Evolution of Aquatic Plants in Rivers and Wetlands across the Andes in northern South America

Ana María Bedoya Ovalle

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Richard G. Olmstead, Chair

Adam D. Leaché

Caroline A. E. Strömberg

Program Authorized to Offer Degree:
Biology

University of Washington

Abstract

Evolution of Aquatic Plants in Rivers and Wetlands across the Andes in northern South America

Ana María Bedoya Ovalle

Chair of the Supervisory Committee:
Professor Richard G. Olmstead
Biology

Northern South America is one of the most biodiverse regions on earth and understanding the causes and processes that have led to the assembly of this rich biota has been central to ecological and evolutionary investigations since Humboldt and Wallace. Given that northern South America is a geologically dynamic region consisting of a complex and heterogeneous landscape matrix, landscape change is viewed as a primary factor promoting species diversification. The Andean Cordillera is deemed responsible for triggering explosive radiations and increasing diversification rates in Neotropical taxa. All botanical investigations that have aimed to understand the role of landscape change on the evolution of the Neotropical flora have been restricted to terrestrial plants. However, fossil and stratigraphic data show that the uplift of the Andes not only resulted in the development of high-elevation tropical ecosystems, but also in changes in drainage basin reconfiguration through time. Aquatic plants are most diverse in the Neotropics. How did Andean uplift and drainage basin reconfiguration shaped the evolution of aquatic plants in rivers and wetlands across the Andes in northern South America?

In this dissertation I aim to address this question and provide a novel perspective on the impact of landscape change on the evolution of the Neotropical flora, by using empirical data obtained from samples collected in the field from which I generate comprehensive genomic datasets and conduct phylogenetic, populations genetics, and biogeographical analyses. To

investigate the impact of Andean uplift and drainage basin reconfiguration on river plant evolution, I use *Marathrum foeniculaceum* and *Marathrum utile* (Podostemaceae) as model systems. These are the only species living strictly in fast-flowing rivers across the Andes in northern South America. To explore how landscape change in northern South America shaped the evolution of plants in standing-water ecosystems such as wetlands, I study selected species of *Ludwigia* (Onagraceae). *Ludwigia* include amphibious and strictly aquatic plants restricted to standing-water ecosystems like flooded savannas, lakes and ponds.

In Chapter 1, I generate plastome data for five species of Podostemaceae and analyze the structure of the chloroplast genomes in a comparative framework within the order Malpighiales. This chapter is an introduction to learning about the evolution of Podostemaceae. The results of this chapter show that plastid genomes in the family are among the smallest in the Malpighiales due to variation in length of the inverted repeats, gene loss, and intergenic region variation. The uncommon loss of a number of genes is reported and we suggest fast rates of evolution in the plastomes of Podostemaceae.

In Chapter 2, I investigate how drainage basin formation linked to Andean uplift shaped the evolution of plants in fast-flowing aquatic ecosystems such as river-rapids and waterfalls. Specifically, I aim to 1) test if the geographical separation of drainage basins interrupted by the Andes and other topological units in northern South America limits gene flow and structures populations according to river drainages, 2) determine if the timing of divergence events in populations of the two species in northern South America correspond with the currently proposed timing of Andean uplift, and 3) provide a hypothesis for the pattern and timing of drainage basin separation. I conduct phylogenetic, population genetics, phylogenetic networks, and divergence dating analyses of populations of *M. foeniculaceum* and *M. utile* located in river drainages across the Andes. Andean uplift isolated populations of *Marathrum* limiting gene flow. At the same time, drainage basin reconfiguration brought species in sympatry where hybridization occurs. I propose that the pattern of divergence of populations reflects the formation of river drainages, which was not complete until <4.1 Ma.

In Chapter 3, I investigate the impact of Andean uplift on the evolution of plants living in standing-water ecosystems across drainage basins interrupted by this mountain system in South America using the genus *Ludwigia* as the study model. I 1) test if divergence times in taxa in northern and central South America took place prior, during, or after current estimates of major recent pulses of Andean uplift Andean uplift and 2) provide a framework to test further hypotheses for the role of landscape change shaping the evolution of plants living in standing-water ecosystems. I conduct divergence dating, diversification, and biogeographic analyses of *Ludwigia* species from South America using existing sanger data. Andean uplift did not cause an increase in the rates of accumulation of *Ludwigia* species through time. I provide hypotheses for why diversification in *Ludwigia* remained constant, as well as the probable context in which *Ludwigia* evolved in South America.

Together, these three projects advance our understanding of aquatic plants living in fast-flowing and standing-water aquatic ecosystems in the Neotropics. Furthermore, in this dissertation I present evidence to fill a gap in our current knowledge base, by complementing the extensive research findings on the role of Andean uplift in shaping the evolution of plants in terrestrial ecosystems, with investigations on aquatic plants.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Over the course of time that it took me to finish this dissertation, there was not a single day that I did not feel endless love and support from my family: Lucero, Fernando, and Maria Camila. "Common sense, rational thinking, and logic shall be the pillars of your life", my father constantly said to me when I was a child. I would not have pursued science if not for those words shaping my existence. *Mamá, papá y Mila, los amo sin fronteras de tiempo ni espacio.*

To my life partner and forever love, Bob, without whom I would not have had the life quality I have had while conducting my research projects. Thank you for supporting my career and for your sacrifices and constant care. Thank you for reading my manuscripts and applications, and providing ideas to improve my figures.

This dissertation is based on specimens collected in the field. Therefore, I am in debt with those who assisted me in my collecting trips: Fernando Bedoya, Maria Paula Contreras, "Moño" Durango, Toribio Durango, Juliana Vélez, Nicolás Rodríguez, Mateo Fernández, David Ocampo, Daniel Ocampo, Jules Dominé, and Adrián Pinzón. I am also grateful to Viviana Londoño for facilitating a sample of *M. foeniculaceum* from Boyacá.

To the friends I made while living in Seattle (especially Amber Hageman, Jennifer Hsiao, Luke Weaver, Will Brightly, Claire Rusch, Diego Alonso, Ethan Linck, and Molly Phillips) thank you for seeing me and for giving me the space to express my full self beyond language and cultural differences. My friends Camila Martínez, Elena Stiles, and Fénix Garcia have been instrumental before and after the pandemic. I am grateful to life for being surrounded by such smart women conducting fascinating research. Our weekly discussions about life in general, navigating obstacles in research, and life as women kept me going despite difficulties. I am looking forward to seeing how our relationship and scientific careers unfold.

I thank my dissertation advisor Richard G. Olmstead for his support and for trusting me

and giving me free rein to explore the research ideas that interested me the most. I thank him for helping me navigate the social aspect of academia as well, and for his constant words of encouragement. I am forever grateful to Adam Leaché who was always incredibly generous with his time and resources. I thank Caroline Strömberg for her comments on my manuscripts and ideas and for serving as a role model for me. To the rest of my dissertation Committee, Lorenz Hauser and John Klicka, thank you for your feedback, time, and enthusiasm. I would also like to thank Santiago Madriñán for his continued support in Colombia since my undergrad. I thank Brad Ruhfel and Thomas Philbrick for our collaborative work on Podostemaceae. An important part of this research builds upon the work of Tom on the family Podostemaceae and I am grateful for his enormous willingness to share his knowledge in this group. I thank David Giblin for his help at the UW herbarium any time I needed it. I was very touched by the artistic talent of Crystal Soojeong Shin when I saw how amazing my voucher specimens looked.

I am very thankful to the Biology department at the University of Washington for support to my research through departmental fellowships and for providing safe conditions for conducting research and for being a foreign student. Specially, I thank Krista Clouser, Dave Hurley, Ron Killman, and Jeannette Takashima. Finally, I would like to thank the following funding sources: The graduate school at the UW for support through the Boeing fellowship. The Biology department at the UW provided financial support to my research through the Giles, WRF Hall, and Denton Endowments. The American Society of Plant Taxonomists supported me through the Graduate student Fellowship and awarded me with the George R. Cooley award for best paper in plant systematics in 2020. The Botanical Society of America and the Explorers Club provided financial support through research grants.

DEDICATION

To Colombia, the country of the three Cordilleras and the territory that has witnessed the processes that I aimed to investigate here. To the Colombian people, who have suffered most but find strength to keep going. In memory of Agnes Arber (1879–1960) for her pioneer work on aquatic plants and for helping pave the way for women in the sciences.

Chapter 1

PLASTID GENOMES OF FIVE SPECIES OF RIVERWEEDS (PODOSTEMACEAE): STRUCTURAL ORGANIZATION AND COMPARATIVE ANALYSIS IN MALPIGHIALES

frontiers in Plant Science

ORIGINAL RESEARCH
published: 20 August 2019
doi: 10.3389/fpls.2019.01035

Plastid Genomes of Five Species of Riverweeds (Podostemaceae): Structural Organization and Comparative Analysis in Malpighiales

Ana M. Bedoya[1], Bradley R. Ruhfel[2], C. Thomas Philbrick[3], Santiago Madriñán[4], Claudia P. Bove[5], Attila Mesterházy[6] and Richard G. Olmstead[1]*

[1] Department of Biology and Burke Museum, University of Washington, Seattle, WA, United States, [2] University of Michigan Herbarium, University of Michigan, Ann Arbor, MI, United States, [3] Department of Biological and Environmental Sciences, Western Connecticut State University, Danbury, CT, United States, [4] Laboratorio de Botánica y Sistemática, Departamento de Ciencias Biológicas, Universidad de los Andes, Bogotá, Colombia, [5] Departamento de Botânica, Museu Nacional, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, [6] Directorate of Hortobágy National Park, Debrecen, Hungary

OPEN ACCESS

Edited by:
Carl J. Rothfels, University of California, Berkley, United States

Reviewed by:
Lachezar A. Nikolov, University of California, Los Angeles, United States
Angela Jean McDonnell, Chicago Botanic Garden, United States

***Correspondence:**
Ana M. Bedoya
ambedoya@uw.edu

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *04 June 2019*
Accepted: *24 July 2019*
Published: *20 August 2019*

Citation:
Bedoya AM, Ruhfel BR, Philbrick CT, Madriñán S, Bove CP, Mesterházy A and Olmstead RG (2019) Plastid Genomes of Five Species of Riverweeds (Podostemaceae): Structural Organization and Comparative Analysis in Malpighiales. Front. Plant Sci. 10:1035. doi: 10.3389/fpls.2019.01035

With the advent of next-generation sequencing technologies, whole-plastome data can be obtained as a byproduct of low-coverage sequencing of the plant genomic DNA. This provides an opportunity to study plastid evolution across groups, as well as testing phylogenetic relationships among taxa. Within the order Malpighiales (~16,000 spp.), the Podostemaceae (~300 spp.) stand out for their unique habit, living attached to rocks in fast-flowing aquatic habitats, and displaying highly modified morphologies that confound our understanding of their classification, biology, and evolution. In this study, we used genome skimming data to assemble the full plastid genome of 5 species within Podostemaceae. We analyzed our data in a comparative framework within Malpighiales to determine the structure, gene content, and rearrangements in the plastomes of the family. The Podostemaceae have one of the smallest plastid genomes reported so far for the Malpighiales, possibly due to variation in length of inverted repeat (IR) regions, gene loss, and intergenic region variation. We also detected a major inversion in the large single-copy region unique to the family. The uncommon loss or pseudogenization of *ycf1* and *ycf2* in angiosperms and in land plants in general is also found to be characteristic of Podostemaceae, but the compensatory mechanisms and implications of this and of the pseudogenization of *accD* and *rpl23* and loss of *rps16* remain to be explained in this group. In addition, we estimated a phylogenetic tree among selected species in Malpighiales. Our findings indicate that the Podostemaceae are a distinct lineage with long branches that suggest faster rates of evolution in the plastome of the group, compared with other taxa in the order. This study lays the foundations for future phylogenomic studies in the family.

Keywords: genome rearrangements, Malpighiales, phylogenomics, plastome, Podostemaceae

INTRODUCTION

The plastids have a relatively small, maternally inherited, haploid genome (Sugiura, 1992). It ranges between 120 and 170 kb in length and is generally composed of a circular structure with two IRs that are mirror images in terms of gene content (IRa and IRb), separated from each other by a large and a small single-copy regions (LSC and SSC, respectively) (Downie and Palmer, 1992; Sugiura, 1992). Because the plastome encodes genes that are essential for fundamental processes such as photosynthesis and its own replication, it has been generally understood that its genome shows a relatively high degree of conservation in size, structure, and gene content within land plants (Palmer, 1985; Wicke et al., 2011). However, structural rearrangements, gene losses, and expansions and contractions in IRs are widely documented across species (Goulding et al., 1996; Krause, 2011; Weng et al., 2014; Schwarz et al., 2015; Xu et al., 2015; Rabah et al., 2019; Shrestha et al., 2019). Such rearrangements have been relevant in a systematic framework when supporting the monophyly of certain groups (Jansen and Palmer, 1987; Downie and Palmer, 1992; Hoot and Palmer, 1994; Cosner et al., 2004).

With the advent of next-generation sequencing technologies, information from whole-genome data is quickly available at a low cost (Metzker, 2009). Given that plastomes exist in high copy numbers in plant cells, even a genome skimming approach where the nuclear genome is sequenced at low-coverage provides a mechanism to obtain a fully assembled plastome as a byproduct (Straub et al., 2012; Olmstead and Bedoya, 2019). Over the past few years, this has provided the advantage of rapidly generating whole-plastid sequences for a large number of taxa (Daniell et al., 2016). This information has been used to disentangle phylogenetic relationships and to study plastid evolution in selected groups of plants (Ruhfel et al., 2014; Cauz-Santos et al., 2017; Firetti et al., 2017; Gitzendanner et al., 2018; Li and Zheng, 2018; Liu et al., 2018; Li et al., 2019; Lloyd Evans et al., 2019).

Malpighiales is a large order with 36 families, more than 700 genera, and ~16,000 species (Wurdack and Davis, 2009; The Angiosperm Phylogeny Group, 2016). Full plastid assemblies for 111 species in the families Chrysobalanaceae, Clusiaceae, Erythroxylaceae, Euphorbiaceae, Linaceae, Malpighiacee, Passifloraceae, Salicaceae, and Violaceae currently reside in the NCBI database. In addition, previous studies using whole-plastome data of *Passiflora edulis* Sims (Cauz-Santos et al., 2017) and of *Byrsonima crassifolia* (L.) Kunth and *Byrsonima coccolobifolia* Kunth (Menezes et al., 2018) have provided insights into plastome evolution in the order Malpighiales, reporting rearrangements that are unique to Passifloraceae (Rabah et al., 2019; Shrestha et al., 2019), identifying regions of high sequence divergence, and helping resolve the phylogeny of the group.

Within the morphologically and ecologically diverse group Malpighiales, the family Podostemaceae stands out for its unusual habit (Xi et al., 2012). Riverweeds (as members of this family are also called) are notable for living attached to rocks in fast-flowing water habitats such as river rapids and waterfalls, with flowers that project above the water surface and fruits that develop and shed seeds only in the dry season when the water level is low (van Royen, 1951; Philbrick and Novelo, 1995; Rutishauser, 1995; Rutishauser, 1997; Philbrick and Novelo, 1998). Much remains to be explored in Podostemaceae despite a number of morphological (van Royen, 1951; Novelo and Philbrick, 1997; Rutishauser et al., 1999; Jäger-Zürn, 2011), developmental (Rutishauser, 1995; Rutishauser, 1997; Jäger-Zürn, 2005, Jäger-Zürn, 2007), and karyological (Oropeza et al., 1998; Oropeza et al., 2002) studies followed by phylogenetic and biogeographical investigations (Kita and Kato, 2001; Ruhfel et al., 2011; Tippery et al., 2011; Koi et al., 2012; Ruhfel et al., 2016).

The extreme conditions experienced by the Podostemaceae have resulted in highly modified vegetative and reproductive morphologies (Eckardt and Baum, 2010). Such forms constitute a taxonomical challenge because the high degree of modification of vegetative and reproductive structures results in a small number of morphological traits that are informative, making the study of the biology and evolution of this group difficult. Given this scenario, genomic data surface as the tool to gain better insight into the evolution of this notable group of plants.

In this study, we present the fully annotated plastid genomes of 5 species of Podostemaceae: *Apinagia riedelii* Tul., *Marathrum capillaceum* (Pulle) P. Royen, *Marathrum utile* Tul., *Monostylis capillacea* Tul., and *Tristicha trifaria* (Bory ex Willd.) Spreng. We analyzed our data in a comparative framework within Malpighiales to detect rearrangements and structural characteristics of the plastome of this distinctive family, taking advantage of the data already available in the order. A phylogenetic tree was inferred with whole-plastid data to test relationships and examine sequence divergence and amount of change within the family and order. Our investigation constitutes the first report of a complete nucleotide sequence and structure of the plastid genome in the Podostemaceae.

MATERIALS AND METHODS

Taxon Sampling, DNA Extraction, and Sequencing

Samples of *A. riedelii*, *M. capillaceum*, *M. utile*, *M. capillacea*, and *T. trifaria* were collected in South America and Africa. Information on collection localities and voucher specimens is shown in **Table 1**. Together, these samples represent 2 of 3 subfamilies within Podostemaceae (Podostemoideae and Tristichoideae). Subfamily Tristichoideae is sister to a clade comprising the Podostemoideae and the monotypic Weddellinoideae (Kita and Kato, 2001). Therefore, any patterns shared between Tristichoideae and Podostemoideae would most likely be synapomorphies of the Podostemaceae. All species included have a distribution restricted to the Neotropics except for the pantropical *T. trifaria*.

Total genomic DNA was extracted from silica-dried leaf tissue using a modified CTAB protocol and purified by isopropanol precipitation, or *via* silica columns (Epoch Life Science, Missouri City, TX, USA) from the aqueous supernatant after chloroform/isoamyl alcohol purification (Neubig et al., 2014). DNA was run on a 1% agarose gel to assess DNA quality, and concentration was measured with a Qubit fluorometer using the dsDNA BR Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA). A volume of 90 μL of total DNA of *M. utile* was used to prepare a library with

TABLE 1 | Provenance, voucher information, and/or GenBank accession numbers of the species in Malpighiales whose plastomes were included in this study.

Species	Family	Voucher (Herbarium)	GenBank accession no.	Collection locality
Hirtella racemosa	Chrysobalanaceae	—	NC_024060	
Garcinia mangostana	Clusiaceae		NC_036341	
Byrsonima crassifolia	Euphorbiaceae	—	NC_037192	
Passiflora edulis	Passifloraceae	—	NC_034285	
*Apinagia riedelii**	Podostemaceae	C.P. Bove 2513 (R)	MN165812	Brazil, South America
*Marathrum utile**	Podostemaceae	AMB 497 (ANDES)	MN165814	Colombia, South America
*Marathrum capillaceum**	Podostemaceae	C.P. Bove 2493 (R)	MN165813	Brazil, South America
*Monostylis capillacea**	Podostemaceae	C.P. Bove 2524 (R)	MN165815	Brazil, South America
*Tristicha trifaria**	Podostemaceae	A. Mesterhazy MLI 128(Z)	MN165816	Mali, Africa
Salix purpurea	Salicaceae	—	NC_026722	
Viola seoulensis	Violaceae	—	NC_026986	

Voucher number and collection locality are provided only for those species whose genome was generated in this study (*).

an average fragment size of 500 bp, using the Kapa Biosystems Hyper prep kit at the QB3 Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley. Whole-genome shotgun sequencing was also performed at the QB3 Sequencing Laboratory, with 150 bp paired-end reads on 1 lane of an Illumina HiSeq4000. For the remaining species, a volume of 50 µL of 50 ng/µL total DNA was used to prepare libraries with average fragment size of 500 bp by Rapid Genomics LLC (Gainesville, FL, USA). Whole-genome sequencing of 150 bp paired-end reads was performed at the same facility by multiplexing samples in 1 lane of an Illumina HiSeqX.

Plastome Assembly and Annotations

Read quality of paired-end Illumina reads was assessed in FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and adapter sequences were removed using Trimmomatic (Bolger et al., 2014). The pipeline GetOrganelle (Jin et al., 2018) was used to select trimmed reads that corresponded to the plastid using the plastome of *Garcinia mangostana* L. (Clusiaceae) as a reference. The pipeline was also used to assemble the filtered reads. The annotations of the plastomes of *G. mangostana*, *Manihot esculenta* Crantz, and *Salix purpurea* L. (see **Table 1** for GenBank accession numbers) were transferred to the final circular plastid consensus sequences of *A. riedelii*, *M. utile*, *M. capillaceum*, *M. capillacea*, and *T. trifaria* with the tool "Annotate from source" in Geneious 9.1.8. (Biomatters Ltd., Auckland, New Zealand). Annotations were manually inspected, and tRNAs were further checked with tRNAscan-SE v2.0 as implemented in GeSeq (Tillich et al., 2017). GC content and boundaries between the IRa IRb, LSC, and SSC regions were determined in Geneious. The diagrams for the circular genomes were obtained with the program OGDRAW (Greiner et al., 2019).

In addition, a second approach to plastome assembly was conducted for *M. utile* to confirm the output of GetOrganelle. In this second assembly method, plastid filtered reads from GetOrganelle were imported in Geneious 9.1.8. The BBDuk tool was used to trim low-quality bases (Q20) and discard short reads (<10 bp). Reads were further normalized and error corrected using the tool BBNorm with target coverage level 30. A total of 225,896 filtered reads were assembled *de novo* using the Medium sensitivity/Fast option in the Geneious Prime *de novo* assembler. The options "Don't merge variants" and "Produce scaffolds" were left unchecked.

In order to obtain a draft circular plastome, the consensus sequence of the largest contig (112,008 bp with 41.9X mean coverage) was generated. The Geneious Prime plugin "Find Repeats" was used in order to find the IRs. The *de novo* assembly of short reads in Geneious does not allow a full assembly of both IRs. Instead, it generates a consensus sequence with 1 full IR and the truncated ends of the second IR. For this reason, the latter were trimmed, and the single instance of the full IR was extracted. This extracted IR was reversed complement and concatenated with the previously trimmed consensus sequence of the largest contig. The generated draft genome was used as a reference to map the trimmed paired reads without normalization. This map-to-reference assembly was used for single nucleotide polimorphism (SNP) variant calling and to generate a final full circular plastid consensus sequence.

Plastome of Podostemaceae in a Comparative Framework

To detect differences in the plastomes of the selected species of Podostemaceae with respect to other Malpighiales, we compared the assembled plastid genomes with six species representing six plant families in the order Malpighiales. The families included for comparison represent all the three major clades in Malpighiales (Xi et al., 2012). Accession numbers for the species included in this comparative analysis are listed in **Table 1**. Visual inspection of rearrangements was performed using progressive Mauve v.2.4.0 with default "seed families" and default values for all other parameters (Darling, 2004). As Mauve cannot handle duplicated regions, one of the IRs of each genome was manually removed following Firetti et al. (2017). The boundaries between the IRa IRb, LSC, and SSC regions in all species were inspected in Geneious using the fully assembled plastids.

We used the software mVista in Shuffle-LAGAN mode to explore variation in gene content within Malpighiales. *Garcinia mangostana* was used as reference in order to detect possible gene losses, gene variation, or gene conservation in Podostemaceae. Genes with <50% similarity were inspected directly in the annotated genomes of Podostemaceae to determine if they were intact, open reading frames. In a separate analysis, *A. riedelii* was used as reference to determine the level of similarity across the whole-plastome sequence in Malpighiales with respect to Podostemaceae.

In order to test relationships and examine sequence divergence and amount of change within both Malpighiales and Podostemaceae, a phylogenetic tree was inferred using the plastid genomes of all studied species. *Averrhoa carambola* L. (Oxalidaceae) was used as an outgroup to root the tree. To generate the alignment, in each species the IRb regions were deleted to remove duplicated genes; protein-coding regions, tRNAs, rRNAs, and noncoding regions were extracted, and all genes located on the reverse strand were reversed complemented. The extracted regions were aligned with MAFFT v7.309 in Geneious and then concatenated. The final alignment was 134,969 bp long. The software PartitionFinder2 (Lanfear et al., 2016) was used to select the best partitioning scheme, using a greedy search (Lanfear et al., 2012) in RAxML (Stamatakis, 2014). In the analysis, the three codon positions for each protein-coding region and each tRNA and rRNA were considered separately. Noncoding regions were analyzed together. Maximum likelihood phylogenetic inference was performed using RAxML v8.2 (Stamatakis, 2014), with the "rapid bootstrap analysis and search for best-scoring ML tree option" and 10,000 bootstrap replicates. Per-partition branch lengths were estimated independently.

RESULTS

Genome Content and Structure in Podostemaceae

After sequencing, trimming, and selecting reads corresponding only to the plastids in GetOrganelle, 1,581,656 paired reads were recovered for *A. riedelii*, 1,443,458 for *M. utile*, 225,344 for *M. capillaceum*, 1,087,996 for *M. capillacea*, and 313,332 for *T. trifaria*. The largest plastome was that of *A. riedelii* with a length of 134,912 bp (1177.6X coverage), followed by *M. capillaceum* with 134,374 bp (190.8X coverage), *M. capillacea* with 133,944 bp (736.3X coverage), *M. utile* with 131,951 bp (1264.2X coverage), and *T. trifaria* with 130,285 bp (217.6X coverage). Assembly of the plastome of *M. utile* using Geneious 9.1.8 yielded the same sequence as with GetOrganelle, but mean coverage was lower (514.9X vs. 1264.2X).

All 5 full plastome assemblies in Podostemaceae showed the typical quadripartite structure characteristic of the plastids (see **Figure 1**). GC content in the IRs is higher than in other regions of the plastid, possibly due to the presence of tRNA genes, as suggested in Dipsacales (Fan et al., 2018). In the 5 species, the 2 IRs span 29.7% to 31.4% of the plastome (**Table 2**).

Gene content was the same across the Podostemaceae species studied, with each genome including 71 protein coding genes, 30 tRNAs, and 4 rRNAs for a total of 105 genes, 13 of which contain 1 intron and 1 (*trnK-UUU*), which contains 2 introns. Of the total number of genes, 77 (~73.33) occur in the LSC, 10 (~9.52%) in the SSC, and 18 (~17.14%) in the IRs. With regard to protein coding genes, 55 (~77.46%) are included in the LSC, 9 (~12.68%) in the SSC, and 7 (~9.86%) in the IRs. Most tRNAs exist in the LSC region with 28 (~73.33%) tRNAs, followed by 7 (~23.33%) in the IRs, and only 1 (~3.33%) in the SSC region. All rRNAs were found in the IRs. A full account of gene content for the Podostemaceae species is listed in **Table 3**.

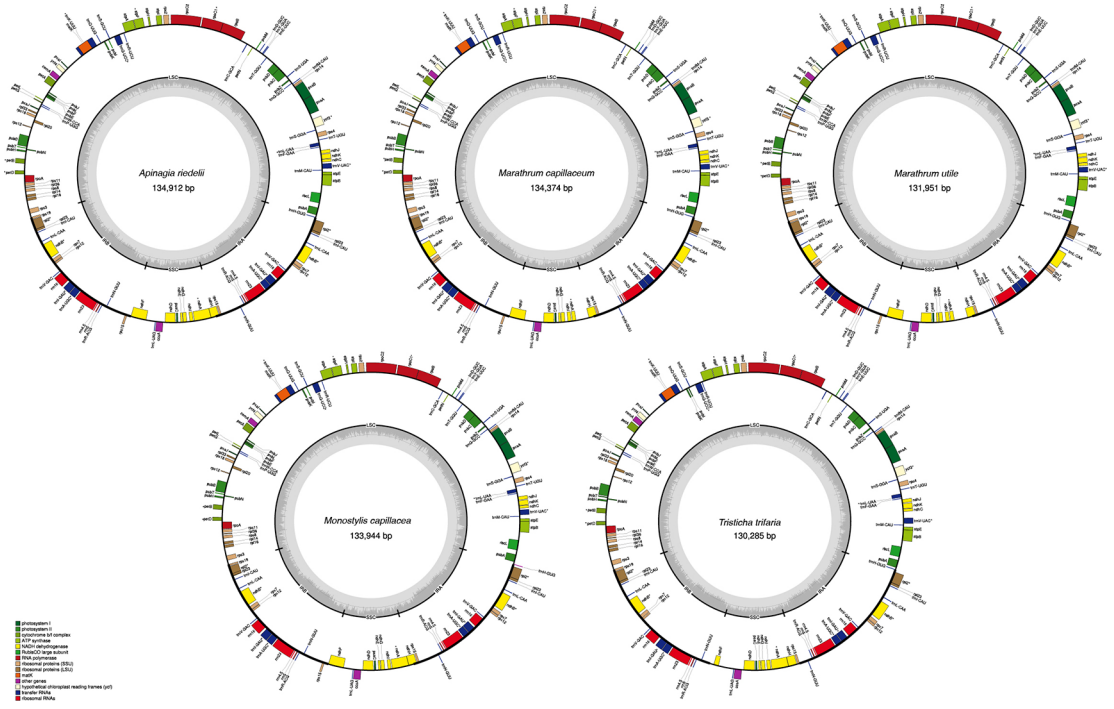


FIGURE 1 | Plastid genomes of the 5 species of Podostemaceae included in this study. Only functional genes are drawn, and GC content graphs are included as dark gray bars toward the center of each diagram. Intron-containing genes are marked with (*).

TABLE 2 | Structural information of the plastid genomes of Podostemaceae, Clusiaceae, Malpighiaceae, Chrysobalanaceae, Violaceae, Passifloraceae, and Salicaceae. The percentages of the total size of the genome that corresponds to each region are included.

Species	Family	Plastome genome size (bp)	IRs length (bp)	SSC length (bp)	LSC length (bp)
Apinagia riedelii	Podostemaceae	134,912	21,049 × 2 (~30.1%)	12,437 (~8.9%)	85,377 (~61%)
Monostylis capillacea	Podostemaceae	133,944	21,026 × 2 (~31.4)	12,395 (~9.3%)	79,497 (~59.4%)
Marathrum utile	Podostemaceae	131,951	19,945 × 2 (~30.2%)	12,283 (~9.3%)	79,778 (~60.5%)
Marathrum capillaceum	Podostemaceae	134,374	21,041 × 2 (~31.3)	12,302 (~9.2%)	79,990 (~59.5%)
Tristicha trifaria	Podostemaceae	130,285	19,349 × 2 (~29.7)	12,662 (~9.7%)	78,925 (~60.6%)
Garcinia mangostana	Clusiaceae	158,179	27,009 × 2 (~34.1%)	17,704 (~11.2%)	86,457 (~54.7%)
Byrsonima crassifolia	Malpighiaceae	160,212	26,975 × 2 (~33.7%)	17,814 (~11.1%)	88,448 (~55.2%)
Hirtella racemosa	Chrysobalanaceae	162,891	26,866 × 2 (~33%)	19,915 (~12.2%)	89,244 (~54.8%)
Viola seoulensis	Violaceae	156,507	26,404 × 2 (~33.7%)	18,008 (~11.5%)	85,691 (~54.8%)
Passiflora edulis	Passifloraceae	151,406	26,152 × 2 (~34.5%)	13,378 (~8.8%)	85,724 (~56.6%)
Salix purpurea	Salicaceae	155,590	27,459 × 2 (~35.3%)	16,220 (~10.4%)	84,452 (~54.3%)

TABLE 3 | Gene content in all Podostemaceae species included in this study.

Gene function	Gene group	Gene name
Self-replication	Ribosomal RNA genes	***rrn 4.5, rrn5, rrn16, rrn23***
	Transfer RNA genes	***trnA-UGC****, *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU*, *trnG-GCC*, *trnG-UCC**, *trnH-GUG*, ***trnI-CAU***, ***trnI-GAU****, *trnK-UUU**, ***trnL-CAA***, *trnL-UAA**, *trnL-UAG*, *trnM-CAU*, ***trnN-GUU***, *trnP-UGG*, *trnQ-UUG*, ***trnR-ACG***, *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, ***trnV-GAC***, *trnV-UAC**, *trnW-CCA*, *trnY-GUA*
	Small subunit of ribosome	*rps2*, *rps3*, *rps4*, ***rps7***, *rps8*, *rps11*, ***rps12***, *rps14*, ***rps15***, *rps18*, ***rps19***
	Large subunit of ribosome	***rpl2****, *rpl14*, *rpl16*, *rpl20*, *rpl33*, *rpl36*
	RNA polymerase subunits	*rpoA*, *rpoB*, *rpoC1**, *rpoC2*
Photosynthesis	Subunits of NADH dehydrogenase	*ndhA**, ***ndhB****, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
	Subunits of photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*, *ycf3**
	Subunits of photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*
	Subunits of cytochrome b/f complex	*petA*, *petB**, *petD**, *oetG*, *petL*, *petN*
	Subunits of ATP synthase	*atpA*, *atpB*, *atpE*, *atpF**, *atpH*, *atpI*
	Large subunit of Rubisco	*rbcL*
Other	Maturase	*matK*
	Envelope membrane protein	*cemA*
	C-type cytochrome synthesis	*ccsA*
	ORFs	*ycf4*

Genes in bold correspond to genes that are located in the IRs and hence are duplicated. Genes that contain introns are marked with asterisk (*).

Plastome of Podostemaceae in a Comparative Framework Within Malpighiales

Information on plastid genome size and size of the IRa, IRb, LSS, and SSC regions in all species shows that the Podostemaceae possess the smallest genome of the species included in this study (**Table 2**). This reduction is relatively uniform across the IRs, LSS, and SSC, as the proportions of each region in the plastid remain fairly similar in Malpighiales. However, in Podostemaceae, the LSC region did not shrink as much as the SSC and IRs regions, occupying a slightly larger percentage of the plastid in Podostemaceae (**Table 2**). Inspection of the plastomes of Podostemaceae and selected members of the Malpighiales with Mauve shows a large inversion of ~49,000 bp in the LSC region. The inversion is located between the genes *rbcL* and *trnK*. This rearrangement is unique in Podostemaceae with respect to the other Malpighiales species inspected (**Figure 2**). Other rearrangements are seen in *P. edulis* as previously reported (Cauz-Santos et al., 2017; Shrestha et al., 2019).

A comparison of border positions of the four plastid regions in the full organelle sequences across the 11 species studied is shown in **Figure 3**. The LSC/IRb border is located within the *rps19* gene, creating a 220-bp truncated copy (pseudogene) in the IRa in all the Podostemaceae species studied, as well as in *G. mangostana* and *Hirtella racemosa* Lam. In *Viola seoulensis* Nakai, this duplicated fragment is only 68 bp, in line with previous work (Menezes et al., 2018). Variations in the length of the IRb in *B. crassifolia*, *S. purpurea*, and *P. edulis* caused the LSC/IRb border to fall within the *rpl22* gene in the former two species, and between *rpl22* and *rps19* in *P. edulis*. This created a pseudogene in the IRa of both *B. crassifolia* and *S. purpurea*. In Podostemaceae and in *G. mangostana*, the boundaries of *trnH* and the truncated copy of *rps19* overlap by 7 bp in the IRa. In all species except in *P. edulis*, *trnH-GUG* is the first gene in the LSC region. This exception has been proposed to be caused by a small inversion at the beginning of the LSC region containing the *psbA* and *trnH-GUG* genes (Cauz-Santos et al., 2017). The SSC/IRa is located within the *ndhH* gene in *A. riedelii*, *M. capillaceum*, *M. capillacea*, and *M. utile*, creating a pseudogene in the IRb.

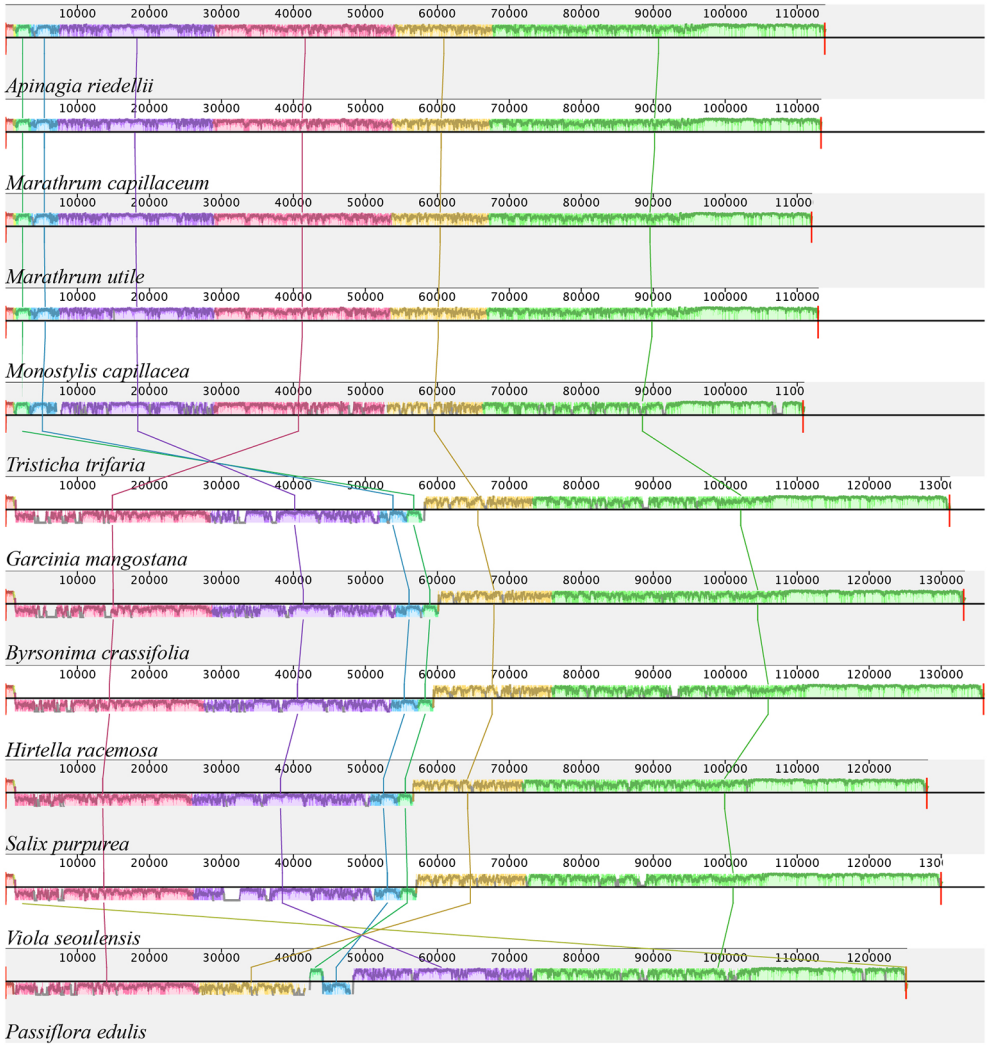


FIGURE 2 | Alignment resulted from Mauve showing a large inversion shared by all Podostemaceae. Color bars indicate syntenic blocks, and connecting lines indicate correspondence of blocks across genomes.

This border is shifted to the *rps15* gene in *T. trifaria* and *P. edulis*, where a small fragment of this gene (< 20 bp long) spans the IRa and is duplicated in the IRb. In the remaining species, the SSC/IRa border falls in the *ycf1* gene, which is located downstream of the *ndhH* and *rps15*. As a consequence, a *ycf1* pseudogene is produced in the IRb. This gene is reduced to a pseudogene in Podostemaceae.

An alignment of 11 species in six families with *G. mangostana* used as reference is shown in **Figure 4**. In this alignment, the large inversion previously identified was reinverted in order to enhance visualization and allow gene content comparison. We found that species in Podostemaceae share the loss of the *rps16* gene with most other Malpighiales, except for *B. crassifolia* (Malpighiaceae), where the gene is present. Similarly, the Podostemaceae are like other Malpighiales in the retention of the *atpF* Group II intron, which is absent only in *P. edulis*. On the contrary, the gene for the subunit of acetyl-Co-A-carboxylase (*accD*) is highly divergenet in the Podostemaceae and not in frame in M. capillacea and in M. capillaceum. The large subunit of ribosome protein (rpl23), and the chloroplast open reading frames *ycf1* and *ycf2* are reduced to pseudogenes only in Podostemaceae and in *P. edulis* (Cauz-Santos et al., 2017) (**Figure 3**).

The analysis performed in mVista using *A. riedelii* as reference is shown in **Figure 5**. *Apinagia riedelii*, *M. capillaceum*, *M. utile*, and *M. capillacea*, all members of the Podostemoideae, show high similarity across their plastome. In fact, the percentage similarity supports that all four species belonging to this subfamily are more similar to each other than any of them are to *Tristicha*, in the subfamily Tristichoideae. As expected, similarity is higher in coding regions than in intergenic sequences.

