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

Simon Uribe-Convers[1,2,3,¶,*], Matthew L. Settles[1,2,¶], David C. Tank[1,2,3]

[1]Department of Biological Sciences, University of Idaho, Moscow, Idaho, USA

[2]Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, Idaho, USA

[3]Stillinger Herbarium, University of Idaho, Moscow, Idaho, USA

[*]Corresponding author: uribe.convers@gmail.com (SUC)

[¶]These authors contributed equally to this work.

Abstract

Advances in high-throughput sequencing (HTS) have allowed researchers to obtain large amounts of biological sequence information at speeds and costs unimaginable only a decade ago. Phylogenetics, and the study of evolution in general, is quickly migrating towards using HTS to generate larger and more complex molecular datasets. In this paper, we present a method that utilizes microfluidic PCR and HTS to generate large amounts of sequence data suitable for phylogenetic analyses. The approach uses a Fluidigm microfluidic PCR array and two sets of PCR primers to simultaneously amplify 48 target regions across 48 samples, incorporating sample-specific barcodes and HTS adapters (2,304 unique amplicons per microfluidic array). The final product is a pooled set of amplicons ready to be sequenced, and thus, there is no need to construct separate, costly genomic libraries for each sample. Further, we present a bioinformatics pipeline to process the raw HTS reads to either generate consensus sequences (with or without ambiguities) for every locus in every sample or—more importantly—recover the separate alleles from heterozygous target regions in each sample. This is important because it adds allelic information that is well suited for coalescent-based phylogenetic analyses that are becoming very common in conservation and evolutionary biology. To test our subgenomic method and bioinformatics pipeline, we sequenced 576 samples across 96 target regions belonging to the South American clade of the genus *Bartsia* L. in the plant family Orobanchaceae. After sequencing cleanup and alignment, the experiment resulted in ~25,300bp across 486 samples for a set of 48 primer pairs targeting the plastome, and ~13,500bp for 363 samples for a set of primers targeting regions in the nuclear genome. Finally, we constructed a combined concatenated matrix from all 96 primer combinations, resulting in a combined aligned length of ~40,500bp for 349 samples.

Introduction

Advances in high-throughput sequencing (HTS) have allowed researchers to obtain large amounts of genomic information at speeds and costs unimaginable only 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 common [1-3]. Driven by the need to generate homogenous, informative, and affordable multilocus datasets, we present a new approach for obtaining affordable 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., 4,5], genotyping of single nucleotide polymorphisms (SNP) [e.g., 6-8], gene expression [e.g., 9-11], and targeted resequencing [e.g., 12,13] but, to our knowledge, has not yet been used to assemble molecular phylogenetic datasets for systematic studies [but see 14 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 simultaneously amplify 48 target regions across 48 samples, incorporating sample-specific barcodes and HTS adapters (2,304 amplicons per microfluidic array). This four-primer PCR approach circumvents the need to construct genomic libraries for every sample, avoiding the high costs and time requirements involved in library preparation. Furthermore, by using a dual barcoding strategy, we are able to multiplex on a single Illumina MiSeq run 24 microfluidic arrays, representing two distinct sets of 48 target regions (plastome and nuclear, in this experiment) across 576 samples (55,296 distinct amplicon sequences) from the South American clade of the plant genus *Bartsia* L. (Fig. 1) and its close relatives in Orobanchaceae, demonstrating the power of this approach for species-level phylogenetics.

Fig.1 Floral diversity in the South American *Bartsia* clade.

The sections from top left to bottom right are: *Strictae, Diffusae, Laxae,* and *Orthocarpiflorae*.

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., 15-18], the straightforwardly interpreted results due to its non-recombining nature, conserved gene order, and gene content [19], and its historical importance since the beginning of the field [e.g., 20]. Large datasets have been produced with approaches that have involved massively parallel sequencing [e.g., 17], a compilation of coding regions from both whole plastome sequences [e.g., 18] and targeted approaches [e.g., 21], transcriptomics [e.g., 22], RNA hybridization or capture probes [e.g., 3], and long range PCR coupled with HTS [e.g., 23]. Because the chloroplast genome evolves relatively slowly, ~3–5 times slower than the nuclear genome in plants [24-26], 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., 17,27] to the level of major clades [e.g., 18,28]. 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 [29]. 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 [30-32].

For the nuclear genome, phylogenomic datasets have been obtained in plant systems using genome skimming [e.g., 33], sequence capture [e.g., 34,35], and restriction-site associated DNA—RADseq [e.g., 36]. Likewise, the field of phylogenomics in animals has advanced with

datasets obtained with targeted amplicon sequencing (TAS) in Pancrustacea [37] and North American tiger salamanders [38], GBS in butterflies [39], fish [40], and beetles [41]. However, genome-scale datasets for animal phylogenetics has been most heavily impacted by sequence capture approaches focused on ultraconserved genomic elements (UCEs) at various taxonomic scales, e.g., vertebrates [1], amniotes [2,42], turtles [43], birds [44], ray-finned fishes [45], and chipmunks [46]. Furthermore, UCEs have been shown to be an important resource for gathering information from museum specimens [47], and to be useful at shallow evolutionary time scales in birds [48].

In plant systems, genome skimming [33] 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 capture bait design, genome skimming requires no pre-existing genomic information. Genome skimming is a reference-guided approach that takes advantage of high copy 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. [33], study introgression in *Fragaria* L. species [49], identify horizontal transfer of DNA from the mitochondrion to the chloroplast in *Asclepias syriaca* L. [50], resolve phylogenetic relationships in the family Chrysobalanaceae [51], recover plastomes across multiple genera [23], and was also used 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 need

to first construct HTS libraries for each sample in the study, a step that greatly increases the time and costs of the experiment. Second, variable regions flanking the UCEs are often captured at much reduced depth as one moves away from the UCE, or the UCE is lost completely if the target taxon is phylogenetically divergent from the one used in the bait design [3,48]. Smith et al. [48] 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 utility 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 [52], 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 [53]. 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., 37,38], but contains major improvements in efficiency. For example, Bybee et al. [37] 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 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. [37] 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.

In this paper, we test the performance and utility of our targeted, subgenomic approach using the Neotropical clade of the plant genus *Bartsia* L. (Orobanchaceae) (Fig. 1). This clade is comprised of approximately 45 closely related species that are part of an ongoing rapid and recent radiation in the páramo ecosystem above tree line (~2900m in elevation) throughout the Andes [54]. Using minimal genomic resources collected via plastome sequencing [23] 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 across hundreds of samples, which allowed us to investigate evolutionary relationships at the species level. While shotgun approaches 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 across a significantly greater number of samples, which is ideal for phylogenetic studies at low taxonomic levels.

Methods

Microfluidics PCR Primer Design and Validation

Preliminary data acquisition

Data used for chloroplast and nuclear microfluidics PCR primer design were generated from two taxa using long PCR to generate Plastome DNA templates for HTS [23] and three taxa (4 samples, Table 1) using genome skimming [33]. DNA was extracted from ~0.02 g of silica gel-dried tissue using a modified 2X CTAB method [55], 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 500bp 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). The long PCR libraries and one genome skimming library were sequenced on an Ilumina HiSeq 2000 at the Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley, whereas the remaining genome skimming libraries were sequenced on an Illumina HiSeq 2000 at the Genomics Core Facility at the University of Oregon (Table 1). Raw reads

were cleaned using SeqyClean v 1.8.10 (https://bitbucket.org/izhbannikov/seqyclean) using defaults settings.

Table 1 Sample and sequencing information for preliminary data acquisition.

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	UO	This study
Bartsia pedicularoides Benth.	Uribe-Convers 2011-64	Illumina HiSeq 2000	100 paired end	52.9	UO	This study
Bartsia serrata Molau	Uribe-Convers 2012-15	Illumina HiSeq 2000	100 paired end	65.1	UO	This study

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; UO = Genomics Core Facility at the University of Oregon.

Chloroplast target selection

For the chloroplast PCR primer set, the cleaned reads were assembled against a reference genome (*Sesamum indicum* L., GenBank accession JN637766) using the Alignreads pipeline v. 2.25 [52], 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 genome

skimming—were aligned using MAFFT v.7.017b in its default settings [56]. To target the putatively most phylogenetically informative regions of the plastome, we developed custom scripts in R [57] to identify the most variable regions in the alignment that were flanked by conserved regions, and spanned between 400bp and 1000bp. This allowed us to rank and prioritize regions in the plastome for primer design (Scripts deposited in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.fh592).

Nuclear target selection

For the nuclear PCR primer set, cleaned reads from the genome skimming samples were compared to two publicly available genomic databases, a list of the pentatricopeptide repeat genes (PPR) and the conserved orthologous set II (COSII), using the BLAST-Like Alignment Tool (BLAT, tileSize=7, minIdentity =80) [58]. 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. [59], and was previously used to successfully infer the phylogenetic relationships of the plant family Verbenaceae and the *Verbena* L. complex [60], and in the plant clade Campanuloideae (Campanulaceae) [61]. Similarly, the COSII genes have been identified to be putatively single-copy and orthologous across the Euasterid plant clade [62], and several loci have been used for phylogenetic reconstructions of closely related species in the plant families Orobanchaceae [63] and Solanaceae [64].

