Convers 2011-099	ID	Ecuador	Chimborazo	-1.7550, -78.8006
Bartsia laticrenata Benth.	2009-105	K09-09	K	Ecuador	Imbabura	0.5833, -77.6667
Bartsia laticrenata Benth.	2011-146	Uribe-Convers 2011-006	ID	Ecuador	Carchi	0.6773, -77.8781
Bartsia laticrenata Benth.	2011-166	Uribe-Convers 2011-027	ID	Ecuador	Imbabura	0.1234, -78.2580
Bartsia laticrenata Benth.	2011-255	Uribe-Convers 2011-115	ID	Ecuador	Azuay	-2.7831, -79.2239
Bartsia laticrenata Benth.	2013-533	Uribe-Convers 2013-153	ID	Colombia	Nariño	1.0961, -77.6910
Bartsia laticrenata Benth.	2014-036	Uribe-Convers 2013-155	ID	Colombia	Nariño	1.0946, -77.7018
Bartsia lydiae S.P. Sylvester	2014-069	Sylvester 1754	Z	Peru	Cusco	-13.1998, -71.8556
Bartsia lydiae S.P. Sylvester	2014-070	Sylvester 878	Z	Peru	Cusco	-13.2431, -71.9796
Bartsia lydiae S.P. Sylvester	2014-071	Sylvester 1730	Z	Peru	Cusco	-13.2026, -71.8544
Bartsia lydiae S.P. Sylvester	2014-072	Sylvester 1649	Z	Peru	Cusco	-13.2693, -71.9820
Bartsia melampyroides (Kunth) Benth.	2011-202	Uribe-Convers 2011-061	ID	Ecuador	Cotopaxi	-0.6266, -78.4747
Bartsia melampyroides (Kunth) Benth.	2011-206	Uribe-Convers 2011-065	ID	Ecuador	Cotopaxi	-0.9038, -78.7247
Bartsia melampyroides (Kunth) Benth.	2011-227	Uribe-Convers 2011-087	ID	Ecuador	Chimborazo	-1.5236, -78.8405
Bartsia melampyroides (Kunth) Benth.	2011-236	Uribe-Convers 2011-096	ID	Ecuador	Chimborazo	-1.7488, -78.7971
Bartsia melampyroides (Kunth) Benth.	2011-262	Uribe-Convers 2011-122	ID	Peru	La Libertad	-7.9966, -78.4230
Bartsia melampyroides (Kunth) Benth.	2012-134	Uribe-Convers 2011-162	ID	Peru	Cajamarca	-7.1960, -78.5661
Bartsia melampyroides (Kunth) Benth.	2012-150	Uribe-Convers 2011-178	ID	Peru	Cajamarca	-6.9167, -78.6143
Bartsia melampyroides (Kunth) Benth.	2013-087	Uribe-Convers 2011-185	ID	Peru	Cajamarca	-6.7562, -78.5819
Bartsia melampyroides (Kunth) Benth.	2013-154	Uribe-Convers 2012-027	ID	Peru	Cusco	-12.4926, -70.0147
Bartsia melampyroides (Kunth) Benth.	2013-158	Uribe-Convers 2012-032	ID	Peru	Cusco	-12.7037, -71.9465
Bartsia melampyroides (Kunth) Benth.	2013-165	Uribe-Convers 2012-046	ID	Peru	Cusco	-12.8237, -71.7115
Bartsia melampyroides (Kunth) Benth.	2009-077	Tank 2005-07	WTU	Peru	Cajamarca	-7.1936, -78.5597
Bartsia melampyroides (Kunth) Benth.	2010-195	Tank 2005-07	WTU	Peru	Cajamarca	-7.1936, -78.5597
Bartsia melampyroides (Kunth) Benth.	2011-271	Uribe-Convers 2011-131	ID	Peru	La Libertad	-8.1036, -78.2947
Bartsia melampyroides (Kunth) Benth.	2012-140	Uribe-Convers 2011-168	ID	Peru	Cajamarca	-7.0470, -78.2729
Bartsia melampyroides (Kunth) Benth.	2013-093	Uribe-Convers 2011-196	ID	Peru	Ancash	-9.0478, -77.6100
Bartsia melampyroides (Kunth) Benth.	2013-105	Uribe-Convers 2011-214	ID	Peru	Ancash	-9.3815, -77.5191
Bartsia melampyroides (Kunth) Benth.	2013-305	Uribe-Convers 2012-040	ID	Peru	Cusco	-12.7099, -71.9486
Bartsia melampyroides (Kunth) Benth.	2013-312	Uribe-Convers 2012-054	ID	Peru	Cusco	-12.5917, -71.9569
Bartsia melampyroides (Kunth) Benth.	2013-317	Uribe-Convers 2012-065	ID	Peru	Apurimac	-12.4166, -71.1683
Bartsia melampyroides (Kunth) Benth.	2013-320	Uribe-Convers 2012-074	ID	Peru	Ayacucho	-12.6720, -73.8046
Bartsia melampyroides (Kunth) Benth.	2013-323	Uribe-Convers 2012-082	ID	Peru	Huancavelica	-11.2433, -73.0906
Bartsia mutica (Kunth) Benth.	2011-168	Uribe-Convers 2011-029	ID	Ecuador	Imbabura	0.2964, -78.3478
Bartsia mutica (Kunth) Benth.	2012-110	Uribe-Convers 2011-137	ID	Peru	Cajamarca	-7.4078, -78.7834
Bartsia mutica (Kunth) Benth.	2012-113	Uribe-Convers 2011-140	ID	Peru	Cajamarca	-7.0022, -78.8468