Phylogenetic analysis was conducted using an optimal scheme with 53 partitions as resulted from PartitionFinder2. Information on partitions and substitution models is included

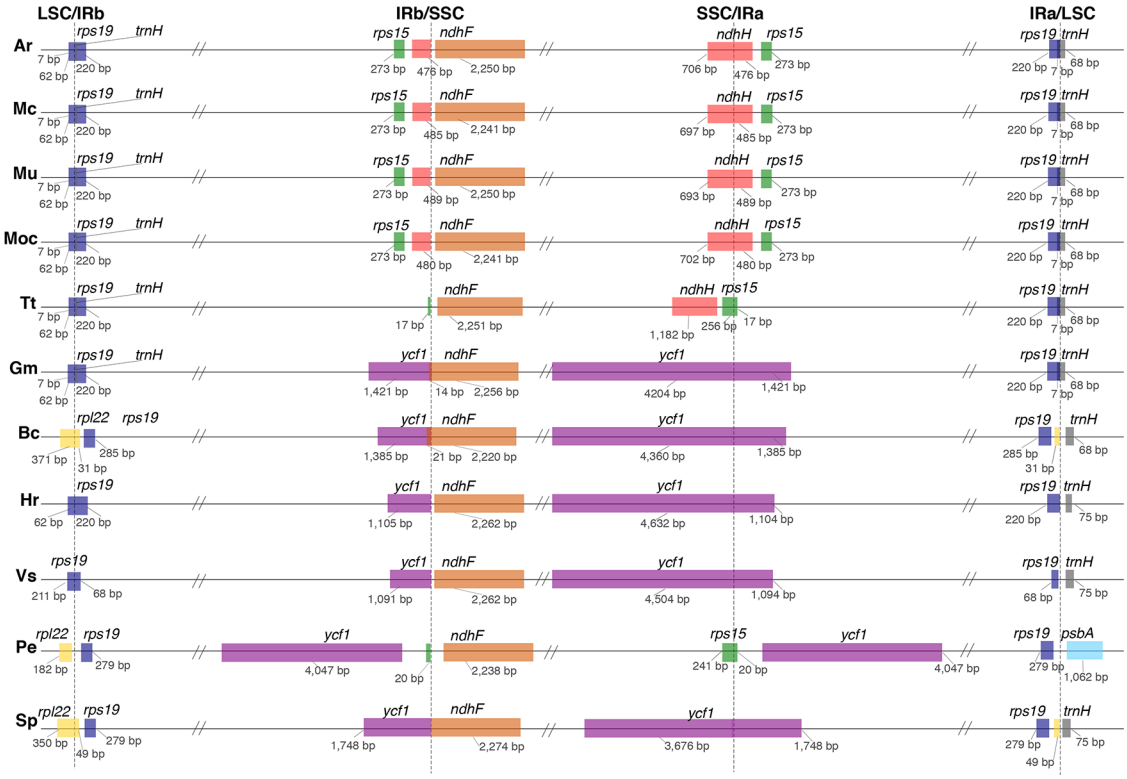


FIGURE 3 | Comparison of border positions of the 4 plastid regions (LSC, IRb, SSC, IRa) among plastomes of Ar, *Apinagia riedelii;* Mc, *Marathrum capillaceum;* Mu, *Marathrum utile*; MoC, *Monostylis capillacea*; Tt, *Tristicha trifaria;* Gm, *Garcinia mangostana;* Bc, *Byrsonima crassifolia;* Hr, *Hirtella racemosa;* Vs, *Viola seoulensis;* Pe, *Passiflora edulis;* Sp, *Salix purpurea*. Functional genes and truncated fragments are shown with the same color. The sizes of fragments in genes that are located in a boundary are shown.

in the **Supplementary Material**. Among the Podostemaceae, the Podostemoideae are supported as monophyletic and sister to *T. trifaria* (**Figure 6**). The phylogeny also shows that the branches leading to taxa in the Podostemaceae from the common ancestor of Malpighiales are much longer than the branches leading to other taxa within the order. *Garcinia mangostana* (Clusiaceae) is supported as sister to Podostemaceae (100% bootstrap), in line with previous work, but this clade was found as sister to *H. racemosa* (Chrysobalanaceae), contrary to previous work (Xi et al., 2012; Menezes et al., 2018) where Chrysobalanaceae is found as more closely related to Malpighiaceae. *Salix purpurea*, *P. edulis*, and *V. seoulensis* are supported as a clade (100% bootstrap), and the relationships among them are in agreement with Xi et al., 2012. However, *B. crassifolia* (Malpighiaceae) is reconstructed as sister to this clade (85% bootstrap), and as mentioned above, this contradicts previous published work (Xi et al., 2012; Menezes et al., 2018).

DISCUSSION

The 130,218- to 134,912-bp size range of the plastome reported in this study for Podostemaceae species falls within the average size of angiosperm plastomes (Sugiura, 1992). However, it is notable that the full plastid genomes generated here for the family are among the smallest reported so far in Malpighiales (Shrestha et al., 2019; https://www.ncbi.nlm.nih.gov/genome).

It has been proposed that plastome size variation could be caused by variation in length of IR regions, gene loss, and intergenic region variation (Palmer et al., 1987; Wolfe et al., 1992; Wakasugi et al., 1994; Chumley et al., 2006; Xiao-Ming et al., 2017). We have reported here that the IRs in the Podostemaceae are ~6 kb smaller than in the other Malpighiales used for comparison (**Table 2**), and we have also reported the loss of *rps16* and reduction to pseudogenes of *accD* (in some species of Podostemaceae), *ycf1,* and *ycf2.* However, the average size of the plastome of Podostemaceae is smaller than the other Malpighiales examined here by 16 to 28 kb, and this difference cannot be explained by a smaller length of the IRs and by gene losses alone. Intergenic region variation as well as intron loss also contribute to this difference in plastome size, considering that the number of introns reported for Podostemaceae is smaller than in *P. edulis* (Cauz-Santos et al., 2017) and that intergenic regions are the most variable in our comparative study (**Figure 4**). Indeed, when calculating the total length of intergenic regions in Podostemaceae and of the other species in Malpighiales analyzed here, the Podostemaceae are shorter by ~5.5 kp on average. This implies that all three processes responsible for genome size variation mentioned above are responsible for the reduction in size of the plastomes in Podostemaceae.

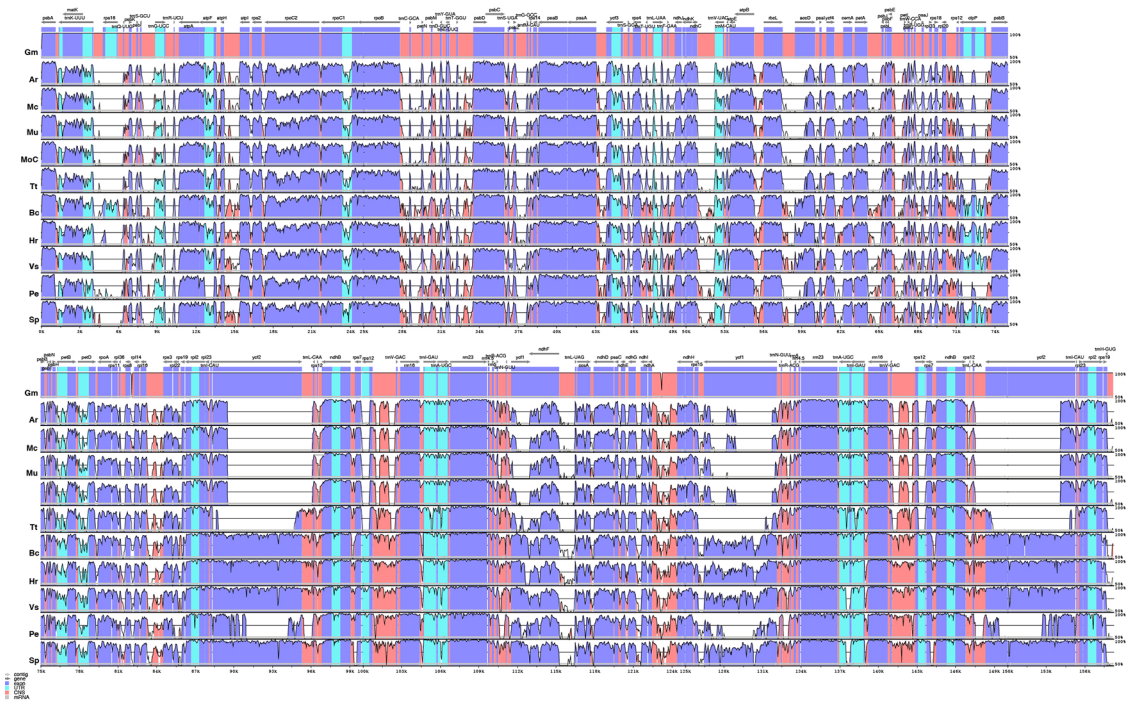


FIGURE 4 | Comparison of percentage identity of plastomes in mVista using *Garcinia mangostana* (Gm) as reference. Ar, *Apinagia riedelii*; Mc, *Marathrum capillaceum*; Mu, *Marathrum utile*; Moc, *Monostylis capillacea*; Tt, *Tristicha trifaria*; Bc, *Byrsonima crassifolia*; Hr, *Hirtella racemosa*; Vs, *Viola seoulensis*; Pe, *Passiflora edulis*; Sp, *Salix purpurea*. The vertical axis corresponds to the percentage identity (50%–100%), while the horizontal axis shows the position of each region within the locus. Arrows indicate the transcription of annotated genes in the reference genome. Genome regions are color coded.

The large rearrangement in the LSC region appears to be a synapomorphy of Podostemaceae, but this observation should be confirmed in more species in the family before this trait is considered to be of any systematic relevance. Other structural rearrangements have been reported in Malpighiales such as the 3 inversions in the LSC region in *P. edulis* (Cauz-Santos et al., 2017), high rates of rearrangements in *Passiflora* (Rabah et al., 2019; Shrestha et al., 2019), and a single small inversion in the LSC region of *Hevea brasiliensis* (Tangphatsornruang et al., 2011). We found no evidence of other structural rearrangements within Podostemaceae.

Evaluation of the boundaries of the 4 plastid regions across all species suggests that the locations of borders of the IRs in the Podostemoideae sampled are fairly conserved, but differ to a small degree in all 5 species studied. This is consistent with the IR boundaries being in a dynamic state in most angiosperms (Goulding et al., 1996). A change in length in the IRs of *T. trifaria*, which are slightly smaller than in Podostemoideae (**Table 2**), could be interpreted as either a contraction of the IRs in *T. trifaria* or an expansion of the regions in Podostemoideae. Either way, expansions and contractions of the IRs have occurred more than once in Malpighiales, creating pseudogenes (Cauz-Santos et al., 2017; Menezes et al., 2018; Shrestha et al., 2019). Podostemaceae are no exception to these variations in length, but as mentioned above, these do not seem to be the sole reason why Podostemaceae have one of the smallest plastomes in Malpighiales.

With regard to gene content, the retention of the *atpF* Group II intron is considered an ancestral condition in land plants with a single gain within the streptophytes, before the origin of land plants, followed by losses in charophytes (Daniell et al., 2008). This intron has also been found to be lost from the plastome of members of Euphorbiacceae, Phyllanthaceae, Elatinaceae, Lophopixidaceae, and Passifloraceae (Daniell et al., 2008). Podostemaceae is a lineage within Malpighiales that retains the ancestral state for presence of the *atpF* group II intron.

Targeted gene disruptions in tobacco have identified four plastid genes with essential functions beyond photosynthesis: *accD*, *clP*, *ycf1*, and *ycf2* (Drescher et al., 2000; Kuroda and Maliga, 2003; Kode et al., 2005; Kikuchi et al., 2013; Parker et al., 2014; Dong et al., 2015). Even though these four genes are retained in the plastid genomes of most angiosperms, including parasitic species that are chlorophyll-deficient (dePamphilis and Palmer, 1990; Funk et al., 2007; Jansen et al., 2007; Parker et al., 2014), there are multiple other parasitic, mycoheterotrophic plants, and taxa outside Malpighiales where these genes are missing from the plastids (Kim, 2004; Magee et al., 2010; Lei et al., 2016; Graham et al., 2017). As reported here, some of these genes may have been reduced to pseudogenes independently in Podostemaceae and in *Passiflora* (Shrestha et al., 2019).

The pseudogenization or loss of genes from the plastids has been reported to be a consequence of it being transferred to the nuclear genome (Jansen et al., 2011; Cauz-Santos et al., 2017). This event of plastid gene transfer remains to be examined in Podostemaceae. The *rps16* gene is considered to be present in

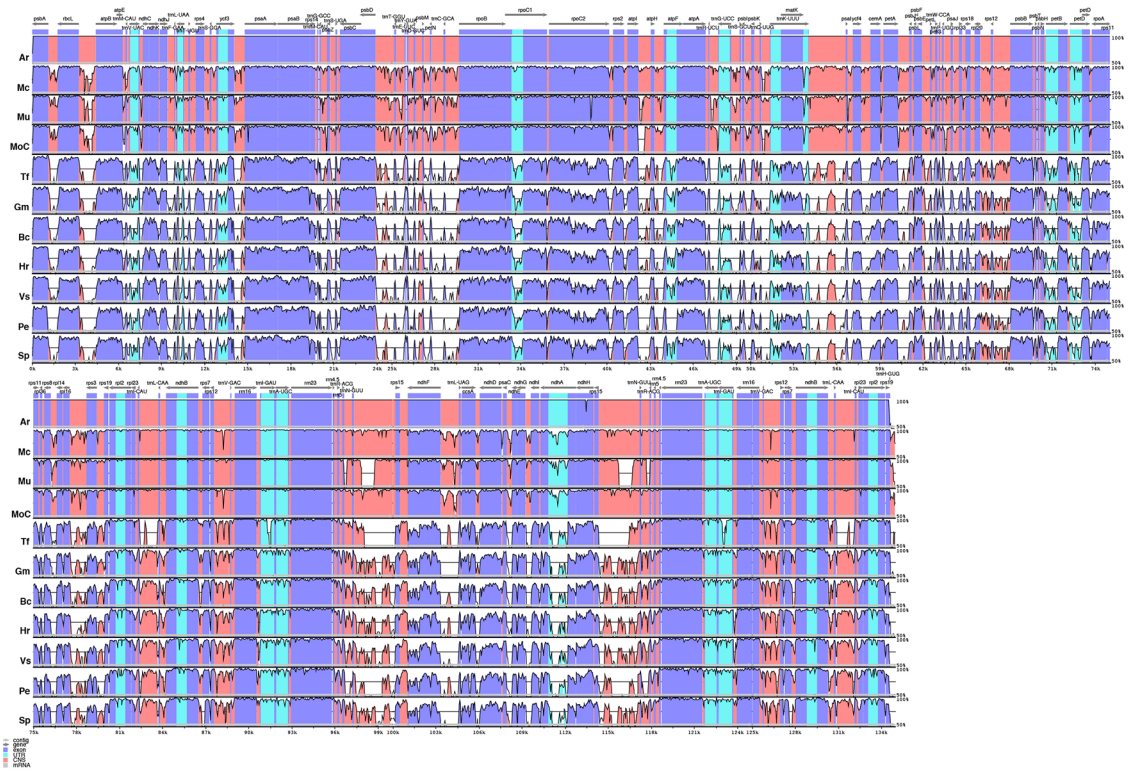


FIGURE 5 | Comparison of percentage identity of plastomes in mVista using *Apinagia riedelii* (Ar) as reference. Mc, *Marathrum capillaceum*; Mu, *Marathrum utile*; Moc, *Monostylis capillacea*; Tt, *Tristicha trifaria*; Gm, *Garcinia mangostana*; Bc, *Byrsonima crassifolia*; Hr, *Hirtella racemosa*; Vs, *Viola seoulensis*; Pe, *Passiflora edulis*; Sp, *Salix purpurea*. The vertical axis corresponds to the percentage identity (50%–100%), while the horizontal axis shows the position of each region within the locus. Arrows indicate the transcription of annotated genes in the reference genome. Genome regions are color coded.

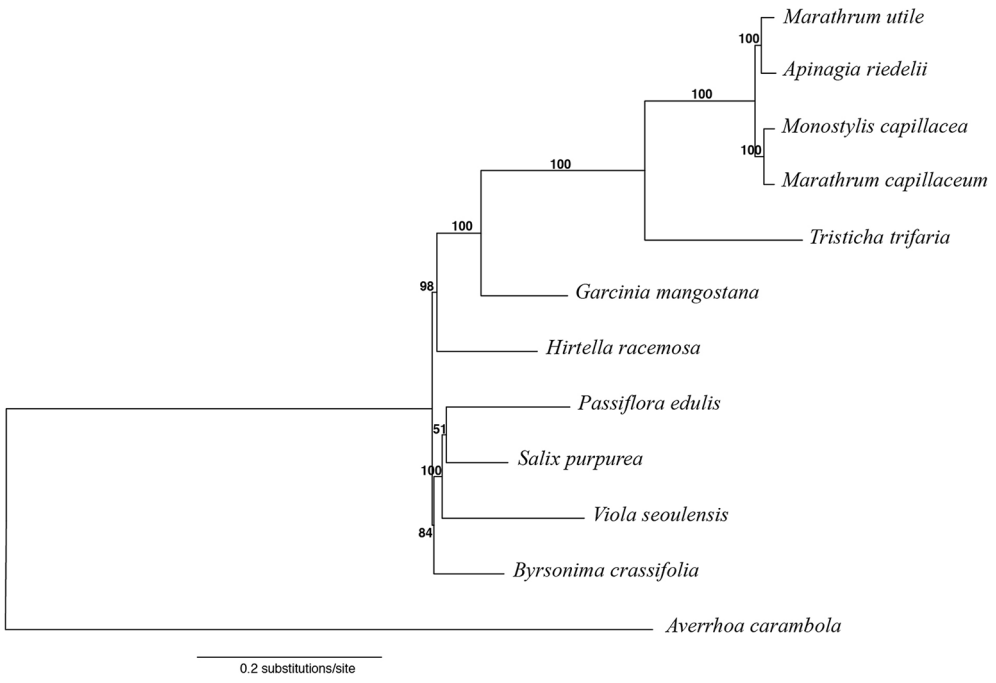


FIGURE 6 | Maximum likelihood tree obtained with RAxML, using *Averrhoa carambola* as outgroup for rooting. Bootstrap support is shown above branches.

the common ancestor of land plants (Daniell et al., 2016) and is found in the plastomes of most angiosperms (Ueda et al., 2008). However, it has been repeatedly reported as lost in Malpighiales (Asif et al., 2010; Daniell et al., 2008; Jansen et al., 2007; Steane, 2005), including our findings of it being missing in Podostemaceae and in other angiosperms (Keller et al., 2017). The multiple losses of *rps16* from the plastids have been explained by the fact that the nuclear encoded *rps16* is dually targeted to the mitochondria and the plastids (Ueda et al., 2008; Keller et al., 2017). This has also been reported to be responsible for the pseudogenization of rpl23 (Bubunenko et al., 1994). Examination of the presence of this gene in the mitochondrial and nuclear DNA would be necessary to test if this explanation also applies to Podostemaceae.

The *ycf1* gene is one of the largest and most variable genes in the plastid genome of land plants, and as mentioned above, it has been proposed to be fundamental for plant function as a key component of the general protein import channel (Dong et al., 2015; Kikuchi et al., 2013). It is rarely missing from the plastome of autotrophic plant lineages, with the exception of Poaceae, some species of *Passiflora*, *Vaccinium macrocarpon*, and some species of *Erodium* (de Vries et al., 2015). However, this gene is more commonly lost from the organellar genome of parasitic, mycoheterotrophic, and carnivorous plant taxa such as *Orobanche purpurea*, species in Droseraceae, and a number of orchids (Guisinger et al., 2010; Parker et al., 2014; Graham et al., 2017; Nevill et al., 2019). Our finding that *ycf1* is pseudogenized in Podostemaceae adds this group to one of the unique autotrophic lineages in angiosperms where this is known to have occurred. However, the mechanisms that compensate for this loss and the implications of it remain to be studied.

The high similarity across the plastome in the subfamily Podostemoideae (**Figure 5**), which are more similar to each other than they are to *T. trifaria*, is explained by the fact that the members of this subfamily share a more recent common ancestor (**Figure 6**). The short branches within Podostemoideae indicate that fewer changes have accumulated since the species diverged, possibly as a consequence of recent speciation events with little subsequent sequence evolution (Soltis et al., 2019). Additionally, the fact that the branches leading to taxa in the Podostemaceae from their common ancestor in Malpighiales are much longer than the branches leading to other taxa within the order is an indicator of faster rates of evolution in the plastome of riverweeds, giving support to previous suggestions (Ruhfel et al., 2016).

Long branches depicting accelerated rates of evolution have been reported in parasitic plants, where multiple changes in the chloroplast respond to a switch from an autotrophic to a heterotrophic metabolism, causing a reduced function of the genome (Young and dePamphilis, 2005; Stefanovic et al., 2007; Lemaire et al., 2011; Givnish et al., 2018). However, the switch from autotrophy to heterotrophy has not occurred in the Podostemaceae. Instead, faster rates of evolution in Podostemaceae could be explained by their rapid life cycle and shorter generation times; most species of Podostemaceae are annual herbs because they depend on the water level to complete their life cycle, dying and shedding seeds in the dry season when the water level is low. This inverse correlation between evolutionary rate and generation time has been suggested for plants as well as for other organisms such as mammals (Bromham et al., 1996; Verdú, 2002; Smith and Donoghue, 2008). Interestingly, the same pattern of long branches observed in Podostemaceae has been found in the Hydrostachyaceae (Cornales) based on phylogenetic analysis using plastid data (Olmstead et al., 2000; Albach et al., 2001; Fan and Xiang, 2003), and the Hydrostachyaceae are the only angiosperm family that shares the unique habit of Podostemaceae (Jäger-Zürn, 1998; Qiu-Yun Xiang, 1999; Rutishauser et al., 2005). However, faster rates of evolution have also been correlated to other life history traits such as plant height, genome size, and age at first reproduction among others (Lehtonen and Lanfear, 2014; Bromham et al., 2015). Which factors are responsible for faster rates of evolution in Podostemaceae and whether they (it) has anything to do with the habit of Podostemaceae and Hydrostachyaceae, remain to be determined.

The phylogenetic relationships found here for the selected species of Malpighiales (**Figure 6**) are in line with previous work where Salicaceae and Passifloraceae are in a clade that shares a more recent common ancestor with Violaceae (and Goupiaceae), whereas Clusiaceae and Podostemaceae are together in a separate clade (Xi et al., 2012; Cai et al., 2019). The relationships within Podostemaceae also follow previous work that suggest that *Marathrum* is paraphyletic (Tippery et al., 2011; Philbrick et al., 2018), calling for a revision of the classification of the genus. Our results (**Figure 6**) also follow a recent study (Cai et al., 2019) in the placement of Chrysobalanaceae, using 5,113 orthology clusters to infer a phylogeny of Malpighiales. These results contradict previous works (Xi et al., 2012; Menezes et al., 2018) that have placed Chrysobalanaceae and Malpighiaceae as more closely related to one another than they are to any of the other families in the Malpighiales included here. The incongruence across data sets is in the deep nodes within the order, reinstating the difficulty in reconstructing deep nodes in Malpighiales (Wurdack and Davis, 2009).

CONCLUSIONS

In this study, we assembled five full plastid genomes of species in Podostemaceae and analyzed them in a comparative framework within Malpighiales. We detected an important inversion in the LSC region that could be of systematic relevance as a synapomorphy of the group and also described slight variations in the length of the IRs in all the species included in the study. The plastomes of the family are among the smallest reported to date in the order Malpighiales, and we suggest that this small size is a result of a combination of variation in length of IR regions, gene loss, and intergenic region variation and intron loss. Gene content is the same within the Podostemaceae, and some of the gene loss and pseudogenization events reported are common in angiosperms (e.g., *rps16*, *rpl23*, and *accD*), whereas others are very rare (e.g., *ycf1* and *ycf2*). The mechanisms that compensate for these losses and the implications of their occurrence in Podostemaceae remain a subject of study. Our results suggest an accelerated rate of evolution for the group and reinstate the difficulty in the inferring relationship in deep nodes in

Malpighiales. Ultimately, this study provides insights into the structure and evolution of plastomes in Podostemaceae and lays the foundations for phylogenomic studies in the family.

DATA AVAILABILITY

Whole-plastid genome sequences are deposited in GenBank, and accession numbers are provided in **Table 1**. The records can be found in GenBank (https://www.ncbi.nlm.nih.gov/genbank/).

AUTHOR CONTRIBUTIONS

AB and RO designed the study and wrote the manuscript. AB analyzed the data and conducted field work in Colombia, collecting *Marathrum utile*. BR and CP contributed tissue samples, sequences, and comments. SM provided help with collecting permits and sampling logistics in the field in Colombia. CB contributed tissue samples of *A. riedelii*, *Marathrum capillaceum* and *Monostylis capillacea* from Brazil, as we all comments to the manuscript. AM provided tissue samples of *Tristica trifaria* from Africa.

FUNDING

AB was supported by the Sargent award from the Biology department at the University of Washington, the Colciencias fellowship for Graduate studies (Doctorados en el Exterior-679), and the BSA and ASPT Research Grants for graduate students. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant. BR was supported by a National Science Foundation Grant (DEB-1754329). Collection of some tissue samples was done with support of the National Science Foundation Grants (DEB-0444589 and DEB-1754199) and Connecticut State University-AAUP research grants to CP, and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) grants PROTAX 562251/2010-3, REFLORA 563534/2010-9, and Productivity Grant (307870/2014-6) to CB. Open Access page charges covered in part by the Universidad de los Andes research results publication fund CI-001, and by the University of Washington and University of Michigan.

ACKNOWLEDGMENTS

We are thankful to Dr. Adam Leaché at the University of Washington for providing laboratory space, use of equipment, and comments on the manuscript. We thank Dr. Kurt Neubig for his generous help with the sequencing of *A. riedelii*, *M. capillaceum*, *M. capillacea*, and *T. trifaria*. Dr. Nuebig and Dr. Lucas Majure offered valuable comments on the manuscript. We also thank Dr. Sasa Stefánovic for comments on chloroplast evolution and loss of the *yfc1* gene. Dr. Carl J. Rothfels and Dr. Lachezar A. Nikolov at UCLA, and Dr. Angela Jean McDonnell at Chicago Botanical Garden provided important comments, reviews and edits that improved the initial version of this manuscript. Maria Paula Contreras provided assistance in the field in Colombia.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01035/full#supplementary-material

REFERENCES

Albach, D. C., Soltis, D. E., Chase, M. W., and Soltis, P. S. (2001). Phylogenetic placement of the enigmatic angiosperm. *Hydrostachys. Taxon* 50, 781. doi: 10.2307/1223707

Asif, M. H., Mantri, S. S., Sharma, A., Srivastava, A., Trivedi, I., Gupta, P., et al. (2010). Complete sequence and organisation of the *Jatropha curcas* (Euphorbiaceae) chloroplast genome. *Tree Genet. Genomes* 6, 941–952. doi: 10.1007/s11295-010-0303-0

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bromham, L., Hua, X., Lanfear, R., and Cowman, P. F. (2015). Exploring the relationships between mutation rates, life history, genome size, environment, and species richness in flowering plants. *Am. Nat.* 185, 507–524. doi: 10.1086/680052

Bromham, L., Rambaut, A., and Harvey, P. H. (1996). Determinants of rate variation in mammalian DNA sequence evolution. *J. Mol. Evol.* 43, 610–621. doi: 10.1007/BF02202109

Bubunenko, M. G., Schmidt, J., and Subramanian, A. R. (1994). Protein substitution in chloroplast ribosome evolution: a eukaryoticcytosolic protein has replaced its organelle homologue(L23) in spinach. *J. Mol. Biol.* 240, 28–41.

Cai, L., Xi, Z., Amorim, A. M., Sugumaran, M., Rest, J. S., Liu, L., et al. (2019). Widespread ancient whole-genome duplications in Malpighiales coincide with Eocene global climatic upheaval. *New Phytol.* 221, 565–576. doi: 10.1111/nph.15357

Cauz-Santos, L. A., Munhoz, C. F., Rodde, N., Cauet, S., Santos, A. A., Penha,H. A., et al. (2017). The chloroplast genome of *Passiflora edulis* (Passifloraceae) assembled from long sequence reads: structural organization and phylogenomic studies in Malpighiales. *Front. Plant. Sci.* 8, 334. doi: 10.3389/fpls.2017.00334

Chumley, T. W., Palmer, J. D., Mower, J. P., Fourcade, H. M., Calie, P. J., Boore, J. L., et al. (2006). The complete chloroplast genome sequence of *Pelargonium* × *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol. Biol. Evol.* 23, 2175–2190. doi: 10.1093/molbev/msl089

Cosner, M. E., Raubeson, L. A., and Jansen, R. K. (2004). Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. *BMC Evol. Biol.* 4, 27. doi: 10.1186/1471-2148-4-27

Daniell, H., Lin, C.-S., Yu, M., and Chang, W.-J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17, 134. doi: 10.1186/s13059-016-1004-2

Daniell, H., Wurdack, K. J., Kanagaraj, A., Lee, S.-B., Saski, C., and Jansen, R. K. (2008). The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of *atpF in* Malpighiales: RNA editing and multiple losses of a group II intron. *Theor. Appl. Genet.* 116, 723–737. doi: 10.1007/s00122-007-0706-y

Darling, A. C. E. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704

de Vries, J., Sousa, F. L., Bölter, B., Soll, J., and Gould, S. B. (2015). YCF1: a green TIC? *Plant Cell.* 27, 1827–1833. doi: 10.1105/tpc.114.135541

dePamphilis, C. W., and Palmer, J. D. (1990). Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. *Nature* 348, 337–339. doi: 10.1038/348337a0

Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S., et al. (2015). Ycf1, the most promising plastid DNA barcode of land plants. *Sci Rep.* 5, 8348. doi: 10.1038/srep08348

Downie, S. R., and Palmer, J. D (1992). "Use of chloroplast DNA rearrangements in reconstructing plant phylogeny," in *Molecular Systematics of Plants*. Eds. P. S. Soltis, D. E. Soltis, and J. J. Doyle (Boston, MA: Springer US), 14–35. doi: 10.1007/978-1-4615-3276-7_2

Drescher, A., Ruf, S., Calsa, T., Carrer, H., and Bock, R. (2000). The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant J.* 22, 97–104. doi: 10.1046/j.1365-313x.2000.00722.x

Eckardt, N. A., and Baum, D. (2010). The Podostemad puzzle: the evolution of unusual morphology in the Podostemaceae. *Plant Cell.* 22, 2104–2104. doi: 10.1105/tpc.110.220711

Fan, C., and Xiang, Q.-Y. J. (2003). Phylogenetic analyses of Cornales based on *26S* rRNA and combined *26S* rDNA-*matK-rbcL* sequence data. *Am. J. Bot.* 90, 1357–1372. doi: 10.3732/ajb.90.9.1357

Fan, W.-B., Wu, Y., Yang, J., Shahzad, K., and Li, Z.-H. (2018). Comparative chloroplast genomics of Dipsacales species: insights into sequence variation, adaptive evolution, and phylogenetic relationships. *Front. Plant Sci.* 9, 689. doi: 10.3389/fpls.2018.00689

Firetti, F., Zuntini, A. R., Gaiarsa, J. W., Oliveira, R. S., Lohmann, L. G., and Van Sluys, M.-A. (2017). Complete chloroplast genome sequences contribute to plant species delimitation: a case study of the *Anemopaegma* species complex. *Am. J. Bot.* 104, 1493–1509. doi: 10.3732/ajb.1700302

Funk, H. T., Berg, S., Krupinska, K., Maier, U. G., and Krause, K. (2007). Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii. BMC Plant Biol.* 7, 45. doi: 10.1186/1471-2229-7-45

Gitzendanner, M. A., Soltis, P. S., Wong, G. K.-S., Ruhfel, B. R., and Soltis, D. E. (2018). Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. *Am. J. Bot.* 105, 291–301. doi: 10.1002/ajb2.1048

Givnish, T. J., Zuluaga, A., Spalink, D., Soto Gomez, M., Lam, V. K. Y., Saarela, J. M., et al. (2018). Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. *Am. J. Bot.* 105, 1888–1910. doi: 10.1002/ajb2.1178

Goulding, S. E., Wolfe, K. H., Olmstead, R. G., and Morden, C. W. (1996). Ebb and flow of the chloroplast inverted repeat. *MGG Mol. Gen. Genet.* 252, 195–206. doi: 10.1007/BF02173220

Graham, S. W., Lam, V. K. Y., and Merckx, V. S. F. T. (2017). Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. *New Phytol.* 214, 48–55. doi: 10.1111/nph.14398

Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res*. 47, W59–W64. doi:10.1093/nar/gkz238

Guisinger, M. M., Chumley, T. W., Kuehl, J. V., Boore, J. L., and Jansen, R. K. (2010). Implications of the plastid genome sequence of *Typha* (Typhaceae, Poales) for understanding genome evolution in Poaceae. *J. Mol. Evol.* 70, 149–166. doi: 10.1007/s00239-009-9317-3

Hoot, S. B., and Palmer, J. D. (1994). Structural rearrangements, including parallel inversions, within the chloroplast genome of *Anemone* and related genera. *J. Mol. Evol.* 38, 274–281. doi: 10.1007/BF00176089

Jäger-Zürn, I. (1998). "Anatomy of Hydrostachyaceae," in *Extreme adaptations in angiospermous hydrophytes*. Eds. E. Landolt, I. Jager-Zurn, and R. A. A. Schnell (Berlin: Borntraeger), 129–196.

Jäger-Zürn, I. (2005). Shoot apex and spathella: two problematical structures of Podostemaceae–Podostemoideae. *Plant Syst. Evol.* 253, 209–218. doi: 10.1007/s00606-005-0297-5

Jäger-Zürn, I. (2007). The shoot apex of Podostemaceae: *de novo* structure or reduction of the conventional type? *Flora* 202, 383–394. doi: 10.1016/j.flora.2007.04.003

Jäger-Zürn, I. (2011). Neglected features of probable taxonomic value in Podostemaceae: the case of *Polypleurum. Flora* 206, 38–46. doi: 10.1016/j.flora.2010.01.008

Jansen, R. K., Cai, Z., Raubeson, L. A., Daniell, H., dePamphilis, C. W., Leebens-Mack, J., et al. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *PNAS* 104, 19369–19374. doi: 10.1073/pnas.0709121104

Jansen, R. K., and Palmer, J. D. (1987). A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *PNAS* 84, 5818–5822. doi: 10.1073/pnas.84.16.5818

Jansen, R. K., Saski, C., Lee, S.-B., Hansen, A. K., and Daniell, H. (2011). Complete plastid genome sequences of three Rosids (*Castanea, Prunus, Theobroma*): evidence for at least two independent transfers of *rpl22* to the nucleus. *Mol. Biol. Evol.* 28, 835–847. doi: 10.1093/molbev/msq261

Jin, J.-J., Yu, W.-B., Yang, J.-B., Song, Y., Yi, T.-S., and Li, D.-Z. (2018). GetOrganelle: a simple and fast pipeline for *de novo* assembly of a complete circular chloroplast genome using genome skimming data. *bioRxiv*. doi: 10.1101/256479

Keller, J., Rousseau-Gueutin, M., Martin, G. E., Morice, J., Boutte, J., Coissac, E., et al. (2017). The evolutionary fate of the chloroplast and nuclear *rps16* genes as revealed through the sequencing and comparative analyses of four novel legume chloroplast genomes from *Lupinus. DNA Res.* 24, 343–358. doi: 10.1093/dnares/dsx006

Kikuchi, S., Bédard, J., Hirano, M., Hirabayashi, Y., Oishi, M., Imai, M., et al. (2013). Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* 339, 571–574. doi: 10.1126/science.1229262

Kim, K.-J. (2004). Complete chloroplast genome sequences from Korean Ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* 11, 247–261. doi: 10.1093/dnares/11.4.247

Kita, Y., and Kato, M. (2001). Infrafamilial phylogeny of the aquatic angiosperm Podostemaceae inferred from the nucleotide sequences of the *matK* gene. *Plant Biol.* 3, 156–163. doi: 10.1055/s-2001-12895

Kode, V., Mudd, E. A., Iamtham, S., and Day, A. (2005). The tobacco plastid *accD* gene is essential and is required for leaf development: essential plastid gene. *Plant J.* 44, 237–244. doi: 10.1111/j.1365-313X.2005.02533.x

Koi, S., Kita, Y., Hirayama, Y., Rutishauser, R., Huber, K. A., and Kato, M. (2012). Molecular phylogenetic analysis of Podostemaceae: implications for taxonomy of major groups. *Bot. J. Linn. Soc.* 169, 461–492. doi: 10.1111/j.1095-8339.2012.01258.x

Krause, K. (2011). Piecing together the puzzle of parasitic plant plastome evolution. *Planta* 234, 647–656. doi: 10.1007/s00425-011-1494-9

Kuroda, H., and Maliga, P. (2003). The plastid *clpP1* protease gene is essential for plant development. *Nature* 425, 86–89. doi: 10.1038/nature01909

Lanfear, R., Calcott, B., Ho, S. Y. W., and 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

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2016). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 43, 772–773. doi: 10.1093/molbev/msw260

Lehtonen, J., and Lanfear, R. (2014). Generation time, life history and the substitution rate of neutral mutations. *Biol. Lett.* 10, 20140801. doi: 10.1098/rsbl.2014.0801

Lei, W., Ni, D., Wang, Y., Shao, J., Wang, X., Yang, D., et al. (2016). Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of *Astragalus membranaceus. Sci. Rep.* 6, 21669. doi: 10.1038/srep21669

Lemaire, B., Huysmans, S., Smets, E., and Merckx, V. (2011). Rate accelerations in nuclear *18S* rDNA of mycoheterotrophic and parasitic angiosperms. *J. Plant Res.* 124, 561–576. doi: 10.1007/s10265-010-0395-5

Li, B., and Zheng, Y. (2018). Dynamic evolution and phylogenomic analysis of the chloroplast genome in Schisandraceae. *Sci. Rep.* 8, 9285. doi: 10.1038/s41598-018-27453-7

Li, Z.-H., Ma, X., Wang, D.-Y., Li, Y.-X., Wang, C.-W., and Jin, X.-H. (2019). Evolution of plastid genomes of *Holcoglossum* (Orchidaceae) with recent radiation. *BMC Evol. Biol.* 19, 63. doi: 10.1186/s12862-019-1384-5

Liu, H., He, J., Ding, C., Lyu, R., Pei, L., Cheng, J., et al. (2018). Comparative analysis of complete chloroplast genomes of *Anemoclema, Anemone, Pulsatilla,* and *Hepatica* revealing structural variations among genera in tribe Anemoneae (Ranunculaceae). *Front. Plant. Sci.* 9, 1097. doi: 10.3389/fpls.2018.01097

Lloyd Evans, D., Joshi, S. V., and Wang, J. (2019). Whole chloroplast genome and gene locus phylogenies reveal the taxonomic placement and relationship of

Tripidium (Panicoideae: Andropogoneae) to sugarcane. *BMC Evol. Biol.* 19, 33. doi: 10.1186/s12862-019-1356-9

Magee, A. M., Aspinall, S., Rice, D. W., Cusack, B. P., Semon, M., Perry, A. S., et al. (2010). Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Res.* 20, 1700–1710. doi: 10.1101/gr.111955.110

Menezes, A. P. A., Resende-Moreira, L. C., Buzatti, R. S. O., Nazareno, A. G., Carlsen, M., Lobo, F. P., et al. (2018). Chloroplast genomes of *Byrsonima* species (Malpighiaceae): comparative analysis and screening of high divergence sequences. *Sci. Rep.* 8, 2210. doi: 10.1038/s41598-018-20189-4

Metzker, M. L. (2009). Sequencing technologies—the next generation. *Nat. Rev. Genet.* 11, 31. doi: 10.1038/nrg2626

Neubig, K., Whitten, W., Abbott, J., Elliott, S., Soltis, D. E., and Soltis, P. S., (2014). "Variables affecting DNA preservation in archival DNA specimens," in *DNA Banking in the 21st Century: Proceedings of the U.S. Workshop on DNA Banking*. Eds. W. L. Applequist and L. M. Campbell (St. Louis: The William L. Brown Center at the Missouri Botanical Garden).