Using a custom R-script (deposited in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.fh592), reads that matched any reference gene from these two databases were kept, binned with their respective reference locus, and aligned using MAFFT in its default settings. We then used the online tool IntronFinder (http://solgenomics.net/, last accessed in January 2014) from the Sol Genomics Network [65] to predict exon/intron junction

positions in the COSII genes. The PPR genes do not contain introns [59] and thus this step was not necessary for these loci. Reference loci that had reads from at least two taxa aligned to them forming conserved 'islands' separated by 400-800bp, including estimated introns with an assumed average length of 100bp, were selected for primer design (Fig. 2A). Additionally, an alignment from nuclear rDNA internal and external transcribed spacers sequences—ITS and ETS, respectively [54]—as well as an alignment of sequences of the *PHOT1* gene and one of the *PHOT2* gene [66] were made in MAFFT with default settings.

Microfluidic PCR primer design and validation

Forward and reverse primers for the selected chloroplast regions and nuclear loci were designed using Primer3 [67-69] following the recommended criteria specified in the Fluidigm 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 taxon 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 S1 Table. 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. 2B and 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, in order to dramatically increases the number of samples that can be multiplexed in one sequencing run (S2 Table).

Primer validation

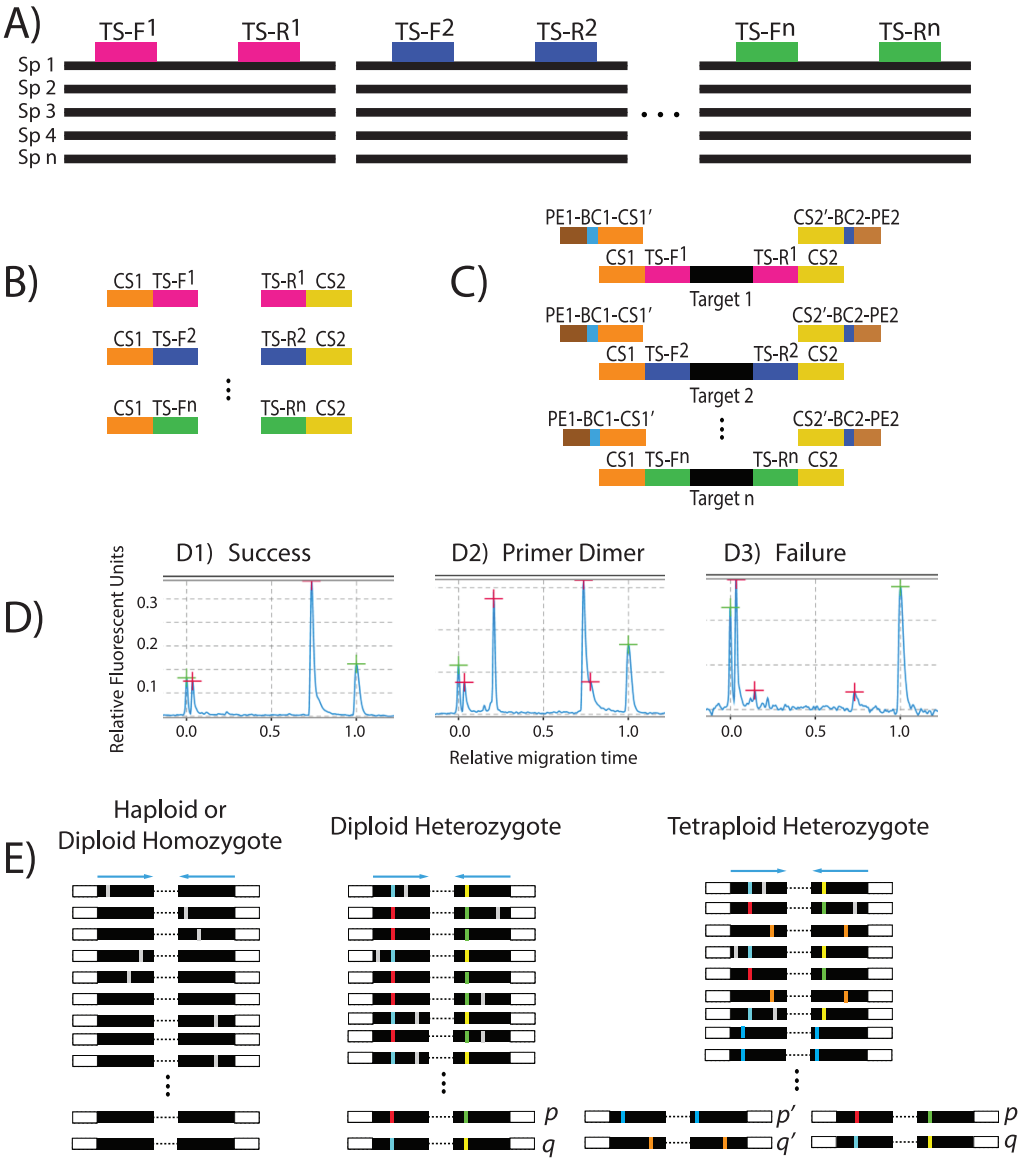
Due to the complexity of simultaneously using two sets of primers in one PCR, it is necessary to validate each set of primers prior to the actual microfluidic PCR amplification. Primer validation is a crucial step to ensure that no primer dimers are formed and that no interaction and/or competition between 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). Resulting 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. 2D, S1 Table).

Fig. 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 performed by simulating the microfluidic amplification reaction 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').

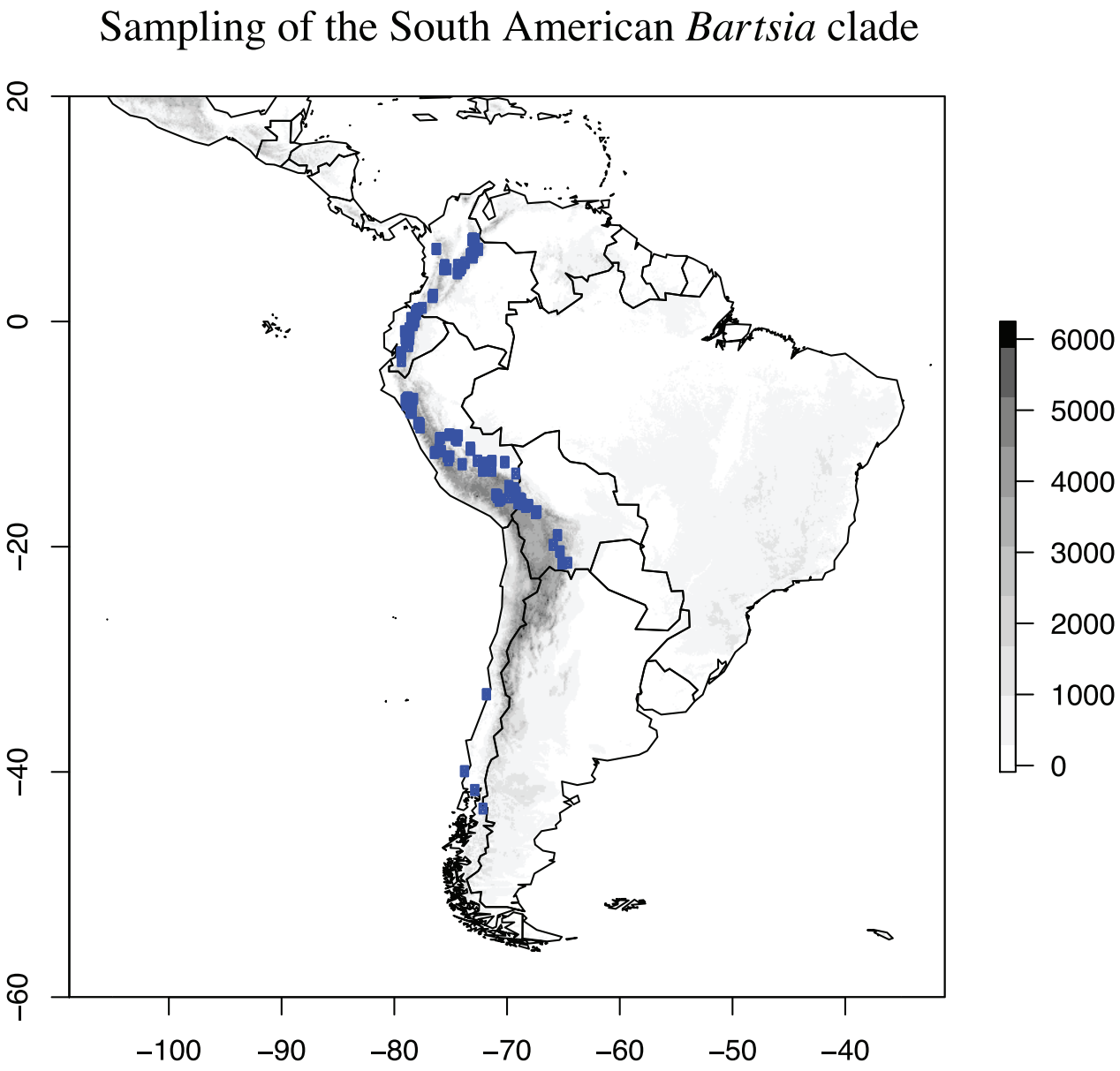


Sampling, microfluidic PCR and sequencing

We were interested in generating data to investigate the evolutionary history of the Neotropical *Bartsia* clade [54], 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 across the 576 samples. These samples encompassed the entire geographic breadth of the South American clade, with samples ranging from northern Colombia to southern Chile (S3 Table, Fig. 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 (S3 Table). For all samples, DNA was extracted from ~0.02 g of silica gel-dried tissue using a modified 2X CTAB method [55], yielding ~30 to 70 ng/µl of DNA per sample.