Bartsia mutica (Kunth) Benth.	2012-127	Uribe-Convers 2011-155	ID	Peru	Cajamarca	-7.2485, -78.4707
Bartsia mutica (Kunth) Benth.	2012-148	Uribe-Convers 2011-176	ID	Peru	Cajamarca	-6.8692, -78.1130
Bartsia mutica (Kunth) Benth.	2013-089	Uribe-Convers 2011-191	ID	Peru	Ancash	-9.0314, -77.7267
Bartsia mutica (Kunth) Benth.	2013-091	Uribe-Convers 2011-193	ID	Peru	Ancash	-9.1028, -77.6772
Bartsia mutica (Kunth) Benth.	2013-104	Uribe-Convers 2011-213 (Ind 1)	ID	Peru	Ancash	-9.3815, -77.5191

Species	DNA Accession	Collector and Voucher No.	Herbarium	Country	State	Latitude, Longitude
Bartsia mutica (Kunth) Benth.	2014-041	Uribe-Convers 2011-213 (Ind 2)	ID	Peru	Ancash	-9.3815, -77.5191
Bartsia orthocarpiflora Benth.	2011-150	Uribe-Convers 2011-011	ID	Ecuador	Carchi	0.7941, -77.8771
Bartsia orthocarpiflora Benth.	2011-163	Uribe-Convers 2011-024	ID	Ecuador	Imbabura	0.1263, -78.2695
Bartsia orthocarpiflora Benth.	2011-172	Uribe-Convers 2011-033	ID	Ecuador	Pichincha	-0.0092, -78.0348
Bartsia orthocarpiflora Benth.	2012-109	Uribe-Convers 2011-046	ID	Ecuador	Pichincha	-0.1886, -78.1161
Bartsia orthocarpiflora Benth.	2011-144	Uribe-Convers 2011-004	ID	Ecuador	Carchi	0.6612, -77.8884
Bartsia orthocarpiflora Benth.	2011-148	Uribe-Convers 2011-008	ID	Ecuador	Carchi	0.6794, -77.8782
Bartsia orthocarpiflora Benth.	2011-151	Uribe-Convers 2011-012	ID	Ecuador	Carchi	0.7971, -77.9111
Bartsia orthocarpiflora Benth.	2011-158	Uribe-Convers 2011-019	ID	Ecuador	Carchi	0.8087, -77.9591
Bartsia orthocarpiflora Benth.	2011-164	Uribe-Convers 2011-025	ID	Ecuador	Imbabura	0.1257, -78.2590
Bartsia orthocarpiflora Benth.	2011-167	Uribe-Convers 2011-028	ID	Ecuador	Imbabura	0.1234, -78.2580
Bartsia orthocarpiflora Benth.	2011-143	Uribe-Convers 2011-003	ID	Ecuador	Carchi	0.6583, -77.8990
Bartsia orthocarpiflora Benth.	2011-207	Uribe-Convers 2011-066	ID	Ecuador	Cotopaxi	-0.9082, -78.7448
Bartsia orthocarpiflora Benth.	2011-230	Uribe-Convers 2011-090	ID	Ecuador	Chimborazo	-1.5502, -78.4787
Bartsia orthocarpiflora Benth.	2011-238	Uribe-Convers 2011-098	ID	Ecuador	Chimborazo	-1.7561, -78.8073
Bartsia orthocarpiflora Benth.	2011-242	Uribe-Convers 2011-102	ID	Ecuador	Chimborazo	-2.1773, -78.5486
Bartsia orthocarpiflora Benth.	2011-245	Uribe-Convers 2011-105	ID	Ecuador	Azuay	-2.7774, -79.1695
Bartsia orthocarpiflora Benth.	2009-104	K09-8a	K	Ecuador	Pichincha-Sucumbios	0.1187, -77.9660
Bartsia orthocarpiflora Benth.	2011-233	Uribe-Convers 2011-093	ID	Ecuador	Chimborazo	-1.5486, -78.4449
Bartsia orthocarpiflora Benth.	2011-216	Uribe-Convers 2011-075	ID	Ecuador	Tungurahua	-1.1165, -78.4564
Bartsia orthocarpiflora Benth. ssp. orthocarpiflora	2011-192	Uribe-Convers 2011-051	ID	Ecuador	Pichincha	-0.3271, -78.1510
Bartsia orthocarpiflora Benth. ssp. orthocarpiflora	2014-039	Uribe-Convers 2013-158 (Ind 1)	ID	Colombia	Nariño	0.9320, -77.8682
Bartsia orthocarpiflora Benth. ssp. orthocarpiflora	2014-040	Uribe-Convers 2013-158 (Ind 2)	ID	Colombia	Nariño	0.9320, -77.8682
Bartsia orthocarpiflora Benth. ssp. villosa Molau	2011-191	Uribe-Convers 2011-050	ID	Ecuador	Pichincha	-0.3271, -78.1510
Bartsia orthocarpiflora Benth. ssp. villosa Molau	2013-512	Uribe-Convers 2013-131 (Ind 1)	ID	Colombia	Antioquia	6.4477, -76.0855
Bartsia orthocarpiflora Benth. ssp. villosa Molau	2013-513	Uribe-Convers 2013-131 (Ind 2)	ID	Colombia	Antioquia	6.4477, -76.0855
Bartsia orthocarpiflora Benth. ssp. villosa Molau	2013-514	Uribe-Convers 2013-132 (Ind 1)	ID	Colombia	Antioquia	6.4425, -76.0827