Nevill, P. G., Howell, K. A., Cross, A. T., Williams, A. V., Zhong, X.,Tonti-Filippini, J., et al. (2019). Plastome-wide rearrangements and gene losses in carnivorous Droseraceae. *Genome Biol. Evol.* 11, 472–485. doi: 10.1093/gbe/evz005

Novelo, R. A., and Philbrick, C. T. (1997). Taxonomy of Mexican Podostemaceae. *Aquat. Bot.* 57, 275–303. doi: 10.1016/S0304-3770(96)01122-9

Olmstead, R. G., and Bedoya, A. M. (2019). Whole genomes: the holy grail. A commentary on molecular phylogenomics of the tribe shoreeae (Dipterocarpaceae) using whole plastid genomes. *Ann. Bot.* 123, iv–v. doi: 10.1093/aob/mcz055

Olmstead, R. G., Kim, K.-J., Jansen, R. K., and Wagstaff, S. J. (2000). The phylogeny of the Asteridae sensu lato based on chloroplast *ndhF* gene sequences. *Mol. Phylogenet. Evol.* 16, 96–112. doi: 10.1006/mpev.1999.0769

Oropeza, N., Mercado-Ruaro, P., Novelo, R. A., and Thomas Philbrick, C. (1998). Karyomorphological studies of Mexican species of *Marathrum* (Podostemaceae). *Aquat. Bot.* 62, 207–211. doi: 10.1016/S0304-3770(98)00086-2

Oropeza, N., Palomino, G., Novelo, R. A., and Philbrick, C. T. (2002). Karyomorphological studies in *Oserya, Vanroyenella* and *Tristicha* (Podostemaceae sensu lato). *Aquat. Bot.* 73, 163–171. doi: 10.1016/S0304-3770(02)00018-9

Palmer, J. D. (1985). "Evolution of chloroplast and mitochondrial DNA in plants and algae," in *Molecular Evolutionary Genetics*. Ed. R. MacIntyre (New York: Plenum Press), 131–240. doi: 10.1007/978-1-4684-4988-4_3

Palmer, J. D., Nugent, J. M., and Herbon, L. A. (1987). Unusual structure of *Geranium* chloroplast DNA: a triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. *PNAS* 84, 769–773. doi: 10.1073/pnas.84.3.769

Parker, N., Wang, Y., and Meinke, D. (2014). Natural variation in sensitivity to a loss of chloroplast translation in *Arabidopsis. Plant Physiol.* 166, 2013–2027. doi: 10.1104/pp.114.249052

Philbrick, C. T., and Novelo, R. A. (1995). New World Podostemaceae: ecological and evolutionary enigmas. *Brittonia* 47, 210. doi: 10.2307/2806959

Philbrick, C. T., and Novelo, R. A. (1998). Flowering phenology, pollen flow, and seed production in *Marathrum rubrum* (Podostemaceae). *Aquat. Bot.* 62, 199–206. doi: 10.1016/S0304-3770(98)00090-4

Philbrick, C. T., Ruhfel, B. R., and Bove, C. P. (2018). Contributions to the taxonomy of *Rhyncholacis* (Podostemaceae): evidence of monophyly, description of a new species, and transfer of the monotypic *Macarenia*. *Phytotaxa* 357, 107. doi: 10.11646/phytotaxa.357.2.3

Qiu-Yun Xiang, J. (1999). Systematic affinities of Grubbiaceae and Hydrostachyaceae within Cornales—insights from *rbcL* sequences. *Harv. Pap. Bot.* 4, 527–541. https://www.jstor.org/stable/41761589

Rabah, S. O., Shrestha, B., Hajrah, N. H., Sabir, M. J., Alharby, H. F., Sabir, M. J., et al. (2019). *Passiflora* plastome sequencing reveals widespread genomic rearrangements. *J. Syst. Evol.* 57, 1–14. doi: 10.1111/jse.12425

Ruhfel, B. R., Bittrich, V., Bove, C. P., Gustafsson, M. H. G., Philbrick, C. T., Rutishauser, R., et al. (2011). Phylogeny of the clusioid clade (Malpighiales): evidence from the plastid and mitochondrial genomes. *Am. J. Bot.* 98, 306–325. doi: 10.3732/ajb.1000354

Ruhfel, B. R., Bove, C. P., Philbrick, C. T., and Davis, C. C. (2016). Dispersal largely explains the Gondwanan distribution of the ancient tropical clusioid plant clade. *Am. J. Bot.* 103, 1117–1128. doi: 10.3732/ajb.1500537

Ruhfel, B. R., Gitzendanner, M. A., Soltis, P. S., Soltis, D. E., and Burleigh, J. (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

Rutishauser, R. (1995). Developmental patterns of leaves in Podostemaceae compared with more typical flowering plants: saltational evolution and fuzzy morphology. *Can. J. Botany* 73, 1305–1317. doi: 10.1139/b95-142

Rutishauser, R. (1997). Structural and developmental diversity in Podostemaceae (river-weeds). *Aquat. Bot.* 57, 29–70. doi: 10.1016/S0304-3770(96)01120-5

Rutishauser, R., Novelo, R. A., and Philbrick, C. T. (1999). Developmental morphology of New World Podostemaceae: *Marathrum* and *Vanroyenella*. *Int. J. Plant Sci.* 160, 29–45. doi: 10.1086/314097

Rutishauser, R., Pfeifer, E., Novelo, R. A., and Thomas Philbrick, C. (2005). *Diamantina lombardii*—an odd Brazilian member of the Podostemaceae. *Flora* 200, 245–255. doi: 10.1016/j.flora.2004.09.004

Schwarz, E. N., Ruhlman, T. A., Sabir, J. S. M., Hajrah, N. H., Alharbi, N. S., Al-Malki, A. L., et al. (2015). Plastid genome sequences of legumes reveal parallel inversions and multiple losses of *rps16* in papilionoids: parallel inversions and *rps16* losses in legumes. *J. Syst. Evol.* 53, 458–468. doi: 10.1111/jse.12179

Shrestha, B., Weng, M.-L., Theriot, E. C., Gilbert, L. E., Ruhlman, T. A., Krosnick, S. E., et al. (2019). Highly accelerated rates of genomic rearrangements and nucleotide substitutions in plastid genomes of *Passiflora* subgenus *Decaloba*. Mol. *Phylogenet. Evol.* 138, 53–64. doi: 10.1016/j.ympev.2019.05.030

Smith, S. A., and Donoghue, M. J. (2008). Rates of molecular evolution are linked to life history in flowering plants. *Science* 322, 86–89. doi: 10.1126/science.1163197

Soltis, P. S., Folk, R. A., and Soltis, D. E. (2019). Darwin review: angiosperm phylogeny and evolutionary radiations. *P. Roy. Soc. B-Biol. Sci.* 286, 20190099. doi: 10.1098/rspb.2019.0099

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Steane, D. A. (2005). Complete nucleotide sequence of the chloroplast genome from the Tasmanian Blue Gum, *Eucalyptus globulus* (Myrtaceae). *DNA Res.* 12, 215–220. doi: 10.1093/dnares/dsi006

Stefanovic, S., Kuzmina, M., and Costea, M. (2007). Delimitation of major lineages within *Cuscuta* subgenus *Grammica* (Convolvulaceae) using plastid and nuclear DNA sequences. *Am. J. Bot.* 94, 568–589. doi: 10.3732/ajb.94.4.568

Straub, S. C. K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R. C., and Liston, A. (2012). Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *Am. J. Bot.* 99, 349–364. doi: 10.3732/ajb.1100335

Sugiura, M. (1992). "The chloroplast genome," in *10 Years Plant Molecular Biology*. Eds. R. A. Schilperoort and L. Dure (Dordrecht: Springer Netherlands), 149–168. doi: 10.1007/978-94-011-2656-4_10

Tangphatsornruang, S., Uthaipaisanwong, P., Sangsrakru, D., Chanprasert, J., Yoocha, T., Jomchai, N., et al. (2011). Characterization of the complete chloroplast genome of *Hevea brasiliensis* reveals genome rearrangement, RNA editing sites and phylogenetic relationships. *Gene* 475, 104–112. doi: 10.1016/j.gene.2011.01.002

The Angiosperm Phylogeny Group. (2016). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* 181, 1–20. doi: 10.1111/boj.12385

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W6–W11. doi: 10.1093/nar/gkx391

Tippery, N. P., Philbrick, C. T., Bove, C. P., and Les, D. H. (2011). Systematics and phylogeny of Neotropical riverweeds (Podostemaceae: Podostemoideae). *Syst. Bot.* 36, 105–118. doi: 10.1600/036364411X553180

Ueda, M., Nishikawa, T., Fujimoto, M., Takanashi, H., Arimura, S.-I., Tsutsumi, N., et al. (2008). Substitution of the gene for chloroplast *rps16* was assisted by generation of a dual targeting signal. *Mol. Biol. Evol.* 25, 1566–1575. doi: 10.1093/molbev/msn102

van Royen, P. (1951). *The Podostemaceae of the New World*. (Harte: Medd. Bot, Museum Utrecht). Available at https://www.jstor.org/stable/41761589. Available at: https://books.google.com/books?id=IeV1SwAACAAJ.

Verdú, M. (2002). Age at maturity and diversification in woody angiosperms. *Evolution* 56, 1352–1361. doi: 10.1111/j.0014-3820.2002.tb01449.x

Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T., and Sugiura, M. (1994). Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *PNAS* 91, 9794–9798. doi: 10.1073/pnas.91.21.9794

Weng, M.-L., Blazier, J. C., Govindu, M., and Jansen, R. K. (2014). Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Mol. Biol. Evol.* 31, 645–659. doi: 10.1093/molbev/mst257

Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Müller, K. F., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant. Mol. Biol* 76, 273–297. doi: 10.1007/s11103-011-9762-4

Wolfe, K. H., Morden, C. W., and Palmer, J. D. (1992). Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *PNAS* 89, 10648–10652. doi: 10.1073/pnas.89.22.10648

Wurdack, K. J., and Davis, C. C. (2009). Malpighiales phylogenetics: gaining ground on one of the most recalcitrant clades in the angiosperm tree of life. *Am. J. Bot.* 96, 1551–1570. doi: 10.3732/ajb.0800207

Xi, Z., Ruhfel, B. R., Schaefer, H., Amorim, A. M., Sugumaran, M., Wurdack, K. J., et al. (2012). Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *PNAS* 109, 17519–17524. doi: 10.1073/pnas.1205818109

Xiao-Ming, Z., Junrui, W., Li, F., Sha, L., Hongbo, P., Lan, Q., et al. (2017). Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. *Sci. Rep.* 7, 1555. doi: 10.1038/s41598-017-01518-5

Xu, J.-H., Liu, Q., Hu, W., Wang, T., Xue, Q., and Messing, J. (2015). Dynamics of chloroplast genomes in green plants. *Genomics* 106, 221–231. doi: 10.1016/j.ygeno.2015.07.004

Young, N. D., and dePamphilis, C. W. (2005). Rate variation in parasitic plants: correlated and uncorrelated patterns among plastid genes of different function. *BMC Evol. Biol.* 5, 16. doi: 10.1186/1471-2148-5-16

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary Material

Table S1. Optimal scheme obtained with PartitionFinder2. Subset (partition) number, per-partition best-fit models and loci included in each partition are included.

Subset	Best Model	Partition names
1	GTR+I+G	psbB_pos1, psbD_pos1, PetG_pos1, psbN_pos1, atpH_pos1, psbC_pos2
2	GTR+I+G	ndhD_pos2, ndhG_pos2, ndhE_pos2, petL_pos2, psbK_pos3, psbI_pos3, psbZ_pos2, PetG_pos2, petN_pos2
3	GTR	PetG_pos3, psbK_pos1, psaI_pos3, psbF_pos3, psbE_pos3, psbM_pos3
4	GTR+I+G	ndhC_pos1, rs12e1_pos2, trnH, petL_pos1, rps4_pos3, psaI_pos1, ycf4_pos2, psbZ_pos1, psaJ_pos1
5	GTR+G	atpI_pos2, psaJ_pos2, ndhC_pos2, psaI_pos2, psbM_pos2
6	GTR+I+G	psaJ_pos3, psaC_pos3, psbI_pos1, rpl14_pos3, rpl36_pos3, rps3_pos2, rps15_pos3, rps8_pos3, rpl16_pos3
7	GTR+G	rps3_pos1, rps18_pos2, rps8_pos2, rpl33_pos1
8	GTR+G	rpoA_pos1, rps8_pos1, rpoC2_pos1, rps3_pos3, rpl16_pos1, rps18_pos1, rpl33_pos2, psbN_pos3
9	GTR+G	psbD_pos3, psaA_pos1, psbB_pos3, psaB_pos3, rpl33_pos3, psbC_pos1, rbcL_pos3
10	GTR+G	rps14_pos3, atpH_pos3, ndhJ_pos3, rps4_pos1, atpI_pos3, psbH_pos3, rps18_pos3, cemA_pos2, atpF_pos3, rpoB_pos2, rps2_pos1, ndhC_pos3
11	GTR+I+G	psbC_pos3, psaB_pos2, psbB_pos2, psaA_pos3, ndhJ_pos2, trnRUCU, rps12e2e3_pos2
12	GTR+G	rpl20_pos1, rpl20_pos3, ccsA_pos1, cemA_pos3, ndhG_pos1, rs12e1_pos3, rpl20_pos2
13	GTR+G	rs12e1_pos1, rpl36_pos1, rpl14_pos1, rps11_pos2
14	GTR+I+G	trnFC, psbT_pos1, trnS, trnD, psbE_pos2, trnE, trnLUAG
15	GTR+G	ndhH_pos2, rpl14_pos2, psbT_pos2, rpoB_pos1, petA_pos2
16	GTR+G	psbA_pos3, petL_pos3, psbT_pos3
17	GTR+I+G	psbH_pos1, petN_pos1, psbM_pos1, psbE_pos1
18	GTR+I+G	rbcL_pos2, psbH_pos2
19	GTR+I+G	psbA_pos2, psbD_pos2, atpH_pos2, psbF_pos2, psbN_pos2
20	GTR+G	psbK_pos2, petB_pos1, petD_pos2, petD_pos1
21	GTR+I+G	ndhF_pos3, petB_pos2, ndhI_pos3, ndhK_pos3, ndhH_pos3, ndhD_pos3, rps11_pos1
22	GTR+G	ndhF_pos1, petB_pos3
23	GTR+G	trnG, petD_pos3
24	GTR+G	rps19_pos3, rps19_pos1, rps19_pos2
25	GTR+I+G	rpoC2_pos2, rpoA_pos2, rpoC1_pos1, ndhI_pos2
26	GTR+G	rpoC2_pos3, atpA_pos3, rpoA_pos3, petA_pos3, ycf4_pos3

27	GTR+I+G	rpl16_pos2, ndhK_pos2, ndhH_pos1, rps11_pos3
28	GTR+I+G	rpl36_pos2, trnfM, trnC, trnQ
29	GTR+G	rpl2_pos2, rpl2_pos1
30	GTR+G	rpl23_pos2, rpl23_pos3, rpl23_pos1, rpl2_pos3
31	GTR+G	ndhB_pos2, ndhB_pos1, ndhB_pos3, trnI
32	GTR+G	rps7_pos3, rps7_pos1, rps7_pos2
33	GTR+G	rps12e2e3_pos1, psbI_pos2, rps14_pos1, ndhE_pos1, rps12e2e3_pos3, rps2_pos3, trnS2, rps14_pos2
34	GTR+I+G	ccsA_pos2, ndhF_pos2
35	GTR+G	ndhA_pos1, ccsA_pos3
36	GTR+I+G	cemA_pos1, rps15_pos1, rps15_pos2
37	GTR+G	ndhA_pos2, ndhA_pos3
38	GTR+G	rps2_pos2, petN_pos3, ndhI_pos1, ycf4_pos1, rps4_pos2, ndhK_pos1
39	GTR+G	ndhE_pos3, psbZ_pos3, ndhG_pos3
40	GTR+I+G	rrn45, psbF_pos1, trnT2, psaC_pos2, trnR, trnN, rrn5, psbA_pos1, trnT, trnV, trnL, psaA_pos2, trnG2, psaC_pos1, trnY
41	GTR+G	ndhD_pos1
42	GTR+I+G	atpA_pos1, petA_pos1, rpoB_pos3, ndhJ_pos1, atpI_pos1
43	GTR+I+G	atpA_pos2
44	GTR+G	rpoC1_pos3, atpF_pos1
45	GTR+G	ycf3_pos3, ycf3_pos1, atpF_pos2, trnV2, ycf3_pos2
46	GTR+I+G	rpoC1_pos2
47	GTR+I+G	trnS3, trnIGAU, psaB_pos1, trnA
48	GTR+I+G	rbcL_pos1
49	GTR+I+G	trnK
50	GTR+I+G	rrn23, rrn16
51	GTR+G	trnW, trnP
52	GTR+I+G	trnL2
53	GTR+I+G	noncoding

Chapter 2

ANDEAN UPLIFT, DRAINAGE BASIN FORMATION, AND THE EVOLUTION OF PLANTS LIVING IN FAST-FLOWING AQUATIC ECOSYSTEMS IN NORTHERN SOUTH AMERICA

MISS ANA M. BEDOYA (Orcid ID : 0000-0002-6245-3481)

Andean uplift, drainage basin formation, and the evolution of plants living in fast-flowing aquatic ecosystems in northern South America

Ana M. Bedoya[1], Adam D. Leaché[2] and Richard G. Olmstead[3]

Department of Biology and Burke Museum, University of Washington, Seattle, WA, United States, 98195

[1] https://orcid.org/0000-0002-6245-3481

[2] https://orcid.org/0000-0001-8929-6300

[3] https://orcid.org/0000-0001-7660-4159

Author for correspondence:

Ana M. Bedoya

Tel: +1 2063104972

Email: ambedoya@uw.edu

Received: *21 June 2021*

Accepted: *20 July 2021*

Summary

- Northern South America is a geologically dynamic and species-rich region. Fossil and stratigraphic data show that mountain uplift in the tropical Andes reconfigured river drainages. These landscape changes shaped the evolution of the flora in the region, yet the impacts on aquatic taxa have been overlooked.
- We explore the role of landscape change on the evolution of plants living strictly in rivers across drainage basins in northern South America by conducting population structure, phylogenetic inference, and divergence-dating analyses for two species in the genus *Marathrum* (Podostemaceae).
- Mountain uplift and drainage basin formation isolated populations of *M. utile* and *M. foeniculaceum* in northern South America and created barriers to gene flow across river drainages. Sympatric species hybridize and the hybrids show the phenotype of one parental line. We propose that the pattern of divergence of populations reflects the formation of river drainages, which was not complete until <4.1 Ma.
- Our study provides a clear picture of the role of landscape change on the evolution of plants living strictly in rivers in northern South America. By shifting the focus to aquatic taxa, we provide a novel perspective on the processes shaping the evolution of the Neotropical flora.

Key words: Andean uplift, Marathrum, Neotropics, next-generation sequencing, northern South America, phylogenetics, rivers, target enrichment

Introduction

Landscape change is recognized as a major driver of lineage diversification (Hoorn *et al.*, 2013). In the Neotropics, the most species-rich region in the world, the uplift of the Andean cordillera not only contributed directly and indirectly to the assembly of the terrestrial biota (Janzen, 1967; Kattan *et al.*, 2004; Antonelli & Sanmartín, 2011; Sklenář *et al.*, 2011; Smith *et al.*, 2014; Hoorn *et al.*, 2018; Quintero & Jetz, 2018), but also impacted the evolution of riverine organisms by shifting the physical locations of watersheds (Albert *et al.*, 2006, 2020; Albert & Crampton, 2010; Picq *et al.*, 2014; Tagliacollo *et al.*, 2015; Ruokolainen *et al.*, 2019). Andean uplift played a major role in the diversification of terrestrial plants in the Neotropics (Hughes & Eastwood, 2006; Madriñán *et al.*, 2013; Nürk *et al.*, 2013; Lagomarsino *et al.*, 2016; Pérez-Escobar *et al.*, 2017; Richardson *et al.*, 2018; Testo *et al.*, 2019). However, the impacts of Andean uplift and landscape changes on plants in rivers remains unaddressed.

Within the Neotropics, northern South America (nSA) stands out for its heterogeneous terrain, with the Andean cordillera splitting into three ranges, *i.e.,* the Western (WC), Central (CC), and Eastern Cordilleras (EC). Another important topographic element in the region is the Sierra Nevada de Santa Marta (SNSM), which is a separate formation from the Andes and provides the highest coastal relief on Earth (Duque-Caro, 1979; Villagómez *et al.*, 2011; Parra *et al.*, 2019). Multiple rivers originate in the SNSM, with evidence for uplift of this mountain system taking place no earlier than 2 Ma (Villagómez *et al.*, 2011; Parra *et al.*, 2019). Rivers on the northern slope of the SNSM drain to the ocean whereas those in the southern slopes are part of the Magdalena drainage basin (Abell *et al.*, 2008), although they are not connected through fast-flowing water currents to the other tributaries of this watershed. Palynological, stratigraphic, and macrofossil data (Latrubesse *et al.*, 2010; Hoorn *et al.*, 2010) support a landscape evolution model in which from the late Miocene (*ca.* 11Ma) to present, exceptionally rapid uplift of the Andes transformed the aquatic ecosystems in the region by causing a shift from a wetland-dominated environment (*Pebas Lake*) to the formation of current drainage basins, *i.e.,* Pacific, Sierra Nevada de Santa Marta (SNSM), Magdalena, Orinoco, and Amazon.

Given the dynamic nature of the landscape in nSA and the role of the Andes in shaping rivers in the region, here we investigate how drainage basin formation linked to Andean uplift shaped the evolution of river plants. We use as model systems two species of aquatic plants that live strictly in rivers: *Marathrum utile* Tul. and *Marathrum foeniculaceum* Bonpl, which are in the family Podostemaceae (riverweeds). This family is unlike any other angiosperms (except for the Hydrostachyaceae that are restricted to Africa and Madagascar), in that they are herbs living attached to rocks partially or completely submerged in tropical fast-flowing aquatic ecosystems like river-rapids and waterfalls (Fig. **1**; van Royen, 1951; Philbrick & Novelo, 1995).

*Marathrum util*e and *M. foeniculaceum* are distributed from Central America to nSA, and they occur in sympatry in part of their range (Fig. **1**). The two species are the only taxa in the Podostemaceae (ca. 270 species; Novelo *et al.*, 2009), and of plants inhabiting fast-flowing aquatic ecosystems, with allopatric populations currently disconnected from each other by the Andes, the SNSM, inter-Andean valleys, and the lowlands between the Andes and the SNSM.

As is characteristic of the Podostemaceae, *M. utile* and *M. foeniculaceum* are mostly annuals and have flowers with a reduced perianth and capsular fruits that are produced and project above the water surface in the dry season (Fritsch, 1903; Willis, 1915; Philbrick & Novelo, 1998; Koi *et al.*, 2015). Hundreds of seeds are shed prior to the onset of the wet season when the water level is still low, and attach to the rock with a sticky

mucilage (Reyes-Ortega *et al.*, 2009). The adult plants remain vegetative and attached to rocks via adhesive hairs and a biofilm of cyanobacteria (Rutishauser *et al.*, 1999; Jäger-Zürn & Grubert, 2000). These life history characteristics suggest a reduced dispersal ability across the entire family (Philbrick *et al.*, 2010).

Marathrum utile and *M. foeniculaceum* are readily recognizable from one another in that they have distinctive leaf types; *Marathrum foeniculaceum* has highly dissected leaves, and *M. util*e is the only species in the genus with entire, laminar leaves (van Royen, 1951) (Fig. **1**). However, the genus *Marathrum* has a high degree of modification of the vegetative organs, which are thalloid in appearance, (Rutishauser, 1995, 1997; Rutishauser *et al.*, 1999) and a reduced number of reproductive characters useful for classification (Philbrick, pers. comm.). This has resulted in high levels of synonymy in the genus (Novelo & Philbrick, 1997; Novelo *et al.*, 2009; Tippery *et al.*, 2011; Table S1), with 30 species initially described (van Royen, 1951) and only 10 currently recognized.

Marathrum utile and *M. foeniculaceum* are an ideal system for investigating the impact of Andean uplift and drainage basin reconfiguration on the evolution of riverine plants in nSA. In particular, these species are restricted to riverine habitats, they have a broad distribution across the northern Andes, their life history should restrict their dispersal ability, and they are broadly sympatric yet readily differentiated based on phenotype. Focusing on these aquatic taxa provides a novel perspective on the processes that are responsible for the assembly of Neotropical plants and on how evolutionary trajectories are shaped by geology and the natural history of species. We specifically aim to 1) test if the geographical separation of drainage basins interrupted by the Andes, the SNSM, and inter-Andean valleys limits gene flow and structures populations according to river drainages, 2) determine if the timing of divergence events in populations of the two species in nSA correspond with the currently proposed timing of Andean uplift, and 3) provide a hypothesis for the pattern and timing of drainage basin separation. To accomplish these goals, we generate target enrichment and ddRADseq data to infer population structure within species and phylogenetic relationships of populations and species, which could include reticulation.

Materials and Methods

Data collection and processing

We collected 118 individuals; 75 individuals with the phenotype of *M. foeniculaceum,* 40 of *M. utile*, and three of *Weddellina squamulosa* Tul. across the Andes and the Sierra Nevada de Santa Marta (SNSM) (Fig. **2**).

Sampling covered three major localities in three drainage basins: 1) Caribe (northern and south-eastern slopes of the SNSM basin), 2) Antioquia (eastern slope of the Central Cordillera in the Magdalena basin), and 3) Boyacá (eastern slope of the Eastern Cordillera in the Orinoco basin) (Fig. **2**, Table S2). Caribe is a coastal region in Colombia that spans multiple political units (departments), whereas Antioquia and Boyacá each corresponds to a single department where plants were collected. Collecting effort in the three major localities included 21 sites (rivers and waterfalls) and individuals were collected from various sections of the river and different rocks.

Samples were preserved as vouchers and leaf material was stored in silica gel. We extracted DNA from silica-dried leaf material using a modified CTAB protocol (Doyle & Doyle, 1987), purified the DNA extracts with isopropanol precipitation, and assessed DNA concentration and fragmentation levels using a Qubit fluorometer and gel electrophoresis (1% agarose gels).

Double digest Restriction Associated DNA sequencing (ddRADseq)

To explore phylogenetic relationships in *M. utile* and *M. foeniculaceum* in nSA, we prepared ddRADseq libraries for 115 samples following Peterson *et al.*, 2012. In brief, we used *PstI* and *MspI* to double digest *ca.* 500 ng high-molecular weight DNA and purified fragments with Sera-Mag SpeedBeads prior to ligation of barcoded Illumina adapters. Libraries were size selected to 415–515 bp on a Pippin prep (Sage Science) using a modified protocol where pooled samples were each run in a separate lane and size selected in reference to one lane running only the R2 marker reagent. Final library amplification used Illumina indexed primers and HF-Phusion TAQ. We determined library size distributions and concentrations on an Agilent 2200 TapeStation. Samples were sent to the QB3 sequencing facility at UC Berkeley, where sequenceable library concentrations were determined prior to multiplexing equimolar amounts of each library and sequencing 150 single-end reads in two lanes of an Illumina HiSeq 4000.

Following the exclusion of 20 samples that failed sequencing or had low coverage, ddRADseq data for the remaining 95 samples were quality checked with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and demultiplexed and assembled in ipyrad v0.9.42 (Eaton & Overcast, 2020) using a clustering threshold of 0.88. We removed consensus sequences that had low coverage (6 reads), excessive undetermined or heterozygous sites (>12), an excess of shared heterozygosity among samples (paralog filter=0.5), and included loci present in at least four individuals. Preliminary inspection of our resulting assemblies showed that there where high levels of missing data in our

ddRADseq dataset with no loci shared across ≥90% of the samples and only 70 loci shared across half of the samples (Table S3). To increase the number of shared loci recovered and allow for direct comparison of our two genomic datasets, we selected the ddRADseq samples that were successfully sequenced in our target enrichment (TE) dataset (n=34) (details of TE experiment below). Reads from this subset where assembled in ipyrad v0.9.42 (Eaton & Overcast, 2020) with parameters specified as described above, and we generated two assemblies that included loci that were present in at least 17 individuals (*min17*; 50% missing data) and at least 4 individuals (*min4*; ~88% missing data).

Low-depth genome of Marathrum and identification of loci for target enrichment

We collected genome-skimming data for one individual of *Marathrum utile* (Table S2) generated at the QB3 sequencing facility (Berkeley, CA). Total DNA was sheared to a target size of 300–500 bp and 150 paired-end reads were sequenced in an Illumina HiSeq 4000. Genome-skimming reads were loaded in FastQC to assess sequence quality and adapter sequences were removed with Trimmomatic v0.39 (Bolger *et al.*, 2014). *De novo* assembly was performed with the CLC Genomics Workbench 8 (https://digitalinsights.qiagen.com).

To select a set of target sequences that are nuclear, single-copy, and have orthologs across *Marathrum*, we applied the methods in Chau *et al.*, 2018. Briefly, the *de novo* assembled genome of *M. utile*, a transcriptome of *Hypericum perforatum* L. (Hypericaceae) (One Thousand Plant Transcriptomes Initiative, 2019; Carpenter *et al.*, 2019), a plastome of *Garcinia mangostana* L. (Calophyllaceae), and a mitochondriome of *Populus tremula* L. (Salicaceae) (Table S2) were used as input for the pipeline Sondovac (Schmickl et al., 2016). We complemented the target loci identified with Sondovac with the *Arabidopsis-Populus-Vitis-Oryza* shared single-copy genes (APVO SSC) and pentatricopeptide repeat genes (PPR) (downloaded from http://www.arabidopsis.org), which are shared across angiosperms and have been found to be useful for resolution of inter-specific relationships in some plant groups (Yuan et al., 2009, 2010; Duarte et al., 2010; Methods S1). Design of probes targeting a loci space of 214,484 bp for 536 exons in 217 genes with 2x tiling density was performed by RapidGenomics (Gainesville, Florida, US).

Target enrichment sequencing and assembly

Taxon sampling for the TE experiment was guided by a preliminary phylogenetic analysis of all the ddRADseq data using SVDquartets with unlinked SNPs (Fig. S1; Chifman & Kubatko, 2014). Of the 118 samples collected, we selected 46 individuals for TE with the aims of 1) including specimens with high-quality genomic DNA and high coverage across the ddRADseq assembly, 2) sampling multiple individuals per populations in

each drainage basin, and 3) providing equal representation of the clades identified by the ddRADseq data. Sequencing was performed by RapidGenomics (Gainesville, Florida, US). Briefly, 250–1000 ng of high-molecular weight DNA was fragmented to a target size of 400 bp. Enriched pools were combined in equimolar ratios, probes were hybridized to pools to enrich for targets, and 150 paired-end reads were sequenced in 25% of a lane of an Illumina HiSeq 3000.

We quality checked the demultiplexed reads with FastQC, removed adapter sequences with Trimmomatic and assembled loci with HybPiper v. 1.3.1 (Johnson *et al.*, 2016). We assessed paralogs by 1) extracting sequences from alternate contigs for exons with paralog warnings with the scripts 'paralog_investigator.py' and 'paralog_retriever.py', 2) aligning sequences for each exon with paralog warning with MAFFT v7.309 (Katoh & Standley, 2013) and inspecting the alignments, and 3) inferring and examining phylogenetic trees for each exon with paralog warnings (Methods S2; Fig S2). Resulting exon sequences from the same gene were concatenated and gene alignments were created with MAFFT using the automatic alignment strategy in Geneious v9.1.8. (Biomatters Ltd., Auckland, New Zealand), followed by manual inspection. This method of concatenation takes the allele chosen by Hybpiper (Methods S2) for each exon. However, without knowledge of linkage across exons within a gene, this could result in chimeric sequences for multi-exon genes. To test the impact of this strategy on phylogenetic inference, we reran all analyses that use the concatenated exons (SVDquartets, ASTRAL, and ML; described below) using only single-exon genes (71 genes). Assembly of genes with introns was attempted but capture of intronic regions was highly uneven across samples, resulting in poorly aligned sequences with a 45% increase in the amount of missing data in the total data matrix and with over 80% missing data for multiple samples across loci.

Population structure and genetic constitution across drainage basins

To examine population structure we ran ADMIXTURE v1.3 (Alexander *et al.*, 2009) and considered clustering scenarios with up to 6 groups (K=1–6). SNPs used as input were extracted from our TE data with dDocent (Puritz *et al.*, 2014) by mapping the sequenced reads to the reference sequences of our targeted loci, assembling the reads, and calling SNPs (clustering threshold (c)= 0.88, default values for remaining parameters). We further filtered our resulting SNPs to include loci with no missing data, only one SNP per locus (--thin 5,348 which is the maximum gene length), and a minor allele frequency ≥0.05 using VCFtools (Danecek *et al.*, 2011). We generated biallelic SNPs with Plink V1.9.0 (Purcell *et al.*, 2007). Given the high levels of missing data in the ddRADseq loci, we did not use these data for population structure inference. Admixture graphs were plotted with structurePlotter (https://zenodo.org/badge/latestdoi/79741087).

To further characterize the genetic constitution of populations, the overall heterozygosity (% of all SNPs that are heterozygous from the unlinked and filtered SNPs extracted with dDocent) was estimated with VCFtools (--het function). We looked for ploidy variation across individuals by calculating the distribution of allele frequencies at biallelic SNPs with nQuire (Weiß *et al.*, 2018). For each individual, we used the binary alignment maps resulting from read assembly with dDocent and reduced noise (mismappings) by removing frequencies below 0.2 (--denoise) as recommended by Corrêa dos Santos *et al.*, 2017. We inspected the distribution of allele frequencies(--histo), calculated the best-fit between the ideal and empirical histograms (--histotest), and estimated delta likelihood values after comparing experimental data against the maximized log-likelihood of the free model (--lrdmodel) (Viruel *et al.*, 2019).

Phylogenetic inference

Bifurcating relationships

A total of 42 samples were successful in the TE experiment (four *Marathrum* samples failed sequencing or recovered relatively short exons of low quality; Fig. S3). To eliminate the conflict that can arise in phylogenetic reconstruction with the inclusion of admixed individuals (Leaché *et al.*, 2014; McVay *et al.*, 2017; García *et al.*, 2017; Zhang *et al.*, 2019; Ginsberg *et al.*, 2019), we removed samples inferred as admixed with ADMIXTURE. We reassembled the ddRADseq dataset with ipyrad, excluding admixed individuals (17 ddRADseq samples retained) with loci shared across minimum 4 individuals (*min4pures*) to maximize the number of SNPs recovered across quartets of samples.

The recovered genetic units from ADMIXTURE were used as the operational taxonomic units for species tree inference with ASTRAL III (Zhang *et al.*, 2018) and SVDquartets (Chifman & Kubatko, 2014). Target enrichment data was used to infer 50% consensus majority rule unrooted gene trees in MrBayes v3.2.7 (Ronquist & Huelsenbeck, 2003) with the TICR pipeline (Stenz *et al.*, 2015). We by-passed model selection and specified the parameter-rich GTR+G model of substitution. MrBayes was run with 3 runs, 4 chains, 10 million generations, a sample frequency of 200, and a burn-in fraction of 0.25. Convergence of chains was assessed in Tracer (ESS>200 for all default parameters) and the average deviation of split frequencies (<0.05). Gene trees were used as input for ASTRAL III and local posterior probabilities were estimated. We ran SVDquartets analyses in PAUP v4.0a (Swofford, 2003) evaluating all quartets and using 1000 bootstraps on the concatenated alignment of our target capture data (specifying genes as separate partitions) and on the unlinked SNPs of the *min4pures* ddRADseq assembly (Methods S3). The latter dataset maximizes the number of shared loci recovered across quartets of species (Eaton *et al.*, 2016). We also ran both SVDquartets analyses

removing invariant sites and non-biallelic-SNPs for comparison with the divergence-dating analysis with SNAPP (see divergence-dating subsection below).

We explored the effect of inclusion of admixed individuals on phylogenetic inference with our datasets by inferring phylogenetic relationships with the full concatenated TE dataset and our *min4* and *min17* ddRADseq assemblies. The *min4* and *min17* assemblies were used for SVDquartets and Maximum Likelihood (ML) inferences respectively. We ran SVDquartets and ASTRAL III as described above but used individuals-as-tips instead of defining genetic clusters from ADMIXTURE as operational taxonomic units, since admixed individuals cannot be assigned to any of the clusters. Maximum likelihood inference was conducted with RAxML-HPC-HYBRID (Stamatakis, 2014) on the CIPRES Science Gateway. We used the parameter-rich GTRCAT model of substitution following Abadi *et al.*, 2019 and defined each gene as a separate partition in our TE dataset. Branch support was estimated using the automatic bootstrap convergence function (-autoMRE; Methods S3) (Pattengale *et al.*, 2010). To distinguish low support from gene tree conflict, we applied the quartet sampling method (Pease *et al.*, 2018) with 1000 replicates per branch (Methods S3). Trees were rooted using *W. squamulosa*. We report the resulting extended majority rule SVDquartets trees.

Phylogenetic networks

We explored non-bifurcating relationships of admixed individuals in *M.utile* and *M. foeniculaceum* in nSA by inferring phylogenetic networks from gene trees under maximum pseudo-likelihood (InferNetwork_MPL) in PhyloNet v.3.8.0 (Than *et al.*, 2008; Wen *et al.*, 2018). We ran PhyloNet on three datasets that each contained the outgroup (*W. squamulosa*), 2–3 randomly selected samples of *M. utile* and *M. foeniculaceum* from each major locality, and two randomly selected admixed individuals from one major locality (Antioquia, Caribe, or Boyacá). Populations from each drainage basin were used as operational taxonomic units, where admixed individuals from Boyacá are assigned to their own population whereas admixed individuals from Antioquia and Caribe are not. We allowed 0–4 reticulation events and contracted edges in gene trees that have <70 bootstrap support (Methods S3). Input gene trees were inferred under ML with RAxML-NG (Kozlov *et al.*, 2019) using the GTR+G model, 2 threads and 1000 bootstraps (Methods S3). We rooted the gene trees with the outgroup with the ape package in R v4.0.2 (Paradis & Schliep, 2019).

The best resulting phylogenetic network was chosen by computing the log-likelihood of phylogenetic networks given the input gene trees (Yu *et al.*, 2012) with the function CalGTProb in PhyloNet, contracting edges in gene trees with <70 bootstrap support. Model selection was performed using the Akaike information criterion

with number of parameters set as the number of branch lengths plus the number of hybridization probabilities being estimated (Morales-Briones *et al.*, 2018).

Divergence times for populations across drainage basins

To determine the time-frame and context in which populations of *M. utile* and *M. foeniculaceum* in nSA diverged across drainage basins and contrast the timing of species and population divergence with the currently proposed timing of recent Andean uplift, we inferred a time-calibrated species tree with SNAPP (Bryant *et al.*, 2012) as implemented in BEAST 2.6.3 (Bouckaert *et al.*, 2014). The TE matrix with only pure individuals was filtered to exclude invariant and non-binary sites with *Phrynomics* in R (Leaché *et al.*, 2015; https://github.com/bbanbury/phrynomics). A secondary calibration was defined for the divergence time of *W. squamulosa* (monotypic genus and species in the subfamily Weddellinoideae) from the subfamily Podostemoideae (a clade sister to *Weddellina*, which includes *Marathrum* and the remaining genera of Podostemaceae except for 6 taxa sister to Weddellinoideae-Podostemoideae; Koi *et al.*, 2012; Ruhfel *et al.*, 2016). A normal prior distribution with a mean of 62 Ma and a standard deviation of 4 Ma was specified. This distribution encompasses the estimates for this node reported in previous published work by Magallón *et al.*, 2015 and Ruhfel *et al.*, 2016 (Table 1) who used multiple fossil age-constraint priors outside Podostemaceae, including suggested alternate placements for them (Ruhfel *et al.*, 2013). Input file for SNAPP was prepared with snapp_prep.rb (https://github.com/mmatschiner/snapp_prep) which uses a combination of priors and operators optimized for divergence-dating with SNAPP (Stange *et al.*, 2018). We ran two chains for 300,000 generations, sampling every 50, and assessed stationarity and mixing of all parameters as well as convergence of both runs with Tracer v1.7.1 (Rambaut *et al.*, 2018). We discarded a 10% burnin fraction, combined the log files and posterior distribution of trees from the two runs with LogCombiner, and summarized the posterior distribution maximum clade credibility (MCC) tree in TreeAnnotator.

Results

Data processing: ddRADseq, genome assembly, and target enrichment

A total of 169,474 loci that are shared in minimum four individuals were recovered across all 95 *Marathrum* and *Weddellina* samples, but only 70 loci were recovered at ~50% missing data) (Table S3). The *min4* (~88% missing data) and *min17* (50% missing data) assemblies of our subset of 34 samples including admixed individuals resulted in 28,023 and 679 recovered loci respectively and the *min4pures* assembly of the subset of 17 non-admixed samples included 11,366 loci (Table S4), 9,907 unlinked SNPs and 3,996 invariant, non-

binary, unlinked SNPs. Although there are many loci recovered in our dataset, a smaller proportion of those are present across most samples, with no loci shared in all of them.

De novo assembly of genome-skimming reads resulted in a draft assembled genome of *M. utile* consisting of 627,733 contigs with a total length of 1.095 Gb and N50=11,454 bp (Bioproject PRJNA673497, SRR12956182). Figure S3 and Table S5 show the locus recovery efficiency and the number of sequences and loci recovered after paralog filtering per sample respectively. Recovery of assembled sequences was lower for the outgroup taxon (*W. squamulosa*) compared to the ingroup. Thirty-eight exons that failed sequencing across ≥50% of the samples or are potential paralogs were removed (Table S6). Most paralog warnings corresponded to allelic variation (Methods S2). After paralog filtering, we recovered 208 genes across samples with an average of 507 exons in 205 genes. The resulting matrix had 201,515 bp, 11,437 SNPs, and an average gene size of 973 bp. This matrix was reduced to 200,946 bp and 8,076 SNPs (11,063 bp and 7,467 SNPs after removal of invariant and non-binary sites) with the exclusion of admixed individuals.

Population structure across drainage basins

Read assembly and SNP calling in dDocent recovered 203 unlinked, biallelic SNPs (11,506 SNPs before selecting only one SNP per locus with minimum allele frequency ≥0.05). Parametric population structure estimation with ADMIXTURE resulted in a lower cross-validation error for *K*=4 and K=5 populations (0.08310 and 0.09195, respectively; Figs. **3**, S4). The best model shows that within each Antioquia and Caribe there is strong population structure with two genetically distinct, non-admixed clusters as well as multiple individuals with admixture proportions from the two clusters. The two distinct, non-admixed genetic clusters correspond to the phenotypes of *M. foeniculaceum* and *M. utile*. All admixed individuals have about equal proportions of ancestry of the non-admixed clusters and their phenotype corresponds to that of *M. foeniculaceum,* with highly dissected leaves (Figs. **1**, **3**). There is no admixture detected in individuals of the same species across drainage basins. The genetic constitution of the two individuals that represent the population from Boyacá was inferred to include relatively equal proportions from *M. utile* from Caribe and *M. foeniculaceum* from Antioquia suggesting a possible hybrid origin for these samples. The only difference between K=4 and K=5 is that the latter detects population substructure in *M. utile* from Antioquia.

Estimates of overall heterozygosity of individuals showed remarkably high inbreeding coefficients in non-admixed individuals (0.93 on average), whereas admixed individuals all have an excess heterozygosity (inbreeding coefficients =-0.73 on average; Table S7). Ploidy estimation from target enrichment (TE) data with

nQuire indicated that non-admixed individuals are polyploids (at least tetraploids), whereas admixed individuals are diploid (Table S8). However, violations to the assumptions of nQuire that prevent estimating the ploidy of individuals in this study are further discussed.

Phylogenetic inference

Bifurcating relationships

Results from phylogenetic analyses excluding admixed individuals are shown in Figs. **4,** S5. The bipartitions recovered in the unrooted ddRADseq species trees are compatible with the rooted species trees inferred from TE data. Rooted species trees recover *M. foeniculaceum* as monophyletic and *M. utile* as either paraphyletic (Fig. **4a,b**) or monophyletic (Fig. S5).

As expected, the inclusion of admixed individuals affects phylogenetic inference and causes admixed individuals to fall into various places on phylogenies reconstructed with different datasets and algorithms, capturing non-bifurcating patterns resulting from gene flow. The inferred topologies using TE data recovered the four non-admixed genetic clusters from ADMIXTURE (Fig. **3**) as strongly supported clades nested in more inclusive clades with admixed individuals, but *M. utile* from Caribe in the ML tree is recovered as paraphyletic (Figs. S6, S7, S8). Tree inference with ASTRAL III and ML results in the clustering of populations of *M. utile* and *M foeniculaceum* and admixed individuals from Antioquia (Figs. S7, S8). However, quarter sampling scores in the ML tree (Fig. S8, top) indicate that this clade has a supported secondary evolutionary relationship with *M. utile* from Antioquia and Caribe being more closely related to one another than to *M. foeniculaceum* in nSA (Table S9). The inferred topologies with ddRADseq data show that the bipartitions including *M. utile* and *M. foeniculaceum* from Antioquia are compatible with the TE rooted trees inferred under the same algorithm (Figs. S6, S8) but the bipartitions for the two species from Caribe differ from the TE rooted trees. QC scores for the ML tree with ddRADseq are low (Fig. S8), with possible presence of an alternate topology at most branches. A strong majority of quartets support an alternate placement of the admixed individual from Boyacá in the ML tree of ddRADseq data (it being more closely related to *M. foeniculaceum* from Caribe than to *Marathrum* from Antioquia; Fig. S8, Table S10). The placement of this clade varies across the SVDquartets, ASTRAL III, and ML inferences and the ddRADseq and TE datasets (Figs. S6, S7, S8). Long branches subtending *Marathrum* and *M. foeniculaceum* in nSA are evident in the TE and ddRADseq ML trees (Fig. S8).