Fig. 3 Sampling effort in the South American *Bartsia* clade.

A total of 576 samples, localities 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.



Microfluidic PCR was performed in an Access Array System (Fluidigm) using 24 (12 for the chloroplast and 12 for the nuclear set) 48.48 Access Array Integrated Fluidic Circuits (Fluidigm) following the manufacturer's protocols. This particular array allows for 48 samples to be simultaneously amplified across 48 distinct primer pairs, resulting in 2,304 isolated and unique PCR amplicons per array. While we chose here to amplify 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 amplification of all 96 primer pairs in a single array (i.e., target specific primers for one chloroplast region and one nuclear locus pooled prior to amplification). 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 350bp, 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 final yield of 21.4 million 300bp 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.

Illumina sequence data processing

Reads from the Illumina MiSeq run were demultiplexed for sample, using the sample-specific dual barcode combinations, and chloroplast region or nuclear locus, using the target specific primers with the python preprocessing application dbcAmplicons (https://github.com/msettles/dbcAmplicons). Representative sequences for each amplicon were then identified using the reduce_amplicons R script (https://github.com/msettles/dbcAmplicons/blob/master/scripts/R/reduce_amplicons.R). This pipeline allows for an optional trimming step, where the number of trimmed bases can be set independently for both the forward and reverse reads (using the “--trim-1 and --trim-2" flags in

the reduce_amplicons R script). This step is especially useful in instances of high sequencing errors rates towards the end of a read, especially in read 2 on the MiSeq platform. In this study 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 dbcAmplicons application allows for target specific primer matching errors less than or equal to four, as long as the final four bases of the 3' end of the target specific primers were exact matches, thus yielding firm ends. The target specific primer sequence is removed and paired reads that overlap by at least 10 bp are joined into a single continuous sequence. Finally, for every sample at every locus, the reduce_amplicons R script produces consensus sequences (using the "-p" flag in the reduce_amplicons R script) with IUPAC ambiguity codes (-p ambiguity) for individual sites represented by more than one base when each variant is present in at least 5 reads and 5% of the total number of reads (these thresholds are adjustable using the "-s" and "–f" flags, respectively), or without ambiguities ("-p consensus"). For allele recovery ("-p occurrence"), reads are reduced to counts of identical pairs or joined paired reads. If the count represents at least 5% of the total number of reads and contains at least 5 reads (again, adjustable by the user), that amplicon is retained as a candidate allele for the sample and target (Fig. 2E). Conversely, if both the sequencing depth and minimum total percentage thresholds are not met, the sample specific target 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 than actual biological variation.

Because the chloroplast genome is haploid, consensus sequences were generated for each of the 48 regions (using the "-p consensus" flag in the reduce_amplicons R script). In cases where the read 1 and read 2 consensus sequences did not overlap, the reads were concatenated

into a single continuous sequence. Each region was aligned with MUSCLE v3.8.31 [70] in its default settings, alignments were cleaned with Phyutility v2.2.4 [71] 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 were discarded. Finally, the 48 chloroplast alignments were concatenated with Phyutility into a single locus.

To accommodate putative heterozygosity at nuclear loci, we generated consensus sequences with IUPAC ambiguity codes (using the "-p ambiguity" flag in the reduce_amplicons R script) for every sample at every locus, as well as individual allelic occurrences for each sample when applicable (using the "-p occurrence" flag in the reduce_amplicons R script). As with the chloroplast set, paired reads that did not overlap were concatenated, and each region was independently aligned using MUSCLE and cleaned with Phyutility at a 50 percent threshold. Because the ITS and ETS regions are physically linked in the rDNA repeat, theses regions were concatenated and treated as a single locus using Phyutility.

This data processing pipeline resulted in alignments for 47 independent nuclear loci – with allelic information, when relevant – and a concatenated chloroplast dataset. To compare alternative strategies for phylogenetic analyses, consensus sequences with IUPAC ambiguity codes for each of the 47 nuclear loci were concatenated into a single dataset (~13,500bp; the concatenated nuclear dataset), and the concatenated nuclear dataset and concatenated chloroplast dataset were combined into a single alignment of more than 40,500bp after cleaning.

Phylogenetic analyses

The concatenated chloroplast dataset was analyzed with PartitionFinder [72] to find the best partitioning scheme while also identifying the best-fit model of sequence evolution for each

possible partition. Using these partitioning schemes and models of sequence evolution, we conducted maximum likelihood (ML) analyses as implemented in GARLI v2.0.1019 [73] with ten independent runs, each with 50 nonparametric bootstrap replicates. Bootstrap support was assessed with the program SumTrees v3.3.1 of the DendroPy v3.12.0 package [74]. Likewise, we analyzed the dataset in a Bayesian framework as implemented in MrBayes v3.2.1 [75] 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 [76]. 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 [77] to ensure that the each locus was indeed single copy. Second, we analyzed the concatenated nuclear 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 (concatenation with monophyly constraints; CMC) [78]. Although not a formal coalescent-based species tree method, comparisons of the CMC approach to coalescent-based species tree approaches have found them comparable and potentially the least sensitive to taxonomic sampling [78]. Furthermore, the 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 79 for a potentially scalable approach]. Finally, the combined dataset (chloroplast and nuclear loci) – with and without monophyly constraints - was analyzed with RAxML. Although we understand the importance of analyzing this type of dataset in a coalescent framework, the scope of the present study is not to infer the species tree or make systematic conclusions for the clade in question, which is the focus of ongoing and future work, but rather to demonstrate the efficacy of this targeted, subgenomic approach for generating large phylogenetic datasets using microfluidic PCR and HTS.

Results

Preliminary Data Acquisition

Low coverage genomes were sequenced for genome skimming from four samples representing three species of *Bartsia*: *B. pedicularoides* Benth. (two samples), *B. santolinifolia* (Kunth) Benth., and *B. serrata* Molau. *Bartsia pedicularoides 1* yielded ~51.6 million 100 bp single-end reads (sequenced at the Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley). The other three samples (sequenced at the Genomics Core Facility at the University of Oregon) yielded an average of ~51.4 million 100bp paired-end reads per library (Table 1; GenBank Sequence Read Archive [SRA]: SRR2045582, SRR2045585, SRR2045588, SRR2045589). Seqyclean v 1.8.10 (https://bitbucket.org/izhbannikov/seqyclean) processing resulted in ~46.8 million reads for *B. pedicularoides 1*, ~52.9 million for *B. pedicularoides 2*, ~46.4 million for *B. santolinifolia*, and ~65.1 million for *B. serrata*. The plastomes assembled with the Alignreads pipeline from these samples had an average sequencing depth of 995x (Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.fh592).

Chloroplast and Nuclear Primer Design and Validation

The *Bartsia* chloroplast alignment of six plastomes (including only one copy of the inverted repeat) had a length of ~125kb (Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.fh592). From this alignment, we were able to design a total of 74 primer pairs that spanned the entire plastome. Following primer 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%) (S1 Table).

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 (aligned lengths of 6,711bp, 578 bp, 1,272bp - Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.fh592) all contained multiple places to design primers based on our criteria. A total of 188 primer pairs were designed from all datasets (S1 Table). 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 (S3 Table) that

represented 46 species of the clade and 28 related taxa as outgroups (included primarily to evaluate how far outside of the target group primers would successfully amplify targeted loci). Microfluidic amplification of the samples using 24 48.48 Access Array Integrated Fluidic Circuits (Fluidigm) resulted in up to 96 amplicons per sample (a total of 55,296 microfluidic reactions). After pooling and normalizing amplicons for each sample, pools were sequenced on an Illumina MiSeq platform with the Reagent Kit version 3, yielding ~20.3 million 300 bp paired-end reads. Raw reads were deposited in the GenBank Sequence Read Archive (SRA SRP058302).