Bartsia orthocarpiflora Benth. ssp. villosa Molau	2013-515	Uribe-Convers 2013-132 (Ind 2)	ID	Colombia	Antioquia	6.4425, -76.0827
Bartsia orthocarpiflora Benth. ssp. villosa Molau	2013-516	Uribe-Convers 2013-133	ID	Colombia	Caldas	4.9965, -75.3310
Bartsia orthocarpiflora Benth. ssp. villosa Molau	2013-531	Uribe-Convers 2013-148	ID	Colombia	Nariño	1.2107, -77.3308
Bartsia orthocarpiflora Benth. ssp. villosa Molau	2013-532	Uribe-Convers 2013-150	ID	Colombia	Nariño	1.0929, -77.6811
Bartsia orthocarpiflora Benth. ssp. villosa Molau	2014-037	Uribe-Convers 2013-156 (Ind 1)	ID	Colombia	Nariño	0.9251, -77.8571
Bartsia orthocarpiflora Benth. ssp. villosa Molau	2014-038	Uribe-Convers 2013-156 (Ind 2)	ID	Colombia	Nariño	0.9251, -77.8571
Bartsia patens Benth.	2011-268	Uribe-Convers 2011-128	ID	Peru	La Libertad	-7.9708, -78.2079
Bartsia patens Benth.	2012-133	Uribe-Convers 2011-161	ID	Peru	Cajamarca	-7.1960, -78.5661
Bartsia patens Benth.	2013-094	Uribe-Convers 2011-199	ID	Peru	Ancash	-9.0478, -77.6100
Bartsia patens Benth.	2013-109	Uribe-Convers 2011-220	ID	Peru	Lima	-11.5971, -76.1924
Bartsia patens Benth.	2013-181	Uribe-Convers 2012-093	ID	Peru	Junín	-10.0740, -74.9365
Bartsia patens Benth.	2013-097	Uribe-Convers 2011-204	ID	Peru	Ancash	-9.0496, -77.5971
Bartsia patens Benth.	2013-327	Uribe-Convers 2012-094	ID	Peru	Junín	-10.0740, -74.9365
Bartsia patens Benth.	2014-042	Uribe-Convers 2011-220 (Ind 2)	ID	Peru	Lima	-11.5971, -76.1924
Bartsia pauciflora Molau	2013-184	Uribe-Convers 2012-101	ID	Peru	Junín	-10.0271, -74.9199
Bartsia pauciflora Molau	2013-209	Uribe-Convers 2013-019	ID	Bolivia	La Paz	-15.5083, -69.0557
Bartsia pauciflora Molau	2013-222	Uribe-Convers 2013-037	ID	Bolivia	La Paz	-17.0280, -67.2929
Bartsia pauciflora Molau	2013-351	Uribe-Convers 2013-021 (Ind 3)	ID	Bolivia	La Paz	-14.8259, -69.2316
Bartsia pauciflora Molau	2014-050	Uribe-Convers 2013-021 (Ind 1)	ID	Bolivia	La Paz	-14.8259, -69.2316
Bartsia pauciflora Molau	2014-051	Uribe-Convers 2013-021 (Ind 2)	ID	Bolivia	La Paz	-14.8259, -69.2316
Bartsia pauciflora Molau	2014-088	Uribe-Convers 2013-019 (Ind 2)	ID	Bolivia	La Paz	-15.5083, -69.0557
Bartsia pauciflora Molau	2014-089	Uribe-Convers 2013-019 (Ind 3)	ID	Bolivia	La Paz	-15.5083, -69.0557
Bartsia pauciflora Molau	2014-106	Uribe-Convers 2013-037 (Ind 2)	ID	Bolivia	La Paz	-17.0280, -67.2929
Bartsia pauciflora Molau	2014-107	Uribe-Convers 2013-037 (Ind 3)	ID	Bolivia	La Paz	-17.0280, -67.2929
Bartsia pedicularoides Benth.	2011-064	Uribe-Convers 2010-066	ID	Colombia	Santander	6.9908, -72.6836
Bartsia pedicularoides Benth.	2011-205	Uribe-Convers 2011-064	ID	Ecuador	Cotopaxi	-0.6266, -78.4747
Bartsia pedicularoides Benth.	2011-217	Uribe-Convers 2011-076	ID	Ecuador	Tungurahua	-1.1148, -78.4549
Bartsia pedicularoides Benth.	2011-234	Uribe-Convers 2011-094	ID	Ecuador	Chimborazo	-1.5499, -78.4424
Bartsia pedicularoides Benth.	2011-241	Uribe-Convers 2011-101	ID	Ecuador	Chimborazo	-1.7573, -78.8010
Bartsia pedicularoides Benth.	2011-243	Uribe-Convers 2011-103	ID	Ecuador	Chimborazo	-2.1868, -78.5355

Species	DNA Accession	Collector and Voucher No.	Herbarium	Country	State	Latitude, Longitude
Bartsia pedicularoides Benth.	2011-248	Uribe-Convers 2011-108	ID	Ecuador	Azuay	-2.7813, -79.1853
Bartsia pedicularoides Benth.	2013-116	Uribe-Convers 2011-231	ID	Peru	Junín	-11.9790, -75.0942
Bartsia pedicularoides Benth.	2013-194	Uribe-Convers 2013-003	ID	Bolivia	La Paz	-16.2874, -68.1284
Bartsia pedicularoides Benth.	2013-199	Uribe-Convers 2013-009	ID	Bolivia	La Paz	-16.2375, -68.4783