Phylogenetic inference with SVDquartets, ASTRAL, and ML with single-exon genes (Figs. S9, S10, S11) shows that relationships inferred across species and non-admixed populations are the same to concatenating

alleles across exons without prior knowledge of their linkage (Figs. **4**, S5, S6, S7, and S8). The placement of admixed individuals from all drainage basins differs across trees inferred with concatenated exons and with single exons, capturing non-bifurcating patterns resulting from gene flow (*e.g.,* as shown in the inferred phylogenetic networks for Boyacá in Fig. **4**).

Phylogenetic networks

Networks with 2 reticulation events across different terminal taxa were chosen as the best models in all three subsets, each including admixed individuals from either Antioquia, Boyacá, or Caribe (Table 2, Fig. **4c**). All chosen networks infer a reticulation event in the branch that subtends the admixed individuals in agreement with our ADMIXTURE results (Fig. **3**). A second reticulation event across terminal taxa is inferred deeper in the phylogeny with *M. foeniculaceum* in nSA forming a clade sister to *M. utile* or originating from it. This is consistent with the recovery of *M. utile* in nSA as paraphyletic with SVDquartets and ASTRAL III including invariant and non-binary sites (Fig. **4**) and as monophyletic with SVDquartets excluding invariant and non-biallelic sites (Fig. S5). Network inference with single-exon genes (Fig. S9c, Table S11) recovered the same non-bifurcating relationships but some inheritance probabilities differed, particularly for the earliest reticulation event detected across subsets.

Divergence times for populations across drainage basins

The SNAPP calibrated phylogeny of *Marathrum* in nSA using a matrix of 11,063 biallelic and variant sites recovered species as clades (Fig. **5**), in agreement with the SVDquartets inference of biallelic SNPs (Fig. S5). Our estimates suggest that *M. utile* and *M. foeniculaceum* diverged at ~17 Ma. Populations of *M. foeniculaceum* in Antioquia and Caribe split at ~12 Ma, whereas *M. utile* from those two same drainage basins split more recently at ~4 Ma (Table S12).

Discussion

Landscape change in northern South America (nSA) through the Miocene and Pliocene shaped the evolution of populations of *M. utile* and *M. foeniculaceum* in the region. During this time period, distinct populations of the two species diversified into different drainage basins currently interrupted by the Andes, the Sierra Nevada de Santa Marta (SNSM), and inter-Andean valleys (Figs. **3, 5**). We show that population divergence occurred in conjunction with currently suggested major recent pulses of Andean uplift (Fig. **5**). River drainage reconfiguration caused the isolation of populations across the Magdalena, SNSM, and Orinoco basins, while

also bringing species in sympatry. Analyses of nuclear data showed high levels of admixture in areas of sympatry, and a bias for admixed individuals to have the phenotype of *M. foeniculaceum*. Whether these evolutionary patterns and processes resulting from landscape change apply to *Marathrum* as a whole, remains an open question. Below, we propose that the phylogenetic relationships of populations across drainage basins may reflect the history of drainage basin separation and discuss our results in the context of the systematics of the two *Marathrum* species studied.

Andean uplift, river basin formation and timing of divergence of populations of *Marathrum* in nSA

The divergence of *Marathrum foeniculaceum* and *M. utile* in nSA (~17 Ma) took place prior to the separation of the westernmost Andes from western Amazonia and the Orinoco (Fig. **5**, Table S12; Hoorn *et al.*, 2010; Montes *et al.*, 2021). Divergence of population of the two species in nSA track events of major uplift in the region. *Marathrum foeniculaceum* and *M. utile* split into populations in Antioquia and Caribe around ~ 12 Ma and ~4 Ma respectively (Fig. **5,** Table S12), in conjunction with previously proposed times for rapid Andean uplift during the late Miocene and Pliocene at *ca.* 12–6 Ma and *ca.* 4.5 Ma (Gregory-Woodzicki, 2000; Garzione *et al.*, 2008; Mora *et al.*, 2010; Hoorn *et al.*, 2010; Anderson *et al.*, 2016). Landscape changes after the second pulse of uplift may have brought divergent populations of the two species in sympatry, facilitating admixture events (Figs. **3, 4c, 5**). The short terminal branches characteristic of *M. utile* and *M. foeniculaceum* from various rivers in Caribe (Fig. S8) suggest their recent divergence in the SNSM. Indeed, surface uplift that gave rise to the current topology of the SNSM was estimated no earlier than 2 Ma (Villagómez *et al.*, 2011; Parra *et al.*, 2019).

We have shown that the population from Boyacá originated from a reticulation event between *M. utile* from Caribe and *M. foeniculaceum* from Antioquia (Figs. **3**, **4c**). This must have occurred ≤4.1 Ma, when the latter two became genetically distinct. (Fig. **5**, Table S12). A lowland inter-Andean fluvial portal that existed from 13–4 Ma (Lundberg & Chernoff, 1992; Montes *et al.*, 2021) or bird-mediated dispersal could have facilitated transport of seeds and roots of the parental lineages or of admixed individuals of *Marathrum* to Boyacá.

The pattern and timing of divergence of populations of *Marathrum* in nSA provides a hypothesis for the order and timing of separation of fast-flowing riverine ecosystems in watersheds across the Andes. However, limited sampling of Boyacá (Orinoco), the possible establishment of this population through long-distance dispersal, and lack of sampling of the Pacific drainage basin do not permit us to address all such events. Fast-flowing rivers in the SNSM finally became disconnected from those in Magdalena at ~4.1 Ma with a prior splitting

event ~12 Ma, possibly followed by river capture events (Fig. **5**). Fast-flowing water ecosystems in the Orinoco could have disconnected from those in the Magdalena and SNSM drainage basins with the closure of the trans-Andean fluvial portal ~4 Ma, but further sampling is needed to support this hypothesis.

The divergence times presented here are a means of exploring plausible evolutionary scenarios, but they should be taken cautiously as they rely on a secondary calibration and are subject to potential bias (Schenk, 2016). No fossils of Podostemaceae have been found to date, but phytoliths for the family have been described and found to be taxonomically informative (da Costa *et al.*, 2018). Future studies aiming to re-examine the divergence dates provided here should rely on fossil calibrations in the Clusioid clade (including somewhat distantly related taxa) or undertake the endeavor of finding phytoliths in dated strata. Also, gene flow can result in underestimates of species divergence times, particularly for nodes that involve a reticulation event (Leaché *et al.*, 2014). It thus may be that *M. foeniculaceum* and *M. utile* in nSA first split earlier than estimated here (ca. 17 Ma; Fig. **5**, Table S12) but that gene flow, as suggested by the one deep reticulation event inferred (Fig. **4c**), caused the estimation of more recent ages for this first split.

Incongruence in phylogenetic reconstruction across datasets

Why is *M. utile* recovered as a clade in species tree inference of biallelic SNPs (Figs. **5**, S5), but as paraphyletic when including all variable and invariant sites (Fig. **4a,b**)? Reticulation may be responsible for the conflict. Separate PhyloNet analyses that included admixed samples from Antioquia, Boyacá, and Caribe each support reticulation between monophyletic species (Fig. **4c**). Phylogenetic analyses that do not allow reticulation assume that relationships are bifurcating, and this can cause phylogenetic errors when gene flow is present (Leaché *et al.*, 2014). The phylogenetic analyses that did not support the monophyly of *M. utile* are two-step (ASTRAL) and coalescent aware concatenation (SVDquartets) summary methods that do not use all of the information in the data or suffer from information loss due to averaging over the whole genome, and this reduces their accuracy compared to full likelihood methods under the species coalescent (Zhu & Yang, 2021). SNAPP also suffers from data reduction problems. Namely, the exclusion of invariant sites results in acquisition bias, which may incorporate errors in phylogenetic inference (Lewis, 2001; Bertels *et al.*, 2014; Leaché *et al.*, 2015a). Divergence dating with SNAPP includes an ascertainment-bias correction for reliable estimation of phylogeny, node-ages, and population size (Stange *et al.*, 2018). However, this acquisition bias correction failed to recover the same topology as was inferred when invariant and non-binary sites were included in species tree inference (Figs. **4a,b, 5,** S5).

The differences across target enrichment and ddRADseq datasets when including admixed individuals (mainly the placement of admixed individuals from Boyacá; Figs. S6, S8) may be explained by missing data, short internal branches, and bioinformatic steps used for the assembly of loci (e.g. clustering threshold; Leaché *et al.*, 2015). These phenomena are characteristic of our ddRADseq data and may also be responsible for the low QS scores in the ML analysis (Fig. S8). The high levels of missing data in our ddRADseq data (Table S4) could be a result of previously suggested high rates of evolution in Podostemaceae (Bedoya *et al.*, 2019), which reduce the number of shared loci due to allele dropout (Wagner *et al.*, 2013; Arnold *et al.*, 2013; Eaton *et al.*, 2016). The long branch subtending *M. foeniculaceum* (Figs. S7, S8) may be an indicator of rapid evolution in this group, even as *M. utile* and *M.foeniculaceum* in nSA diverged relatively recently in geologic time (Fig. **5**). High levels of missing data resulting from low quality DNA, low sequencing depth, or poor read quality are unlikely in our dataset given that DNA quality inspection in agarose-gels showed no sign of considerable DNA degradation and read quality and sequencing depth were good after inspection with FastQC and ipyrad. The differences between the ML and species tree analyses including admixed individuals would most likely be due to incomplete lineage sorting and gene flow (Leaché *et al.*, 2014).

Limited gene flow across the Andes, dispersal, and admixture in *Marathrum* in nSA

The strong population structure consistent with geography and excess homozygosity (Table S7) found in non-admixed *M. foeniculaceum* and *M. utile* in nSA suggests a history of population isolation with little to no gene flow across drainage basins (Fig. **3**). This may be explained by 1) large effect of genetic drift in populations of the two species, which are small, discontinuous and with short generation times (Willis, 1914; van Royen, 1951; Rutishauser, 1995; Cook, 1996; Philbrick *et al.*, 2010; Koi *et al.*, 2015; Cheek & Lebbie, 2018). 2) Watershed boundaries (mountains and lowlands) forming barriers to gene flow as seen in fish (Albert & Crampton, 2010; Lujan & Armbruster, 2011; Musilová *et al.*, 2013; Picq *et al.*, 2014; Roxo *et al.*, 2014; Tagliacollo *et al.*, 2015; Lujan *et al.*, 2019; Albert *et al.*, 2020). 3) Limited pollen and seed dispersal ability in Podostemaceae resulting in inbreeding or selfing.

High levels of differentiation among populations of Podostemaceae are linked to their limited dispersal ability (Baggio *et al.*, 2013; Katayama *et al.*, 2016). However, long-distance dispersal has been invoked as a possible explanation for the pantropical distribution of some groups in the family (Kita & Kato, 2004; Koi *et al.*, 2015; Ruhfel *et al.*, 2016). Birds may be vectors for seed dispersal if seeds get attached to their feet via a sticky mucilage in the seed coat (Willis, 1902; Philbrick & Novelo, 1995, 1997; Leleeka *et al.*, 2016) and root

fragments were found to re-attach to rocks, therefore providing a possible means for vegetative dispersal following the flow of rivers (Philbrick *et al.*, 2015). How seed and pollen dispersal is facilitated or hindered in *Marathrum* within and across rivers remains to be addressed, but our results confirm that long-distance dispersal events are rare.

PhyloNet results support the hypothesis that populations from each drainage basin result from an ancient reticulation event (Fig. **4c**). This pattern is replicated between river basins, where admixed individuals have similar ancestry proportions and inheritance probabilities from *M. utile* and *M. foeniculaceum* occurring in sympatry (Figs. **3, 4c**). This, together with the results that non-admixed individuals have an excess homozygosity (whereas admixed individuals have an excess heterozygosity; Table S7), is consistent with secondary contact of previously isolated populations. The dispersal strategies mentioned above for Podostemaceae together with multiple river capture events could be responsible for the co-occurrence and subsequent admixture of *M. foeniculaceum* and *M. utile* in some rivers across the Andes in nSA.

The uniformity in admixture proportions across admixed individuals (Fig. **3**) could stem from either very recent and frequent admixture events (*i.e.*, F1 hybrids with no further mixing) or a past hybridization event followed by genome duplication resulting in allopolyploids. Ploidy of our samples could not be estimated from sequence data alone, because the ratios at which alleles are assumed to occur at different ploidy levels by nQuire hold true only for hybrids from heterozygous genome sources (Weiß *et al.*, 2018). Non-admixed individuals here are highly homozygous, causing hybrids to be classified as diploids with nQuire (Table S8). The dual ancestry of the two individuals from Boyacá stems from *M. utile* from Caribe and *M. foeniculaceum* from Antioquia in more or less equal proportions (Fig. **3**) but our limited sampling in this drainage basins revealed no non-admixed individuals of either species.

Our results present an interesting case of inheritance of entire vs. highly dissected leaves, with admixed individuals (putative hybrids) having highly dissected leaves, the phenotype of one parental species, *M. foeniculaceum* (Figs. **1**, **3**). Common garden experiments should be conducted to further explore the genetic basis of leaf morphology in *M. utile* and *M. foeniculaceum*

Systematics of *M. utile* and *M. foeniculaceum*

Marathrum foeniculaceum and *M. utile* in nSA are recovered as traditional taxonomic units, either being an ancestor-descendant species pair (Fig. **4,a,b**) or both monophyletic (Figs. **5**, S5). However, the recognition of

the two species as valid taxonomic units requires assessing phylogenetic relationships across their entire distribution. The high levels of admixture and reticulation events reported here, and the potential hybrid origin of individuals from Boyacá, shed light on a more complex reality to delimitation of species boundaries in *Marathrum* than could have been envisioned in the original monograph of the group (van Royen, 1951). Delimitation of *M. foeniculaceum*, excluding those admixed individuals of putative hybrid origin, is not possible by morphology alone. The long branch subtending *Marathrum* from nSA (Figs. **5**, S8) is likely a product of sparse taxon sampling since *Weddellina* is sister to subfamily Podostemoideae, where >40 genera are placed including *Marathrum*.

Conclusions

This study constitutes a significant advance in our understanding of the impact of landscape change in the evolution of plants in rivers in northern South America and provides a comprehensive genomic dataset for the only plant species that inhabit fast-flowing aquatic ecosystems across mountains in the region. Our study highlights the interrelatedness of geology and biology by demonstrating how tightly constrained the evolution of *M. utile* and *M. foeniculaceum* is to river reconfiguration in northern South America. We conclude that the fate of rivers and of these plants is linked and that this relationship has impacted the genetic constitution of populations and species across watersheds. Whether this applies to the entire genus *Marathrum* remains to be investigated. Future research avenues include studying the process of formation and maintenance of genetically and phenotypically distinct units, the impact of high levels of admixture in the genetic make-up of populations, the biogeographic history of *M. utile* and *M. foeniculaceum* across their entire range, and the history of the maternally inherited plastome in this interesting group of aquatic plants.

Acknowledgements

We thank two anonymous reviewers for their insightful and careful comments that helped improve earlier versions of this manuscript. We also thank to C. Thomas Philbrick for his helpful and thorough insights into the biology and classification of Podostemaceae, access to voucher specimens, and help with determination of samples. Special thanks to Santiago Madriñán at the Universidad de los Andes and the Jardín Botánico de Cartagena “Guillermo Piñeres”, Turbaco, Colombia for facilitating field work logistics and providing facilities for specimen processing in Colombia. We thank Caroline Strömberg and Camila Martínez for comments on the manuscript, Rodrigo Bernal and Saul Hoyos for information on conducting field work in Antioquia, and Camilo Montes for insightful conversations on the geological background of nSA. Viviana Londoño provided a

sample of *M. foeniculaceum* and Maria Paula Contreras, David Ocampo, Daniel Ocampo, Jules Dominé, Fernando Bedoya, and Adrián Pinzón provided assistance in the field. Voucher specimens are stored in ANDES and deposited at the UW Herbarium. Samples were collected under the MARCO collecting permit Res. No. 1177 October 9, 2014 of the Universidad de los Andes, export permit No. 02248. Funding was provided by the Botanical Society of America and American Society of Plant Taxonomists Research Awards, Explorers Club Exploration Fund Grant, Department of Biology at the University of Washington Giles Award, WRF Hall International Endowed Fellowship, Denton writing Fellowship, UW Herbarium Endowment, UW Graduate School Boeing International Fellowship, and the Colciencias fellowship for Graduate studies (Doctorados en el Exterior-679) to A.M.B. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant.

Author Contribution

A.M.B. designed the study advised by R.G.O. A.M.B. collected and analyzed the data. R.G.O. and A.D.L. provided guidance on the methods for data analyses and interpretation of results, as well as reagents, equipment, and laboratory space. A.M.B. wrote the manuscript with significant contributions from R.G.O and A.D.L.

Data Availability

All raw sequence reads from the low-depth genome, target enrichment experiment, and ddRADseq datasets are deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive under Bioproject PRJNA673497 (Table S2). Fasta sequences of targeted loci, input files and trees resulted from phylogenetic inference, population structure, phylogenetic networks, and divergence dating are available at https://figshare.com/collections/Andean_uplift_drainage_basin_formation_and_the_evolution_of_plants_living_in_fast-flowing_aquatic_ecosystems_in_northern_South_America-_Data_and_results/5280572.

References

Abadi S, Azouri D, Pupko T, Mayrose I. **2019**. Model selection may not be a mandatory step for phylogeny reconstruction. *Nature Communications* **10**: 934.

Abell R, Thieme ML, Revenga C, Bryer M, Kottelat M, Bogutskaya N, Coad B, Mandrak N, Balderas SC, Bussing W, *et al.* 2008. Freshwater Ecoregions of the World: A New Map of Biogeographic Units for Freshwater Biodiversity Conservation. *BioScience* **58**: 403–414.

Albert JS, Crampton WGR. **2010**. The geography and ecology of diversification in Neotropical freshwaters. *Nature Education Knowledge* **3**: 13.

Albert JS, Lovejoy NR, Crampton WGR. **2006**. Miocene tectonism and the separation of cis- and trans-Andean river basins: Evidence from Neotropical fishes. *Journal of South American Earth Sciences* **21**: 14–27.

Albert JS, Tagliacollo VA, Dagosta F. **2020**. Diversification of Neotropical Freshwater Fishes. *Annual Review of Ecology, Evolution, and Systematics* **51**: annurev-ecolsys-011620-031032.

Alexander DH, Novembre J, Lange K. **2009**. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**: 1655–1664.

Anderson VJ, Horton BK, Saylor JE, Mora A, Tesón E, Breecker DO, Ketcham RA. **2016**. Andean topographic growth and basement uplift in southern Colombia: Implications for the evolution of the Magdalena, Orinoco, and Amazon river systems. *Geosphere* **12**: 1235–1256.

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

Antonelli A, Sanmartín I. **2011**. Why are there so many plant species in the Neotropics? *TAXON* **60**: 403–414.

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K. **2013**. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology* **22**: 3179–3190.

Baggio RA, Firkowski CR, Boeger MRT, Boeger WA. **2013**. Differentiation within and between river basins of *Podostemum irgangii* (Podostemaceae), a rapid-water macrophyte of southern Brazil. *Aquatic Botany* **107**: 33–38.

Bedoya AM, Ruhfel BR, Philbrick CT, Madriñán S, Bove CP, Mesterházy A, Olmstead RG. **2019**. Plastid Genomes of Five Species of Riverweeds (Podostemaceae): Structural Organization and Comparative Analysis in Malpighiales. *Frontiers in Plant Science* **10**: 1035.

Bertels F, Silander OK, Pachkov M, Rainey PB, van Nimwegen E. **2014**. Automated Reconstruction of Whole-Genome Phylogenies from Short-Sequence Reads. *Molecular Biology and Evolution* **31**: 1077–1088.

Bolger AM, Lohse M, Usadel B. **2014**. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. **2014**. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology* **10**: e1003537.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. **2012**. Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis. *Molecular Biology and Evolution* **29**: 1917–1932.

Carpenter EJ, Matasci N, Ayyampalayam S, Wu S, Sun J, Yu J, Jimenez Vieira FR, Bowler C, Dorrell RG, Gitzendanner MA, *et al.* 2019. Access to RNA-sequencing data from 1,173 plant species: The 1000 Plant transcriptomes initiative (1KP). *GigaScience* **8**: giz126.

Chau JH, Rahfeldt WA, Olmstead RG. **2018**. Comparison of taxon-specific versus general locus sets for targeted sequence capture in plant phylogenomics. *Applications in Plant Sciences* **6**: e1032.

Cheek M, Lebbie A. **2018**. *Lebbiea* (Podostemaceae-Podostemoideae), a new, nearly extinct genus with foliose tepals, in Sierra Leone. *PLOS ONE* **13**: e0203603.

Chifman J, Kubatko L. **2014**. Quartet Inference from SNP Data Under the Coalescent Model. *Bioinformatics* **30**: 3317–3324.

Cook CDK. **1996**. *Aquatic plant book*. The Hague, Netherlands: SPB Academic.

Corrêa dos Santos RA, Goldman GH, Riaño-Pachón DM. **2017**. ploidyNGS: visually exploring ploidy with Next Generation Sequencing data. *Bioinformatics* **33**: 2575–2576.

da Costa FGCM, Klein DE, Philbrick CT, Bove CP. **2018**. Silica bodies in leaves of neotropical Podostemaceae: taxonomic and phylogenetic perspectives. *Annals of Botany* **122**: 1187–1201.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, *et al.* 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

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

Duarte JM, Wall PK, Edger PP, Landherr LL, Ma H, Pires JC, Leebens-Mack J, dePamphilis CW. **2010**. Identification of shared single copy nuclear genes in *Arabidopsis, Populus, Vitis and Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology* **10**: 61.

Duque-Caro H. **1979**. Major structural elements and evolution of northwestern Colombia. In: Geological and geophysical investigations of continental margins: AAPG Memoir **29:** 329–351.

Eaton DAR, Overcast I. **2020**. ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* **36**: 2592–2594.

Eaton DAR, Spriggs EL, Park B, Donoghue MJ. **2016**. Misconceptions on Missing Data in RAD-seq Phylogenetics with a Deep-scale Example from Flowering Plants. *Systematic Biology* **66**: 399–412.

García N, Folk RA, Meerow AW, Chamala S, Gitzendanner MA, Oliveira RS de, Soltis DE, Soltis PS. **2017**. Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae). *Molecular Phylogenetics and Evolution* **111**: 231–247.

Garzione CN, Hoke GD, Libarkin JC, Withers S, MacFadden B, Eiler J, Ghosh P, Mulch A. **2008**. Rise of the Andes. *Science* **320**: 1304–1307.

Ginsberg PS, Humphreys DP, Dyer KA. **2019**. Ongoing hybridization obscures phylogenetic relationships in the *Drosophila subquinaria* species complex. *Journal of Evolutionary Biology* **32**: 1093–1105.

Gregory-Woodzicki KM. **2000**. Uplift history of the Central and Northern Andes: A Review. *GSA Bulletin* **112**: 1091–1105.

Hoorn C, Mosbrugger V, Mulch A, Antonelli A. **2013**. Biodiversity from mountain building. *Nature Geoscience* **6**: 154–154.

Hoorn C, Perrigo A, Antonelli A. **2018**. *Mountains, climate and biodiversity*. Hoboken: Wiley.

Hoorn C, Wesselingh FP, ter Steege H, Bermudez MA, Mora A, Sevink J, Sanmartin I, Sanchez-Meseguer A, Anderson CL, Figueiredo JP, *et al.* 2010. Amazonia Through Time: Andean Uplift, Climate Change, Landscape Evolution, and Biodiversity. *Science* **330**: 927–931.

Hughes C, Eastwood R. **2006**. Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences* **103**: 10334–10339.

Jäger-Zürn I, Grubert M. **2000**. Podostemaceae Depend on Sticky Biofilms with Respect to Attachment to Rocks in Waterfalls. *International Journal of Plant Sciences* **161**: 599–607.

Janzen DH. **1967**. Why Mountain Passes are Higher in the Tropics. *The American Naturalist* **101**: 233–249.

Johnson MG, Gardner EM, Liu Y, Medina R, Goffinet B, Shaw AJ, Zerega NJC, Wickett NJ. **2016**. HybPiper: Extracting Coding Sequence and Introns for Phylogenetics from High-Throughput Sequencing Reads Using Target Enrichment. *Applications in Plant Sciences* **4**: 1600016.

Katayama N, Kato M, Imaichi R. **2016**. Habitat specificity enhances genetic differentiation in two species of aquatic Podostemaceae in Japan. *American Journal of Botany* **103**: 317–324.

Katoh K, Standley DM. **2013**. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**: 772–780.

Kattan G, Franco P, Rojas V, Morales G. **2004**. Biological diversification in a complex region: a spatial analysis of faunistic diversity and biogeography of the Andes of Colombia. *Journal of Biogeography* **31**: 1829–1839.

Kita Y, Kato M. **2004**. Phylogenetic relationships between disjunctly occurring groups of *Tristicha trifaria* (Podostemaceae): Phylogeny of *Tristicha trifaria*. *Journal of Biogeography* **31**: 1605–1612.

Koi S, Ikeda H, Rutishauser R, Kato M. **2015**. Historical biogeography of river-weeds (Podostemaceae). *Aquatic Botany* **127**: 62–69.

Koi S, Kita Y, Hirayama Y, Rutishauser R, Huber KA, Kato M. **2012**. Molecular phylogenetic analysis of Podostemaceae: implications for taxonomy of major groups: molecular phylogenetic analysis of Podostemaceae. *Botanical Journal of the Linnean Society* **169**: 461–492.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. **2019**. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**: 4453–4455.

Lagomarsino LP, Condamine FL, Antonelli A, Mulch A, Davis CC. **2016**. The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytologist* **210**: 1430–1442.

Latrubesse EM, Cozzuol M, da Silva-Caminha SAF, Rigsby CA, Absy ML, Jaramillo C. **2010**. The Late Miocene paleogeography of the Amazon Basin and the evolution of the Amazon River system. *Earth-Science Reviews* **99**: 99–124.

Leaché AD, Banbury BL, Felsenstein J, de Oca A nieto-Montes, Stamatakis A. **2015a**. Short Tree, Long Tree, Right Tree, Wrong Tree: New Acquisition Bias Corrections for Inferring SNP Phylogenies. *Systematic Biology* **64**: 1032–1047.

Leaché AD, Chavez AS, Jones LN, Grummer JA, Gottscho AD, Linkem CW. **2015b**. Phylogenomics of Phrynosomatid Lizards: Conflicting Signals from Sequence Capture versus Restriction Site Associated DNA Sequencing. *Genome Biology and Evolution* **7**: 706–719.

Leaché AD, Harris RB, Rannala B, Yang Z. **2014**. The Influence of Gene Flow on Species Tree Estimation: A Simulation Study. *Systematic Biology* **63**: 17–30.

Leleeka M, Sanavar D, Tandon R, Uniyal PL. **2016**. Features of seeds of Podostemaceae and their survival strategy in freshwater ecosystems. *Rheedea* **26**: 29–36.

Lewis PO. **2001**. A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data. *Systematic Biology* **50**: 913–925.

Lujan NK, Armbruster JW. **2011**. The Guiana Shield. In: Albert J, Reis R, eds. Historical Biogeography of Neotropical Freshwater Fishes. University of California Press, Berkeley., 211–224.

Lujan NK, Armbruster JW, Werneke DC, Teixeira TF, Lovejoy NR. **2020**. Phylogeny and biogeography of the Brazilian–Guiana Shield endemic Corymbophanes clade of armoured catfishes (Loricariidae). *Zoological Journal of the Linnean Society* **188**: 1213–1235.

Lundberg JG, Chernoff B. **1992**. A Miocene Fossil of the Amazonian Fish Arapaima (Teleostei, Arapaimidae) from the Magdalena River Region of Colombia--Biogeographic and Evolutionary Implications. *Biotropica* **24**: 2.

Madriñán S, Cortés AJ, Richardson JE. **2013**. Páramo is the world's fastest evolving and coolest biodiversity hotspot. *Frontiers in Genetics* **4:** 192.

Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, Hernández-Hernández T. **2015**. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* **207**: 437–453.

McVay JD, Hipp AL, Manos PS. **2017**. A genetic legacy of introgression confounds phylogeny and biogeography in oaks. *Proceedings of the Royal Society B* **284:** 20170300.

Montes C, Silva CA, Bayona GA, Villamil R, Stiles E, Rodriguez-Corcho AF, Beltran-Triviño A, Lamus F, Muñoz-Granados MD, Pérez-Angel LC, *et al.* 2021. A Middle to Late Miocene Trans-Andean Portal: Geologic Record in the Tatacoa Desert. *Frontiers in Earth Science* **8**: 587022.

Mora A, Baby P, Roddaz M, Parra M, Brusset S, Hermoza W, Espurt N. **2010**. Tectonic history of the Andes and sub-Andean zones: implications for the development of the Amazon drainage basin. In: Hoorn C, Wesselingh F, eds. Amazonia, Landscape and Species Evolution. Wiley, Oxford, 38–60.

Morales-Briones DF, Liston A, Tank DC. **2018**. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). *New Phytologist* **218**: 1668–1684.

Musilová Z, Kalous L, Petrtýl M, Chaloupková P. **2013**. Cichlid Fishes in the Angolan Headwaters Region: Molecular Evidence of the Ichthyofaunal Contact between the Cuanza and Okavango-Zambezi Systems. *PLoS ONE* **8**: e65047.

Novelo RA, Philbrick CT. **1997**. Taxonomy of Mexican Podostemaceae. *Aquatic Botany* **57**: 275–303.

Novelo RA, Philbrick CT, Crow GE. **2009**. Podostemaceae. Flora Mesoamericana 3: 1–7. First published on the Flora Mesoamericana website 7 July 2009. <http://www.tropicos.org/docs/meso/podostemaceae.pdf>.

Nürk NM, Scheriau C, Madriñán S. **2013**. Explosive radiation in high Andean Hypericum—rates of diversification among New World lineages. *Frontiers in Genetics* **4:** 1–14.

One Thousand Plant Transcriptomes Initiative. **2019**. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**: 679–685.

Paradis E, Schliep K. **2019**. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: 526–528.

Parra M, Echeverri S, Patiño AM, Ramírez JC, Mora A, Sobel ER, Almendral A, Pardo-Trujillo A. **2019**. Cenozoic evolution of the Sierra Nevada de Santa Marta, Colombia. In: Gómez J, Mateus-Zabala D, eds. The geology of Colombia, Volume 3 Paleogene-Neogene. Bogotá: Servicio Geológico Colombiano, Publicaciones Geológicas Especiales, 259–297.

Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. **2010**. How Many Bootstrap Replicates Are Necessary? *Journal of Computational Biology* **17**: 337–354.

Pease JB, Brown JW, Walker JF, Hinchliff CE, Smith SA. **2018**. Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. *American Journal of Botany* **105**: 385–403.

Pérez-Escobar OA, Chomicki G, Condamine FL, Karremans AP, Bogarín D, Matzke NJ, Silvestro D, Antonelli A. **2017**. Recent origin and rapid speciation of Neotropical orchids in the world's richest plant biodiversity hotspot. *New Phytologist* **215**: 891–905.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. **2012**. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE* **7**: e37135.

Philbrick CT, Bove CP, Stevens HI. **2010**. Endemism in Neotropical Podostemaceae. *Annals of the Missouri Botanical Garden* **97**: 425–456.

Philbrick CT, Novelo RA. **1995**. New World Podostemaceae: Ecological and Evolutionary Enigmas. *Brittonia* **47**: 210.

Philbrick CT, Novelo A. **1997**. Ovule number, seed number and seed size in Mexican and North American species of Podostemaceae. *Aquatic Botany* **57**: 183–200.

Philbrick CT, Novelo RA. **1998**. Flowering phenology, pollen flow, and seed production in *Marathrum* rubrum (Podostemaceae). *Aquatic Botany* **62**: 199–206.

Philbrick CT, Philbrick PKB, Lester BM. **2015**. Root Fragments as Dispersal Propagules in the Aquatic Angiosperm *Podostemum ceratophyllum* Michx. (Hornleaf Riverweed, Podostemaceae). *Northeastern Naturalist* **22**: 643–647.

Picq S, Alda F, Krahe R, Bermingham E. **2014**. Miocene and Pliocene colonization of the Central American Isthmus by the weakly electric fish *Brachyhypopomus occidentalis* (Hypopomidae, Gymnotiformes). *Journal of Biogeography* **41**: 1520–1532.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, *et al.* 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**: 559–575.

Puritz JB, Hollenbeck CM, Gold JR. **2014**. dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* **2**: e431.

Quintero I, Jetz W. **2018**. Global elevational diversity and diversification of birds. *Nature* **555**: 246–250.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. **2018**. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology* **67**: 901–904.

Reyes-Ortega I, Sánchez-Coronado ME, Orozco-Segovia A. **2009**. Seed germination in *Marathrum schiedeanum* and *M. rubrum* (Podostemaceae). *Aquatic Botany* **90**: 13–17.

Richardson JE, Madriñán S, Gómez-Gutiérrez MC, Valderrama E, Luna J, Banda-R. K, Serrano J, Torres MF, Jara OA, Aldana AM, *et al.* 2018. Using dated molecular phylogenies to help reconstruct geological, climatic, and biological history: Examples from Colombia. *Geological Journal* **53**: 2935–2943.

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

Roxo FF, Albert JS, Silva GSC, Zawadzki CH, Foresti F, Oliveira C. **2014**. Molecular Phylogeny and Biogeographic History of the Armored Neotropical Catfish Subfamilies Hypoptopomatinae, Neoplecostominae and Otothyrinae (Siluriformes: Loricariidae). *PLoS ONE* **9**: e105564.

van Royen P. **1951**. *The Podostemaceae of the New World*. Mededelingen van het Botanisch Museum en Herbarium van de Rijks universiteit te Utrecht. Harte.

Ruhfel BR, Bove CP, Philbrick CT, Davis CC. **2016**. Dispersal largely explains the Gondwanan distribution of the ancient tropical clusioid plant clade. *American Journal of Botany* **103**: 1117–1128.

Ruhfel BR, Stevens PF, Davis CC. **2013**. Combined Morphological and Molecular Phylogeny of the Clusioid Clade (Malpighiales) and the Placement of the Ancient Rosid Macrofossil *Paleoclusia*. *International Journal of Plant Sciences* **174**: 910–936.

Ruokolainen K, Moulatlet GM, Zuquim G, Hoorn C, Tuomisto H. **2019**. Geologically recent rearrangements in central Amazonia river network and their importance for the riverine barrier hypothesis. *Frontiers of Biogeography* **11**: e45046.

Rutishauser R. **1995**. Developmental patterns of leaves in Podostemaceae compared with more typical flowering plants: saltational evolution and fuzzy morphology. *Canadian Journal of Botany* **73**: 1305–1317.

Rutishauser R. **1997**. Structural and developmental diversity in Podostemaceae (river-weeds). *Aquatic Botany* **57**: 29–70.

Rutishauser R, Novelo RA, Philbrick CT. **1999**. Developmental Morphology of New World Podostemaceae: *Marathrum* and *Vanroyenella*. *International Journal of Plant Sciences* **160**: 29–45.

Schenk JJ. **2016**. Consequences of Secondary Calibrations on Divergence Time Estimates. *PLOS ONE* **11**: e0148228.

Schmickl R, Liston A, Zeisek V, Oberlander K, Weitemier K, Straub SCK, Cronn RC, Dreyer LL, Suda J. **2016**. Phylogenetic marker development for target enrichment from transcriptome and genome skim data: the pipeline and its application in southern African *Oxalis* (Oxalidaceae). *Molecular Ecology Resources* **16**: 1124–1135.

Sklenář P, Dušková E, Balslev H. **2011**. Tropical and Temperate: Evolutionary History of Páramo Flora. *The Botanical Review* **77**: 71–108.

Smith BT, McCormack JE, Cuervo AM, Hickerson MichaelJ, Aleixo A, Cadena CD, Pérez-Emán J, Burney CW, Xie X, Harvey MG, *et al.* 2014. The drivers of tropical speciation. *Nature* **515**: 406–409.

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

Stange M, Sánchez-Villagra MR, Salzburger W, Matschiner M. **2018**. Bayesian Divergence-Time Estimation with Genome-Wide Single-Nucleotide Polymorphism Data of Sea Catfishes (Ariidae) Supports Miocene Closure of the Panamanian Isthmus. *Systematic Biology* **67**: 681–699.

Stenz NWM, Larget B, Baum DA, Ané C. **2015**. Exploring Tree-Like and Non-Tree-Like Patterns Using Genome Sequences: An Example Using the Inbreeding Plant Species *Arabidopsis thaliana* (L.) Heynh. *Systematic Biology* **64**: 809–823.

Swofford D. **2003**. *PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4*. Sianuer Associates, Sunderland.

Tagliacollo VA, Roxo FF, Duke-Sylvester SM, Oliveira C, Albert JS. **2015**. Biogeographical signature of river capture events in Amazonian lowlands. *Journal of Biogeography* **42**: 2349–2362.

Testo WL, Sessa E, Barrington DS. **2019**. The rise of the Andes promoted rapid diversification in Neotropical *Phlegmariurus* (Lycopodiaceae). *New Phytologist* **222**: 604–613.

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

Tippery NP, Philbrick CT, Bove CP, Les DH. **2011**. Systematics and Phylogeny of Neotropical Riverweeds (Podostemaceae: Podostemoideae). *Systematic Botany* **36**: 105–118.

Villagómez D, Spikings R, Mora A, Guzmán G, Ojeda G, Cortés E, van der Lelij R. **2011**. Vertical tectonics at a continental crust-oceanic plateau plate boundary zone: Fission track thermochronology of the Sierra Nevada de Santa Marta, Colombia: thermochronology of northern Colombia. *Tectonics* **30**: TC4004.

Viruel J, Conejero M, Hidalgo O, Pokorny L, Powell RF, Forest F, Kantar MB, Soto Gomez M, Graham SW, Gravendeel B, *et al.* 2019. A Target Capture-Based Method to Estimate Ploidy From Herbarium Specimens. *Frontiers in Plant Science* **10**: 937.

Wagner CE, Keller I, Wittwer S, Selz OM, Mwaiko S, Greuter L, Sivasundar A, Seehausen O. **2013**. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* **22**: 787–798.

Weiß CL, Pais M, Cano LM, Kamoun S, Burbano HA. **2018**. nQuire: a statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinformatics* **19**: 122.

Wen D, Yu Y, Zhu J, Nakhleh L. **2018**. Inferring Phylogenetic Networks Using PhyloNet. *Systematic Biology* **67**: 735–740.

Willis JC. **1902**. Studies in the morphology and ecology of the Podostemaceae of Ceylon and India. *Annals of the Royal Botanic Gardens, Peradeniya* **1**: 267–465.

Willis JC. **1914**. On the lack of adaptation in the Tristichaceae and Podostemaceae. *Annals of Botany* **87**: 532–550.

Willis JC. **1915**. The origin of the Tristichaceae and Podostemaceae. *Annals of Botany* **29**: 299–306.

Yu Y, Degnan JH, Nakhleh L. **2012**. The Probability of a Gene Tree Topology within a Phylogenetic Network with Applications to Hybridization Detection. *PLoS Genetics* **8**: e1002660.

Yuan Y-W, Liu C, Marx HE, Olmstead RG. **2009**. The pentatricopeptide repeat (PPR) gene family, a tremendous resource for plant phylogenetic studies: Methods. *New Phytologist* **182**: 272–283.

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.

Zhang L-N, Ma P-F, Zhang Y-X, Zeng C-X, Zhao L, Li D-Z. **2019**. Using nuclear loci and allelic variation to disentangle the phylogeny of Phyllostachys (Poaceae, Bambusoideae). *Molecular Phylogenetics and Evolution* **137**: 222–235.

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

Zhu T, Yang Z. **2021**. Complexity of the simplest species tree problem. *Molecular Biology and Evolution*: msab009.