Data Processing

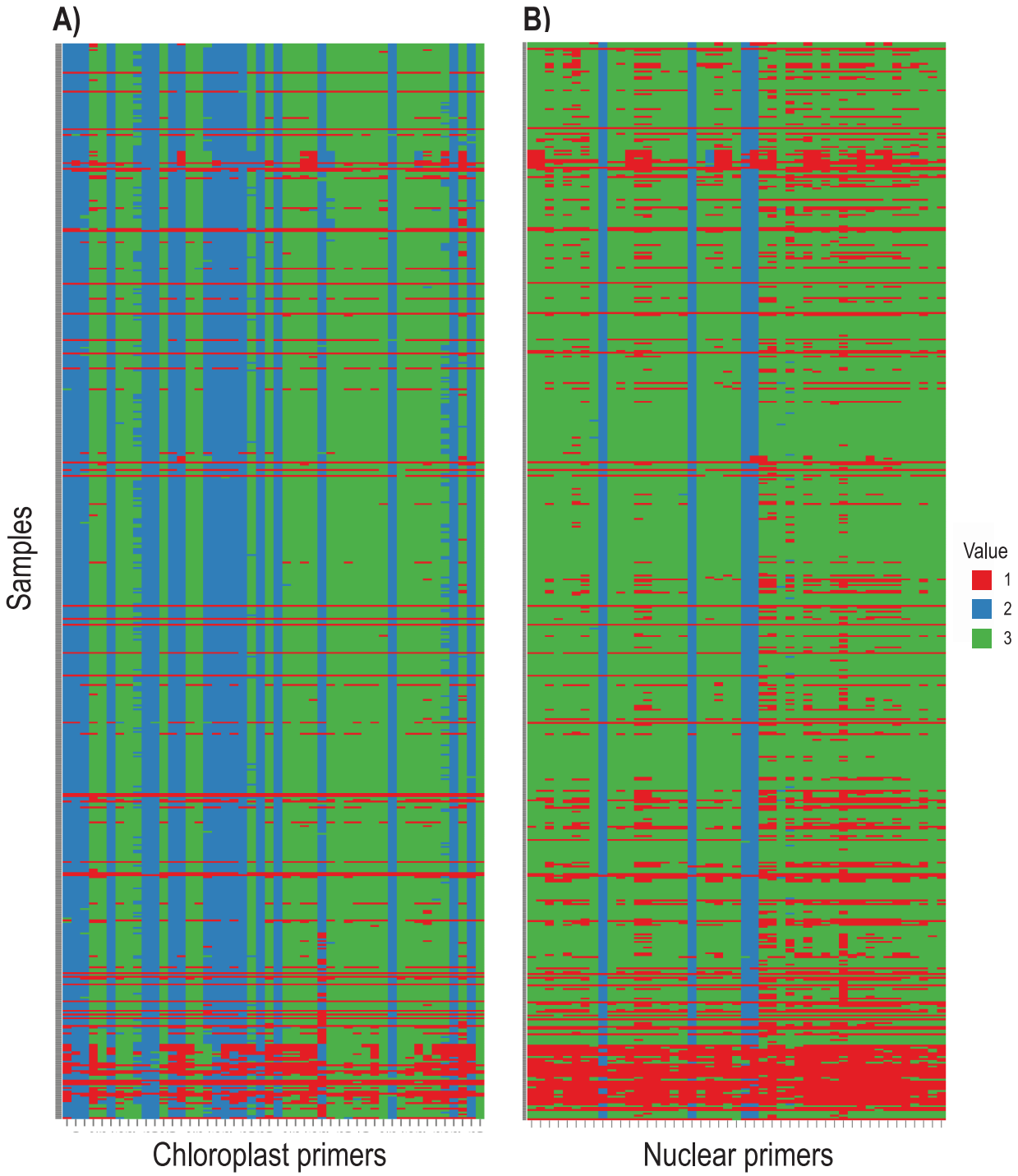
Following processing with dbcAmplicons, ~16.9 million reads (77.7%) were sufficiently matched to both barcodes (sample specific) and primers (target specific). Discarded reads (~4.5 million reads) were a combination of PhiX Control v3 (Illumina; ~3.2 million reads, or 15%) and reads that did not pass our criteria for matching both barcodes and the primers (~1.3 million reads, or 7.3%).

Chloroplast set. Of the 576 samples used in this study, 528 (91.7%) amplified at least one chloroplast DNA amplicon and 486 (84.0%) produced more than 40 amplicons (>21,300bp) (Fig. 4A). The majority of the samples that did not amplify efficiently belonged to outgroup taxa that are distantly related to the South American clade of *Bartsia*, suggesting that the designed primers were too specific to work efficiently outside this clade. This highlights the importance of careful primer design and validation that is in line with the taxonomic breath of the intended study. Because our primary focus was on the South American *Bartsia* clade, and primers were designed

with plastomes from this clade, these results were not surprising. Following processing with reduce_amplicons, multiple sequence alignment, and cleaning, the final chloroplast dataset included 486 samples and had an aligned length of ~25,300bp. The majority 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).

Fig. 4 Bird's-eye view heat map showing the target coverage for each sample

Target 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 paired reads overlapped by at least 10bp 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).



Nuclear set. For the nuclear DNA set, 47 out the 48 regions were amplified from the majority of samples (Fig. 4B), and only one region did not satisfy our amplicon demultiplexing criteria, i.e., did not meet both barcode and primer matching criteria. From these 47 regions, we were able to

recover sequence data from an average of 442.7 samples (76.0%), ranging from 520 (90.3%) to 318 (55.2%). Allele recovery (from processing with reduce_amplicons) resulted in 85.41% of these samples having one allele, 13.67 % having two, 0.82% having three, and 0.10% having four alleles (S4 Table). After summarizing across every sample for each species, 17 taxa presented a combination of one and two alleles across loci (diploid as the minimum ploidy level), while the remaining 33 taxa had a combination of three and four alleles (tetraploid as the minimum ploidy level) (S4 Table). Following alignment cleaning of consensus sequences with ambiguities (from processing with reduce_amplicons), the nuclear gene regions had an average aligned length of 332bp (from 267 to 459bp). Preliminary ML phylogenetic analyses in RAxML revealed three loci where paralogous copies were amplified, and another three loci that were too variable to be unambiguously aligned (due to non-specific PCR amplification). These six loci were removed prior to downstream phylogenetic analyses, resulting in a concatenated nuclear dataset of 41 regions (nrDNA plus 40 single copy nuclear gene regions) with an aligned, cleaned length of ~13,500bp from 363 samples. Finally, we constructed a combined concatenated matrix (nuclear and chloroplast) with an aligned, cleaned length of ~40,500bp, including 349 samples (all sequences deposited in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.fh592).

Phylogenetic analyses

Individual nuclear gene trees were largely unresolved, and paralogous loci were identified in three of the 48 loci. The concatenated chloroplast dataset was first analyzed with PartitionFinder to identify 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,

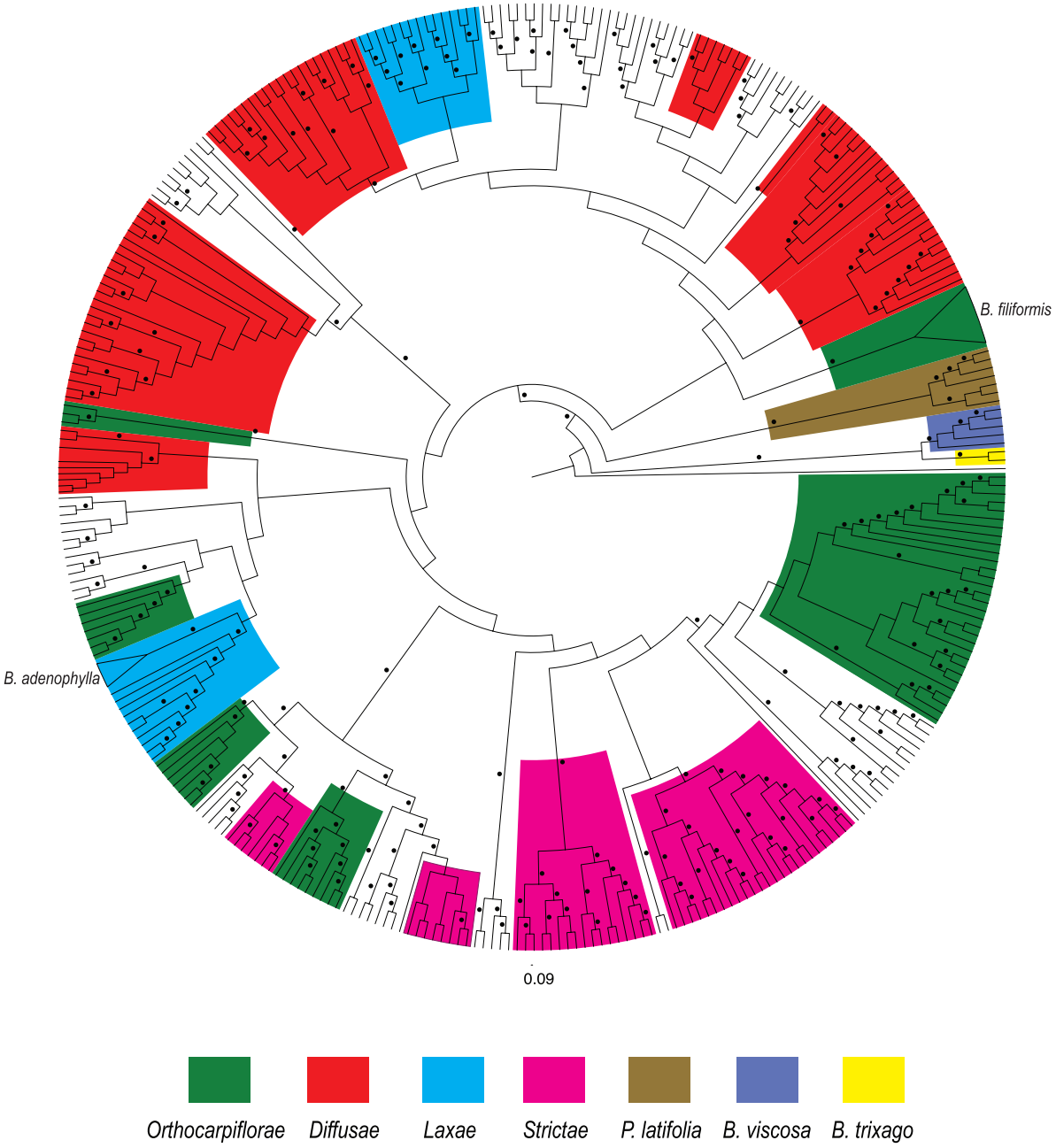
F81, K80+I, TVMef+I+G, TVMef+I+G, F81, TVM+I+G, TVM+I+G (data partitions and corresponding models can be found in Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.fh592). Analyses in ML and Bayesian frameworks, in GARLI and MrBayes, respectively, resulted in the same overall phylogenetic relationships 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 a very similar tree, the results and discussion will be based on the combined concatenated dataset, with and without constraints (all tree files deposited in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.fh592).