Bartsia pedicularoides Benth.	2013-210	Uribe-Convers 2013-020	ID	Bolivia	La Paz	-14.8259, -69.2316
Bartsia pedicularoides Benth.	2010-201	K 09-12	K	Ecuador	Azuay	-1.1167, -78.7000
Bartsia pedicularoides Benth.	2011-071	Uribe-Convers 2010-073	ID	Colombia	Boyacá	6.1283, -72.8051
Bartsia pedicularoides Benth.	2011-220	Uribe-Convers 2011-079	ID	Ecuador	Tungurahua	-1.0768, -78.4296
Bartsia pedicularoides Benth.	2011-222	Uribe-Convers 2011-081	ID	Ecuador	Tungurahua	-1.0768, -78.4296
Bartsia pedicularoides Benth.	2011-256	Uribe-Convers 2011-116	ID	Ecuador	Azuay	-2.7831, -79.2239
Bartsia pedicularoides Benth.	2013-354	Uribe-Convers 2013-052 (Ind 1)	ID	Colombia	Santander	7.2245, -72.8980
Bartsia pedicularoides Benth.	2013-360	Uribe-Convers 2013-059 (Ind 1)	ID	Colombia	Santander	7.3302, -72.8497
Bartsia pedicularoides Benth.	2013-364	Uribe-Convers 2013-070 (Ind 1)	ID	Colombia	Santander	6.9931, -72.6823
Bartsia pedicularoides Benth.	2013-371	Uribe-Convers 2013-085	ID	Colombia	Boyacá	6.3805, -72.3409
Bartsia pedicularoides Benth.	2013-373	Uribe-Convers 2013-088	ID	Colombia	Boyacá	6.3624, -72.3373
Bartsia pedicularoides Benth.	2013-495	Uribe-Convers 2013-112	ID	Colombia	Boyacá	5.9271, -73.0862
Bartsia pedicularoides Benth.	2013-517	Uribe-Convers 2013-136	ID	Colombia	Caldas	4.9950, -75.3317
Bartsia pedicularoides Benth.	2013-524	Uribe-Convers 2013-141	ID	Colombia	Tolima	4.6354, -75.1255
Bartsia pedicularoides Benth.	2013-529	Uribe-Convers 2013-146 (Ind 1)	ID	Colombia	Cauca	2.3633, -76.3495
Bartsia pedicularoides Benth.	2013-530	Uribe-Convers 2013-146 (Ind 2)	ID	Colombia	Cauca	2.3633, -76.3495
Bartsia pedicularoides Benth.	2014-052	Uribe-Convers 2013-052 (Ind 2)	ID	Colombia	Santander	7.2245, -72.8980
Bartsia pedicularoides Benth.	2014-054	Uribe-Convers 2013-059 (Ind 2)	ID	Colombia	Santander	7.3302, -72.8497
Bartsia pedicularoides Benth.	2014-075	Uribe-Convers 2013-003 (Ind 2)	ID	Bolivia	La Paz	-16.2874, -68.1284
Bartsia pedicularoides Benth.	2014-081	Uribe-Convers 2013-009 (Ind 2)	ID	Bolivia	La Paz	-16.2375, -68.4783
Bartsia pedicularoides Benth.	2014-082	Uribe-Convers 2013-009 (Ind 3)	ID	Bolivia	La Paz	-16.2375, -68.4783
Bartsia pedicularoides Benth.	2014-090	Uribe-Convers 2013-020 (Ind 2)	ID	Bolivia	La Paz	-14.8259, -69.2316
Bartsia pedicularoides Benth.	2014-091	Uribe-Convers 2013-020 (Ind 3)	ID	Bolivia	La Paz	-14.8259, -69.2316
Bartsia pedicularoides Benth.	2014-119	Uribe-Convers 2013-070 (Ind 2)	ID	Colombia	Santander	6.9931, -72.6823
Bartsia pedicularoides Benth.	2011-075	Antonelli 574	GB	Ecuador	Azuay	-2.9781, -78.6884
Bartsia peruviana Walp.	2013-195	Uribe-Convers 2013-004	ID	Bolivia	La Paz	-16.3082, -68.0282
Bartsia peruviana Walp.	2013-198	Uribe-Convers 2013-007	ID	Bolivia	La Paz	-16.1999, -68.4754
Bartsia peruviana Walp.	2013-201	Uribe-Convers 2013-011	ID	Bolivia	La Paz	-15.7351, -68.6573
Bartsia peruviana Walp.	2013-235	Uribe-Convers 2012-013	ID	Peru	Moquegua	-15.3971, -70.7752
Bartsia peruviana Walp.	2013-296	Uribe-Convers 2012-020	ID	Peru	Moquegua	-15.0129, -69.2928
Bartsia peruviana Walp.	2013-297	Uribe-Convers 2012-021	ID	Peru	Puno	-15.7964, -68.5346
Bartsia peruviana Walp.	2014-076	Uribe-Convers 2013-004 (Ind 2)	ID	Bolivia	La Paz	-16.3082, -68.0282
Bartsia peruviana Walp.	2014-079	Uribe-Convers 2013-007 (Ind 2)	ID	Bolivia	La Paz	-16.1999, -68.4754
Bartsia pyricarpa Molau	2011-266	Uribe-Convers 2011-126	ID	Peru	La Libertad	-7.9879, -78.2496
Bartsia pyricarpa Molau	2011-275	Uribe-Convers 2011-135	ID	Peru	La Libertad	-8.1036, -78.2947