Supporting Information

Fig. S1 SVDquartets extended majority-rule consensus tree for all 95 ddRADseq samples

Fig. S2 Examples of maximum likelihood trees from exons with paralog warnings

Fig. S3 Target enrichment recovery efficiency across samples

Fig. S4 Cross-validation error and number of clusters from ADMIXTURE

Fig. S5 SVDquartets extended majority-rule consensus trees from non-admixed individuals and excluding invariant and non-biallelic SNPs

Fig. S6 SVDquartets extended majority-rule consensus trees including admixed individuals and inferred from target enrichment and ddRADseq data

Fig. S7 ASTRAL III tree including admixed individuals and inferred from target enrichment and ddRADseq data

Fig. S8 Maximum likelihood tree inferred from target enrichment and ddRADseq data and including admixed individuals

Fig. S9 Phylogenetic trees inferred with target enrichment data using single-exon genes for comparison with the results for the same analyses in Fig. 4 using concatenated multi-exon genes

Fig. S10 Phylogenetic trees inferred with target enrichment data using single-exon genes and including admixed individuals for comparison with the results for the same analyses using concatenated multi-exon genes in Figs. S5, S6, and S7

Fig. S11 Phylogenetic tree inferred with target enrichment data and under maximum likelihood using single-exon genes for comparison with the results from the same analysis in Fig. S8 using concatenated multi-exon genes

Methods S1. Selection of sequences for target enrichment

Methods S2. Selection of assembled contigs with Hybpiper and paralog filtering

Methods S3. Specifications for maximum likelihood, SVDquartets, Quartet Sampling, and PhyloNet analyses

Table S1. Accepted species of *Marathrum* and *Marathrum foeniculaceum* synonyms

Table S2. Voucher and accession numbers of samples in this study

Table S3. Ipyrad loci recovered across all ddRADseq samples (*min4*)

Table S4. Ipyrad statistics for all ddRADseq assemblies

Table S5. Target enrichment recovery efficiency

Table S6. Hybpiper statistics for the number of paralog copies detected for each sample (see separate file)

Table S7. Overall heterozygosity of each individual as estimated with VCFtools

Table S8. nQuire results for ploidy level estimation

Table S9. Number of quartet sampling replicates for the concordant (count 0), and the two discordant (count 1 and count 2) quartet arrangements for the maximum likelihood tree inferred with target enrichment data (Fig. S8)

Table S10. Number of quartet sampling replicates for the concordant (count 0), and the two discordant (count 1 and count 2) quartet arrangements for the maximum likelihood tree inferred with ddRADseq data (Fig. S8)

Table S11. AIC model selection results for phylogenetic networks with 0–3 reticulations including admixed individuals from Antioquia, Boyacá and Caribe and using single-exon genes

Table S12. Nodes ages estimated with SNAPP

Table 1. Summary of ages for the stem node of the subfamily Podostemoideae estimated from previous studies and in this study.

	Bedoya et al., 2020	**Magallón et al., 2015**	**Ruhfel et al., 2016**	
95% HPD Min (Ma)	54.10	53.53	49.9	55.4
95% HPD Mean (Mya)	61.77	67.82	62	66.1
95% HPD Max (Mya)	69.44	77.29	71.9	76.8

To calibrate the species tree, we assumed a normal prior distribution with a mean of 62 Ma and a standard deviation of 4 Ma. Ruhfel et al. (2016) presented two dating results for alternative placements of the fossil *Paleoclusia chevalieri.*

Table 2. AIC model selection results for phylogenetic networks with 0–3 reticulations including admixed individuals from Antioquia (Magdalena drainage basin), Boyacá (Orinoco drainage basin) and Caribe (Sierra Nevada de Santa Marta drainage basin).

Drainage basin	# Reticulations	Parameters (k)	Log likelihood	AIC	ΔAIC	Relative likelihood	wAIC	Cumulative wAIC
Antioquia	**3**	**14**	**-2331.34**	**4690.68**	**0**	**1**	**1**	**1**
	2	12	-2379.82	4783.64	92.96	6.53-21	0	1
	1	10	-2450.89	4921.79	231.11	6.52E-51	0	1
	0	8	-2455.99.	4927.99	237.31	2.94E-52	0	1
	4	16	-2837.21	5706.42	922.78	4.17E-201	0	1
Boyacá	**3**	**12**	**-2853.80**	**5731.60**	**0.00**	**1.00**	**0.96**	**0.96**
	2	10	-2858.90	5737.79	6.19	0.05	0.04	1.00
	1	8	-2884.89	5785.79	54.19	0.00	0.00	1.00
	4	14	-2982.84	5993.69	262.09	0.00	0.00	1.00
	0	6	-3031.70	6075.39	343.79	0.00	0.00	1.00
Caribe	**2**	**12**	**-2643.92**	**5309.84**	**0**	**1**	**1**	**1**
	1	10	-2838.19	5694.39	384.55	3.13E-84	0	1
	0	8	-2940.83	5895.67	585.82	6.15E-128	0	1
	4	16	-2961.20	5954.40	642.55	2.95E-140	0	1
	3	14	-3030.17	6086.35	776.50	2.42E-169	0	1

Significant results with wAIC values ≥ 0.95 are shown in bold. The number of parameters (k) is (n-3) +(2h), where n is the number of tips in the network and h is the number of reticulation events. In Boyacá, admixed individuals are assigned to one population.

Figures

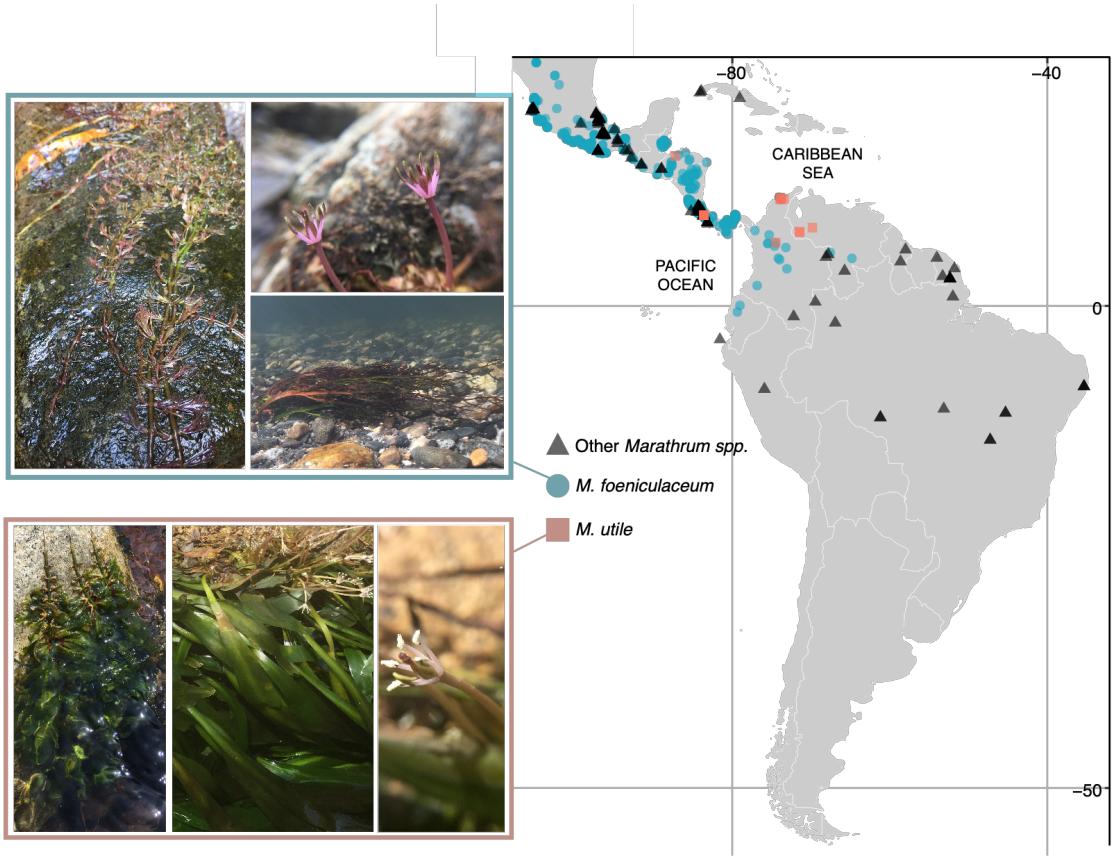


Figure 1. Distribution of *Marathrum*, *M. foeniculaceum*, and *M. utile*. The vegetative and reproductive morphology of *M. utile* and *M. foeniculaceum* is shown. *Marathrum utile* has laminar, entire leaves and *M. foeniculaceum* has highly dissected leaves.

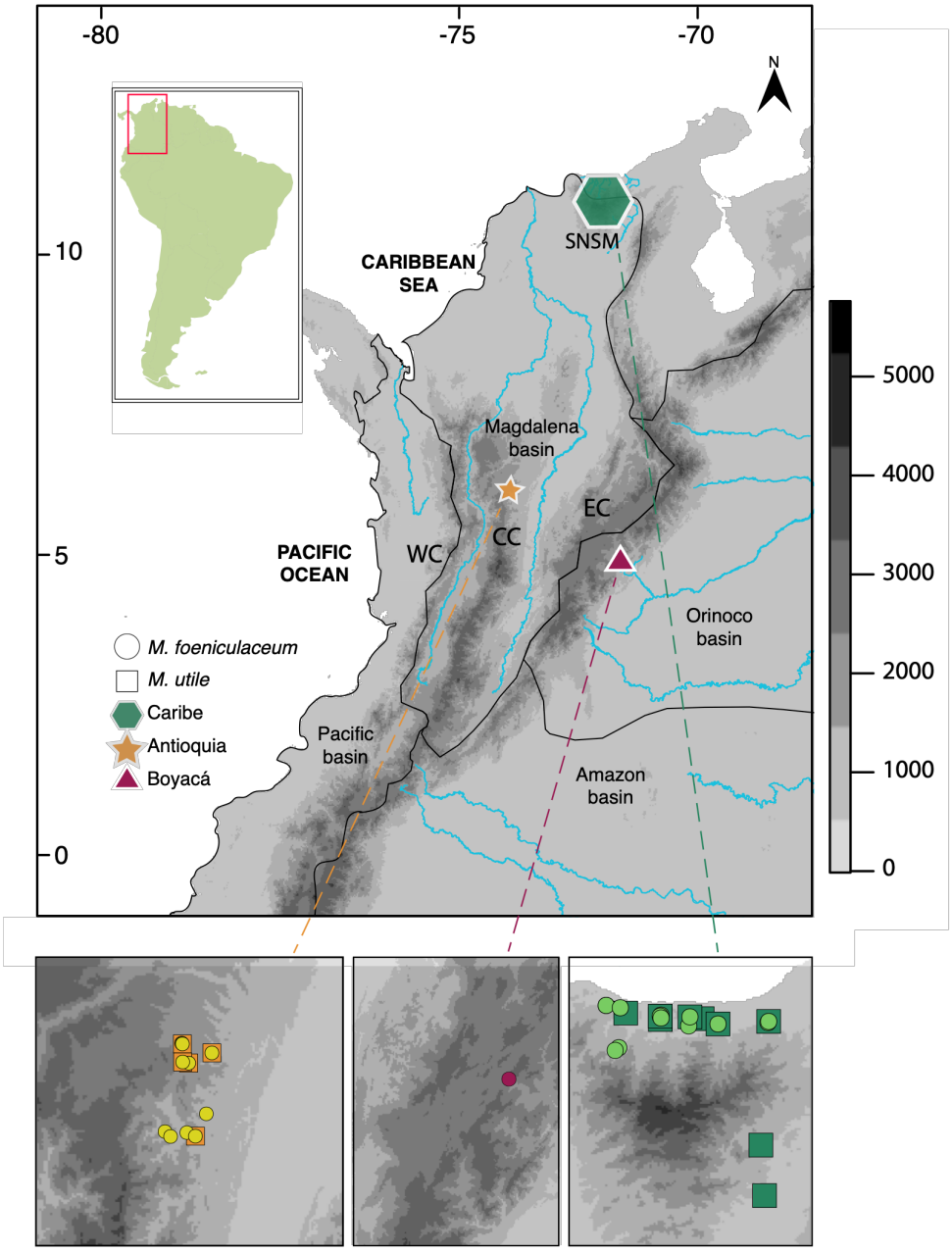


Figure 2. Top: Map of the Andean region in northern South America (nSA). The three Andean Cordilleras and the Sierra Nevada de Santa Marta (SNSM) are shown. WC= Western Cordillera, CC=Central cordillera, EC=Eastern Cordillera. Drainage basins are delimited. Major rivers are shown in blue. Elevation is indicated in m.a.s.l. The three major localities where *Marathrum* was collected (*i.e.* Caribe, Antioquia, and Boyacá) are indicated with symbols. Bottom: Sample sites where *M. foeniculaceum* and *M. utile* were collected. Each square and circle represents a single site within the major localities. *M. foeniculaceum* refers to plants with that phenotype.

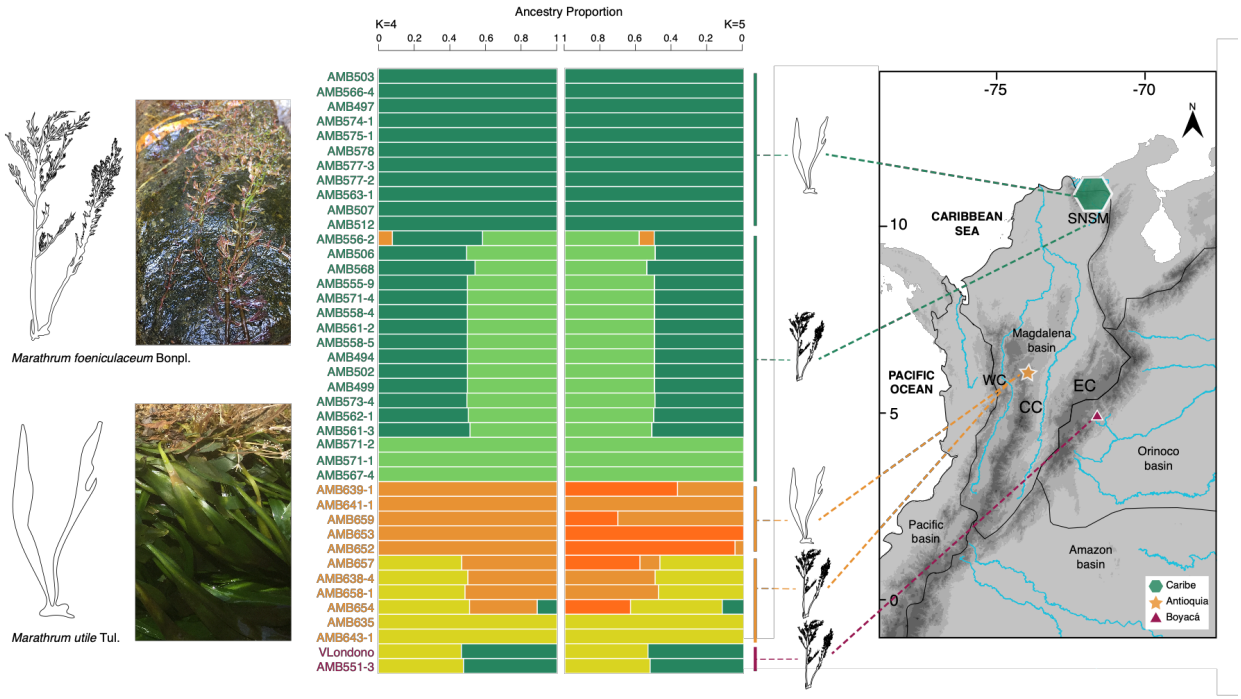


Figure 3. ADMIXTURE results for K=4 (the optimal value) and K=5 (to illustrate population substructure in *M. utile* from Antioquia) inferred from target enrichment data. The phenotype and provenance of samples is shown.

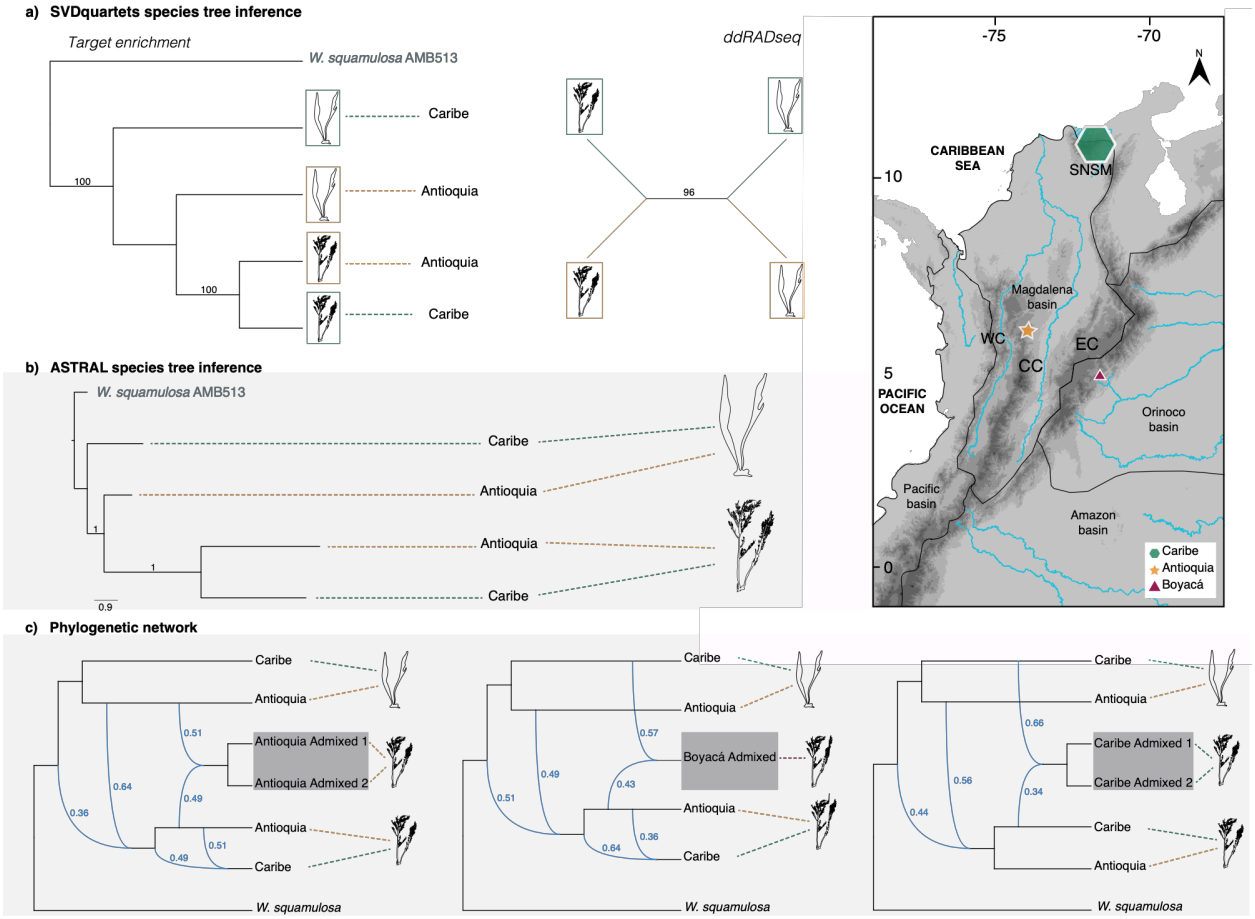


Figure 4. a) SVDquartets extended majority-rule consensus trees inferred with target enrichment and ddRADseq data (admixed individuals are excluded). Bootstrap support >70 is shown above branches. b) ASTRAL III tree inferred from target enrichment data (admixed individuals are excluded). Local posterior probabilities >0.7 are shown above branches. c) Phylogenetic networks resulted from maximum pseudo-likelihood inference with PhyloNet. Inferred reticulation events (curved, blue lines) are accompanied by inheritance probabilities. The phenotype and provenance of operational taxonomic units is shown.

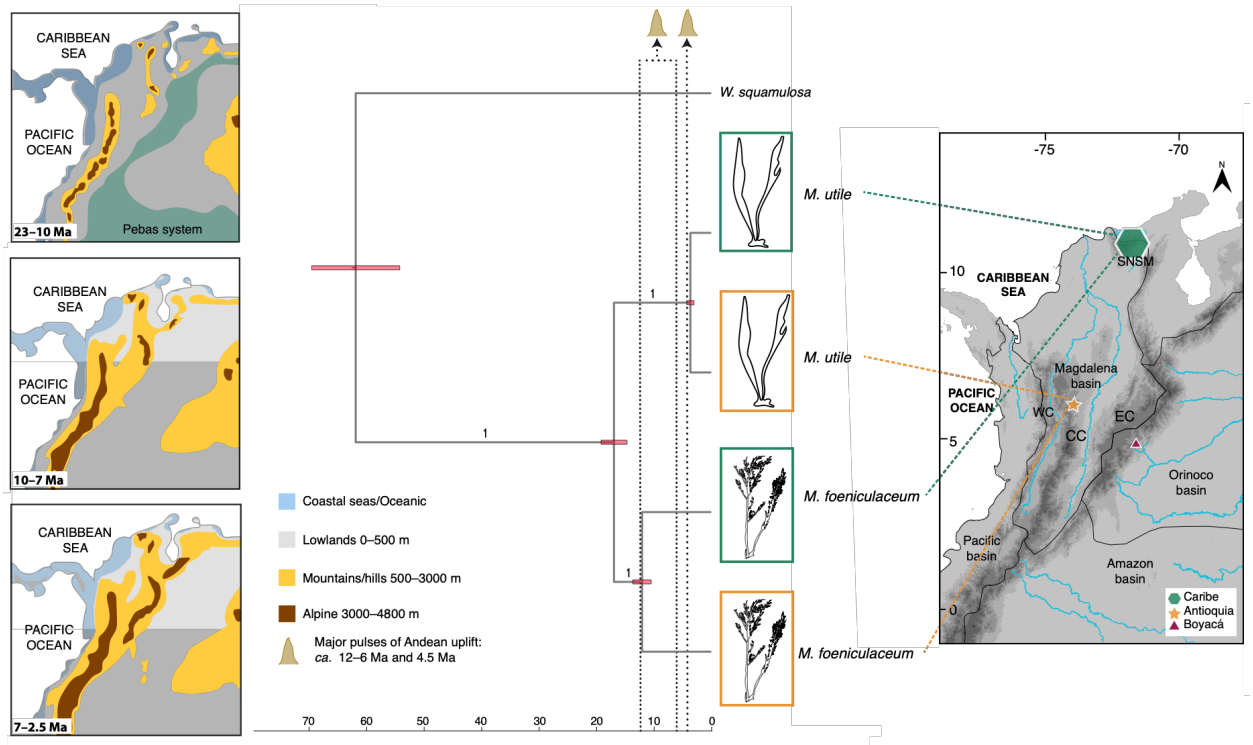


Figure 5. Time-calibrated species tree inferred with SNAPP. Posterior probability values are shown on branches. A summary of landscape changes in northern South America at different time periods is shown (modified from Hoorn *et al.,* 2010).

New Phytologist

***New Phytologist* Supporting Information**

Article title: Andean uplift, drainage basin formation, and the evolution of plants living in fast-flowing aquatic ecosystems in northern South America

Authors: Ana M. Bedoya, Adam D. Leaché and Richard G. Olmstead

Article acceptance date: 20 July 2021

The following Supporting Information is available for this article:

Fig. S1 SVDquartets extended majority-rule consensus tree for all 95 ddRADseq samples

Fig. S2 Examples of maximum likelihood trees from exons with paralog warnings

Fig. S3 Target enrichment recovery efficiency across samples

Fig. S4 Cross-validation error and number of clusters from ADMIXTURE

Fig. S5 SVDquartets extended majority-rule consensus trees from non-admixed individuals and excluding invariant and non-biallelic SNPs

Fig. S6 SVDquartets extended majority-rule consensus trees including admixed individuals and inferred from target enrichment and ddRADseq data

Fig. S7 ASTRAL III tree including admixed individuals and inferred from target enrichment and ddRADseq data

Fig. S8 Maximum likelihood tree inferred from target enrichment and ddRADseq data and including admixed individuals

Fig. S9 Phylogenetic trees inferred with target enrichment data using single-exon genes for comparison with the results for the same analyses in Fig. 4 using concatenated multi-exon genes

Fig. S10 Phylogenetic trees inferred with target enrichment data using single-exon genes and including admixed individuals for comparison with the results for the same analyses using concatenated multi-exon genes in Figs. S4, S5, and S6

Fig. S11 Phylogenetic tree inferred with target enrichment data and under maximum likelihood using single-exon genes for comparison with the results from the same analysis in Fig. S7 using concatenated multi-exon genes

Table S1 Accepted species of *Marathrum* and *Marathrum foeniculaceum* synonyms

Table S2 Voucher and accession numbers of samples in this study

Table S3 Ipyrad loci recovered across all ddRADseq samples (*min4*)

Table S4 Ipyrad statistics for all ddRADseq assemblies

Table S5 Target enrichment recovery efficiency

Table S6 Hybpiper statistics for the number of paralog copies detected for each sample (see separate file)

Table S7 Overall heterozygosity of each individual as estimated with VCFtools

Table S8 nQuire results for ploidy level estimation

Table S9 Number of quartet sampling replicates for the concordant (count 0), and the two discordant (count 1 and count 2) quartet arrangements for the maximum likelihood tree inferred with target enrichment data (Fig. S7)

Table S10 Number of quartet sampling replicates for the concordant (count 0), and the two discordant (count 1 and count 2) quartet arrangements for the maximum likelihood tree inferred with ddRADseq data (Fig. S7)

Table S11 AIC model selection results for phylogenetic networks with 0–3 reticulations including admixed individuals from Antioquia, Boyacá and Caribe and using single-exon genes

Table S12 Nodes ages estimated with SNAPP

Methods S1 Selection of sequences for target enrichment

Methods S2 Selection of assembled contigs with Hybpiper and paralog filtering

Methods S3 Specifications for maximum likelihood, SVDquartets, Quartet Sampling, and PhyloNet analyses.

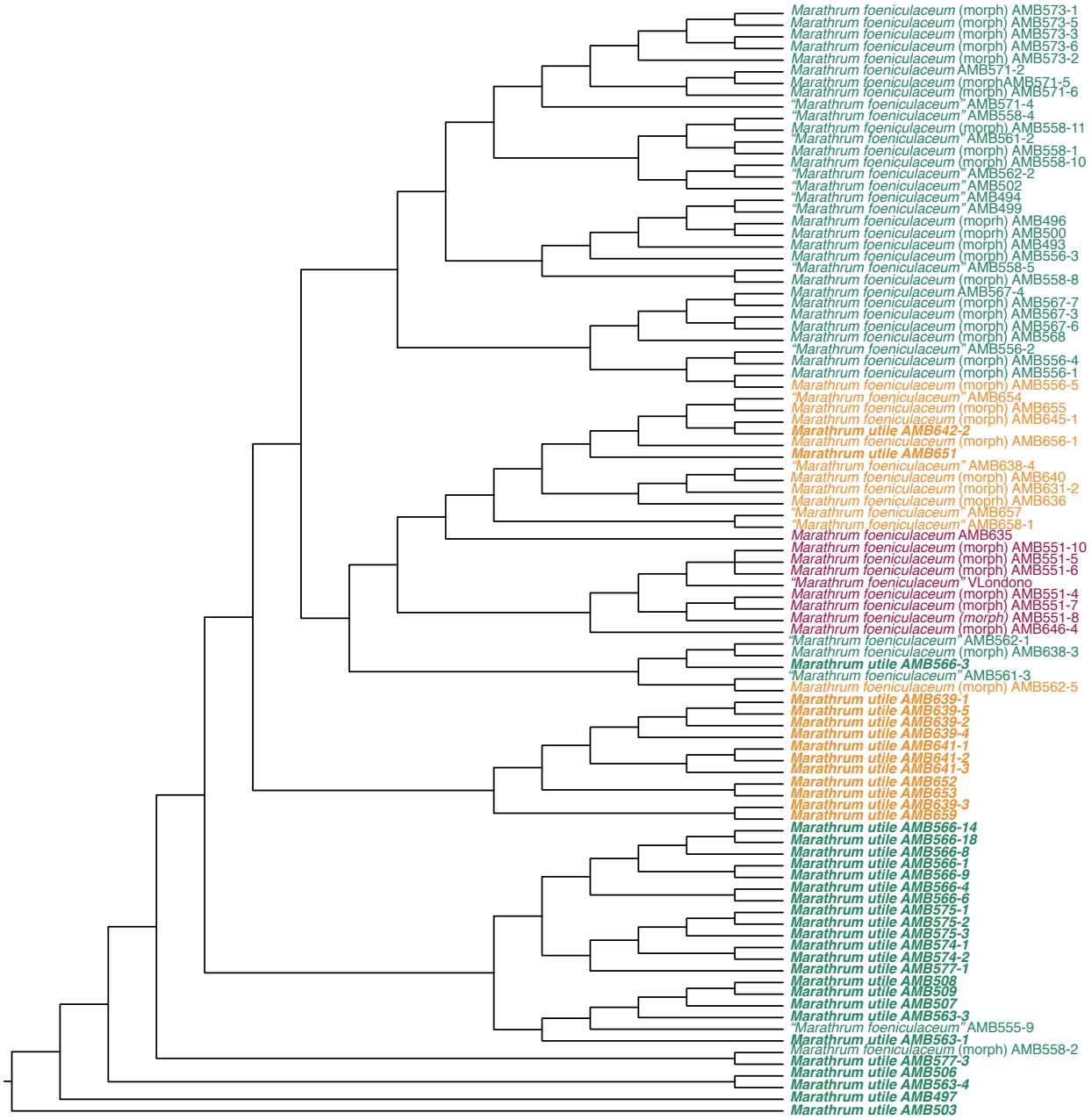


Figure S1. Species tree inferred with SVDquartets for 95 ddRADseq samples successfully sequenced in this study. Individuals of *M. foeniculaceum* and *M. utile* (bold) from Antioquia (yellow), Boyacá (red) and Caribe (green) are shown. Tree was inferred as unrooted, but rooting is presented to be consistent with posterior analyses that were rooted with the outgroup *W. squamulosa*. Admixed individuals as inferred with ADMIXTURE are indicated with quotation marks. Individuals for which no information on genetic constitution is available but have the phenotype of *M. foeniculaceum* are indicated with "(morph)".

Figure S2. Examples of maximum likelihood trees inferred from exons with paralog warnings. Bootstrap support shown above branches. Trees are rooted with the outgroup (*W. squamulosa*). Colors correspond to Antioquia (yellow), Caribe (green), and Boyacá (red). Non-admixed individuals as inferred with ADMIXTURE are in bold, whereas admixed individuals are indicated with quotation marks. Bootstrap support >70 shown with red circles. Alternate contigs for admixed individuals cluster with *M. utile* and *M. foeniculaceum,* indicating that the alternate contigs correspond to allelic variation. Phylogenetic trees also show lack of high support, resolution, and complete sorting of alleles across populations and species as expected for population-level phylogenetic inference of closely related and recently diverged taxa, and for the use of relatively short sequences.

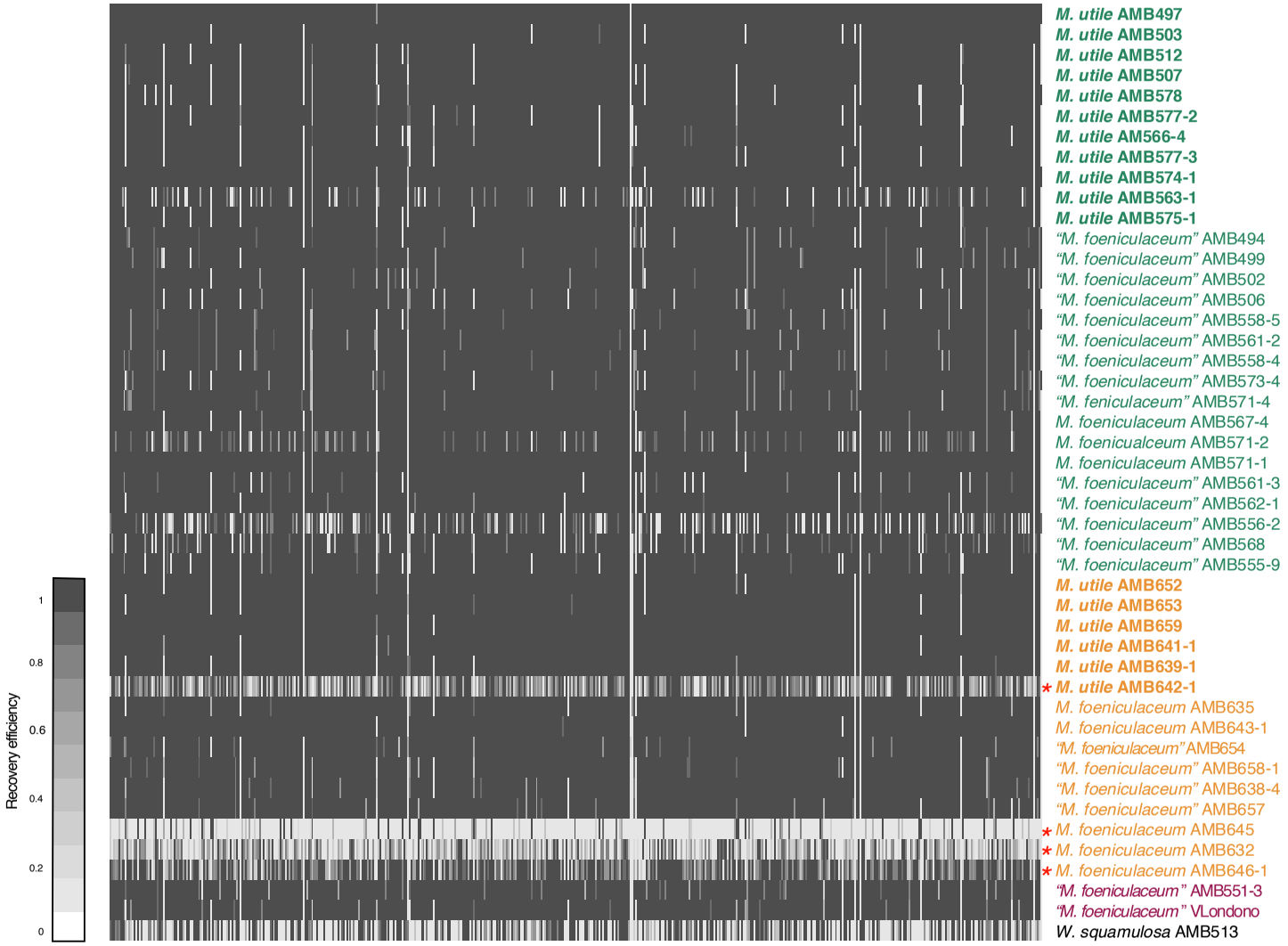


Figure S3. Target enrichment loci recovery efficiency for 42 samples (rows) across exons(columns). Recovery efficiency is measured as the length of the sequence recovered as a percentage of the length of the targeted sequence. Darker gray: The full length of the targeted sequence was recovered. White: no sequence was recovered. (*) Individuals that failed sequencing or recovered short exons of low quality. Individuals of *M. foeniculaceum* and *M. utile* (bold) from Antioquia (yellow), Boyacá (red) and Caribe (green) are shown. Admixed individuals as inferred with ADMIXTURE are indicated with quotation marks.

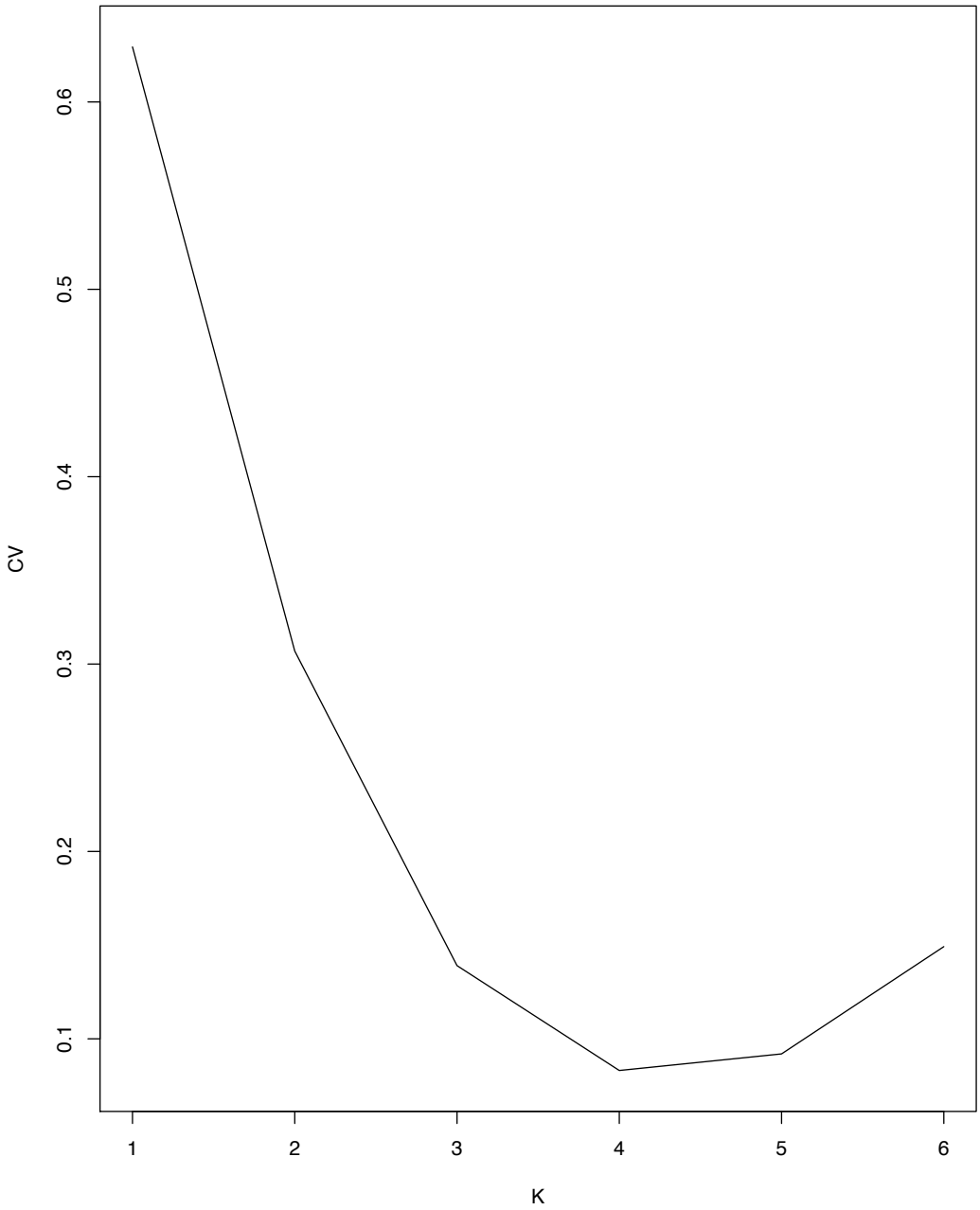


Figure S4. Cross-validation error and number of clusters (K) resulted from genetic structure analysis with ADMIXTURE using target enrichment data.

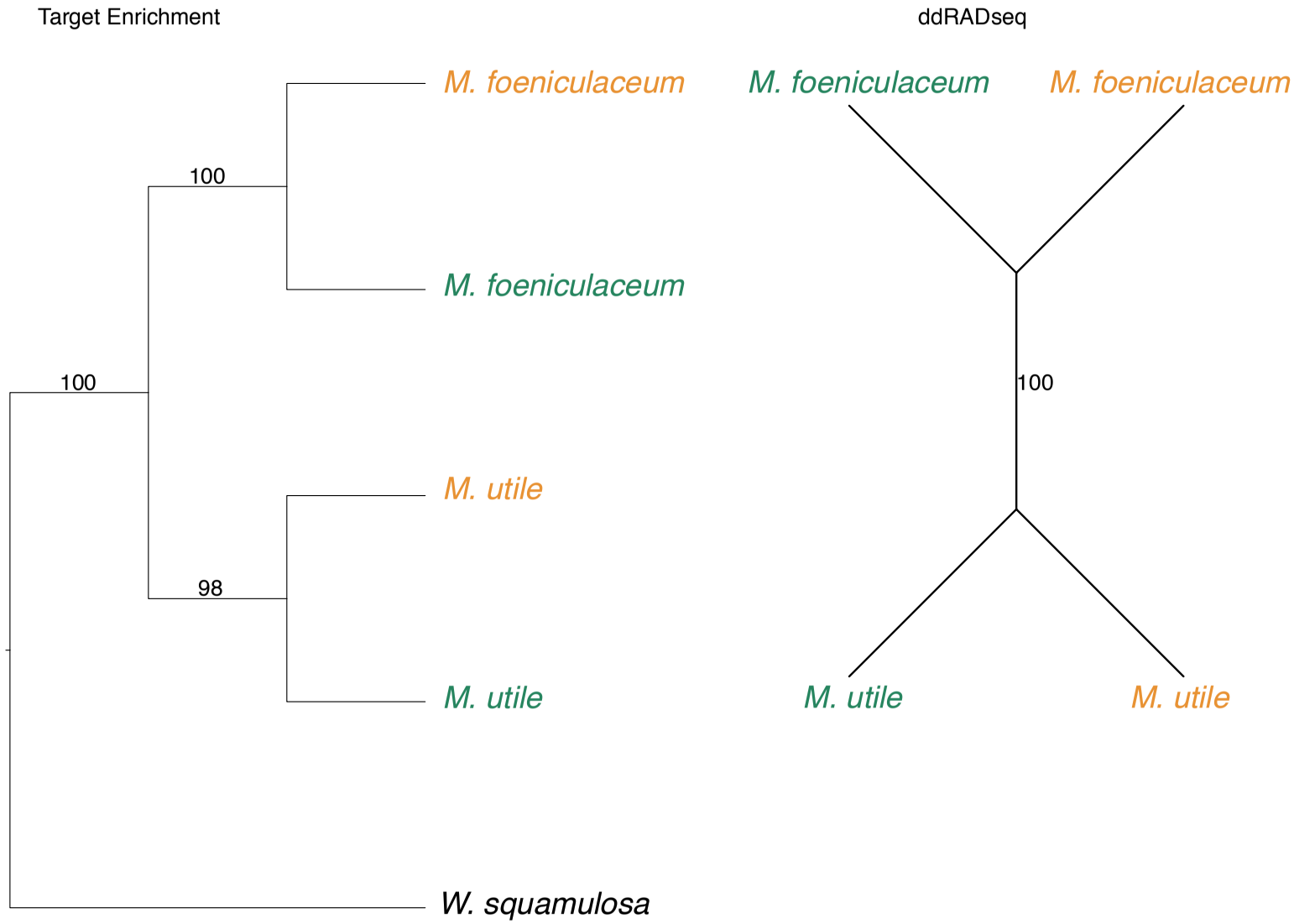


Figure S5. SVDquartets extended majority-rule consensus trees inferred from target enrichment and ddRADseq after exclusion of invariant and non-binary sites. Colors correspond to Antioquia (yellow) and Caribe (green). ddRADseq tree is unrooted. Bootstrap support >70 is shown on branches.