Several clades that correspond to relationships between outgroup taxa and South American *Bartsia* species were recovered with 100% bootstrap support (BS) and a posterior probability (PP) of 1.0. First, all individuals included for the outgroup taxa *Bellardia trixago, Bellardia viscosa*, and *Parentucellia latifolia* were reciprocally monophyletic. Second, all accessions of South American *Bartsia* species (the large majority of the sampling in this study) were monophyletic, and *P. latifolia* is the sister group to this clade. Third, *Bellardia trixago* and *B. viscosa* formed a distinct clade, and this clade is sister to the *P. latifolia* plus South American *Bartsia* clade (Fig. 5).

Fig. 5 Cladogram of the phylogenetic relationships of the South American *Bartsia* species and closely related taxa

Cladogram based on the maximum likelihood analysis in RAxML on the combined (chloroplast and nuclear) unconstrained dataset. Circles above the branches represent maximum likelihood

bootstrap support (BS) higher than 50. Clades containing individuals from species from only one of the four morphological sections of the South American clade [sensu 80] 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.



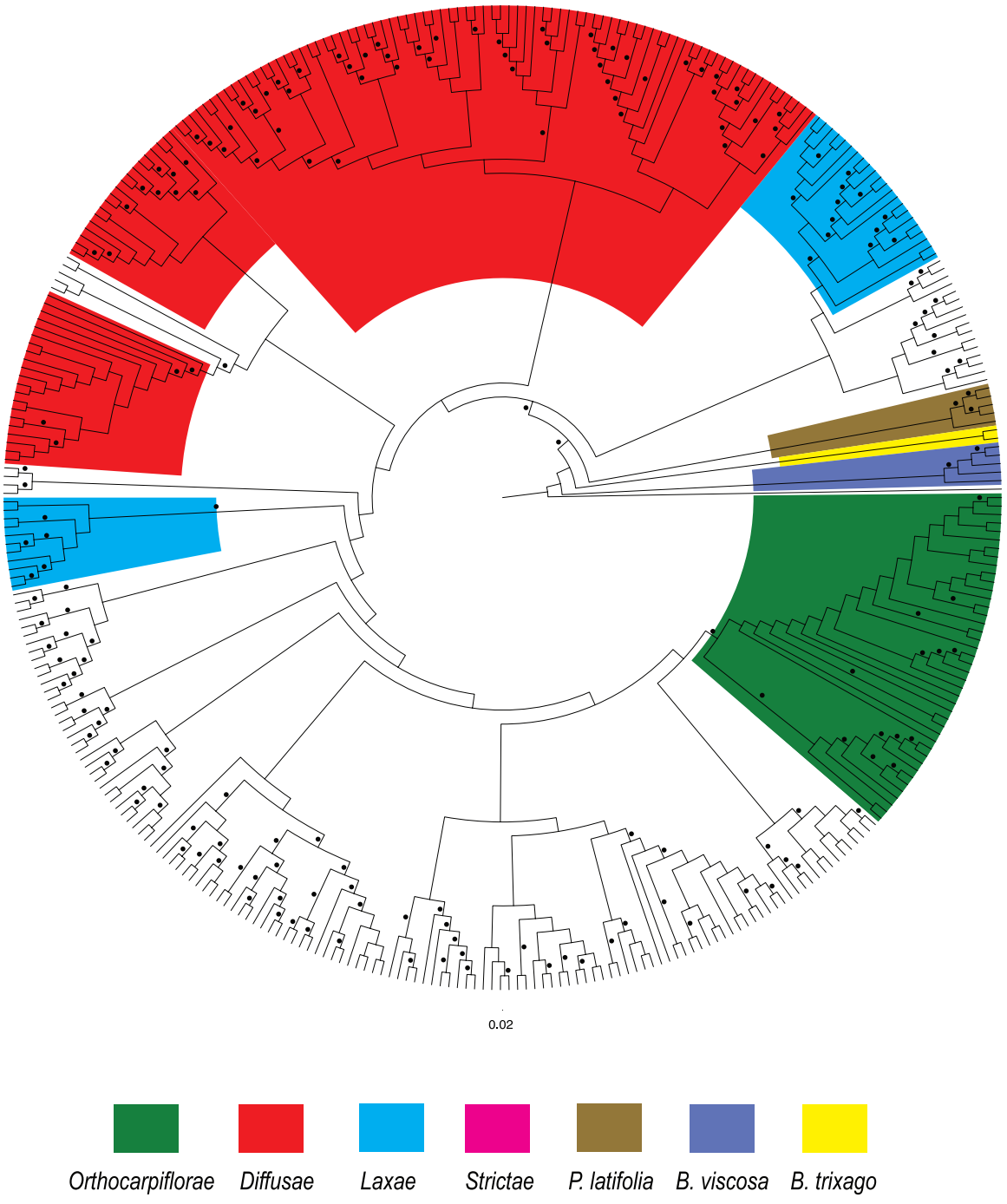
Support for backbone relationships in the South American *Bartsia* clade is low, and thus, very few systematic conclusions can be made at this point. First, we did not recover four monophyletic groups corresponding to the four morphological sections [sensu 80]. However, there are several clades that do contain multiple species from the same section with moderate support (Fig. 5). Furthermore, individuals of most of the species were recovered in multiple different clades, and in fact only two species (*B. filiformis* Wedd. 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 [54], 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 species tree (e.g., coalescent-based) and network analyses to confidently elucidate relationships among species 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 of individuals belonging to named species reduced BS support of the backbone relationships even further (Fig. 6), indicating that sequences from some of the individuals that were constrained to be monophyletic clearly violate this assumption.

Fig. 6 Cladogram of the phylogenetic relationships of the South American *Bartsia* species and closely related taxa

Cladogram 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. Circles above the branches represent maximum likelihood bootstrap support (BS) higher than 50. Clades containing at least two species from one of the four morphological

section of the South American clade [sensu 80] 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, it is clear that the field of phylogenetics, and the study of evolution in general, are 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 expensive genomic shotgun libraries for each sample in the experiment. This step greatly increases the time and cost of any HTS approach, effectively reducing the sample size possible for any experiment. The approach presented here takes advantage of a four-primer PCR amplification to efficiently tag multiple genomic targets 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 PCR primers. 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 barcoded primers (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 smaller 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 30-60ng/μL genomic DNA per sample. By conducting a simultaneous four-primer reaction, one avoids the necessity of performing multiple rounds of manual PCR to incorporate barcodes and adapters – a limitation of the Targeted Amplicon Sequencing (TAS) strategy of Bybee et al. [37]. For example, following the TAS approach, to produce tagged amplicons for the 576 samples and 96 gene regions targeted in this study, it would have required >1,100 96-well plates (or >275 384-well plates) of PCR to produce the same number of barcoded amplicons. 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 also need to be performed. Nevertheless, with high levels of taxon-by-gene region samplings, TAS becomes unpractical. Using the microfluidic PCR 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., 37] 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. 2E), it is possible to distinguish individual alleles at heterozygous nuclear loci without requiring additional assembly or mapping steps. Because targeted loci were specifically amplified one locus per reaction per sample, when paired-end sequences from each specific barcode and primer combination are

identified, heterozygous loci in diploid species appear as high frequency amplicons, allowing us to straightforwardly determine both alleles. 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 81 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., [1] or by selecting only one allele and discarding others e.g., [2,48]). 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. [38] used a haplotype phasing strategy to computationally determine individual alleles. For statistical phasing approaches, the number of individuals per population present in a sample is a critical factor in determining how well haplotype phase can be estimated [82], 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 [83], fish [84], amphibians [85], and reptiles [86], 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., 87], 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 may be possible to calculate the frequencies of alleles across all loci and not only recover individual alleles, but potentially estimate ploidy level – or at least a minimum ploidy level depending on levels of heterozygosity. In plants, this could be 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.