Bartsia pyricarpa Molau	2012-135	Uribe-Convers 2011-163	ID	Peru	Cajamarca	-7.1960, -78.5661
Bartsia pyricarpa Molau	2012-141	Uribe-Convers 2011-169	ID	Peru	Cajamarca	-7.0429, -78.2686
Bartsia pyricarpa Molau	2012-149	Uribe-Convers 2011-177	ID	Peru	Cajamarca	-6.9167, -78.6143
Bartsia pyricarpa Molau	2013-166	Uribe-Convers 2012-047	ID	Peru	Cusco	-12.8561, -71.7115
Bartsia pyricarpa Molau	2013-187	Uribe-Convers 2012-106	ID	Peru	Junín	-10.3913, -74.3631
Bartsia pyricarpa Molau	2013-189	Uribe-Convers 2012-108 (Ind. 1)	ID	Peru	Junín	-10.4462, -74.3734
Bartsia pyricarpa Molau	2009-081	Tank 2005-36	WTU	Peru	La Libertad	-8.1386, -78.2744
Bartsia pyricarpa Molau	2010-199	Tank 2005-36	WTU	Peru	La Libertad	-8.1386, -78.2744
Bartsia pyricarpa Molau	2013-309	Uribe-Convers 2012-049	ID	Peru	Cusco	-12.8559, -71.7287
Bartsia pyricarpa Molau	2013-310	Uribe-Convers 2012-051	ID	Peru	Cusco	-12.8579, -71.7063
Bartsia pyricarpa Molau	2013-321	Uribe-Convers 2012-075	ID	Peru	Ayacucho	-12.6720, -73.8046
Bartsia pyricarpa Molau	2013-324	Uribe-Convers 2012-083	ID	Peru	Huancavelica	-11.2433, -73.0906
Bartsia pyricarpa Molau	2014-044	Uribe-Convers 2012-108 (Ind 2)	ID	Peru	Junín	-10.4462, -74.3734
Bartsia ramosa Molau	2011-073	Uribe-Convers 2010-075	ID	Colombia	Cundinamarca	4.5821, -74.0273

Species	DNA Accession	Collector and Voucher No.	Herbarium	Country	State	Latitude, Longitude
Bartsia ramosa Molau	2011-147	Uribe-Convers 2011-007	ID	Ecuador	Carchi	0.6773, -77.8781
Bartsia ramosa Molau	2011-175	Uribe-Convers 2011-036	ID	Ecuador	Pichincha	0.0018, -78.0274
Bartsia ramosa Molau	2011-019	Uribe-Convers 2010-012	ID	Colombia	Cundinamarca	5.0115, -74.2020
Bartsia ramosa Molau	2011-020	Uribe-Convers 2010-013	ID	Colombia	Cundinamarca	5.0115, -74.2020
Bartsia ramosa Molau	2013-497	Uribe-Convers 2013-116 (Ind 1)	ID	Colombia	Cundinamarca	5.2161, -73.5264
Bartsia ramosa Molau	2013-498	Uribe-Convers 2013-116 (Ind 2)	ID	Colombia	Cundinamarca	5.2161, -73.5264
Bartsia ramosa Molau	2013-525	Uribe-Convers 2013-143	ID	Colombia	Cauca	2.1851, -76.4776
Bartsia ramosa Molau	2013-526	Uribe-Convers 2013-144 (Ind 1)	ID	Colombia	Huila	2.1706, -76.3920
Bartsia ramosa Molau	2013-527	Uribe-Convers 2013-144 (Ind 2)	ID	Colombia	Huila	2.1706, -76.3920
Bartsia ramosa Molau	2013-528	Uribe-Convers 2013-145	ID	Colombia	Cauca	2.3544, -76.3354
Bartsia ramosa Molau	2014-035	Uribe-Convers 2013-154	ID	Colombia	Nariño	1.0961, -77.6910
Bartsia rigida Molau	2013-112	Uribe-Convers 2011-225	ID	Peru	Junín	-11.5222, -75.6409
Bartsia rigida Molau	2013-185	Uribe-Convers 2012-104	ID	Peru	Junín	-10.0271, -74.9199
Bartsia santolinifolia (Kunth) Benth.	2011-037	Uribe-Convers 2010-034	ID	Colombia	Boyacá	5.6934, -72.8266
Bartsia santolinifolia (Kunth) Benth.	2011-042	Uribe-Convers 2010-041	ID	Colombia	Boyacá	5.9170, -73.0841
Bartsia santolinifolia (Kunth) Benth.	2011-052	Uribe-Convers 2010-052	ID	Colombia	Santander	7.3375, -72.8694
Bartsia santolinifolia (Kunth) Benth.	2011-063	Uribe-Convers 2010-065	ID	Colombia	Santander	6.9910, -72.6833
Bartsia santolinifolia (Kunth) Benth.	2011-068	Uribe-Convers 2010-070	ID	Colombia	Boyacá	6.1729, -72.7584
Bartsia santolinifolia (Kunth) Benth.	2013-122	Uribe-Convers 2012-003	ID	Colombia	Boyacá	5.9273, -72.9132
Bartsia santolinifolia (Kunth) Benth.	2009-106	K09-10	K	Bolivia	Cochabamba	-16.6728, -65.2056
Bartsia santolinifolia (Kunth) Benth.	2011-040	Uribe-Convers 2010-039	ID	Colombia	Boyacá	5.9125, -73.0769
Bartsia santolinifolia (Kunth) Benth.	2011-072	Uribe-Convers 2010-074	ID	Colombia	Cundinamarca	4.5816, -74.0269