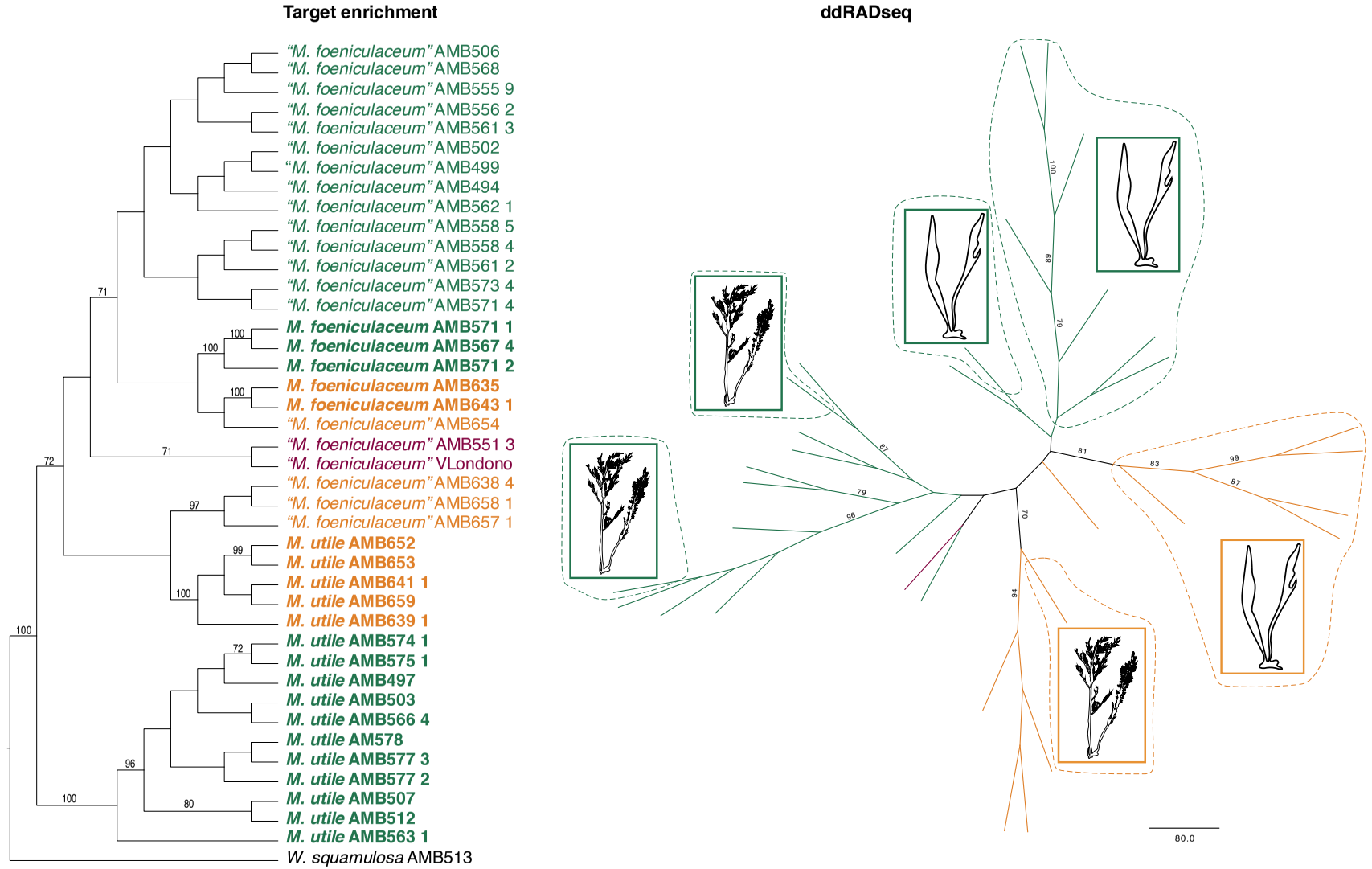


Figure S6. SVDquartets extended majority-rule consensus trees inferred from our target enrichment and ddRADseq data including admixed individuals and using species-as-tips. Bootstrap support >70 is shown above branches. Non-admixed individuals of *M. foeniculaceum* and *M. utile* (bold and inside dashed lines) from Antioquia (yellow), Boyacá (red) and Caribe (green) are shown. Admixed individuals as inferred with ADMIXTURE are indicated with quotation marks in the target enrichment tree.

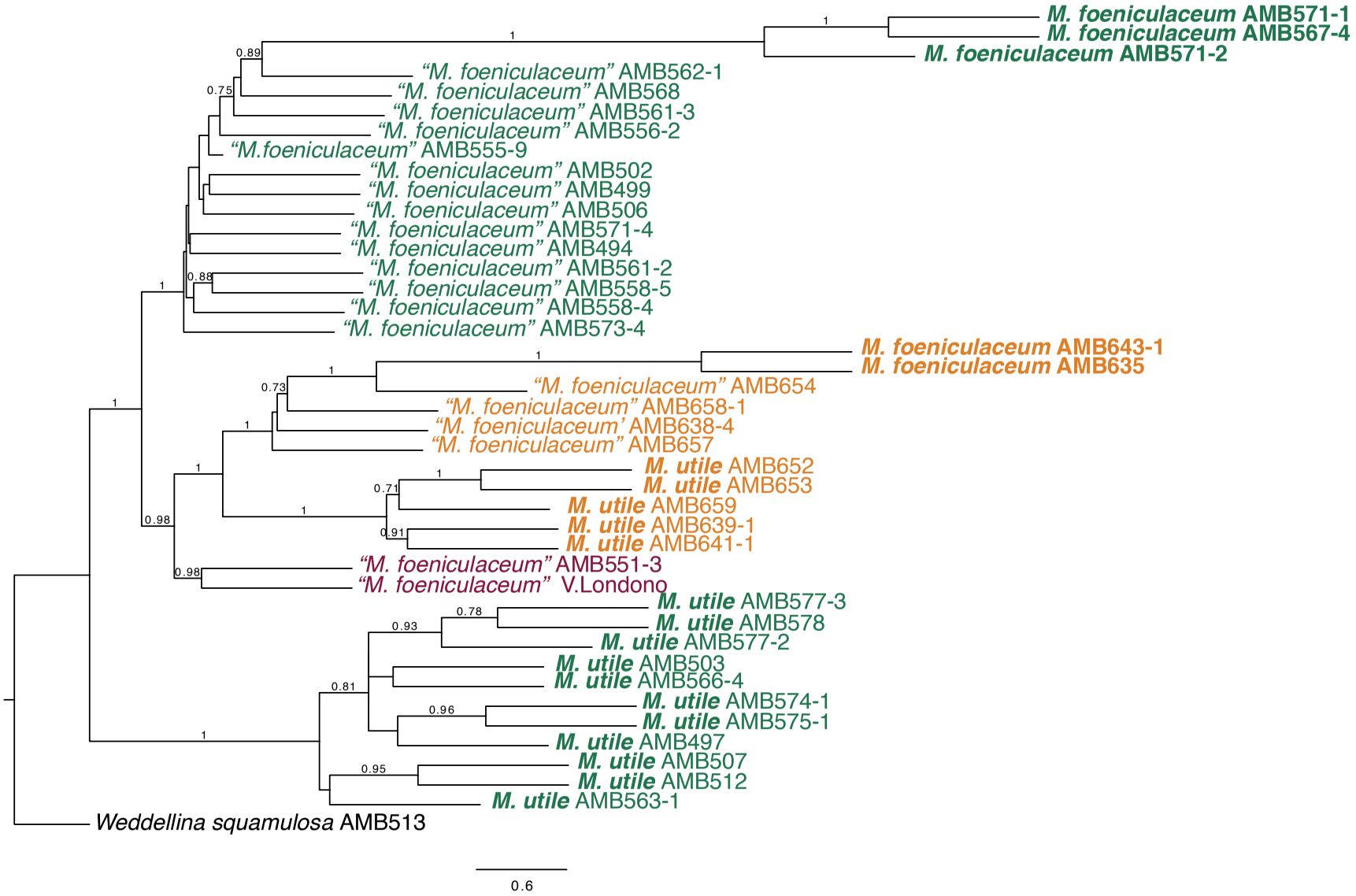


Figure S7. ASTRAL III tree inferred from target enrichment data and including admixed individuals using species-as-tips. Local posterior probabilities > 0.7 are shown above branches. Individuals of *M. foeniculaceum* and *M. utile* (bold) from Antioquia (yellow), Boyacá (red) and Caribe (green) are shown. Admixed individuals as inferred with ADMIXTURE are indicated with quotation marks

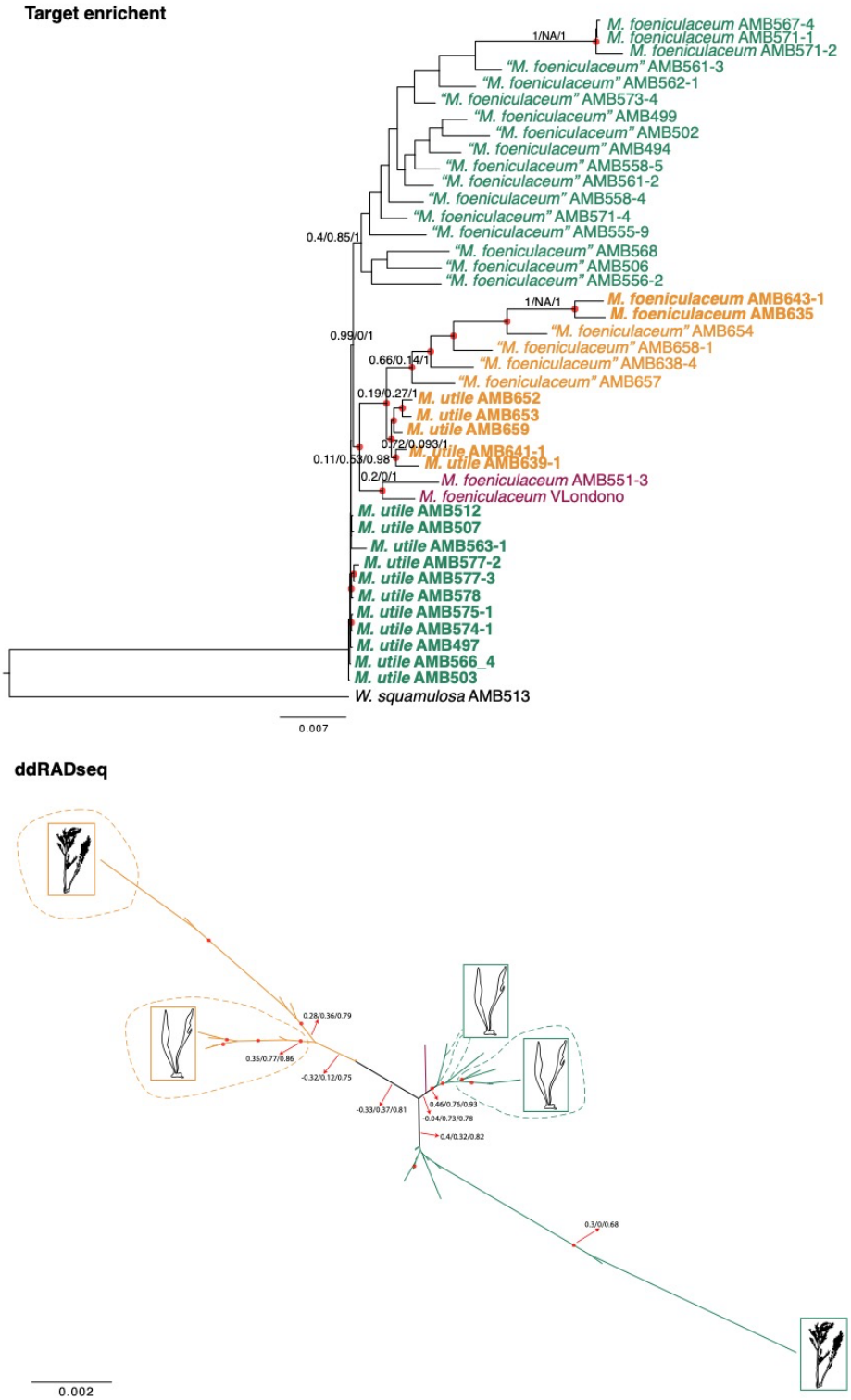


Figure S8. Maximum likelihood trees inferred from target enrichment and ddRADseq data and including admixed individuals. Non-admixed individuals of *M. foeniculaceum* and *M. utile* (bold and inside dashed lines) from Antioquia (yellow), Boyacá (red) and Caribe (green) are shown. Admixed individuals as inferred with ADMIXTURE are indicated with quotation marks. Bootstrap support >70 shown with red circles. Quartet sampling scores (Quartet Concordance/Quartet Differential/Quartet Informativeness) for the main clades (drainage basins and species) are shown above branches.

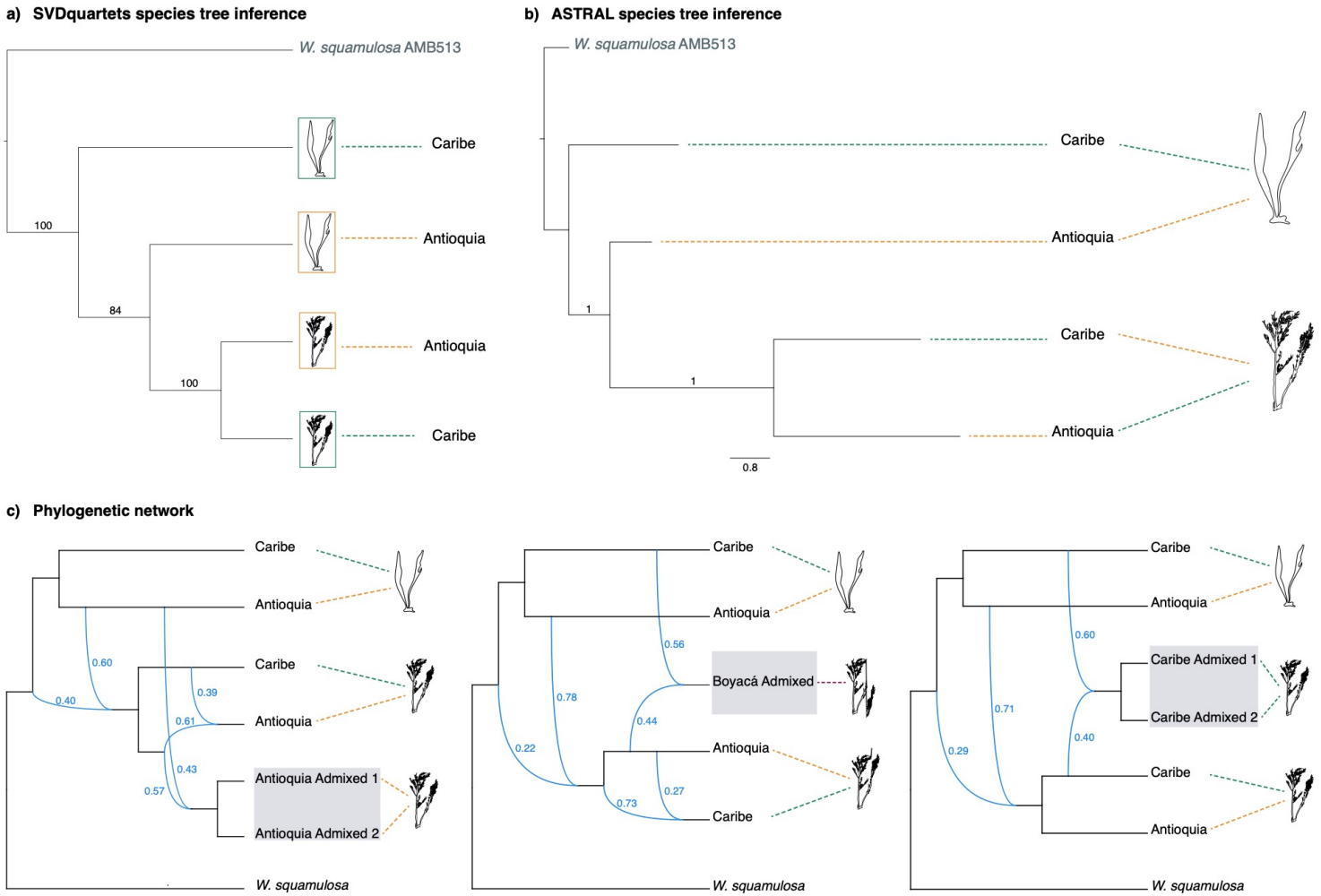


Figure S9. Phylogenetic trees inferred using single-exon genes for comparison with the same analyses using concatenated multi-exon genes in Fig. 4. a) SVDquartets extended majority-rule consensus trees inferred with target enrichment data (admixed individuals are excluded). Bootstrap support >70 is shown above branches. b) ASTRAL III tree inferred from target enrichment data (admixed individuals are excluded). Local posterior probabilities >0.7 are shown above branches. c) Phylogenetic networks resulted from maximum pseudo-likelihood inference with PhyloNet. Inferred reticulation events (curved, blue lines) are accompanied by inheritance probabilities. The phenotype and provenance of operational taxonomic units is shown.

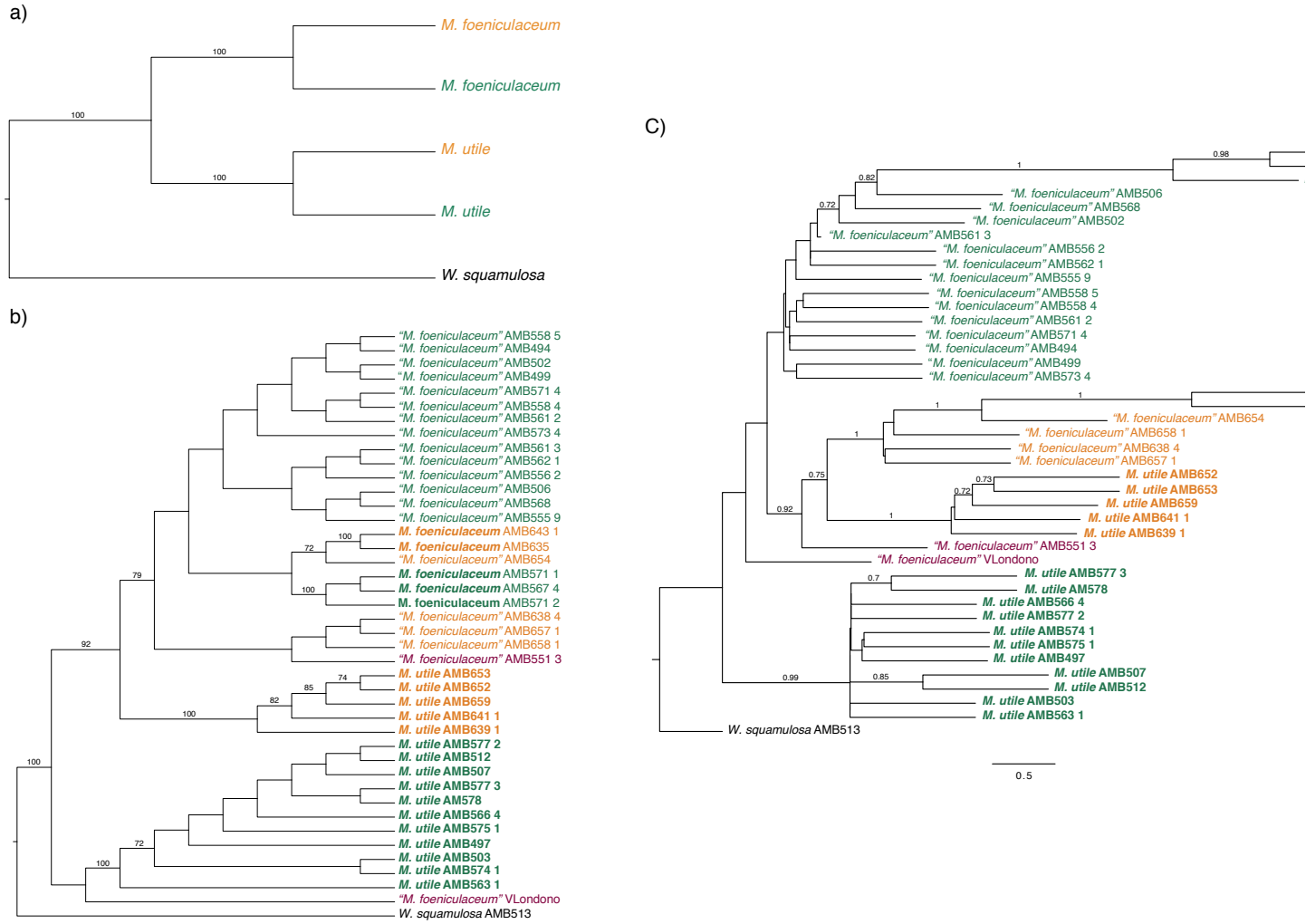


Fig S10. Phylogenetic trees inferred with target enrichment data using single-exon genes for comparison with the same analyses using concatenated multi-exon genes in Figs. S5, S6, and S7. Colors correspond to Antioquia (yellow), Caribe (green), and Boyacá (red). a) SVDquartets extended majority-rule consensus trees inferred after exclusion admixed individuals and of invariant and non-binary sites. Bootstrap support >70 is shown above branches. b) SVDquartets extended majority-rule consensus trees including admixed individuals and using species-as-tips. Bootstrap support >70 is shown above branches. Admixed individuals are indicated with quotation marks. c) ASTRAL III tree inferred including admixed individuals using species-as-tips. Local posterior probabilities > 0.7 are shown above branches.

Figure S11. Phylogenetic tree inferred with target enrichment data and under maximum likelihood using single-exon genes for comparison with the results from the same analysis in Fig. S8 using concatenated multi-exon genes. Colors correspond to Antioquia (yellow), Caribe (green), and Boyacá (red). Non-admixed individuals as inferred with ADMIXTURE are in bold, whereas admixed individuals are indicated with quotation marks. Bootstrap support >70 shown with red circles. Quartet sampling scores (Quartet Concordance/Quartet Differential/Quartet Informativeness) for the main clades (drainage basins and species) are shown above branches

New Phytologist

Table S1. List of accepted species of *Marathrum* and previously described species in the genus that are synonymous to *Marathrum foeniculaceum* sensu Novelo & Philbrick 1997, Novelo 2009, and Tippery 2011

Species	Status
Marathrum aeruginosum	Accepted
Marathrum allenii Woodson	Synonymous to *Marathrum foeniculaceum*
Marathrum azarensis	Accepted
Marathrum cheiriferum P.Royen	Synonymous to *Marathrum foeniculaceum*
Marathrum cubanum	Accepted
Marathrum elegans P.Royen	Synonymous to *Marathrum foeniculaceum*
Marathrum foeniculaceum	Accepted
Marathrum haenkeanum Engl.	Synonymous to *Marathrum foeniculaceum*
Marathrum indifferens P.Royen	Synonymous to *Marathrum foeniculaceum*
Marathrum leptophyllum P.Royen	Synonymous to *Marathrum foeniculaceum*
Marathrum minutiflorum Engl.	Synonymous to *Marathrum foeniculaceum*
Marathrum modestum Nash	Synonymous to *Marathrum foeniculaceum*
Marathrum oxycarpum Tul.	Synonymous to *Marathrum foeniculaceum*
Marathrum pauciflorum	Accepted
Marathrum plumosum	Accepted
Marathrum pusillum P.Royen	Synonymous to *Marathrum foeniculaceum*
Marathrum rubrum Novelo & C.T.Philbrick	Synonymous to *Marathrum foeniculaceum*
Marathrum schiedeanum Cham.	Synonymous to *Marathrum foeniculaceum*
Marathrum squamosum	Accepted
Marathrum stenocarpum (wedd. Ex DC) P.Royen	Synonymous to *Marathrum foeniculaceum*
Marathrum striatifolium	Accepted
Marathrum tenue	Accepted
Marathrum utile	Accepted

Table S2. Voucher numbers, NCBI accessions and locations for the samples collected and included in this study. Samples are arranged by species. Admixed individuals are indicated in quotations marks, and individuals that have the phenotype of *M. foeniculaceum* but whose genetic constitution was not assessed with admixture are indicated as morphs since they could correspond to *M. foeniculaceum* or to hybrids. The sample used for generating genome-skimming data is in bold (SRR12956182 for WGS read data). Individuals used in PhyloNet analyses are indicated with (*). Genomic data are deposited in the NCBI Sequence Read Archive under Bioproject PRJNA673497.

Collector No	**SRA accession # ddRADseq**	**SRA accession # target enrichment**	**GenBank accession numbers**	**Species**	**Coordinates**	**Basin/River**
AMB 497	**SRR12971533**	**SRR12975804**	**-**	***Marathrum utile***	**11°11'47.07''N, 73°27'35.53''W**	**Caribe/Ancho**
AMB 503	SRR12989095	SRR12983242	-	*Marathrum utile*	11°11'33.46"N, 73°34'59.01"W	Caribe/Chorro Catalina
AMB 507	SRR12971532	SRR12975803	-	*Marathrum utile*	11°6'13.93"N, 73°52'19.28"W	Caribe/Buritaca
AMB 508	SRR12971530	-	-	*Marathrum utile*	11°6'13.93"N, 73°52'19.28"W	Caribe/Buritaca
AMB 509	SRR12971529	-	-	*Marathrum utile*	11°6'13.93"N, 73°52'19.28"W	Caribe/Buritaca
*AMB 512	FAILED	SRR12975802	-	*Marathrum utile*	11°5'34.16"N, 73°53'12.51"W	Caribe/Buritaca
AMB 563-1	SRR12971527	SRR12975801	-	*Marathrum utile*	11°13'30"N, 73°41'50"W	Caribe/Don Diego
AMB 563-3	SRR12971526	-	-	*Marathrum utile*	11°13'30"N, 73°41'50"W	Caribe/Don Diego
AMB 563-4	SRR12971525	-	-	*Marathrum utile*	11°13'30"N, 73°41'50"W	Caribe/Don Diego
AMB 566-1	SRR12971524	-	-	*Marathrum utile*	11°12'2"N, 73°27'42"W	Caribe/Ancho
AMB 566-3	SRR12971523	-	-	*Marathrum utile*	11°12'2"N, 73°27'42"W	Caribe/Ancho
AMB 566-4	SRR12971522	SRR12975800	-	*Marathrum utile*	11°12'2"N, 73°27'42"W	Caribe/Ancho
AMB 566-6	SRR12971521	-	-	*Marathrum utile*	11°12'2"N, 73°27'42"W	Caribe/Ancho
AMB 566-8	SRR12971519	-	-	*Marathrum utile*	11°12'2"N, 73°27'42"W	Caribe/Ancho
AMB 566-9	SRR12971518	-	-	*Marathrum utile*	11°12'2"N, 73°27'42"W	Caribe/Ancho
AMB 566-14	SRR12971517	-	-	*Marathrum utile*	11°12'2"N, 73°27'42"W	Caribe/Ancho
AMB 566-18	SRR12971516	-	-	*Marathrum utile*	11°12'2"N, 73°27'42"W	Caribe/Ancho
*AMB 574-1	SRR12971515	SRR12975799	-	*Marathrum utile*	11°12'39"N, 73°15'8"W	Caribe/Jerez
AMB 574-2	SRR12971514	-	-	*Marathrum utile*	11°12'39"N, 73°15'8"W	Caribe/Jerez
AMB 575-1	SRR12971513	SRR12975798	-	*Marathrum utile*	11°12'39"N, 73°15'8"W	Caribe/Jerez
AMB 575-2	SRR12971512	-	-	*Marathrum utile*	11°12'39"N, 73°15'8"W	Caribe/Jerez
AMB 575-3	SRR12971511	-	-	*Marathrum utile*	11°12'39"N, 73°15'8"W	Caribe/Jerez
AMB 577-1	SRR12971510	-	-	*Marathrum utile*	10°42'25"N, 73°17'4"W	Caribe/Jerez
AMB 577-2	FAILED	SRR12975797	-	*Marathrum utile*	10°42'25"N, 73°17'4"W	Caribe/Badillo
AMB 577-3	SRR12971507	SRR12975796	-	*Marathrum utile*	10°42'25"N, 73°17'4"W	Caribe/Badillo
*AMB 578	-	SRR12983243	-	*Marathrum utile*	10°30'5"N, 73°16'9"W	Caribe/Guatapurí
AMB 639-1	SRR12971506	SRR12975795	-	*Marathrum utile*	6°16'0"N, 75°1'45"W	Antioquia/Arenal
AMB 639-2	SRR12971505	-	-	*Marathrum utile*	6°16'0"N, 75°1'45"W	Antioquia/Arenal
AMB 639-3	SRR12971504	-	-	*Marathrum utile*	6°16'0"N, 75°1'45"W	Antioquia/Arenal
AMB 639-4	SRR12971503	-	-	*Marathrum utile*	6°16'0"N, 75°1'45"W	Antioquia/Arenal
AMB 639-5	SRR12971502	-	-	*Marathrum utile*	6°16'0"N, 75°1'45"W	Antioquia/Arenal
*AMB 641-1	SRR12971501	SRR12975793	-	*Marathrum utile*	6°13'57"N, 74°54'58"W	Antioquia/Arenal
AMB 641-2	SRR12971500	-	-	*Marathrum utile*	6°13'57"N, 74°54'58"W	Antioquia/Arenal
AMB 641-3	SRR12971499	-	-	*Marathrum utile*	6°13'57"N, 74°54'58"W	Antioquia/Arenal
AMB 642-1	FAILED	FAILED	-	*Marathrum utile*	6°13'57"N, 74°54'58"W	Antioquia/Arenal
AMB 642-2	SRR12971497	-	-	*Marathrum utile*	6°13'57"N, 74°54'58"W	Antioquia/Arenal
AMB 651	SRR12971596	-	-	*Marathrum utile*	5°55'58"N, 75°5'45"W	Antioquia/Santo Domingo
*AMB 652	SRR12971595	SRR12975792	-	*Marathrum utile*	5°55'58"N, 75°5'45"W	Antioquia/Santo Domingo
AMB 653	SRR12971594	SRR12975791	-	*Marathrum utile*	5°54'51"N, 75°4'28"W	Antioquia/Santo Domingo

AMB 659	SRR129715930	SRR12975790	-	*Marathrum utile*	6°0'2"N, 74°56'15"W	Antioquia/Samaná Norte
AMB 567-4	SRR12971578	SRR12975808	-	*Marathrum foeniculaceum*	11°12'2"N, 73°27'42"W	Caribe/Ancho
*AMB 571-1	FAILED	SRR12975807	-	*Marathrum foeniculaceum*	11°13'46"N, 73°34'40"W	Caribe/Palomino
*AMB 571-2	SRR12971576	SRR12975806	-	*Marathrum foeniculaceum*	11°13'46"N, 73°34'40"W	Caribe/Palomino
*AMB 635	SRR13013619	SRR12983247	-	*Marathrum foeniculaceum*	6°16'4"N, 75°1'49"W	Antioquia/Arenal
*AMB 643-1	-	SRR12983246		*Marathrum foeniculaceum*	6°13'57"N, 74°54'58"W	Antioquia/Arenal
AMB 494	SRR12971587	SRR12975817	-	*"Marathrum foeniculaceum"*	11°13'7.63"N, 73°42'0.14"W	Caribe/Don Diego
AMB 499	SRR12971586	SRR12975816	-	*"Marathrum foeniculaceum"*	11°12'22.95"N, 73°15'17.26"W	Caribe/Jerez
AMB 502	SRR12971575	SRR12975805	-	*"Marathrum foeniculaceum"*	11°11'33.46"N, 73°34'59.01"W	Caribe/Chorro Catalina
AMB 506	SRR13013618	SRR12983248	-	*"Marathrum foeniculaceum"*	11°16'33.96"N, 73°55'26.57"W	Caribe/ Piedras
*AMB 551-3	-	SRR12983244		*"Marathrum foeniculaceum"*	4°53'29"N, 73°16'31"W	Boyacá/Batá
AMB 555-9	SRR12971564	SRR12975794	-	*"Marathrum foeniculaceum"*	11°16'33.96"N, 73°55'26.57"W	Caribe/Piedras
AMB 556-2	SRR12971553	SRR12975789	-	*"Marathrum foeniculaceum"*	11°16'0"N, 73°51'56"W	Caribe/Mendihuaca
AMB 558-4	SRR12971542	SRR12975788	-	*"Marathrum foeniculaceum"*	11°14'41"N, 73°50'35"W	Caribe/Guachaca
AMB 558-5	SRR12971531	SRR12975787	-	*"Marathrum foeniculaceum"*	11°14'41"N, 73°50'35"W	Caribe/Guachaca
AMB 561-2	SRR12971520	SRR12975786	-	*"Marathrum foeniculaceum"*	11°14'11"N, 73°41'51"W	Caribe/Don Diego
AMB 561-3	SRR12971509	SRR12975785	-	*"Marathrum foeniculaceum"*	11°14'11"N, 73°41'51"W	Caribe/Don Diego
AMB 562-1	SRR12971498	SRR12975784	-	*"Marathrum foeniculaceum"*	11°13'56"N, 73°42'2"W	Caribe/Don Diego
AMB 568	SRR12971585	SRR12975815	-	*"Marathrum foeniculaceum"*	11°12'2"N, 73°27'42"W	Caribe/Ancho
*AMB 571-4	SRR12971584	SRR12975814	-	*"Marathrum foeniculaceum"*	11°13'46"N, 73°34'40"W	Caribe/Palomino
*AMB 573-4	FAILED	SRR12975813	-	*"Marathrum foeniculaceum"*	11°12'39"N, 73°15'8"W	Caribe/Jerez
*AMB 638-4	SRR12971582	SRR12975812	-	*"Marathrum foeniculaceum"*	6°16'0"N, 75°1'45"W	Antioquia/Arenal
AMB 654	SRR12971581	SRR12975811	-	*"Marathrum foeniculaceum"*	5°54'51"N, 75°4'28"W	Antioquia/Santo Domingo
*AMB 657	SRR13013620	SRR12983245	-	*"Marathrum foeniculaceum"*	5°54'56"N, 74°58'45"W	Antioquia/Samaná Norte
AMB 658-1	SRR12971580	SRR12975810	-	*"Marathrum foeniculaceum"*	6°0'2"N, 74°56'15"W	Antioquia/Samaná Norte
*V.Londono	SRR12971579	SRR12975809	-	*"Marathrum foeniculaceum"*	4°53'29"N, 73°16'31"W	Boyacá/Batá
AMB 492	FAILED	-	-	*Marathrum foeniculaceum* (morph)	11°13'7.63"N, 73°42'0.14"W	Caribe/Don Diego
AMB 493	SRR12971574	-	-	*Marathrum foeniculaceum* (morph)	11°13'7.63"N, 73°42'0.14"W	Caribe/Don Diego
AMB 495	FAILED	-	-	*Marathrum foeniculaceum* (morph)	11°13'19.31"N, 73°31'38.01"W	Caribe/San Salvador
AMB 496	SRR12971573	-	-	*Marathrum foeniculaceum* (morph)	11°13'19.31"N, 73°31'38.01"W	Caribe/San Salvador
AMB 498	FAILED	-	-	*Marathrum foeniculaceum* (morph)	11°11'47.07"N, 73°27'35.53"W	Caribe/Ancho
AMB 500	SRR12971572	-	-	*Marathrum foeniculaceum* (morph)	11°12'22.95"N, 73°15'17.26"W	Caribe/Jerez
AMB 551-4	SRR12971571	-	-	*Marathrum foeniculaceum* (morph)	4°53'29"N, 73°16'31"W	Boyacá/Batá
AMB 551-5	SRR12971570	-	-	*Marathrum foeniculaceum* (morph)	4°53'29"N, 73°16'31"W	Boyacá/Batá
AMB 551-6	SRR12971569	-	-	*Marathrum foeniculaceum* (morph)	4°53'29"N, 73°16'31"W	Boyacá/Batá
AMB 551-7	SRR12971568	-	-	*Marathrum foeniculaceum* (morph)	4°53'29"N, 73°16'31"W	Boyacá/Batá
AMB 551-8	SRR12971567	-	-	*Marathrum foeniculaceum* (morph)	4°53'29"N, 73°16'31"W	Boyacá/Batá
AMB 551-10	SRR12971566	-	-	*Marathrum foeniculaceum* (morph)	4°53'29"N, 73°16'31"W	Boyacá/Batá
AMB 556-1	SRR12971565	-	-	*Marathrum foeniculaceum* (morph)	11°16'0"N, 73°51'56"W	Caribe/Mendihuaca
AMB 556-3	SRR12971563	-	-	*Marathrum foeniculaceum* (morph)	11°16'0"N, 73°51'56"W	Caribe/Mendihuaca
AMB 556-4	SRR12971562	-	-	*Marathrum foeniculaceum* (morph)	11°16'0"N, 73°51'56"W	Caribe/Mendihuaca
AMB 556-5	SRR12971561	-	-	*Marathrum foeniculaceum* (morph)	11°16'0"N, 73°51'56"W	Caribe/Mendihuaca
AMB 558-1	SRR12971560	-	-	*Marathrum foeniculaceum* (morph)	11°14'41"N, 73°50'35"W	Caribe/Guachaca
AMB 558-2	SRR12971559	-	-	*Marathrum foeniculaceum* (morph)	11°14'41"N, 73°50'35"W	Caribe/Guachaca
AMB 558-8	SRR12971558	-	-	*Marathrum foeniculaceum* (morph)	11°14'41"N, 73°50'35"W	Caribe/Guachaca
AMB 558-10	SRR12971557	-	-	*Marathrum foeniculaceum* (morph)	11°14'41"N, 73°50'35"W	Caribe/Guachaca
AMB 558-11	SRR12971556	-	-	*Marathrum foeniculaceum* (morph)	11°14'41"N, 73°50'35"W	Caribe/Guachaca
AMB 562-2	SRR12971555	-	-	*Marathrum foeniculaceum* (morph)	11°13'56"N, 73°42'2"W	Caribe/Don Diego

AMB 562-5	SRR12971554	-	-	*Marathrum foeniculaceum* (morph)	11°13'56"N, 73°42'2"W	Caribe/Don Diego
AMB 567-3	SRR12971552	-	-	*Marathrum foeniculaceum* (morph)	11°12'2"N, 73°27'42"W	Caribe/Ancho
AMB 567-5	FAILED	-	-	*Marathrum foeniculaceum* (morph)	11°12'2"N, 73°27'42"W	Caribe/Ancho
AMB 567-6	SRR12971551	-	-	*Marathrum foeniculaceum* (morph)	11°12'2"N, 73°27'42"W	Caribe/Ancho
AMB 567-7	SRR12971550	-	-	*Marathrum foeniculaceum* (morph)	11°12'2"N, 73°27'42"W	Caribe/Ancho
AMB 571-3	FAILED	-	-	*Marathrum foeniculaceum* (morph)	11°13'46"N, 73°34'40"W	Caribe/Palomino
AMB 571-5	SRR12971549	-	-	*Marathrum foeniculaceum* (morph)	11°13'46"N, 73°34'40"W	Caribe/Palomino
AMB 571-6	SRR12971548	-	-	*Marathrum foeniculaceum* (morph)	11°13'46"N, 73°34'40"W	Caribe/Palomino
AMB 573-1	SRR12971547	-	-	*Marathrum foeniculaceum* (morph)	11°12'39"N, 73°15'8"W	Caribe/Jerez
AMB 573-2	SRR12971546	-	-	*Marathrum foeniculaceum* (morph)	11°12'39"N, 73°15'8"W	Caribe/Jerez
AMB 573-3	SRR12971545	-	-	*Marathrum foeniculaceum* (morph)	11°12'39"N, 73°15'8"W	Caribe/Jerez
AMB 573-5	SRR12971544	-	-	*Marathrum foeniculaceum* (morph)	11°12'39"N, 73°15'8"W	Caribe/Jerez
AMB 573-6	SRR12971543	-	-	*Marathrum foeniculaceum* (morph)	11°12'39"N, 73°15'8"W	Caribe/Jerez
AMB 631-2	SRR12971541	-	-	*Marathrum foeniculaceum* (morph)	6°16'11"N, 75°1'54"W	Antioquia/Arenal
AMB 631-4	FAILED	-	-	*Marathrum foeniculaceum* (morph)	6°16'11"N, 75°1'54"W	Antioquia/Arenal
AMB 632	FAILED	FAILED	-	*Marathrum foeniculaceum* (morph)	6°16'11"N, 75°1'54"W	Antioquia/Arenal
AMB 636	SRR12971540	-	-	*Marathrum foeniculaceum* (morph)	6°16'0"N, 75°1'45"W	Antioquia/Arenal
AMB 638-3	SRR12971539	-	-	*Marathrum foeniculaceum* (morph)	6°16'0"N, 75°1'45"W	Antioquia/Arenal
AMB 640	SRR12971538	-	-	*Marathrum foeniculaceum* (morph)	6°16'0"N, 75°1'45"W	Antioquia/Arenal
AMB 642-3	FAILED	-	-	*Marathrum foeniculaceum* (morph)	6°13'57"N, 74°54'58"W	Antioquia/Arenal
AMB 645-1	SRR12971537	FAILED	-	*Marathrum foeniculaceum* (morph)	6°11'38"N, 75°0'18"W	Antioquia/Bañadero San Carlos
AMB 646-1	FAILED	FAILED	-	*Marathrum foeniculaceum* (morph)	6°11'52"N, 75°1'41"W	Antioquia/
AMB 646-4	SRR12971536	-	-	*Marathrum foeniculaceum* (morph)	6°11'52"N, 75°1'41"W	Antioquia/
AMB 646-5	FAILED	-	-	*Marathrum foeniculaceum* (morph)	6°11'52"N, 75°1'41"W	Antioquia/
AMB 646-6	FAILED	-	-	*Marathrum foeniculaceum* (morph)	6°11'52"N, 75°1'41"W	Antioquia/
AMB 655	SRR12971535	-	-	*Marathrum foeniculaceum* (morph)	5°55'44"N, 75°0'44"W	Antioquia/Samaná Norte
AMB 656-1	SRR12971534	-	-	*Marathrum foeniculaceum* (morph)	5°54'56"N, 74°58'45"W	Antioquia/Samaná Norte
AMB 658-2	FAILED	-	-	*Marathrum foeniculaceum* (morph)	6°0'2"N, 74°56'15"W	Antioquia/Samaná Norte
*AMB 513	FAILED	SRR1298324	-	*Weddellina squamulosa*	0°51'42.9"N, 71°0'47.5"W	Amazonas/Vaupés
AMB 514	FAILED	-	-	*Weddellina squamulosa*	0°51'42.9"N, 71°0'47.5"W	Amazonas/Vaupés
AMB 515	FAILED	-	-	*Weddellina squamulosa*	0°51'42.9"N, 71°0'47.5"W	Amazonas/Vaupés
-	-	-	NC_036341	*Garcinia mangostana*	-	-
BNDE	-	-		*Hypericum perforatum*	-	-
-	-	-	KT337313.1	*Populus tremula*	-	-

Table S3. Number of loci for which N taxa have data. Ipyrad results for the assembly of all 95 ddRADseq samples and for minimum 4 individuals per loci.