Within the South American *Bartsia* clade, only 23 of 45 species have published chromosome numbers, and seven of these have been characterized as tetraploids based on these counts (S4 Table) [80]. Likewise, chromosome counts have been published for the European *Bartsia alpina* (diploid), and the Mediterranean species *Bellardia viscosa* (tetraploid) and *Parentucellia latifolia* (tetraploid) [80]. The reduce_amplicons pipeline employed here recovered at least three alleles in one locus for six of these species, suggesting that their minimum ploidy level might be tetraploid. Although we only recovered one or two alleles in the remaining four species (suggesting that the taxa may be diploid), it is important to keep in mind that these species comprise a very recently diverged clade, and most species occur in small, isolated

populations [54], where low sequence divergence, autopolyploidy, and homozygosity may mask true ploidy levels. Something to keep in mind, however, is the number of samples (individuals) per species that were recovered as having more than two alleles. For some of these species, e.g., *Bartsia camporum* Diels, *Bartsia serrata* Molau, and *Bellardia viscosa*, the majority of individuals (more than 65%) presented at least three alleles—supporting that these taxa are indeed tetraploid. On the other hand, species such as *Bartsia pyricarpa* Molau, *Bartsia pedicularoides* Benth., and *Bartsia patens* Benth.—all of which are suggested to be tetraploid based on chromosome counts (S4 Table) [80]—were recovered with most of their individuals presenting one or two alleles and only 2% to 15% of the individuals having at least three alleles. This highlights the necessity of including more than one individual from more than one population when assessing levels of ploidy. To fully investigate the utility of this approach for bioinformatically estimating ploidy levels, chromosome counts and/or flow cytometry data from these same samples would be necessary, but this was beyond the scope of this study. Nevertheless, these results are encouraging as they open the door for future comparisons between 'bioinformatically karyotyped' samples and traditional ploidy estimation experiments.

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 [88] and ETS [89] 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. [59] and shown to be phylogenetically informative at the family level in Verbenaceae [90] and at the subfamily level in Campanuloideae [61], and 4) intron-spanning regions from within the COSII gene set developed by Wu et al. [62] and used within Orobanchaceae [63], and the Phototropin 2 gene used at the interspecific level in *Glandularia, Junellia* and *Verbena* in the Verbenaceae [66]. 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, the genome skimming approach [33,52] 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 [91], and Genome 10k for animals [92] – as well as a recently published bioinformatics pipeline for identifying single-copy nuclear genes and designing target specific primers for phylogenomic analyses using existing transcriptome data [93].

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, phylogenetic species tree methods that explicitly incorporate mechanisms

that lead to gene tree-species tree discordance (e.g., coalescent stochasticity, divergence with gene flow) are needed. However, these detailed analyses are beyond the scope of this study (which is focused on data collection approaches), and therefore, the phylogenetic results presented here are preliminary. Nevertheless, the monophyly of the group is highly supported by all analyses, and is in agreement with recent biogeographic study of the clade [54]. Interspecific relationships, however, have very little support – a pattern commonly seen in rapid radiations like this – and only two species were found to be monophyletic. These two species are taxa with small and restricted geographic distributions with likely small effective population sizes. Given that the time to coalescence is directly linked to effective population size [30], it is not unexpected that individuals from these species were monophyletic in our unconstrained analyses. Conversely, when we look at species with a large geographic distributions, and larger effective population sizes (e.g., *B. pedicularoides* Benth.), we see that the individuals are recovered in multiple different groups across the tree (Fig. 5).

Enforcing the monophyly of species has been used as a 'pseudo-species tree' method with good results [78], and some relationships recovered here make evolutionary sense – *Bartsia sericea* Molau and *B. crisafullii* N. Holmgren were recovered as sister species (Fig. 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. 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, each of these (or any combination of) are

potential mechanisms that increase the phylogenetic complexity of this group, and incorporating these processes into species tree estimation in this clade is the focus of ongoing systematic studies using these data.

Conclusion

We presented a novel approach to generate large multilocus phylogenomic datasets for a large number of samples and species using microfluidic PCR and HTS. This approach allows for more control in targeting informative regions of the genome to be sequenced, resulting in datasets that are tailored to address the specific questions being asked, and that are orthologous across samples. Additionally, this method is both cost effective and time efficient, as it does not require genomic shotgun 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 48 nuclear gene regions representing a range of both coding and non-coding data. This targeted, subgenomic strategy for the collection of multilocus data for phylogenetic studies provided us with a large, but modest, set of loci that will be appropriate for sophisticated species tree inference methods (e.g., coalescent-based, networks, concordance analyses), and provided us with the first species level phylogeny for the South American *Bartsia* clade. Furthermore, the bioinformatic approaches employed here allowed for the recovery of individual alleles in heterozygote individuals (without the need for statistical phasing), and opened the door for the exploration of bioinformatic approaches to estimating ploidy levels—an important and often overlooked consideration at low taxonomic levels.

Acknowledgements

We would like to thank Dan New, Tamara Max, and the Institute for Bioinformatics and Evolutionary Studies (IBEST - NIH/NCRR P20RR16448 and P20RR016454) at the University of Idaho for technical assistance and bioinformatic resources. Jonathan Eastman was instrumental in R scripting and data processing. Jack Sullivan, Luke J. Harmon, Eric H. Roalson, Diego F. Morales-Briones, Matthew W. Pennell, Tracy C. Peterson, and XX anonymous reviewers offered helpful comments on the manuscript.

References

1. Lemmon AR, Emme SA, Lemmon EM (2012) Anchored Hybrid Enrichment for Massively High-Throughput Phylogenomics. Systematic Biology 61: 727–744. doi:10.1093/sysbio/sys049.

2. Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, et al. (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Systematic Biology 61: 717–726. doi:10.1093/sysbio/sys004.

3. Stull GW, Moore MJ, Mandala VS, Douglas NA, Kates H-R, et al. (2013) A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes. Applications in Plant Sciences 1: 1–7. doi:10.3732/apps.1200497.

4. Walter K, Holcomb T, Januario T, Du P, Evangelista M, et al. (2012) DNA Methylation Profiling Defines Clinically Relevant Biological Subsets of Non-Small Cell Lung Cancer. Clinical Cancer Research 18: 2360–2373. doi:10.1158/1078-0432.CCR-11-2635-T.

5. Gaedcke J, Grade M, Camps J, Sokilde R, Kaczkowski B, et al. (2012) The Rectal Cancer microRNAome - microRNA Expression in Rectal Cancer and Matched Normal Mucosa. Clinical Cancer Research 18: 4919–4930. doi:10.1158/1078-0432.CCR-12-0016.

6. Byers RL, Harker DB, Yourstone SM, Maughan PJ, Udall JA (2012) Development and mapping of SNP assays in allotetraploid cotton. Theor Appl Genet 124: 1201–1214. doi:10.1007/s00122-011-1780-8.

7. Bhat S, Polanowski AM, Double MC, Jarman SN, Emslie KR (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. doi:10.1371/journal.pone.0039181.t004.

8. Lu X, Wang L, Chen S, He L, Yang X, et al. (2012) Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. Nature Publishing Group 44: 890–894. doi:10.1038/ng.2337.

9. Moignard V, Macaulay IC, Swiers G, Buettner F, Schütte J, et al. (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. doi:10.1038/ncb2709.

10. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, et al. (2013) Single-cell transcriptomics reveals bimodality inexpression and splicing in immune cells. Nature: 1–5. doi:10.1038/nature12172.

11. Dominguez MH, Chattopadhyay PK, Ma S, Lamoreaux L, McDavid A, et al. (2013) Journal of Immunological Methods. Journal of Immunological Methods 391: 133–145. doi:10.1016/j.jim.2013.03.002.

12. Lohr JG, Stojanov P, Lawrence MS, Auclair D, Chapuy B, et al. (2012) Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proc Natl Acad Sci USA 109: 3879–3884. doi:10.1073/pnas.1121343109/-/DCSupplemental.

13. Moonsamy PV, Williams T, Bonella P, Holcomb CL, Höglund BN, et al. (2013) High throughput HLA genotyping using 454 sequencing and the Fluidigm Access Array™ system for simplified amplicon library preparation. Tissue Antigens 81: 141–149. doi:10.1111/tan.12071.

14. Godden GT, Jordon-Thaden IE, Chamala S, Crowl AA, García N, et al. (2012) Making next-generation sequencing work for you: approaches and practical considerations for marker development and phylogenetics. Plant Ecology & Diversity 5: 427–450. doi:10.1080/17550874.2012.745909.

15. Graham SW, Olmstead RG (2000) Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. American Journal of Botany 87: 1712–1730.

16. Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proceedings of the National Academy of Sciences 104: 19363–19368. doi:10.1073/pnas.0708072104.

17. 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. doi:10.1186/1741-7007-7-84.

18. Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE (2010) Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proceedings of the National Academy of Sciences 107: 4623–4628. doi:10.1073/pnas.0907801107.

19. Downie SR, Palmer JD (1992) Use of chloroplast DNA rearrangements in reconstructing

plant phylogeny. In: Soltis PS, Soltis DE, Doyle JJ, editors. Molecular systematics of plants. Chapman and Hall, New York, NY. pp. 14–35.

20. Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, et al. (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. Annals of the Missouri Botanical Garden: 528–580.

21. Moore MJ, Hassan N, Gitzendanner MA, Bruenn RA, Croley M, et al. (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.

22. Xi Z, Ruhfel BR, Schaefer H, Amorim AM, Sugumaran M, et al. (2012) Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. Proceedings of the National Academy of Sciences 109: 17519–17524. doi:10.1073/pnas.1205818109.

23. Uribe-Convers S, Duke JR, Moore MJ, Tank DC (2014) A Long PCR–Based Approach for DNA Enrichment Prior to Next-Generation Sequencing for Systematic Studies. Applications in Plant Sciences 2: 1300063. doi:10.3732/apps.1300063.

24. Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84: 9054–9058.

25. Wolfe KH, Sharp PM, Li W-H (1989) Rates of synonymous substitution in plant nuclear genes. J Mol Evol 29: 208–211.

26. 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. doi:10.1016/j.ympev.2008.09.009.

27. 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. doi:10.1016/j.ympev.2012.08.026.

28. Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG (2014) From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol Biol 14: 23. doi:10.1186/1471-2148-14-23.

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

30. 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.

31. Edwards SV (2009) Is A New And General Theory Of Molecular Systematics Emerging. Evolution 63: 1–19. doi:10.1111/j.1558-5646.2008.00549.x.

32. Knowles LL, Kubatko LS (2010) Estimating species trees: practical and theoretical aspects. Wiley-Blackwell.

33. Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, et al. (2012) Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. American Journal of Botany 99: 349–364. doi:10.3732/ajb.1100335.

34. Tennessen JA, 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. doi:10.1534/g3.113.006288.

35. Mandel JR, Dikow RB, Funk VA, Masalia RR, Staton SE, et al. (2014) A Target Enrichment Method for Gathering Phylogenetic Information from Hundreds of Loci: An Example from the Compositae. Applications in Plant Sciences 2: 1300085. doi:10.3732/apps.1300085.

36. Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). Systematic Biology 62: 689–706. doi:10.1093/sysbio/syt032.

37. Bybee SM, Bracken-Grissom H, Haynes BD, Hermansen RA, Byers RL, et al. (2011) Targeted Amplicon Sequencing (TAS): A Scalable Next-Gen Approach to Multilocus, Multitaxa Phylogenetics. Genome Biology and Evolution 3: 1312–1323. doi:10.1093/gbe/evr106.

38. O'Neill EM, Schwartz R, Bullock CT, Williams JS, Shaffer HB, et al. (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. doi:10.1111/mec.12049.

39. Nadeau NJ, Martin SH, Kozak KM, Salazar C, Dasmahapatra KK, et al. (2012) Genome-wide patterns of divergence and gene flow across a butterfly radiation. Molecular Ecology 22: 814–826. doi:10.1111/j.1365-294X.2012.05730.x.

40. Jones JC, 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. doi:10.1111/mec.12269.

41. Cruaud A, Gautier M, Galan M, Foucaud J, Sauné L, et al. (2014) Empirical Assessment of RAD Sequencing for Interspecific Phylogeny. Mol Biol Evol 31: 1272–1274. doi:10.1093/molbev/msu063.

42. McCormack JE, Faircloth BC, Crawford NG, Gowaty PA, Brumfield RT, et al. (2012) Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. Genome Research 22: 746–754. doi:10.1101/gr.125864.111.

43. Crawford NG, Faircloth BC, MCCORMACK JE, BRUMFIELD RT, Winker K, et al. (2012) More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. Biology Letters 8: 783–786. doi:10.1101/gr.125864.111.

44. McCormack JE, Harvey MG, Faircloth BC, Crawford NG, Glenn TC, et al. (2013) A Phylogeny of Birds Based on Over 1,500 Loci Collected by Target Enrichment and High-Throughput Sequencing. PLoS ONE 8: e54848. doi:10.1371/journal.pone.0054848.s004.

45. Faircloth BC, Sorenson L, Santini F, Alfaro ME (2013) A Phylogenomic Perspective on the Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved Elements (UCEs). PLoS ONE 8: e65923. doi:10.1371/journal.pone.0065923.s001.

46. Bi K, Vanderpool D, Singhal S, Linderoth T, Moritz C, et al. (2012) Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC Genomics 13: 1–1. doi:10.1186/1471-2164-13-403.

47. Bi K, Linderoth T, Vanderpool D, Good JM, Nielsen R, et al. (2013) Unlocking the vault: next-generation museum population genomics. Molecular Ecology 22: 6018–6032. doi:10.1111/mec.12516.

48. Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT (2014) Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Systematic Biology 63: 83–95. doi:10.1093/sysbio/syt061.

49. Salamone I, Govindarajulu R, Falk S, Parks M, Liston A, et al. (2013) Bioclimatic, ecological, and phenotypic intermediacy and high genetic admixture in a natural hybrid of octoploid strawberries. American Journal of Botany 100: 939–950. doi:10.3732/ajb.1200624.

50. Straub SCK, Cronn RC, 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. doi:10.1093/gbe/evt140.

51. Malé P-JG, Bardon L, Besnard G, Coissac E, Delsuc F, et al. (2014) Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. Molecular Ecology Resources. doi:10.1111/1755-0998.12246.

52. Straub SCK, 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. doi:10.1186/1471-2164-12-211.

53. Lemmon AR, Brown JM, Stanger-Hall K, Lemmon EM (2009) The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. Systematic Biology 58: 130–145. doi:10.1093/sysbio/syp017.

54. Uribe-Convers S (2014) Phylogenomic Insights into the Radiation of an Andean Group of

Plants. Ph.D. Dissertation, University of Idaho, Moscow, Idaho, USA. Available: http://digital.lib.uidaho.edu/cdm/ref/collection/etd/id/585.

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

56. Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30: 772–780. doi:10.1093/molbev/mst010.

57. 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://wwwR-projectorg.

58. Kent WJ (2002) BLAT−−The BLAST-Like Alignment Tool. Genome Research 12: 656–664. doi:10.1101/gr.229202.

59. 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. doi:10.1111/j.1469-8137.2008.02739.x.

60. Yuan Y-W, Liu C, Marx HE, Olmstead RG (2010) An empirical demonstration of using pentatricopeptide repeat (PPR) genes as plant phylogenetic tools: phylogeny of Verbenaceae and the Verbena complex. Molecular Phylogenetics and Evolution 54: 23–35. doi:10.1016/j.ympev.2009.08.029.

61. Crowl AA, Mavrodiev E, Mansion G, Haberle R, Pistarino A, et al. (2014) Phylogeny of Campanuloideae (Campanulaceae) with Emphasis on the Utility of Nuclear Pentatricopeptide Repeat (PPR) Genes. PLoS ONE 9: e94199. doi:10.1371/journal.pone.0094199.s022.

62. Wu F, Mueller LA, Crouzillat D, Pétiard V, Tanksley SD (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. doi:10.1534/genetics.106.062455.

63. Li M, Wunder J, Bissoli G, Scarponi E, Gazzani S, et al. (2008) Development of COS genes as universally amplifiable markers for phylogenetic reconstructions of closely related plant species. Cladistics 24: 727–745. doi:10.1111/j.1096-0031.2008.00207.x.

64. Tepe EJ, Farruggia FT, 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. doi:10.3732/ajb.1000516.

65. Bombarely A, Menda N, Tecle IY, Buels RM, Strickler S, et al. (2011) The Sol Genomics Network (solgenomics.net): growing tomatoes using Perl. Nucleic Acids Research 39: D1149–D1155. doi:10.1093/nar/gkq866.

66. 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. doi:10.3732/ajb.0800133.

67. Rozen S, Skaletsky H (2000) Primer3 on the WWW for General Users and for Biologist Programmers. In: Misener S, Krawetz SA, editors. Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, Vol. 132. pp. 365–386.

68. Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. Bioinformatics 23: 1289–1291. doi:10.1093/bioinformatics/btm091.

69. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, et al. (2012) Primer3 - new capabilities and interfaces. Nucleic Acids Research 40: e115–e115. doi:10.1093/nar/gks596.

70. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792–1797. doi:10.1093/nar/gkh340.

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

72. Lanfear R, Calcott B, Ho SYW, Guindon S (2012) Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol 29: 1695–1701. doi:10.1093/molbev/mss020.

73. Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD dissertation, The University of Texas at Austin.

74. Sukumaran J, Holder MT (2010) DendroPy: a Python library for phylogenetic computing. Bioinformatics 26: 1569–1571. doi:10.1093/bioinformatics/btq228.

75. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, et al. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61: 539–542. doi:10.1093/sysbio/sys029.

76. Rambaut A, Drummond AJ (2004) Tracer. University of Edinburgh, Edinburgh, UK Available at http://treebioedacuk/software/tracer/.