Bartsia santolinifolia (Kunth) Benth.	2013-121	Uribe-Convers 2012-002	ID	Colombia	Boyacá	5.9146, -72.9181
Bartsia santolinifolia (Kunth) Benth.	2013-344	Uribe-Convers 2012-129	ID	Colombia	Boyacá	5.9275, -72.9123
Bartsia santolinifolia (Kunth) Benth.	2013-355	Uribe-Convers 2013-053	ID	Colombia	Santander	7.2493, -72.8978
Bartsia santolinifolia (Kunth) Benth.	2013-358	Uribe-Convers 2013-057	ID	Colombia	Santander	7.3338, -72.8541
Bartsia santolinifolia (Kunth) Benth.	2013-365	Uribe-Convers 2013-072	ID	Colombia	Santander	6.9931, -72.6823
Bartsia santolinifolia (Kunth) Benth.	2013-367	Uribe-Convers 2013-074	ID	Colombia	Santander	6.9557, -72.6860
Bartsia santolinifolia (Kunth) Benth.	2013-374	Uribe-Convers 2013-093	ID	Colombia	Boyacá	6.1767, -72.7653
Bartsia santolinifolia (Kunth) Benth.	2013-380	Uribe-Convers 2013-105	ID	Colombia	Boyacá	6.0295, -72.9654
Bartsia santolinifolia (Kunth) Benth.	2013-492	Uribe-Convers 2013-110	ID	Colombia	Boyacá	5.9276, -73.0885
Bartsia santolinifolia (Kunth) Benth.	2013-500	Uribe-Convers 2013-119	ID	Colombia	Cundinamarca	5.2190, -73.5343
Bartsia santolinifolia (Kunth) Benth.	2013-502	Uribe-Convers 2013-122	ID	Colombia	Cundinamarca	4.5604, -74.0069
Bartsia sericea Molau	2011-252	Uribe-Convers 2011-112	ID	Ecuador	Azuay	-2.7831, -79.2239
Bartsia sericea Molau	2012-119	Uribe-Convers 2011-146	ID	Peru	Cajamarca	-6.8860, -78.7396
Bartsia sericea Molau	2012-139	Uribe-Convers 2011-167	ID	Peru	Cajamarca	-7.0470, -78.2729
Bartsia sericea Molau	2012-152	Uribe-Convers 2011-180	ID	Peru	Cajamarca	-6.9167, -78.6143
Bartsia sericea Molau	2009-076	Tank 2005-06	WTU	Peru	Cajamarca	-7.1936, -78.5597
Bartsia sericea Molau	2010-194	Tank 2005-06	WTU	Peru	Cajamarca	-7.1936, -78.5597
Bartsia sericea Molau	2010-202	Tank 2005-06	WTU	Peru	Cajamarca	-7.1936, -78.5597
Bartsia serrata Molau	2013-128	Uribe-Convers 2012-009	ID	Peru	Arequipa	-15.7934, -70.3482
Bartsia serrata Molau	2013-150	Uribe-Convers 2012-011	ID	Peru	Arequipa	-15.4627, -70.6427
Bartsia serrata Molau	2013-152	Uribe-Convers 2012-018	ID	Peru	Moquegua	-15.0411, -69.1431
Bartsia serrata Molau	2013-234	Uribe-Convers 2012-012	ID	Peru	Arequipa	-15.4516, -70.6564
Bartsia serrata Molau	2012-192	Uribe-Convers 2012-015	ID	Peru	Moquegua	-15.3971, -70.7752
Bartsia serrata Molau	2013-293	Uribe-Convers 2012-016	ID	Peru	Moquegua	-15.3710, -70.8298
Bartsia cf sericea Molau	2009-078	Tank 2005-25	WTU	Peru	Cajamarca	-7.4162, -78.6727
Bartsia cf sericea Molau	2010-196	Tank 2005-25	WTU	Peru	Cajamarca	-7.4162, -78.6727
Bartsia cf sericea Molau	2010-203	Tank 2005-25	WTU	Peru	Cajamarca	-7.4162, -78.6727
Bartsia sp No.6	2009-079	Tank 2005-28	WTU	Peru	La Libertad	-7.9900, -78.5348
Bartsia sp No.6	2010-197	Tank 2005-28	WTU	Peru	La Libertad	-7.9900, -78.5348
Bartsia cf inaequalis Benth. ssp. Duripilis (Edwin) Molau	2009-080	Tank 2005-29	WTU	Peru	La Libertad	-7.9900, -78.5348
Bartsia cf inaequalis Benth. ssp. Duripilis (Edwin) Molau	2010-198	Tank 2005-29	WTU	Peru	La Libertad	-7.9900, -78.5348
Bartsia stricta (Kunth) Benth.	2011-028	Uribe-Convers 2010-024	ID	Colombia	Cundinamarca	4.2890, -74.2107
Bartsia stricta (Kunth) Benth.	2011-050	Uribe-Convers 2010-050	ID	Colombia	Santander	7.3335, -72.8540
Bartsia stricta (Kunth) Benth.	2011-062	Uribe-Convers 2010-063	ID	Colombia	Santander	6.9943, -72.6818
Bartsia stricta (Kunth) Benth.	2011-067	Uribe-Convers 2010-069	ID	Colombia	Santander	6.9401, -72.6943
Bartsia stricta (Kunth) Benth.	2011-179	Uribe-Convers 2011-039	ID	Ecuador	Pichincha	0.0091, -78.0102