N	Locus coverage	Sum coverage	N	Locus coverage	Sum coverage	N	Locus coverage	Sum coverage
1	0	0	**38**	154	168201	**75**	0	169462
2	0	0	**39**	136	168337	**76**	1	169463
3	0	0	**40**	123	168460	**77**	2	169465
4	46519	46519	**41**	105	168565	**78**	1	169466
5	28200	74719	**42**	86	168651	**79**	4	169470
6	20962	95681	**43**	73	168724	**80**	2	169472
7	13485	109166	**44**	75	168799	**81**	1	169473
8	10892	120058	**45**	89	168888	**82**	0	169473
9	7606	127664	**46**	68	168956	**83**	1	169474
10	6563	134227	**47**	75	169031	**84**	0	169474
11	5050	139277	**48**	70	169101	**85**	0	169474
12	4442	143719	**49**	49	169150	**86**	0	169474
13	3471	147190	**50**	44	169194	**87**	0	169474
14	2973	150163	**51**	55	169249	**88**	0	169474
15	2310	152473	**52**	32	169281	**89**	0	169474
16	2112	154585	**53**	37	169318	**90**	0	169474
17	1721	156306	**54**	36	169354	**91**	0	169474
18	1510	157816	**55**	23	169377	**92**	0	169474
19	1210	159026	**56**	21	169398	**93**	0	169474
20	1131	160157	**57**	17	169415	**94**	0	169474
21	993	161150	**58**	16	169431	**95**	0	169474
22	868	162018	**59**	6	169437			
23	762	162780	**60**	8	169445			
24	663	163443	**61**	2	169447			
25	563	164006	**62**	1	169448			
26	555	164561	**63**	1	169449			
27	454	165015	**64**	1	169450			
28	499	165514	**65**	2	169452			
29	391	165905	**66**	4	169456			
30	375	166280	**67**	0	169456			
31	349	166629	**68**	0	169456			
32	314	166943	**69**	1	169457			
33	254	167197	**70**	0	169457			
34	245	167442	**71**	1	169458			
35	202	167644	**72**	1	169459			
36	211	167855	**73**	3	169462			
37	192	168047	**74**	0	169462			

Table S4. Ipyrad assembly statistics for our total 95, subset of 34 admixed and non-admixed, and 17 non-admixed ddRADseq samples. Statistics are shown for loci shared across minimum 4 and 17 of the samples in our subset of 34 samples, and minimum 4 samples in our subset of 17 non-admixed individuals.

	95 samples			**34 Samples**						**17 samples (non-admixed)**		
	Min4			*Min4* (~88% missing data)			*Min17* (50% missing data)			*Min4* (~23% missing data)		
	Total filters	Applied order	Retained loci	Total filters	Applied order	Retained loci	Total filters	Applied order	Retained loci	Total filters	Applied order	Retained loci
Total prefiltered loci	0	0	613888	0	0	78080	0	0	78080	0	0	86753
Filtered by rm duplicates	9604	9604	604284	2348	2348	75732	2348	2348	75732	2251	2251	84502
Filtered by max indels	1092	1092	603192	172	172	75560	12	12	75720	66	66	84436
Filtered by max SNPs	2809	2765	600427	330	327	75233	1	1	75719	76	76	84360
Filtered by max shared heterozygosity	48451	47836	552591	9820	9733	65500	304	301	75418	5514	5497	78863
Filtered by min sample	383117	383117	169474	37664	37477	28023	76930	74739	679	67497	67497	11366
Total filtered loci	**445073**	**444414**	**169474**	**50334**	**50057**	**28023**	**79595**	**77401**	**679**	**75404**	**75387**	**11366**

Table S5. Target enrichment recovery efficiency. The number of sequenced reads, exons and genes recovered out of the total 536 and 217 targeted respectively is shown. Target length (TL). The table is organized in descending order from the sample with most exons recovered at 75% of the TL. *Marathrum* samples with relatively short exons were excluded from downstream analyses but the outgroup was included (bold). The first sample listed is the one used for selection of sequences to be targeted with the pipeline Sondovac. Admixed individuals with the phenotype of *M. foeniculaceum* are indicated with quotation marks.

Sample	# Reads	% reads on target	# Exons recovered					% of the total number of targeted exons	Genes recovered after paralog filtering	% of the total number of targeted genes	# paralog warnings
			>25% TL	% of the total number of targeted exons	>50% TL	% of the total number of targeted exons	>75% TL				
M. utile AMB497	18152507	89.8	535	99.81	535	99.81	534	99.63	208	95.41	33
M. utile AMB574-1	6015552	87.7	529	98.69	527	98.32	527	98.32	208	95.41	32
M. utile AMB503	20726495	89.3	526	98.13	526	98.13	526	98.13	208	95.41	27
M. foeniculaceum AMB643-1	4444374	85.6	526	98.13	526	98.13	526	98.13	208	95.41	45
M. foeniculaceum AMB571-1	6308716	84.2	527	98.32	526	98.13	526	98.13	208	95.41	31
M. utile AMB659	6645871	86.9	526	98.13	526	98.13	526	98.13	208	95.41	38
M. utile AMB641-1	6598011	86.9	526	98.13	525	97.95	524	97.76	208	95.41	27
M. utile AMB653	11275084	88.6	523	97.57	522	97.39	522	97.39	208	95.41	38
M. utile AMB512	4921124	86.7	525	97.95	523	97.57	522	97.39	208	95.41	25
M. utile AMB652	10640199	88.1	523	97.57	522	97.39	521	97.20	208	95.41	39
"M. foeniculaceum" AMB494	8038995	89.1	529	98.69	527	98.32	521	97.20	203	93.12	154
"M. foeniculaceum" AMB499	9799316	88.3	531	99.07	529	98.69	521	97.20	208	95.41	173
M. foeniculaceum AMB567-4	845956	48.5	524	97.76	522	97.39	520	97.01	208	95.41	20
M. utile AMB639-1	4951653	88.2	521	97.20	521	97.20	519	96.83	208	95.41	26
M. utile AMB507	843464	68.6	519	96.83	519	96.83	519	96.83	208	95.41	29
M. utile AMB577-3	9935840	86.8	520	97.01	519	96.83	519	96.83	208	95.41	16
M. foeniculaceum AMB635	4098778	86.3	522	97.39	520	97.01	518	96.64	208	95.41	38
"M. foeniculaceum" AMB562-1	871580	48.3	522	97.39	521	97.20	518	96.64	202	92.66	87
M. utile AMB577-2	15127071	89.7	519	96.83	518	96.64	517	96.46	208	95.41	29
"M. foeniculaceum" AMB571-4	6618910	84.5	530	98.88	527	98.32	517	96.46	203	93.12	132
"M. foeniculaceum" AMB502	11814928	90.4	526	98.13	524	97.76	516	96.27	203	93.12	194
"M. foeniculaceum" AMB561-2	2350936	80.9	528	98.51	526	98.13	516	96.27	203	93.12	129
M. utile AMB575-1	1971791	76.1	517	96.46	517	96.46	516	96.27	208	95.41	16
"M. foeniculaceum" AMB654	10076885	87	526	98.13	524	97.76	515	96.08	203	93.12	130
M. utile AMB578	9519370	92.2	516	96.27	516	96.27	515	96.08	208	95.41	17
M. utile AMB566-4	1589623	75.4	518	96.64	517	96.46	515	96.08	208	95.41	20
"M. foeniculaceum" AMB558-5	3386292	84.7	528	98.51	523	97.57	513	95.71	203	93.12	138
"M. foeniculaceum" AMB658-1	4293245	84.7	521	97.20	517	96.46	513	95.71	203	93.12	110
"M. foeniculaceum" AMB573-4	8038288	83.7	525	97.95	524	97.76	512	95.52	203	93.12	147
"M. foeniculaceum" AMB638-4	1878381	77.7	521	97.20	518	96.64	512	95.52	203	93.12	114
"M. foeniculaceum" AMB558-4	2851635	80.3	525	97.95	522	97.39	512	95.52	203	93.12	135
"M. foeniculaceum" AMB561-3	591426	34.9	514	95.90	513	95.71	510	95.15	203	93.12	70
"M. foeniculaceum" AMB657	2993130	82	520	97.01	516	96.27	510	95.15	203	93.12	117
"M. foeniculaceum" AMB506	11255969	92.3	518	96.64	513	95.71	507	94.59	203	93.12	171
"M. foeniculaceum" VLondono	7047222	92.7	515	96.08	512	95.52	506	94.40	208	95.41	123
"M. foeniculaceum" AMB555-9	507428	67	506	94.40	506	94.40	503	93.84	203	93.12	78

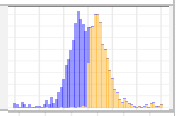
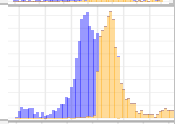
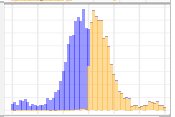
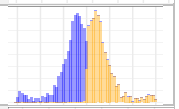
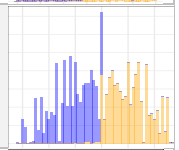
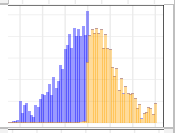
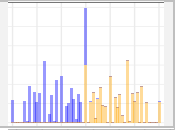
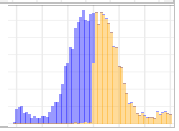
"M. foeniculaceum" AMB551-3	10403228	85.6	521	97.20	514	95.90	500	93.28	208	95.41	154
M. foeniculaceum AMB571-2	1658509	93.7	517	96.46	506	94.40	479	89.37	208	95.41	15
"M. foeniculaceum" AMB568	576529	19.6	498	92.91	492	91.79	478	89.18	202	92.66	23
M. utile AMB563-1	497235	16.3	489	91.23	484	90.30	470	87.69	206	94.50	16
"M. foeniculaceum" AMB556-2	447726	12.2	445	83.02	431	80.41	405	75.56	201	92.20	13
***M. foeniculaceum* AMB646-1**	**528950**	**83.1**	**481**	**89.74**	**439**	**81.90**	**341**	**63.62**	**--**	**--**	**6**
***M. utile* AMB642-1**	**269227**	**76.2**	**450**	**83.96**	**387**	**72.20**	**263**	**49.07**	**--**	**--**	**0**
***M. foeniculaceum* AMB632**	**125422**	**57.9**	**305**	**56.90**	**204**	**38.06**	**119**	**22.20**	**--**	**--**	**0**
***M. foeniculaceum* AMB645**	**364733**	**10.5**	**90**	**16.79**	**60**	**11.19**	**47**	**8.77**	**--**	**--**	**22**
W. squamulosa AMB513	893791	37.5	318	59.33	285	53.17	232	43.28	203	93.12	47

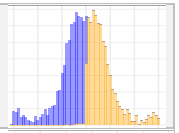
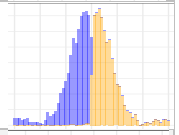
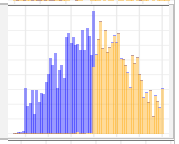
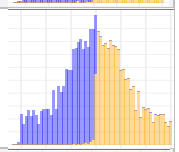
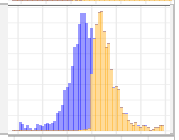
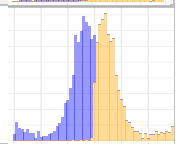
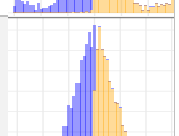
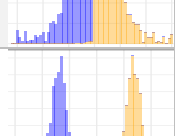
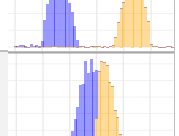
Table S7. Overall heterozygosity of each individual as estimated with VCFtools (--het function). Admixed individuals are indicated with quotation marks and non-admixed individuals are indicated with "(pure)"

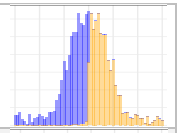
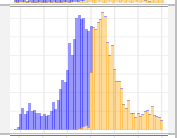
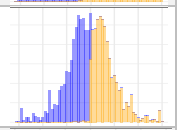
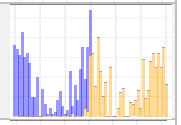
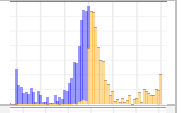
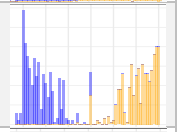
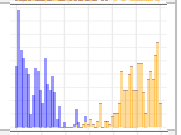
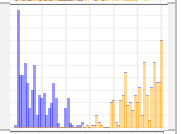
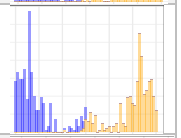
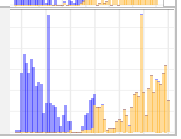
Individual	Observed homozygosity	Expected homozygosity	# Biallelic SNPs	Inbreeding coefficient (F)
"M.foeniculaceum" AMB499	64	129	193	-1.01621
"M.foeniculaceum" AMB502	64	129	193	-1.01621
"M.foeniculaceum" AMB506	64	129	193	-1.01621
"M.foeniculaceum" AMB5513	116	129	193	-0.20348
"M.foeniculaceum" AMB5559	65	129	193	-1.00058
"M.foeniculaceum" AMB5562	121	129	193	-0.12533
"M.foeniculaceum" AMB5584	63	129	193	-1.03184
"M.foeniculaceum" AMB5585	63	129	193	-1.03184
"M.foeniculaceum" AMB5612	64	129	193	-1.01621
"M.foeniculaceum" AMB5613	70	129	193	-0.92244
"M.foeniculaceum" AMB5621	66	129	193	-0.98495
M.foeniculaceum AMB5674 **(pure)**	185	129	193	0.87496
"M.foeniculaceum" AMB568	89	129	193	-0.62547
M.foeniculaceum AMB5711 **(pure)**	191	129	193	0.96874
M.foeniculaceum AMB5712 **(pure)**	190	129	193	0.95311
"M.foeniculaceum" AMB5714	64	129	193	-1.01621
"M.foeniculaceum" AMB5734	63	129	193	-1.03184
M.foeniculaceum AMB635 **(pure)**	187	129	193	0.90622
"M.foeniculaceum" AMB6384	107	129	193	-0.34414
M.foeniculaceum AMB6431 **(pure)**	191	129	193	0.96874
"M.foeniculaceum" AMB654	109	129	193	-0.31288
"M.foeniculaceum" AMB657	105	129	193	-0.3754
"M.foeniculaceum" AMB6581	108	129	193	-0.32851
"M.foeniculaceum" VLondono	116	129	193	-0.20348
"M.foeniculaceum" AMB494	64	129	193	-1.01621
M.utile AMB497 **(pure)**	192	129	193	0.98437
M.utile AMB503 **(pure)**	193	129	193	1
M.utile AMB507 **(pure)**	191	129	193	0.96874
M.utile AMB512 **(pure)**	192	129	193	0.98437
M.utile AMB5631 **(pure)**	188	129	193	0.92185
M.utile AMB5664 **(pure)**	193	129	193	1
M.utile AMB5741 **(pure)**	193	129	193	1
M.utile AMB5751 **(pure)**	193	129	193	1
M.utile AMB5772 **(pure)**	192	129	193	0.98437
M.utile AMB5773 **(pure)**	193	129	193	1
M.utile AMB578 **(pure)**	193	129	193	1
M.utile AMB6391 **(pure)**	181	129	193	0.81245
M.utile AMB6411 **(pure)**	184	129	193	0.85933
M.utile AMB652 **(pure)**	179	129	193	0.78119
M.utile AMB653 **(pure)**	180	129	193	0.79682
M.utile AMB659 **(pure)**	181	129	193	0.81245

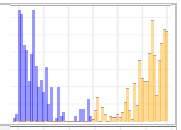
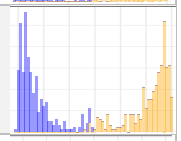
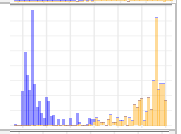
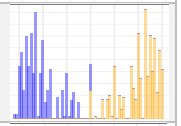
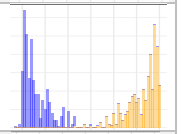
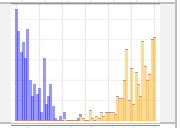
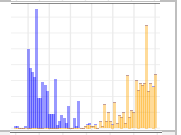
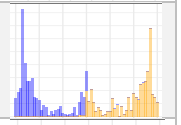
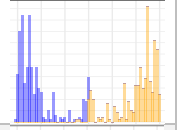
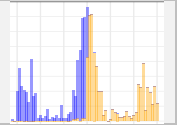
New Phytologist

Table S8. nQuire results for ploidy level estimation for each individual. Coverage (Cov.), percentage of data kept, log-likelihood of the free model, estimated ploidy level (M), and log-likelihood of the experimental data under the ploidy model (*logL*) are shown. The fit of M to the data is estimated through the delta likelihood (ΔLog*L*) against the free model. The fit between the ideal and empirical histograms for each sample and model is determined based on a low sum of squared residuals (SSR), a positive slope (y-y-slope) with low standard error (Std.Error) and high R^2. Chosen models and the values that support them are shown in red. Values supporting more than one model are shown in bold. Histograms of allele frequency and read counts for the first (blue) and second (orange) alleles for each individual are shown.

Individual	ADMIX	Cov.	% kept	Log*L* free	M	log*L*	ΔLog*L*	SSR	y-y slope	Std.Err	R^2	Histogram
M. foeniculaceum AMB494	Admixed	3,465	89.2	8344.6	2x	7891.9	452.6	0.004	0.756	0.031	0.907	
					3x	2384.3	5960.2	0.068	-0.644	0.124	0.310	
					4x	4665.8	3678.7	0.031	0.122	0.233	0.004	
M. foeniculaceum AMB499	Admixed	3,802	92.3	8025.7	2x	7181.2	844.4	0.007	0.669	0.037	0.844	
					3x	2542.2	5483.4	0.063	-0.611	0.112	0.332	
					4x	4575.0	5018.6	0.027	0.113	0.214	0.004	
M. foeniculaceum AMB502	Admixed	3,804	92.6	7702.7	2x	7105.3	597.3	0.007	0.642	0.029	0.888	
					3x	2561.0	5141.5	0.058	-0.566	0.105	0.326	
					4x	4371.0	3331.6	0.025	0.075	0.200	0.002	
M. foeniculaceum AMB506	Admixed	3,741	91.0	7560.1	2x	7003.0	556.9	0.008	0.618	0.035	0.840	
					3x	2541.3	5018.6	0.057	-0.548	0.105	0.312	
					4x	4226.2	3333.8	0.026	-0.000	0.198	4.771e-08	
M. foeniculaceum AMB568	Admixed	1,727	53.9	2034.6	2x	1954.3	80.2	0.032	0.178	0.049	0.179	
					3x	1295.5	739.0	**0.026**	-0.011	0.079	0.000	
					4x	994.4	1040.1	**0.020**	-0.347	0.11	0.132	
M. foeniculaceum AMB555-9	Admixed	3,348	84.7	4684.2	2x	4577.7	106.4	**0.018**	0.361	0.029	0.713	
					3x	2292.1	2392.0	0.034	-0.230	0.075	0.137	
					4x	2469.5	2214.7	**0.018**	-0.243	0.122	0.063	
M. foeniculaceum AMB556-2	Admixed	727	34.4	779.6	2x	683.2	96.4	0.048	0.152	0.091	0.045	
					3x	540.3	239.3	0.036	0.083	0.135	0.006	
					4x	405.0	374.5	0.031	-0.167	0.210	0.010	
M. foeniculaceum AMB558-4	Admixed	3,817	96.4	7767.2	2x	6463.3	505.9	0.010	0.545	0.032	0.830	
					3x	2562.8	4406.4	0.051	-0.488	0.093	0.315	
					4x	3841.9	3127.3	0.023	-0.073	0.176	0.002	

M. foeniculaceum AMB558-5	Admixed	3,630	88.8	7142.6	2x	6579.4	563.2	0.008	0.606	0.029	0.876	
					3x	2469.3	4673.3	0.055	-0.528	0.101	0.314	
					4x	4022.7	3119.9	0.024	-0.000	0.191	1.39075e-08	
M. foeniculaceum AMB561-2	Admixed	3,515	88.2	7673.1	2x	7228.2	444.9	0.006	0.676	0.029	0.901	
					3x	2421.5	5251.5	0.060	-0.574	0.112	0.307	
					4x	4285.3	3387.8	0.027	0.036	0.210	0.000	
M. foeniculaceum AMB561-3	Admixed	3,508	93.1	3683.0	2x	3486.9	196.0	0.029	0.161	0.024	**0.421**	
					3x	2515.5	1167.5	0.022	-0.036	0.046	0.010	
					4x	1945.5	1737.5	0.014	-0.254	0.065	**0.203**	
M. foeniculaceum AMB562-1	Admixed	3,571	87.1	4427.3	2x	4230.9	196.3	0.021	0.293	0.023	0.731	
					3x	2443.4	1983.8	0.031	-0.204	0.059	0.167	
					4x	2345.6	2081.6	**0.014**	-0.155	0.099	0.039	
M. foeniculaceum AMB571-4	Admixed	3,549	88.3	7767.2	2x	7494.3	272.8	0.006	0.684	0.031	0.885	
					3x	2447.7	5319.4	0.061	-0.569	0.116	0.289	
					4x	4354.8	3412.3	0.028	0.033	0.214	0.000	
M. foeniculaceum AMB573-4	Admixed	3,780	92.6	7831.3	2x	7001.6	829.7	0.007	0.657	0.031	0.879	
					3x	2549.7	5281.6	0.061	-0.600	0.107	0.346	
					4x	4491.1	3340.2	0.025	0.118	0.206	0.005	
M. foeniculaceum AMB638-4	Admixed	3,386	85.8	6659.0	2x	6472.1	186.9	0.007	0.623	0.025	0.912	
					3x	2319.5	4339.4	0.053	-0.472	0.107	0.246	
					4x	3740.0	2918.9	0.025	-0.036	0.192	0.000	
**M. foeniculaceum* AMB654	Admixed	2,403	61.1	6798.9	2x	2642.5	4156.3	0.096	-0.301	0.110	0.111	
					3x	4757.7	2041.1	0.020	0.823	0.132	**0.395**	
					4x	1232.8	5566.0	0.065	-1.646	**0.156**	**0.652**	
M. foeniculaceum AMB657	Admixed	3,374	85.1	7305.3	2x	7095.5	209.8	0.005	0.699	0.024	0.930	
					3x	2320.1	4985.2	0.060	-0.543	0.117	0.265	
					4x	4150.2	3155.0	0.027	0.076	0.213	0.002	

M. foeniculaceum AMB658-1	Admixed	3,353	84.3	6528.3	2x	6213.9	314.3	0.008	0.610	0.030	0.873	
					3x	2268.6	4259.6	0.055	-0.519	0.103	0.298	
					4x	3682.6	2845.6	0.025	-0.019	0.192	0.000	
M. foeniculaceum VLondono	Admixed	3,899	97.8	6083.4	2x	5659.0	424.3	**0.014**	0.453	0.029	0.801	
					3x	2605.0	3478.3	0.044	-0.404	0.080	0.301	
					4x	3368.8	2714.	**0.018**	-0.048	0.149	0.001	
M. foeniculaceum AMB551-3	Admixed	2,979	78.5	5769.8	2x	5618.1	151.6	0.009	0.596	0.033	0.845	
					3x	2019.0	3750.7	0.053	-0.456	0.107	0.234	
					4x	3172.3	2597.5	0.026	-0.087	0.191	0.003	
M. foeniculaceum AMB643-1	"Pure"	507	47	646.7	2x	217.5	429.1	0.037	0.162	**0.062**	0.103	
					3x	260.3	386.3	0.045	-0.427	**0.078**	**0.336**	
					4x	600.6	46.0	0.005	0.809	0.105	**0.497**	
**M. foeniculaceum* AMB635	"Pure"	1,196	62	2335.9	**2x**	1506.7	829.1	0.004	**0.741**	0.030	0.908	
					3x	703.4	1632.4	0.068	-0.666	0.119	0.346	
					4x	1547.3	788.5	0.022	**0.631**	0.214	0.128	
M. foeniculaceum AMB567-4	"Pure"	494	52.8	601.6	2x	49.0	552.5	0.079	-0.320	0.063	**0.304**	
					3x	347.4	254.1	0.030	0.043	0.110	0.002	
					4x	543.7	57.8	0.014	0.422	0.162	0.102	
**M. foeniculaceum* AMB571-1	"Pure"	344	47.4	469.9	2x	59.7	410.1	0.078	-0.331	0.057	**0.361**	
					3x	315.3	154.5	0.022	**0.217**	0.100	0.073	
					4x	389.4	80.4	0.015	**0.276**	0.158	0.049	
M. foeniculaceum AMB571-2	"Pure"	301	31.6	398.1	2x	36.4	361.6	0.082	-0.338	0.067	**0.300**	
					3x	228.7	169.4	0.032	0.045	0.116	0.002	
					4x	323.7	74.4	0.018	0.278	0.178	0.039	
M. utile AMB497	"Pure"	426	64.1	667.3	2x	53.3	614.0	0.076	-0.240	**0.080**	0.132	
					3x	300.5	366.8	0.039	-0.086	0.124	0.008	
					4x	623.4	43.9	0.009	0.993	0.146	**0.438**	
M. utile AMB503	"Pure"	466	67.8	583.4	2x	112.7	470.6	0.064	-0.177	**0.062**	0.119	
					3x	345.8	237.5	0.026	0.077	**0.096**	0.010	
					4x	511.3	72.0	0.010	0.530	0.134	**0.208**	

M. utile AMB507	"Pure"	566	52.6	842.5	2x	22.0	820.5	0.081	-0.305	0.075	**0.214**	
					3x	356.7	485.8	0.040	-0.105	0.123	0.012	
					4x	782.6	59.8	0.012	0.782	0.165	**0.275**	
M. utile AMB512	"Pure"	387	54.7	620.5	2x	14.7	605.8	0.081	-0.276	**0.082**	0.159	
					3x	226.5	394.0	0.046	-0.233	0.127	0.053	
					4x	579.8	40.7	0.012	0.961	0.161	**0.374**	
M. utile AMB578	"Pure"	350	52.9	619.2	2x	6.0	613.1	0.093	-0.309	0.102	0.133	
					3x	218.5	400.7	0.056	-0.234	0.157	0.036	
					4x	571.5	47.7	0.020	1.00	**0.212**	**0.273**	
M. utile AMB563-1	"Pure"	393	65.6	506.5	2x	72.0	434.5	0.078	-0.274	0.074	**0.187**	
					3x	306.3	200.2	0.036	-0.035	0.120	0.001	
					4x	490.1	16.3	0.014	0.620	0.168	**0.186**	
M. utile AMB566-4	"Pure"	384	56.9	658.5	2x	1.2	657.3	0.090	-0.366	0.082	**0.250**	
					3x	255.6	402.9	0.045	-0.142	0.137	0.017	
					4x	563.9	94.5	0.018	0.694	0.196	0.175	
M. utile AMB574-1	"Pure"	410	53.2	651.2	2x	12.6	638.6	0.087	-0.366	0.074	**0.292**	
					3x	294.3	356.8	0.036	0.012	0.128	0.000	
					4x	552.5	98.6	0.017	0.574	0.185	0.139	
M. utile AMB575-1	"Pure"	471	64.1	654.8	2x	38.2	616.5	0.083	-0.324	0.075	**0.238**	
					3x	330.7	324.1	0.040	-0.104	0.125	0.011	
					4x	628.9	25.8	0.014	0.705	0.172	**0.220**	
M. utile AMB577-2	"Pure"	420	52.8	634.0	2x	94.1	539.8	0.065	-0.082	0.087	0.014	
					3x	254.4	379.5	0.048	-0.288	0.123	0.084	
					4x	619.7	14.3	0.009	1.056	0.146	0.469	
M. utile AMB577-3	"Pure"	361	53.6	554.8	2x	32.3	522.4	0.076	-0.202	0.087	0.082	
					3x	207.1	347.6	0.047	-0.242	0.129	0.055	
					4x	526.3	28.4	0.011	1.018	0.160	0.406	
M. utile AMB639-1	"Pure"	696	64.5	1208.7	2x	552.5	656.2	**0.016**	0.544	0.059	0.587	
					3x	368.8	839.8	0.060	-0.573	0.111	0.308	
					4x	992.8	215.8	**0.012**	0.973	0.166	0.366	

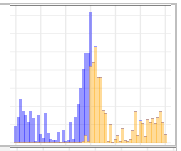
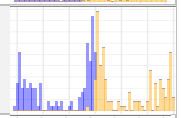
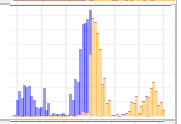
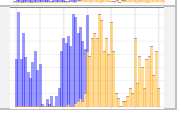
M. utile AMB641-1	"Pure"	634	68.6	1088.6	2x	587.6	501.0	**0.014**	0.560	0.051	0.664	
					3x	363.0	725.6	0.055	-0.475	0.114	0.225	
					4x	853.5	235.1	**0.014**	0.811	0.173	0.271	
M. utile AMB652	"Pure"	134	94.4	212.7	2x	84.5	128.2	0.026	0.436	0.075	**0.360**	
					3x	68.0	144.7	0.058	-0.498	0.121	0.221	
					4x	193.0	19.7	0.014	0.902	0.179	**0.300**	
M. utile AMB653	"Pure"	179	97.8	1550.4	2x	836.4	713.9	0.009	0.659	0.045	0.783	
					3x	455.6	1094.7	0.062	-0.555	0.121	0.262	
					4x	1141.9	408.4	0.016	0.871	0.188	0.266	
M. utile AMB659	"Pure"	512	45.8	635.7	2x	401.9	233.8	0.027	**0.229**	**0.042**	0.333	
					3x	307.8	327.8	0.042	-0.447	**0.047**	0.597	
					4x	438.5	197.2	0.009	**0.326**	0.109	0.130	

Table S9. Number of quartet sampling replicates for the concordant (count 0), and the two discordant (count 1 and count 2) quartet arrangements for the maximum likelihood tree inferred with target enrichment data (Fig. S8). The node label corresponds to the label for each branch. The quartet trees spanning the node for each count are included (topo0 topo1, topo2). One representative taxon for each subset in a quartet is shown. The nodes discussed in the text are highlighted in bold.

node_label	count0	count1	count2	topo0	topo1	topo2
QS2	12	0	21	((MutileAMB503,WsquamulosaAMB513),(MutileAMB653,MutileAMB5664));	((MutileAMB503,MutileAMB653),(WsquamulosaAMB513,MutileAMB5664));	((MutileAMB503,MutileAMB5664),(MutileAMB653,WsquamulosaAMB513));
QS3	98	2	95	((MutileAMB5664,MutileAMB503),(MfoeniculaceumVLondono,MutileAMB497));	((MutileAMB5664,MfoeniculaceumVLondono),(MutileAMB503,MutileAMB497));	((MutileAMB5664,MutileAMB497),(MfoeniculaceumVLondono,MutileAMB503));
QS4	147	228	370	((MutileAMB5741,WsquamulosaAMB513),(MfoeniculaceumAMB5734,MutileAMB5773));	((MutileAMB5741,MfoeniculaceumAMB5734),(WsquamulosaAMB513,MutileAMB5773));	((MutileAMB5741,MutileAMB5773),(MfoeniculaceumAMB5734,WsquamulosaAMB513));
QS5	494	138	258	((MutileAMB5773,MutileAMB5664),(MfoeniculaceumAMB502,MutileAMB512));	((MutileAMB5773,MfoeniculaceumAMB502),(MutileAMB5664,MutileAMB512));	((MutileAMB5773,MutileAMB512),(MfoeniculaceumAMB502,MutileAMB5664));
QS6	998	0	1	((MutileAMB512,MutileAMB5772),(MfoeniculaceumAMB5734,MutileAMB653));	((MutileAMB512,MfoeniculaceumAMB5734),(MutileAMB5772,MutileAMB653));	((MutileAMB512,MutileAMB653),(MfoeniculaceumAMB5734,MutileAMB5772));
QS7	788	88	120	((MfoeniculaceumAMB6581,MutileAMB578),(MfoeniculaceumAMB5585,MfoeniculaceumAMB568));	((MfoeniculaceumAMB6581,MfoeniculaceumAMB5585),(MutileAMB578,MfoeniculaceumAMB568));	((MfoeniculaceumAMB6581,MfoeniculaceumAMB568),(MfoeniculaceumAMB5585,MutileAMB578));
QS8	338	345	279	((MfoeniculaceumAMB568,MfoeniculaceumAMB6384),(MfoeniculaceumAMB5674,MfoeniculaceumAMB5559));	((MfoeniculaceumAMB568,MfoeniculaceumAMB5674),(MfoeniculaceumAMB6384,MfoeniculaceumAMB5559));	((MfoeniculaceumAMB568,MfoeniculaceumAMB5559),(MfoeniculaceumAMB5674,MfoeniculaceumAMB6384));
QS9	151	147	37	((MfoeniculaceumAMB5559,MutileAMB5631),(MfoeniculaceumAMB5711,MfoeniculaceumAMB5714));	((MfoeniculaceumAMB5559,MfoeniculaceumAMB5711),(MutileAMB5631,MfoeniculaceumAMB5714));	((MfoeniculaceumAMB5559,MfoeniculaceumAMB5714),(MfoeniculaceumAMB5711,MutileAMB5631));
QS10	148	8	161	((MfoeniculaceumAMB5714,MutileAMB578),(MfoeniculaceumAMB499,MfoeniculaceumAMB5584));	((MfoeniculaceumAMB5714,MfoeniculaceumAMB499),(MutileAMB578,MfoeniculaceumAMB5584));	((MfoeniculaceumAMB5714,MfoeniculaceumAMB5584),(MfoeniculaceumAMB499,MutileAMB578));
QS11	316	211	369	((MfoeniculaceumAMB5584,MutileAMB6391),(MfoeniculaceumAMB5711,MfoeniculaceumAMB5612));	((MfoeniculaceumAMB5584,MfoeniculaceumAMB5711),(MutileAMB6391,MfoeniculaceumAMB5612));	((MfoeniculaceumAMB5584,MfoeniculaceumAMB5612),(MfoeniculaceumAMB5711,MutileAMB6391));
QS12	111	435	219	((MfoeniculaceumAMB499,MfoeniculaceumAMB5714),(MfoeniculaceumAMB5711,MfoeniculaceumAMB5734));	((MfoeniculaceumAMB499,MfoeniculaceumAMB5711),(MfoeniculaceumAMB5714,MfoeniculaceumAMB5734));	((MfoeniculaceumAMB499,MfoeniculaceumAMB5734),(MfoeniculaceumAMB5711,MfoeniculaceumAMB5714));
QS13	124	1	17	((MfoeniculaceumAMB5734,WsquamulosaAMB513),(MfoeniculaceumAMB5711,MfoeniculaceumAMB5621));	((MfoeniculaceumAMB5734,MfoeniculaceumAMB5711),(WsquamulosaAMB513,MfoeniculaceumAMB5621));	((MfoeniculaceumAMB5734,MfoeniculaceumAMB5621),(MfoeniculaceumAMB5711,WsquamulosaAMB513));

QS14	24	81	6	((MfoeniculaceumAMB5621,MfoeniculaceumAMB654),(MfoeniculaceumAMB5711,MfoeniculaceumAMB5613));	((MfoeniculaceumAMB5621,MfoeniculaceumAMB5711),(MfoeniculaceumAMB654,MfoeniculaceumAMB5613));	((MfoeniculaceumAMB5621,MfoeniculaceumAMB5613),(MfoeniculaceumAMB5711,MfoeniculaceumAMB654));
QS15	76	0	0	((MfoeniculaceumAMB5613,MfoeniculaceumAMB5513),(MfoeniculaceumAMB5711,MfoeniculaceumAMB5712));	((MfoeniculaceumAMB5613,MfoeniculaceumAMB5711),(MfoeniculaceumAMB5513,MfoeniculaceumAMB5712));	((MfoeniculaceumAMB5613,MfoeniculaceumAMB5712),(MfoeniculaceumAMB5711,MfoeniculaceumAMB5513));
QS16	7	3	9	((MfoeniculaceumAMB5712,MfoeniculaceumAMB494),(MfoeniculaceumAMB5674,MfoeniculaceumAMB5711));	((MfoeniculaceumAMB5712,MfoeniculaceumAMB5674),(MfoeniculaceumAMB494,MfoeniculaceumAMB5711));	((MfoeniculaceumAMB5712,MfoeniculaceumAMB5711),(MfoeniculaceumAMB5674,MfoeniculaceumAMB494));
QS17	253	431	54	((MfoeniculaceumAMB5711,MfoeniculaceumAMB5584),(MfoeniculaceumAMB494,MfoeniculaceumAMB5612));	((MfoeniculaceumAMB5711,MfoeniculaceumAMB494),(MfoeniculaceumAMB5584,MfoeniculaceumAMB5612));	((MfoeniculaceumAMB5711,MfoeniculaceumAMB5612),(MfoeniculaceumAMB494,MfoeniculaceumAMB5584));
QS18	53	0	55	((MfoeniculaceumAMB5612,MfoeniculaceumAMB6384),(MfoeniculaceumAMB499,MfoeniculaceumAMB5585));	((MfoeniculaceumAMB5612,MfoeniculaceumAMB499),(MfoeniculaceumAMB6384,MfoeniculaceumAMB5585));	((MfoeniculaceumAMB5612,MfoeniculaceumAMB5585),(MfoeniculaceumAMB499,MfoeniculaceumAMB6384));
QS19	64	1	11	((MfoeniculaceumAMB5585,MutileAMB5751),(MfoeniculaceumAMB502,MfoeniculaceumAMB494));	((MfoeniculaceumAMB5585,MfoeniculaceumAMB502),(MutileAMB5751,MfoeniculaceumAMB494));	((MfoeniculaceumAMB5585,MfoeniculaceumAMB494),(MfoeniculaceumAMB502,MutileAMB5751));
QS20	30	8	1	((MfoeniculaceumAMB494,MfoeniculaceumAMB5714),(MfoeniculaceumAMB499,MfoeniculaceumAMB502));	((MfoeniculaceumAMB494,MfoeniculaceumAMB499),(MfoeniculaceumAMB5714,MfoeniculaceumAMB502));	((MfoeniculaceumAMB494,MfoeniculaceumAMB502),(MfoeniculaceumAMB499,MfoeniculaceumAMB5714));
QS21	235	337	124	((MfoeniculaceumAMB494,MutileAMB507),(MfoeniculaceumAMB506,MfoeniculaceumAMB5562));	((MfoeniculaceumAMB494,MfoeniculaceumAMB506),(MutileAMB507,MfoeniculaceumAMB5562));	((MfoeniculaceumAMB494,MfoeniculaceumAMB5562),(MfoeniculaceumAMB506,MutileAMB507));
QS22	29	7	3	((MfoeniculaceumAMB5562,MfoeniculaceumAMB5585),(MfoeniculaceumAMB568,MfoeniculaceumAMB506));	((MfoeniculaceumAMB5562,MfoeniculaceumAMB568),(MfoeniculaceumAMB5585,MfoeniculaceumAMB506));	((MfoeniculaceumAMB5562,MfoeniculaceumAMB506),(MfoeniculaceumAMB568,MfoeniculaceumAMB5585));
QS23	**501**	**126**	**352**	**((MfoeniculaceumAMB5674,MutileAMB5751),(MutileAMB659,MfoeniculaceumVLondono));**	**((MfoeniculaceumAMB5674,MutileAMB659),(MutileAMB5751,MfoeniculaceumVLondono));**	**((MfoeniculaceumAMB5674,MfoeniculaceumVLondono),(MutileAMB659,MutileAMB5751));**
QS24	**487**	**443**	**68**	**((MfoeniculaceumAMB5513,MutileAMB497),(MfoeniculaceumAMB6581,MutileAMB659));**	**((MfoeniculaceumAMB5513,MfoeniculaceumAMB6581),(MutileAMB497,MutileAMB659));**	**((MfoeniculaceumAMB5513,MutileAMB659),(MfoeniculaceumAMB6581,MutileAMB497));**
QS25	691	6	78	((MutileAMB653,MfoeniculaceumAMB5714),(MfoeniculaceumAMB654,MfoeniculaceumAMB657));	((MutileAMB653,MfoeniculaceumAMB654),(MfoeniculaceumAMB5714,MfoeniculaceumAMB657));	((MutileAMB653,MfoeniculaceumAMB657),(MfoeniculaceumAMB654,MfoeniculaceumAMB5714));
QS26	76	48	20	((MfoeniculaceumAMB657,MfoeniculaceumAMB5621),(MfoeniculaceumAMB654,MfoeniculaceumAMB6384));	((MfoeniculaceumAMB657,MfoeniculaceumAMB654),(MfoeniculaceumAMB5621,MfoeniculaceumAMB6384));	((MfoeniculaceumAMB657,MfoeniculaceumAMB6384),(MfoeniculaceumAMB654,MfoeniculaceumAMB5621));
QS27	83	3	25	((MfoeniculaceumAMB6384,MfoeniculaceumAMB5513),(Mfoeniculaceum	((MfoeniculaceumAMB6384,MfoeniculaceumAMB6431),(MfoeniculaceumAMB5513,MfoeniculaceumAMB6581));	((MfoeniculaceumAMB6384,MfoeniculaceumAMB6581),(MfoeniculaceumAMB6431,MfoeniculaceumAMB5513));