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

78. Linnen C, Farrell B (2008) Comparison of Methods for Species-Tree Inference in the Sawfly Genus Neodiprion (Hymenoptera: Diprionidae). Systematic Biology 57: 876–890. doi:10.1080/10635150802580949.

79. Chifman J, Kubatko L (2014) Quartet Inference from SNP Data Under the Coalescent

Model. Bioinformatics in press. doi:10.1093/bioinformatics/btu530.

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

81. Bao S, Jiang R, Kwan W, Wang B, Ma X, et al. (2011) Evaluation of next-generation sequencing software in mapping and assembly. J Hum Genet 56: 406–414. doi:10.1038/jhg.2011.43.

82. Browning SR, Browning BL (2011) Haplotype phasing: existing methods and new developments. Nature Reviews Genetics 12: 703–714. doi:10.1038/nrg3054.

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

84. Leggatt RA, Iwama GK (2003) Occurrence of polyploidy in the fishes. Reviews in Fish Biology and Fisheries 13: 237–246.

85. Cannatella DC, De Sá RO (1993) Xenopus laevis as a model organism. Systematic Biology 42: 476–507.

86. Bogart JP (1980) Evolutionary implications of polyploidy in amphibians and reptiles. Basic Life Sci 13: 341–378. doi:10.1007/978-1-4613-3069-1_18.

87. Judd WS, Soltis DE, Soltis PS, 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.

88. Baldwin BG (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.

89. Baldwin BG, 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.

90. Marx HE, O'Leary N, Yuan Y-W, Lu-Irving P, Tank DC, et al. (2010) A molecular phylogeny and classification of Verbenaceae. American Journal of Botany 97: 1647–1663. doi:10.3732/ajb.1000144.

91. Ilut DC, Doyle JJ (2012) Selecting Nuclear Sequences for Fine Detail Molecular Phylogenetic Studies in Plants: A Computational Approach and Sequence Repository. Syst Bot 37: 7–14. doi:10.1600/036364412X616576.

92. 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. doi:10.1093/jhered/esp086.

93. Chamala S, García N, Godden GT, Krishnakumar V, Jordon-Thaden IE, et al. (2015)

MarkerMiner 1.0: A new application for phylogenetic marker development using angiosperm transcriptomes. Applications in Plant Sciences 3. doi:10.3732/apps.1400115.

S1 Table. List of the primers designed and used in this study.

Sequences are written in 5'-3' direction. Variability of the region is based on substitution per site.

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	CGCTTTGCAGAGATATAAGGGTG	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	CGGTTTGATAACCTGCCA	ndhA_115487R	CAATCCAGAGTATGTTCC	0.013461538	ndhA exon - ndhA intron

		CT		TATCCA		
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	TTTGGGAATTGGTTGGAATGT	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_C	ATGCGAGTATTCCTGCCA	At5g11480_BAR_C	CCTTGGTTCGTGACGCTAC		At5g11480 - At5g11480

	OSII_F	AG	OSII_R	T		
27	At5g12200_BAR_C OSII_F	AGCCCTATGGGATCCTGACT	At5g12200_BAR_C OSII_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	AATGAACCTGGTGCTTTGATG	At5g14520_BAR_C OSII_R	CAGGGATGATGAACACTAACGA		At5g14520 - At5g14520
29	At5g17990_BAR_C OSII_F	TGAGGGACTCGATGAGATGA	At5g17990_BAR_C OSII_R	CGTTTAAGCACTTCAGCATTGT		At5g17990 - At5g17990
30	At5g23120_BAR_C OSII_F	TGCTGTGCAGGAGACTGTTT	At5g23120_BAR_C OSII_R	CGGCTTGAGACAGCAACATA		At5g23120 - At5g23120
31	At5g23240_BAR_C OSII_F	TTGAAATGTGCCTTGTTTGC	At5g23240_BAR_C OSII_R	CCCGTGGTTGCTTAGACATT		At5g23240 - At5g23240
32	At5g41760_BAR_C OSII_F	TCACAAAGCAACGGTGGATA	At5g41760_BAR_C OSII_R	TCGACGGGATCGGATATAAT		At5g41760 - At5g41760
33	At5g42740_BAR_C OSII_F	AGGTGAAGTGGTCAGCAACC	At5g42740_BAR_C OSII_R	CCTTCAAGCGGTTCTCCTTT		At5g42740 - At5g42740
34	At5g47390_BAR_C OSII_F	GGACGTGTCCGAATCGAG	At5g47390_BAR_C OSII_R	TCATCTGCCACAATATCAACA		At5g47390 - At5g47390
35	ETS-2-F	GCACATGGTGTTGTTTGGTT	ETS-2-R	AATGAGCCATTCGCAGTTTC		External Transcribed Spacer (ETS) - External Transcribed Spacer (ETS)
36	ITS-2-F	AATGGTCCGGTGAAGTGTTC	ITS-2-R	GTTCGCTCGCCGTTACTAAG		Internal Transcribed Spacer (ITS) - Internal Transcribed Spacer (ITS)
37	Photo2-Exon14-F	CATGGCAAACAGATGACCAC	Photo2-Exon12-R	TTCTCAACAGGTATTCCCTAAACC		Phototropin 2 - Phototropin 2
38	AT1G68930_Bar_PPR_F	GTGTGGGATGGTGGAAGACT	AT1G68930_Bar_PPR_R	ATCGAGAATTGGCTTGATCC		AT1G68930 - AT1G68930
39	AT3G18020_Bar_PPR_F	TTCCACCGAGCTCCTAAATG	AT3G18020_Bar_PPR_R	TTCTCCATGTCACAGCATCC		AT3G18020 - AT3G18020
40	AT5G18475_Bar_PPR_F	CCCAATACGTGCATCTTCAA	AT5G18475_Bar_PPR_R	CTCGTTCGACGTCCCAAA		AT5G18475 - AT5G18475
41	AT1G20300_Bar_PPR_F	GGTGGTGTCGTGCGAATAAT	AT1G20300_Bar_PPR_R	TTTCCAAAGCTTGATCACCA		AT1G20300 - AT1G20300
42	AT1G53330-2_Bar_PPR_F	AATGCCATGAAGGTCTTGGA	AT1G53330-2_Bar_PPR_R	CTTCCTCAACCTTCCCTTCC		AT1G53330 - AT1G53330
43	AT2G22410_Bar_PPR_F	GTCGTGTGAGGGACTTGGTT	AT2G22410_Bar_PPR_R	GGCACTAGACCAGCTTCGAC		AT2G22410 - AT2G22410
44	At1g54390_BAR_C OSII_F	AAACGTCTTGATGAGGATCTGAA	At1g54390_BAR_C OSII_R	ATTCACCTCCTTGGCACTGT		At1g54390 - At1g54390
45	At5g13420_BAR_C OSII_F	GGAAATGGAAGCATTGTCAAG	At5g13420_BAR_C OSII_R	AAGTTTGCATGCATCTTGGA		At3g56460 - At3g56460
46	At5g13710_BAR_C OSII_F	TTTGCTGCATACGAGTGGTG	At5g13710_BAR_C OSII_R	CAACAAGTCCTTCTGCAGCTT		At5g13420 - At5g13420
47	At5g38530_BAR_C OSII_F	CCTGTTATTAGAGCGGTTGAGC	At5g38530_BAR_C OSII_R	TCCACACACTGCCATTAGGA		At5g13710 - At5g13710
48	At3g56460_BAR_C OSII_F	GATTTGGTTGCAGCTGGTG	At3g56460_BAR_C OSII_R	GCAATGTTTGCTGGGATTACA		At5g38530 - At5g38530

S2 Table. Sequences for conserved sequences 1 and 2, barcodes, and sequencing adapters.

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, in order to dramatically increases the number of samples that can be multiplexed in one sequencing run. The sequencing adapters are the Illumina standard adapters.

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	CAGAGAG G	
	P7	GCTACGCT	
	P7	CGAGGCT G	
	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		

S3 Table. Samples used in this study.

Herbarium codes follow the Index Herbariorum.

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

Species	DNA Accession	Collector and Voucher No.	Herbarium	Country	State	Latitude, Longitude
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
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

Species	DNA Accession	Collector and Voucher No.	Herbarium	Country	State	Latitude, Longitude
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
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

Species	DNA Accession	Collector and Voucher No.	Herbarium	Country	State	Latitude, Longitude
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
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

Species	DNA Accession	Collector and Voucher No.	Herbarium	Country	State	Latitude, Longitude
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
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	Collector and Voucher No.	Herbariu	Country	State	Latitude,

	Accession		m			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			

S4 Table. Allele occurrences found for each species in this study

Tab 1: Summary of the number of alleles in every sample for each species, and a percentage of how many of these were recovered with more than two alleles. An estimated ploidy level is also given. A reference table of *Bartsia* species and closely related taxa with published chromosome counts from Molau (1990) is also included.

Tab 2: Summary of the number of alleles found in each locus for every sample, and a percentage of how many loci had this specific number of alleles.

Tab 3: Raw data of the number of alleles found in each locus for every sample.