Bartsia stricta (Kunth) Benth.	2011-203	Uribe-Convers 2011-062	ID	Ecuador	Cotopaxi	-0.6266, -78.4747

Species	DNA Accession	Collector and Voucher No.	Herbarium	Country	State	Latitude, Longitude
Bartsia stricta (Kunth) Benth.	2011-180	Uribe-Convers 2011-040	ID	Ecuador	Pichincha	0.0085, -78.0136
Bartsia stricta (Kunth) Benth.	2013-352	Uribe-Convers 2013-050	ID	Colombia	Santander	7.2137, -72.8958
Bartsia stricta (Kunth) Benth.	2013-356	Uribe-Convers 2013-054	ID	Colombia	Santander	7.2485, -72.8965
Bartsia stricta (Kunth) Benth.	2013-357	Uribe-Convers 2013-055	ID	Colombia	Santander	7.3338, -72.8541
Bartsia stricta (Kunth) Benth.	2013-362	Uribe-Convers 2013-066 (Ind 1)	ID	Colombia	Norte De Santander	7.2915, -72.6397
Bartsia stricta (Kunth) Benth.	2013-379	Uribe-Convers 2013-104	ID	Colombia	Boyacá	6.0295, -72.9654
Bartsia stricta (Kunth) Benth.	2013-522	Uribe-Convers 2013-140 (Ind 1)	ID	Colombia	Quindío	4.6463, -75.4272
Bartsia stricta (Kunth) Benth.	2013-523	Uribe-Convers 2013-140 (Ind 2)	ID	Colombia	Quindío	4.6463, -75.4272
Bartsia stricta (Kunth) Benth.	2014-055	Uribe-Convers 2013-066 (Ind 2)	ID	Colombia	Norte De Santander	7.2915, -72.6397
Bartsia stricta (Kunth) Benth.	2011-076	Antonelli 582	GB	Ecuador	Azuay	-2.7808, -79.2253
Bartsia tenuis Molau	2009-075	Tank 2005-02	WTU	Peru	Cajamarca	-7.2458, -78.4694
Bartsia tenuis Molau	2010-193	Tank 2005-02	WTU	Peru	Cajamarca	-7.2458, -78.4694
Bartsia tenuis Molau	2013-332	Uribe-Convers 2012-109	ID	Peru	Junín	-10.4462, -74.3734
Bartsia thiantha Diels	2011-201	Uribe-Convers 2011-060	ID	Ecuador	Cotopaxi	-0.6266, -78.4747
Bartsia thiantha Diels	2013-157	Uribe-Convers 2012-030	ID	Peru	Cusco	-12.6965, -71.9502
Bartsia thiantha Diels	2013-160	Uribe-Convers 2012-034	ID	Peru	Cusco	-12.7099, -71.9486
Bartsia thiantha Diels	2013-163	Uribe-Convers 2012-043	ID	Peru	Cusco	-12.8239, -71.7101
Bartsia thiantha Diels	2009-109	RGO 2009-23	WTU	Peru	Apurimac	-13.5794, -72.8196
Bartsia thiantha Diels	2010-192	RGO 2009-23	WTU	Peru	Apurimac	-13.5794, -72.8196
Bartsia thiantha Diels	2013-155	Uribe-Convers 2012-028	ID	Peru	Cusco	-12.4926, -70.0147
Bartsia thiantha Diels	2013-307	Uribe-Convers 2012-045	ID	Peru	Cusco	-12.8239, -71.7101
Bartsia thiantha Diels	2013-308	Uribe-Convers 2012-048	ID	Peru	Cusco	-12.8559, -71.7287
Bartsia thiantha Diels	2013-311	Uribe-Convers 2012-053	ID	Peru	Cusco	-12.5917, -71.9569
Bartsia tomentosa Molau	2013-346	D.N. Smith 11237	LPB	Bolivia		
Bartsia trichophylla Wedd.	2013-167	Uribe-Convers 2012-052	ID	Peru	Cusco	-12.8690, -71.6789
Bartsia trichophylla Wedd.	2013-167	Uribe-Convers 2012-052	ID	Peru	Cusco	-12.8690, -71.6789
Bartsia trichophylla Wedd.	2013-283	Beck St. G. 28575	LPB	Bolivia		
Bartsia trichophylla Wedd.	2013-348	J.C. Solomon, Bruce Stein 11665	LPB	Bolivia		
Bartsia trichophylla Wedd.	2013-349	J.C. Solomon 18192	LPB	Bolivia		
Bartsia trichophylla Wedd.	2013-350	Beck St. G. 19978	LPB	Bolivia		
Bartsia tricolor Molau	2013-175	Uribe-Convers 2012-078	ID	Peru	Huancavelica	-11.3902, -73.0452
Bartsia tricolor Molau	2013-177	Uribe-Convers 2012-084	ID	Peru	Huancavelica	-11.2433, -73.0906
Bartsia tricolor Molau	2013-345	D.N. Smith 10800	LPB	Bolivia		
Bartsia weberbaueri Diels	2013-174	Uribe-Convers 2012-077	ID	Peru	Huancavelica	-11.3902, -73.0452
Bartsia weberbaueri Diels	2013-176	Uribe-Convers 2012-081	ID	Peru	Huancavelica	-11.2158, -73.0597
Bellardia trixago L.	2009-102	K09-7	K	Ethiopia		12.5000, 37.0083
Bellardia trixago L.	2010-182	Bennett 60	FHO	Spain	Andalucia	36.4167, -6.1333
Bellardia trixago L.	2014-056	M.J.E. Coode & B.M.G. Jones 605	A	Turkey	Hatay	36.2554, 36.3041
Castilleja miniata Douglas ex Hook.	2009-021	Tank 1054	ID	USA		
Euphrasia alsa F.Muell.	2010-191	Zich 220	GH			
Euphrasia collina R.Br.	2010-190	Zich 209	GH			
Euphrasia mollis (Ledeb.) Wettst.	2009-117	Muncuso 107	ID	USA		
Euphrasia regelii Wettst.	2010-188	Ho et al. 1741	GH			
Euphrasia stricta D. Wolff	2009-118	Lytton Musselman 4872	ID	Netherlands		
Euphrasia stricta D. Wolff	2010-189	N/A	N/A			
Hedbergia abyssinica (Benth.) Molau	2009-094	Etuge M. 3488	K	Cameroon	West Region	
Hedbergia abyssinica (Benth.) Molau	2009-093	Pollard, B.J. 364	K	Cameroon	West Region	
Hedbergia abyssinica (Benth.) Molau var. petitiana (A.Rich.) Skan	2009-096	A.J. Paton	K	Tanzania		-8.9864, 33.8811
Hedbergia abyssinica (Benth.) Molauvar. nykiensis R.E. Fries	2009-095	Carter, Abdallah, & Newton 2386	K	Tanzania		-7.2500, 33.5500
Hedbergia decurva Benth.	2009-103	K09-8	K	Uganda	Western Province	1.1183, 34.5250
Hedbergia longiflora Benth. ssp. longiflora	2010-200	K 09-6	K	Uganda	Western Province	1.1183, 34.5250
Lathraea squamaria L.	2010-185	Frajman s.n.	LJU			
Melampyrum carstiense Fritsch	2010-187	Krajsek s.n.	LJU			
Melampyrum lineare Desr.	2009-120	Michael Hays 1889	ID	USA		
Melampyrum sylvaticum L.	2010-186	Krajsek s.n.	LJU			

Species	DNA Accession	Collector and Voucher No.	Herbarium	Country	State	Latitude, Longitude
Odontites aucheri Boiss.	2009-097	M. Oganesian, H. Ter-Voskanian, E. Vitek 03-1575	K	Armenia		39.8597, 44.9653
Odontites maroccanus Bolliger	2009-099	J. Gattefose	K	Morocco		
Odontites vulgaris Moench	2009-101	S. Kharkevich, T. Buch	K	Russia	Primorsky Territory	
Parentucellia latifolia (L.) Caruel	2014-057	Vincent & Freid 8180	MU	USA	California	
Parentucellia latifolia (L.) Caruel	2014-058	Jack & Betty Guggolz 1012	JEPS	USA	California	38.3940, -122.4550
Parentucellia latifolia (L.) Caruel	2014-059	Wetherwax & Martin 2190	JEPS	USA	California	38.1550, -122.8600
Parentucellia latifolia (L.) Caruel	2014-060	Wetherwax & Pendleton 2443	JEPS	USA	California	38.0910, -122.7470
Parentucellia latifolia (L.) Caruel	2014-061	J. Greenhouse, D. Smith s.n.	JEPS	USA	California	38.2350, -122.9120
Parentucellia latifolia (L.) Caruel	2014-062	David Gowen 847	JEPS	USA	California	38.0460, -122.6210
Parentucellia latifolia (L.) Caruel	2014-063	David Gowen 847	JEPS	USA	California	38.0460, -122.6210
Parentucellia latifolia (L.) Caruel	2014-064	David Gowen 847	JEPS	USA	California	38.0460, -122.6210
Parentucellia latifolia (L.) Caruel	2014-065	Wetherwax & Martin 2190	JEPS	USA	California	38.1550, -122.8600
Parentucellia viscosa (L.) Caruel	2009-113	Richard Halse 2249	ID	USA	Oregon	44.8906, -123.2249
Parentucellia viscosa (L.) Caruel	2009-113	Richard Halse 2249	ID	USA	Oregon	44.8906, -123.2249
Parentucellia viscosa (L.) Caruel	2013-290	Espíndola 12-004	ID	Chile		-39.9398, -73.5829
Parentucellia viscosa (L.) Caruel	2013-291	Espíndola 12-005	ID	Chile		-41.6068, -72.6769
Parentucellia viscosa (L.) Caruel	2013-292	Espíndola 12-006	ID	Chile		-43.2578, -71.9485
Parentucellia viscosa (L.) Caruel	2014-121	Espíndola 12-008	ID	Chile		-39.9398, -73.5829
Parentucellia viscosa (L.) Caruel	2014-122	Espíndola 12-012	ID	Chile		-41.6068, -72.6769
Physocalyx major Mart.	2012-194	GOR 2444	ID			
Rhinanthus crista-galli L.	2009-116	Curtis Bjork 6656	ID	USA		
Rhinanthus freynii Fiori	2010-184	Bennett 88	GH			
Rhinanthus serotinus (Schönh. ex Halácsy & Heinr.Braun) Oborny -	2009-114	Lytton Musselman 4871	ID	Netherlands		
Tozzia alpina L.	2010-183	Bennett 87	GH			