				AMB6431,MfoeniculaceumAMB6581));		
QS28	70	0	6	((MfoeniculaceumAMB6581,MutileAMB5741),(MfoeniculaceumAMB6431,MfoeniculaceumAMB654));	((MfoeniculaceumAMB6581,MfoeniculaceumAMB6431),(MutileAMB5741,MfoeniculaceumAMB654));	((MfoeniculaceumAMB6581,MfoeniculaceumAMB654),(MfoeniculaceumAMB6431,MutileAMB5741));
QS29	39	0	0	((MfoeniculaceumAMB654,MutileAMB653),(MfoeniculaceumAMB6431,MfoeniculaceumAMB635));	((MfoeniculaceumAMB654,MfoeniculaceumAMB6431),(MutileAMB653,MfoeniculaceumAMB635));	((MfoeniculaceumAMB654,MfoeniculaceumAMB635),(MfoeniculaceumAMB6431,MutileAMB653));
QS30	913	82	4	((MfoeniculaceumAMB6581,MfoeniculaceumAMB5711),(MutileAMB653,MutileAMB6391));	((MfoeniculaceumAMB6581,MutileAMB653),(MfoeniculaceumAMB5711,MutileAMB6391));	((MfoeniculaceumAMB6581,MutileAMB6391),(MutileAMB653,MfoeniculaceumAMB5711));
QS31	100	20	25	((MutileAMB6411,MfoeniculaceumAMB657),(MutileAMB652,MutileAMB659));	((MutileAMB6411,MutileAMB652),(MfoeniculaceumAMB657,MutileAMB659));	((MutileAMB6411,MutileAMB659),(MutileAMB652,MfoeniculaceumAMB657));
QS32	39	0	0	((MutileAMB659,MfoeniculaceumAMB6431),(MutileAMB652,MutileAMB653));	((MutileAMB659,MutileAMB652),(MfoeniculaceumAMB6431,MutileAMB653));	((MutileAMB659,MutileAMB653),(MutileAMB652,MfoeniculaceumAMB6431));
QS33	48	37	21	((MutileAMB659,MfoeniculaceumAMB6581),(MutileAMB6411,MutileAMB6391));	((MutileAMB659,MutileAMB6411),(MfoeniculaceumAMB6581,MutileAMB6391));	((MutileAMB659,MutileAMB6391),(MutileAMB6411,MfoeniculaceumAMB6581));
QS34	238	79	0	((MfoeniculaceumAMB6384,MfoeniculaceumAMB5559),(MfoeniculaceumAMB5513,MfoeniculaceumVLondono));	((MfoeniculaceumAMB6384,MfoeniculaceumAMB5513),(MfoeniculaceumAMB5559,MfoeniculaceumVLondono));	((MfoeniculaceumAMB6384,MfoeniculaceumVLondono),(MfoeniculaceumAMB5513,MfoeniculaceumAMB5559));
QS35	297	165	2	((MutileAMB652,MutileAMB5741),(MutileAMB507,MutileAMB5631));	((MutileAMB652,MutileAMB507),(MutileAMB5741,MutileAMB5631));	((MutileAMB652,MutileAMB5631),(MutileAMB507,MutileAMB5741));
QS36	10	18	5	((MutileAMB5631,MfoeniculaceumAMB506),(MutileAMB512,MutileAMB507));	((MutileAMB5631,MutileAMB512),(MfoeniculaceumAMB506,MutileAMB507));	((MutileAMB5631,MutileAMB507),(MutileAMB512,MfoeniculaceumAMB506));
QS37	314	30	23	((MfoeniculaceumAMB502,MutileAMB497),(MutileAMB5772,MutileAMB578));	((MfoeniculaceumAMB502,MutileAMB5772),(MutileAMB497,MutileAMB578));	((MfoeniculaceumAMB502,MutileAMB578),(MutileAMB5772,MutileAMB497));
QS38	32	0	5	((MutileAMB578,MfoeniculaceumAMB499),(MutileAMB5772,MutileAMB5773));	((MutileAMB578,MutileAMB5772),(MfoeniculaceumAMB499,MutileAMB5773));	((MutileAMB578,MutileAMB5773),(MutileAMB5772,MfoeniculaceumAMB499));
QS39	166	22	10	((MfoeniculaceumAMB5734,WsquamulosaAMB513),(MutileAMB5751,MutileAMB497));	((MfoeniculaceumAMB5734,MutileAMB5751),(WsquamulosaAMB513,MutileAMB497));	((MfoeniculaceumAMB5734,MutileAMB497),(MutileAMB5751,WsquamulosaAMB513));
QS40	33	2	2	((MutileAMB497,MfoeniculaceumAMB568),(MutileAMB5751,MutileAMB5741));	((MutileAMB497,MutileAMB5751),(MfoeniculaceumAMB568,MutileAMB5741));	((MutileAMB497,MutileAMB5741),(MutileAMB5751,MfoeniculaceumAMB568));

Table S10. Number of quartet sampling replicates for the concordant (count 0), and the two discordant (count 1 and count 2) quartet arrangements for the maximum likelihood tree inferred with ddRADseq data (Fig. S8). The node label corresponds to the label for each branch. The quartet trees spanning the node for each count are included (topo0 topo1, topo2). One representative taxon for each subset in a quartet is shown. The nodes discussed in the text are highlighted in bold.

node_label	count0	count1	count2	topo0	topo1	topo2
QS2	2	13	0	((Marathrum_utile_AMB497,Marathrum_utile_AMB503_),(Marathrum_foeniculaceum_AMB555_9,Marathrum_utile_AMB577_3));	((Marathrum_utile_AMB497,Marathrum_foeniculaceum_AMB555_9),(Marathrum_utile_AMB503_,Marathrum_utile_AMB577_3));	((Marathrum_utile_AMB497,Marathrum_utile_AMB577_3),(Marathrum_foeniculaceum_AMB555_9,Marathrum_utile_AMB503_));
QS3	82	11	1	((Marathrum_utile_AMB577_3,Marathrum_utile_AMB497),(Marathrum_utile_AMB641_1,Marathrum_utile_AMB574_1));	((Marathrum_utile_AMB577_3,Marathrum_utile_AMB641_1),(Marathrum_utile_AMB497,Marathrum_utile_AMB574_1));	((Marathrum_utile_AMB577_3,Marathrum_utile_AMB574_1),(Marathrum_utile_AMB641_1,Marathrum_utile_AMB497));
QS4	175	17	53	((Marathrum_utile_AMB574_1,Marathrum_utile_AMB497),(Marathrum_foeniculaceum_AMB654,Marathrum_foeniculaceum_AMB555_9));	((Marathrum_utile_AMB574_1,Marathrum_foeniculaceum_AMB654),(Marathrum_utile_AMB497,Marathrum_foeniculaceum_AMB555_9));	((Marathrum_utile_AMB574_1,Marathrum_foeniculaceum_AMB555_9),(Marathrum_foeniculaceum_AMB654,Marathrum_utile_AMB497));
QS5	101	214	96	((Marathrum_foeniculaceum_AMB555_9,Marathrum_utile_AMB575_1),(Marathrum_foeniculaceum_AMB568,Marathrum_foeniculaceumAMB506));	((Marathrum_foeniculaceum_AMB555_9,Marathrum_foeniculaceum_AMB568),(Marathrum_utile_AMB575_1,Marathrum_foeniculaceumAMB506));	((Marathrum_foeniculaceum_AMB555_9,Marathrum_foeniculaceumAMB506),(Marathrum_foeniculaceum_AMB568,Marathrum_utile_AMB575_1));
QS6	367	51	31	((Marathrum_foeniculaceumAMB506,Marathrum_utile_AMB575_1),(Marathrum_foeniculaceum_AMB635,Marathrum_foeniculaceum_VLondono));	((Marathrum_foeniculaceumAMB506,Marathrum_foeniculaceum_AMB635),(Marathrum_utile_AMB575_1,Marathrum_foeniculaceum_VLondono));	((Marathrum_foeniculaceumAMB506,Marathrum_foeniculaceum_VLondono),(Marathrum_foeniculaceum_AMB635,Marathrum_utile_AMB575_1));
QS7	**198**	**373**	**213**	**((Marathrum_foeniculaceum_VLondono, Marathrum_utile_AMB566_4),(Marathrum_foeniculaceumAMB502,Marathrum_foeniculaceum_AMB658_1));**	**((Marathrum_foeniculaceum_VLondono,Marathrum_foeniculaceumAMB502),(Marathrum_utile_AMB566_4,Marathrum_foeniculaceum_AMB658_1));**	**((Marathrum_foeniculaceum_VLondono, Marathrum_foeniculaceum_AMB658_1),(Marathrum_foeniculaceumAMB502,Marathrum_utile_AMB566_4));**
QS8	624	167	32	((Marathrum_utile_AMB659,Marathrum_foeniculaceum_VLondono),(Marathrum_foeniculaceum_AMB567_4,Marathrum_foeniculaceumAMB499));	((Marathrum_utile_AMB659,Marathrum_foeniculaceum_AMB567_4),(Marathrum_foeniculaceum_VLondono,Marathrum_foeniculaceumAMB499));	((Marathrum_utile_AMB659,Marathrum_foeniculaceumAMB499),(Marathrum_foeniculaceum_AMB567_4,Marathrum_foeniculaceum_VLondono));
QS9	356	155	69	((Marathrum_foeniculaceumAMB494,Marathrum_foeniculaceum_AMB635),(Marathrum_foeniculaceum_AMB571_2,Marathrum_foeniculaceum_AMB556_2));	((Marathrum_foeniculaceumAMB494,Marathrum_foeniculaceum_AMB571_2),(Marathrum_foeniculaceum_AMB635,Marathrum_foeniculaceum_AMB556_2));	((Marathrum_foeniculaceumAMB494,Marathrum_foeniculaceum_AMB556_2),(Marathrum_foeniculaceum_AMB571_2,Marathrum_foeniculaceum_AMB635));
QS10	41	26	126	((Marathrum_foeniculaceum_AMB567_4,Marathrum_utile_AMB641_1),(Marathrum_foeniculaceum_AMB558_4,Marathrum_foeniculaceum_AMB562_1));	((Marathrum_foeniculaceum_AMB567_4,Marathrum_foeniculaceum_AMB558_4),(Marathrum_utile_AMB641_1,Marathrum_foeniculaceum_AMB562_1));	((Marathrum_foeniculaceum_AMB567_4,Marathrum_foeniculaceum_AMB562_1),(Marathrum_foeniculaceum_AMB558_4,Marathrum_utile_AMB641_1));
QS11	8	26	18	((Marathrum_foeniculaceum_AMB562_1,Marathrum_utile_AMB659),(Marathrum_foeniculaceum_AMB561_2,Marathrum_foeniculaceum_AMB571_4));	((Marathrum_foeniculaceum_AMB562_1,Marathrum_foeniculaceum_AMB561_2),(Marathrum_utile_AMB659,Marathrum_foeniculaceum_AMB571_4));	((Marathrum_foeniculaceum_AMB562_1,Marathrum_foeniculaceum_AMB571_4),(Marathrum_foeniculaceum_AMB561_2,Marathrum_utile_AMB659));

QS12	27	5	28	((Marathrum_foeniculaceum_AMB571_4,Marathrum_utile_AMB575_1),(Marathrum_foeniculaceum_AMB571_2,Marathrum_foeniculaceum_AMB561_2));	((Marathrum_foeniculaceum_AMB571_4,Marathrum_foeniculaceum_AMB571_2),(Marathrum_utile_AMB575_1,Marathrum_foeniculaceum_AMB561_2));	((Marathrum_foeniculaceum_AMB571_4,Marathrum_foeniculaceum_AMB561_2),(Marathrum_foeniculaceum_AMB571_2,Marathrum_utile_AMB575_1));
QS13	4	3	17	((Marathrum_foeniculaceum_AMB561_2,Marathrum_utile_AMB574_1),(Marathrum_foeniculaceumAMB502,Marathrum_foeniculaceum_AMB558_4));	((Marathrum_foeniculaceum_AMB561_2,Marathrum_foeniculaceumAMB502),(Marathrum_utile_AMB574_1,Marathrum_foeniculaceum_AMB558_4));	((Marathrum_foeniculaceum_AMB561_2,Marathrum_foeniculaceum_AMB558_4),(Marathrum_foeniculaceumAMB502,Marathrum_utile_AMB574_1));
QS14	17	4	0	((Marathrum_foeniculaceum_AMB558_4,Marathrum_foeniculaceum_AMB558_5),(Marathrum_foeniculaceum_AMB571_2,Marathrum_foeniculaceumAMB502));	((Marathrum_foeniculaceum_AMB558_4,Marathrum_foeniculaceum_AMB571_2),(Marathrum_foeniculaceum_AMB558_5,Marathrum_foeniculaceumAMB502));	((Marathrum_foeniculaceum_AMB558_4,Marathrum_foeniculaceumAMB502),(Marathrum_foeniculaceum_AMB571_2,Marathrum_foeniculaceum_AMB558_5));
QS15	16	4	9	((Marathrum_foeniculaceum_AMB561_2,Marathrum_utile_AMB577_3),(Marathrum_foeniculaceum_AMB561_3,Marathrum_foeniculaceum_AMB571_4));	((Marathrum_foeniculaceum_AMB561_2,Marathrum_foeniculaceum_AMB561_3),(Marathrum_utile_AMB577_3,Marathrum_foeniculaceum_AMB571_4));	((Marathrum_foeniculaceum_AMB561_2,Marathrum_foeniculaceum_AMB571_4),(Marathrum_foeniculaceum_AMB561_3,Marathrum_utile_AMB577_3));
QS16	93	52	18	((Marathrum_foeniculaceum_AMB558_4,Marathrum_utile_AMB659),(Marathrum_foeniculaceum_AMB567_4,Marathrum_foeniculaceum_AMB558_5));	((Marathrum_foeniculaceum_AMB558_4,Marathrum_foeniculaceum_AMB567_4),(Marathrum_utile_AMB659,Marathrum_foeniculaceum_AMB558_5));	((Marathrum_foeniculaceum_AMB558_4,Marathrum_foeniculaceum_AMB558_5),(Marathrum_foeniculaceum_AMB567_4,Marathrum_utile_AMB659));
QS17	13	0	1	((Marathrum_foeniculaceum_AMB558_5,Marathrum_utile_AMB566_4),(Marathrum_foeniculaceum_AMB567_4,Marathrum_foeniculaceum_AMB556_2));	((Marathrum_foeniculaceum_AMB558_5,Marathrum_foeniculaceum_AMB567_4),(Marathrum_utile_AMB566_4,Marathrum_foeniculaceum_AMB556_2));	((Marathrum_foeniculaceum_AMB558_5,Marathrum_foeniculaceum_AMB556_2),(Marathrum_foeniculaceum_AMB567_4,Marathrum_utile_AMB566_4));
QS18	11	0	0	((Marathrum_foeniculaceum_AMB556_2,Marathrum_utile_AMB497),(Marathrum_foeniculaceum_AMB567_4,Marathrum_foeniculaceum_AMB568));	((Marathrum_foeniculaceum_AMB556_2,Marathrum_foeniculaceum_AMB567_4),(Marathrum_utile_AMB497,Marathrum_foeniculaceum_AMB568));	((Marathrum_foeniculaceum_AMB556_2,Marathrum_foeniculaceum_AMB568),(Marathrum_foeniculaceum_AMB567_4,Marathrum_utile_AMB497));
QS19	76	22	10	((Marathrum_foeniculaceum_AMB571_4,Marathrum_foeniculaceum_AMB657),(Marathrum_foeniculaceumAMB494,Marathrum_foeniculaceumAMB499));	((Marathrum_foeniculaceum_AMB571_4,Marathrum_foeniculaceumAMB494),(Marathrum_foeniculaceum_AMB657,Marathrum_foeniculaceumAMB499));	((Marathrum_foeniculaceum_AMB571_4,Marathrum_foeniculaceumAMB499),(Marathrum_foeniculaceumAMB494,Marathrum_foeniculaceum_AMB657));
QS20	**373**	**8**	**426**	**((Marathrum_foeniculaceum_AMB562_1,Marathrum_utile_AMB577_3),(Marathrum_foeniculaceum_AMB635,Marathrum_foeniculaceum_AMB657));**	**((Marathrum_foeniculaceum_AMB562_1,Marathrum_foeniculaceum_AMB635),(Marathrum_utile_AMB577_3,Marathrum_foeniculaceum_AMB657));**	**((Marathrum_foeniculaceum_AMB562_1,Marathrum_foeniculaceum_AMB657),(Marathrum_foeniculaceum_AMB635,Marathrum_utile_AMB577_3));**
QS21	**95**	**16**	**251**	**((Marathrum_foeniculaceum_AMB657,Marathrum_foeniculaceum_AMB567_4),(Marathrum_utile_AMB652,Marathrum_foeniculaceum_AMB635));**	**((Marathrum_foeniculaceum_AMB657,Marathrum_utile_AMB652),(Marathrum_foeniculaceum_AMB567_4,Marathrum_foeniculaceum_AMB635));**	**((Marathrum_foeniculaceum_AMB657,Marathrum_foeniculaceum_AMB635),(Marathrum_utile_AMB652,Marathrum_foeniculaceum_AMB567_4));**
QS22	257	32	51	((Marathrum_foeniculaceum_AMB638_4,Marathrum_foeniculaceumAMB502),(Marathrum_utile_AMB653,Marathrum_utile_AMB659));	((Marathrum_foeniculaceum_AMB638_4,Marathrum_utile_AMB653),(Marathrum_foeniculaceumAMB502,Marathrum_utile_AMB659));	((Marathrum_foeniculaceum_AMB638_4,Marathrum_utile_AMB659),(Marathrum_utile_AMB653,Marathrum_foeniculaceumAMB502));
QS23	68	4	22	((Marathrum_utile_AMB659,Marathrum_foeniculaceum_AMB635),(Marathrum_utile_AMB653,Marathrum_utile_AMB641_1));	((Marathrum_utile_AMB659,Marathrum_utile_AMB653),(Marathrum_foeniculaceum_AMB635,Marathrum_utile_AMB641_1));	((Marathrum_utile_AMB659,Marathrum_utile_AMB641_1),(Marathrum_utile_AMB653,Marathrum_foeniculaceum_AMB635));

QS24	36	2	5	((Marathrum_utile_AMB641_1,Marathrum_utile_AMB575_1),(Marathrum_utile_AMB653,Marathrum_utile_AMB652));	((Marathrum_utile_AMB641_1,Marathrum_utile_AMB653),(Marathrum_utile_AMB575_1,Marathrum_utile_AMB652));	((Marathrum_utile_AMB641_1,Marathrum_utile_AMB652),(Marathrum_utile_AMB653,Marathrum_utile_AMB575_1));
QS25	19	9	14	((Marathrum_utile_AMB652,Marathrum_foeniculaceum_AMB658_1),(Marathrum_utile_AMB641_1,Marathrum_utile_AMB639_1));	((Marathrum_utile_AMB652,Marathrum_utile_AMB641_1),(Marathrum_foeniculaceum_AMB658_1,Marathrum_utile_AMB639_1));	((Marathrum_utile_AMB652,Marathrum_utile_AMB639_1),(Marathrum_utile_AMB641_1,Marathrum_foeniculaceum_AMB658_1));
QS26	198	18	82	((Marathrum_utile_AMB659,Marathrum_utile_AMB503_),(Marathrum_foeniculaceum_AMB638_4,Marathrum_foeniculaceum_AMB658_1));	((Marathrum_utile_AMB659,Marathrum_foeniculaceum_AMB638_4),(Marathrum_utile_AMB503_,Marathrum_foeniculaceum_AMB658_1));	((Marathrum_utile_AMB659,Marathrum_foeniculaceum_AMB658_1),(Marathrum_foeniculaceum_AMB638_4,Marathrum_utile_AMB503_));
QS27	11	0	11	((Marathrum_foeniculaceum_AMB658_1,Marathrum_foeniculaceumAMB506),(Marathrum_foeniculaceum_AMB635,Marathrum_foeniculaceum_AMB638_4));	((Marathrum_foeniculaceum_AMB658_1,Marathrum_foeniculaceum_AMB635),(Marathrum_foeniculaceumAMB506,Marathrum_foeniculaceum_AMB638_4));	((Marathrum_foeniculaceum_AMB658_1,Marathrum_foeniculaceum_AMB638_4),(Marathrum_foeniculaceum_AMB635,Marathrum_foeniculaceumAMB506));
QS28	15	0	4	((Marathrum_foeniculaceum_AMB638_4,Marathrum_utile_AMB652),(Marathrum_foeniculaceum_AMB635,Marathrum_foeniculaceum_AMB654));	((Marathrum_foeniculaceum_AMB638_4,Marathrum_foeniculaceum_AMB635),(Marathrum_utile_AMB652,Marathrum_foeniculaceum_AMB654));	((Marathrum_foeniculaceum_AMB638_4,Marathrum_foeniculaceum_AMB654),(Marathrum_foeniculaceum_AMB635,Marathrum_utile_AMB652));
QS29	45	13	85	((Marathrum_foeniculaceum_AMB571_2,Marathrum_utile_AMB497),(Marathrum_foeniculaceumAMB506,Marathrum_utile_AMB563_1));	((Marathrum_foeniculaceum_AMB571_2,Marathrum_foeniculaceumAMB506),(Marathrum_utile_AMB497,Marathrum_utile_AMB563_1));	((Marathrum_foeniculaceum_AMB571_2,Marathrum_utile_AMB563_1),(Marathrum_foeniculaceumAMB506,Marathrum_utile_AMB497));
QS30	7	0	0	((Marathrum_utile_AMB563_1,Marathrum_foeniculaceum_AMB555_9),(Marathrum_foeniculaceumAMB506,Marathrum_utileAMB507));	((Marathrum_utile_AMB563_1,Marathrum_foeniculaceumAMB506),(Marathrum_foeniculaceum_AMB555_9,Marathrum_utileAMB507));	((Marathrum_utile_AMB563_1,Marathrum_utileAMB507),(Marathrum_foeniculaceumAMB506,Marathrum_foeniculaceum_AMB555_9));
QS31	31	8	27	((Marathrum_foeniculaceum_AMB571_4,Marathrum_utile_AMB575_1),(Marathrum_utile_AMB566_4,Marathrum_foeniculaceum_AMB555_9));	((Marathrum_foeniculaceum_AMB571_4,Marathrum_utile_AMB566_4),(Marathrum_utile_AMB575_1,Marathrum_foeniculaceum_AMB555_9));	((Marathrum_foeniculaceum_AMB571_4,Marathrum_foeniculaceum_AMB555_9),(Marathrum_utile_AMB566_4,Marathrum_utile_AMB575_1));
QS32	49	13	5	((Marathrum_foeniculaceum_AMB556_2,Marathrum_utile_AMB497),(Marathrum_utile_AMB574_1,Marathrum_utile_AMB575_1));	((Marathrum_foeniculaceum_AMB556_2,Marathrum_utile_AMB574_1),(Marathrum_utile_AMB497,Marathrum_utile_AMB575_1));	((Marathrum_foeniculaceum_AMB556_2,Marathrum_utile_AMB575_1),(Marathrum_utile_AMB574_1,Marathrum_utile_AMB497));

Table S11. AIC model selection results for phylogenetic networks with 0–3 reticulations including admixed individuals from Antioquia, Boyacá and Caribe using single-exon genes. These results are presented for comparison with results from the analysis of the full dataset including multi-exon genes for which exons where concatenated. Best models are shown in bold. The number of parameters (k) is (n-3) +(2h), where n is the number of tips in the network and h is the number of reticulation events. In Boyacá, admixed individuals are assigned to one population.

Drainage basin	# Reticulations	Parameters (k)	Log likelihood	AIC	ΔAIC	Relative likelihood	wAIC	Cumulative wAIC
Antioquia	**3**	**14**	**-750.67**	**1529.34**	**0**	**1**	**0.59**	**0.59**
	2	**12**	**-753.04**	**1530.08**	**0.74**	**0.68**	**0.41**	**1**
	4	16	-766.49	1564.98	35.64	1.81E-08	0	1
	0	8	-770.03	1556.07	26.73	1.56E-06	0	1
	1	10	-775.35	1570.71	41.37	1.03E-09	0	1
Boyacá	**3**	**12**	**-1006.47**	**2036.95**	**0.00**	**0.94**	**0.94**	**0.94**
	2	10	-1011.15	2042.30	5.35	0.06	0.06	1.00
	1	8	-1022.51	2061.02	24.07	5.93E-06	0.00	1.00
	0	6	-1097.37	2206.75	169.79	1.34E-37	0.00	1.00
	4	14	-1104.80	2237.59	200.64	2.69E-44	0.00	1.00
Caribe	**2**	**12**	**-954.15**	**1932.30**	**0**	**1**	**1**	**1**
	4	16	-985.32	2002.63	70.34	5.33E-16	0	1
	3	14	-1013.63	2055.26	122.96	1.99E-27	0	1
	1	10	-1024.19	2068.38	136.09	2.81E-30	0	1
	0	8	-1073.95	2163.90	231.60	5.10E-51	0	1

Table S12. Nodes ages estimated with SNAPP in this study. The minimum, mean, and maximum 95% High Posterior Density values are shown.

Node	**95% HPD Min (Ma)**	**95% HPD Mean (Ma)**	**95% HPD Max (Ma)**
M. foeniculaceum Antioquia–*M. foeniculaceum* Caribe	10.32	11.91	13.54
M. utile Antioquia–*M. utile* Caribe	2.87	3.50	4.14
Marathrum nSA crown	14.52	16.81	19.05
Marathrum–Weddellina	54.10	61.77	69.44

Methods S1. Specifications for target selection with Sondovac and post-hoc filtering.

Sondovac removes reads that map to the plastome or mitochondriome and duplicated transcripts from the transcriptome, and assembles genome reads matching the remaining unique transcripts. The program is a two-part pipeline. The first part of the pipeline takes as input a transcriptome, plastome, mitochondriome (optional) and a paired-end genome skimming input file to select single-copy, nuclear coding regions. We used the first part of the pipeline to remove reads form our draft assembled genome that mapped to the mitochondriome of *Populus tremula* L. (Salicaceae), the plastome of *Garcinia mangostana L.* (Calophyllaceae), and duplicated transcripts from the transcriptome of *Hypericum perforatum* L. (Hypericaceae). The output of the first part is a fasta file (*_blat_unique_transcripts_versus_genome_skim_data-no_missing_fin.fsa) with a filtered BLAT output (*i.e.* transcripts with >1000 BLAT hits and BLAT hits containing masked nucleotides were removed).

The fasta file resulting from the first part of the pipeline was used in Geneious v9.1.8. to *de novo* assemble hits into larger contigs. We removed probe sequences sharing ≥90% sequence similarity and retained the contigs that comprised exons of minimum ≥600 bp for all contigs for a transcript, and ≥120 bp for each contig. The resulting target sequences were complemented with APVO SSC and PPR loci. We conducted searches of the APVO SSC and PPR loci against the *M. utile* contigs in the draft genome with BLASTN. Up to five of the top hits with a bit score >70 were retained. When hits for the same locus overlapped with different contigs, we assembled the contigs using *de novo* assembly in Geneious v9.1.8. We removed all hits matching the same locus with <95% pairwise identity as they could represent paralogs. We retained individual sequences of at least 120 bp, and loci with at least 600 bp length. The resulting datasets were filtered once more for sequence similarity (≥90%) to remove sequences that overlapped across the APVO+PPR, and the set of sequences resulting from Sondovac.

Methods S2. Selection of assembled contigs with Hybpiper and paralog filtering.

To assess paralogs, phylogenetic trees for each exon with paralog warnings were inferred under maximum likelihood with RaxML-NG as follows: ./raxml-ng --all --msa Gene_n.paralogs.phy --model GTR+G --prefix Gene_n_paralogs --seed 2 --threads 10 --bs-metric fbp, where Gene_n corresponds to each single exon alignment. We inspected the resulting alignments and trees and removed 38 whole exons that corresponded to regions with low coverage across loci, regions for which Hybpiper identified more than two contigs across more than 10% of the samples and where a high number of SNPs occurred, and regions for which the exon trees showed possible evidence of gene duplication. Common tree topologies recovered for exons where multiple contigs are explained by allelic variation are shown in Fig S2. Further inspection of alignments for each exon was conducted to determine if the recovered contigs for samples with paralog warnings were different or varied only in length. We found two instances (exons 181 and 328; Table S6) for which even though multiple contigs were assembled for several species, the contigs were not paralogs but differed only in length with the longest contig including SNPs that were outside the sequences in shorter contigs.

In addition to inspecting gene trees to rule out a pattern of gene duplication to determine if paralog warnings corresponded to allelic variation, we inspected the contig sequences for all samples across exons and we calculated the number of exons with 1 and >1 contig per sample (Table S6). For non-admixed (highly homozygous) individuals (see Population structure across drainage basins subsection in results and Table S7), only 1 contig was assembled for most exons, whereas for admixed (and highly heterozygous) individuals the proportion of exons for which >1 contig was assembled was higher compared to non-admixed individuals. This is in line with the fact that more paralog warnings were raised by Hybpiper for admixed than for non-admixed individuals (Table S6) and provides further evidence that paralog warnings are mostly due to allelic variation.

In Hybpiper, contigs for each sample are chosen based on sequencing depth and when multiple full-length contigs with the same depth are recovered, the algorithm chooses the contig with the highest percent identity with the sequence that was used as reference. We further checked that the sequences chosen per individual and per exon by Hybpiper did not show evidence of duplication in the gene trees and highly divergent sequences in the alignments. Given our ADMIXTURE results (Fig. 3) and heterozygosity calculations (Table S7) where non-admixed individuals are genetically structured and highly homozygous and admixed individuals are highly heterozygous, we kept the contigs chosen by Hybpiper instead of choosing the contigs (alleles) ourselves, to avoid biasing the phylogenetic inference of admixed individuals. We assessed how the concatenation of unconfirmed orthologous exon sequences affected our inference based on concatenated exons data as stated in the text (See Materials and Methods section in the main text).

Methods S3. Specifications applied in maximum likelihood, SVDquartets, Quartet sampling, and PhyloNet analyses.

SVDquartets inference on the target enrichment data in Figs. 4, S1, S5, and S6 were conducted in PAUP, evaluating all quartets, and conducting 1000 bootstrap support. Each gene was specified as a separate partition and *W. squamulosa* was defined as the outgroup. Quartet analysis using the min4 ddRADseq assembly in Figures 4, S1 and S6 was conducted similarly, but neither partitions nor outgroups were specified. We did not conduct bootstrapping for the analysis in Fig. S1. The analyses were run as follows: svdq evalq=all bootstrap=multilocus nreps=1000 loci=combined nthreads=ncpus; where *multilocus bootstrap* is conducted to resample with replacement both loci and sites and *combined* is the partition scheme defining genes as partitions. For ddRADseq analyses where partitions weren't specified, we used bootstrap=standard instead.

Maximum likelihood inference of the concatenated target enrichment datasets was conducted calling RAxML from Cipres Science Gateway as follows: raxmlHPC-HYBRID -T 4 -s infile -N autoMRE -n result -q part -k -f a -p 12345 -x 12345 -m GTRCAT -o W_squamulosa. For maximum likelihood inference with ddRADseq data, the script was modified so that neither partitions nor an outgroup was specified.

Quartet sampling scores were calculated treating the entire alignment as one partition for all quartet searches as follows: python quartet_sampling.py --align quartet_sampling_relaxed.phylip --reps 1000 --tree seqcap_raxml_all.nwk --lnlike 2 --threads 2

Phylogenetic Networks with maximum pseudolikelihood was conducted in PhyloNet, optimizing branch lengths and inheritance probabilities for each proposed network. The program was run as follows: InferNetwork_MPL (all) n -a (taxa association map) 5 -x 100 -pl 10 -di resultOutputFile -b 0.7 -o, where n corresponds to the number of reticulations. We evaluated networks with 0–4 reticulations. Maximum likelihood input gene trees with Felsenstein's bootstrap support for Phylonet were inferred with RAxML-NG as follows: ./raxml-ng --all --msa Gene_n.phy --model GTR+G --prefix Gene_n --seed 2 --threads 10 --bs-metric fbp, where Gene_n corresponds to each single gene alignment.
The log likelihood of the best networks inferred with reticulation events 0–4 was calculated given the collection of trees used as input. The command CalGTProb was used as follows: CalGTProb net (gt0, gt1…) -a <taxon map> -b 70 -x 100 -pl 10, where net refers to the network under evaluation, gt refers to each gene tree and the taxon map links individuals to species/populations to sample multiple alleles per species/population. Files used for calculations including networks and gene trees are available in figshare (https://figshare.com/collections/Andean_uplift_drainage_basin_formation_and_the_evolution_of_plants_living_in_fast-flowing_aquatic_ecosystems_in_northern_South_America-_Data_and_results/5280572)

References

Novelo RA, Philbrick CT. **1997**. Taxonomy of Mexican Podostemaceae. *Aquatic Botany* **57**: 275–303.

Novelo RA, Philbrick CT, Crow GE. **2009**. Podostemaceae. Flora Mesoamericana 3: 1–7. First published on the Flora Mesoamericana website 7 July 2009. <http://www.tropicos.org/docs/meso/podostemaceae.pdf>.

Tippery NP, Philbrick CT, Bove CP, Les DH. **2011**. Systematics and Phylogeny of Neotropical Riverweeds (Podostemaceae: Podostemoideae). *Systematic Botany* **36**: 105–118.

Chapter 3

EVOLUTION OF *LUDWIGIA* (ONAGRACEAE) IN STANDING-WATER ECOSYSTEMS ACROSS THE ANDES IN SOUTH AMERICA

Evolution of *Ludwigia* (Onagraceae) in standing-water ecosystems across the Andes in South America

Ana M. Bedoya[1,2] and Richard G. Olmstead[1]

[1]Department of Biology and Burke Museum, University of Washington, Seattle, Washington, 98195, U.S.A.

Manuscript received _____; revision accepted _____

[2]Author for correspondence: ambedoya@uw.edu

Author for correspondence:

Ana M. Bedoya

Tel: +1 2063104972

Email: ambedoya@uw.edu

Abstract

* *Premise of the study:* Andean uplift triggered diversification in montane and lowland terrestrial plant taxa and shaped the evolution of aquatic plants restricted to fast-flowing aquatic ecosystems. However, the role of the uplift of the Andes on the evolution of aquatic plants in standing-water ecosystems remains unaddressed. *Ludwigia* is a genus of aquatic plants with center of diversity in the Neotropics. In this study we aim to test if divergence times in *Ludwigia* species in northern and central South America are concurrent with timing of Andean uplift and provide a framework to test further hypotheses for the role of landscape change shaping the evolution of plants living in standing-water ecosystems.

**Methods:* Using data for one nuclear and 4 chloroplast loci, we inferred a time-calibrated tree of *Ludwigia* distributed in northern and central South America. Lineage-through-time plots and ancestral range reconstruction were performed to investigate if diversification and range evolution were shaped by Andean uplift.

**Key results:* The species of *Ludwigia* studied diverged prior, during, and after Andean uplift without significant change in diversification in the evolutionary history of species that currently inhabit Andean foreland basins and drainages in central South America. These results contrast with studies on terrestrial plants living in mountains and lowlands, as well as in aquatic plants in rivers.

**Conclusions:* We provide hypotheses for why diversification in *Ludwigia* remained constant through the Miocene onwards, as well as the probable context in which *Ludwigia* evolved in South America. This study provides the framework for future research aiming to understand how landscape change shaped the evolution of aquatic plants in standing-water ecosystems in order to gain a better understanding of the processes responsible for the assembly of the Neotropical flora.

Key words: Andean uplift, aquatic plants, biogeography, diversification, *Ludwigia*, Neotropics, South America, wetlands

The Neotropics are the most biodiverse region on earth with ca. 37% of the world's species of seed plants in this region (Antonelli and Sanmartín, 2011b; Antonelli et al., 2018). The study of how this rich flora was assembled has been the focus of extensive botanical research (Gentry, 1982; Hughes et al., 2013; Antonelli et al., 2018; Dick and Pennington, 2019). Previous studies have concluded that landscape heterogeneity from the Miocene onwards, climate fluctuations in the Pleistocene, and pulses of Andean uplift during the Miocene and Pliocene are among the factors that greatly contributed to increased diversification and assembly of the Neotropical flora (Pennington et al., 2004; Hughes and Eastwood, 2006; Antonelli and Sanmartín, 2011b)

Landscape change models developed from fossil pollen, sediment, and macrofossil data, indicate that a vast wetland system, referred to as Lake Pebas was in place in northern and central South America from ca. 23 to 10 Ma (Hoorn et al., 2009, 2010). This wetland system consisted of lentic (standing-water) ecosystems like lacustrine deltas, ponds, wetlands, and lakes where organisms typical of wetlands today (such as a hyperdiverse assemblage of co-occurring crocodilian species) thrived (Wesselingh et al., 2001; Salas-Gismondi et al., 2015; Cadena et al., 2020). Landscape models further indicate that rapid and recent pulses of Andean uplift took place in the Neogene (12–6 Ma and 4.5 Ma; (Gregory-Woodzicki, 2000; Garzione et al., 2008; Hoorn et al., 2010; Mora et al., 2010; Anderson et al., 2016; Montes et al., 2019; Boschman, 2021). These rapid pulses caused a shift from lentic to lotic (fast-flowing water) ecosystems such as fast-flowing rivers and streams in northern South America. As a consequence, the current drainage basins (Pacific, Magdalena, Orinoco, Maracaibo-Caribbean, and Amazon) were established (Hoorn et al., 2010; Latrubesse et al., 2010).

Although many species of aquatic vascular plants have a widespread distribution, aquatic plants overall are most diverse in the Neotropics, with >900 species, of which about 61% are endemic to the region (Chambers et al., 2008). Despite the great changes to aquatic ecosystems due to Andean uplift from the Miocene to present, the impact of mountain building and establishment of drainage basins on the evolution of aquatic plants in the Neotropics only started to be addressed recently (Bedoya et al., 2021). Phylogenetic and population genetic studies indicated that divergence of populations of plants restricted to lotic ecosystems and inhabiting separate drainage basins in northern South America, occurred in conjunction with rapid pulses of Andean uplift (Bedoya et al., 2021). This pattern has been explained by the fact that dispersal in plants living strictly in fast-flowing aquatic ecosystems are known to be constrained by physical connections across rivers (Bedoya et al., 2021), as is the case in other riverine taxa (Waters et al., 2000; Albert et al., 2006; Albert and Crampton, 2010; Tagliacollo et al., 2015).

Taxa in lentic ecosystems occupy wider ranges, whereas organisms bound to lotic ecosystems tend to have a narrower distribution (Hof et al., 2006; Arribas et al., 2012). It has been suggested that the fact that lentic ecosystems are mote temporary and fragmented than lotic ecosystems selects for individuals with a more effective dispersal ability (Paradis et al., 1998; Karrenberg et al., 2002; Hof et al., 2006; Soons, 2006; Soons et al., 2017). The same reasoning predicts that lotic ecosystems are more predictable in space and time and more homogeneous than lentic ecosystems, which are on average young and short lived and immersed in a matrix of terrestrial habitats where the water level significantly fluctuates (Dobson and Frid, 2008; Ribera et al., 2008; Arribas et al., 2012).

Given the mentioned differences in dispersal ability in lentic and lotic ecosystems, the hypothesized shift from predominantly lentic ecosystems to lotic ecosystems, and the consequent fragmentation of lentic ecosystems into the current drainage basins due to Andean uplift, here we aim to investigate the impact of Andean uplift on the evolution of plants living in standing-water ecosystems across drainage basins interrupted by this mountain system in northern South America. We address this question using *Ludwigia* (Onagraceae), a genus of aquatic plants with a pantropical distribution and especially well represented in South and North America (Wagner et al., 2007). The genus includes 82 species of which 39 are endemic or centered in South America and 21 are distributed in lentic ecosystems across the Andes in northern South America (Appendix S1; Wagner et al., 2007). *Ludwigia* species are herbs, shrubs, and small trees that can be classified as hydrophytes (strictly aquatic plants living entirely in the water as emergent, submerged, or floating) or helophytes (amphibious plants which live in seasonally flooded environments like savannas and ponds; Fig. 1). The species present adaptations to the anoxic conditions typical of water-logged environments (Bedoya and Madriñán, 2015). Previous phylogenetic work in the group inferred that Neotropical *Ludwigia* forms a clade sister to North American *Ludwigia*, with some African taxa nested in the former. The monophyly of Neotropical *Ludwigia* has also been confirmed and it has been proposed that the genus originated in South America around 90 Ma (Sytsma et al., 2004; Berger et al., 2016; Freyman and Höhna, 2019).

The distribution of *Ludwigia* across the Andes in South America and the fact that the group inhabits lentic ecosystems, make it an appropriate model to 1) test if divergence times in taxa in northern and central South America took place prior, during, or after current estimates of major recent pulses of Andean uplift and 2) provide a framework to test further hypotheses for the role of landscape change in shaping the evolution of plants living in lentic ecosystems. We infer phylogenetic relationships, divergences times, and ancestral ranges for a nearly comprehensive dataset of *Ludwigia* species. Finally, we compare our

results to previous findings of the role of Andean uplift and drainage basin reconfiguration on the evolution of plants in rivers and in terrestrial ecosystems.

MATERIALS AND METHODS

Sampling– To provide a phylogenetic framework and estimate divergence times of *Ludwigia* in Andean foreland basins and basins in central South America, we mined DNA sequences for 49 Onagraceae (39 *Ludwigia* species) and two Lythraceae species from GenBank (Table 1). *Ludwigia* species included in this study include 28 out of the 39 species endemic or centered in South America. In total, one nuclear and four chloroplast regions were used (*ITS, rpl16, rps16, trnG, trnL*; Table 1). Some chloroplast sequences were extracted from plastome data (Table 1) by aligning a plastome sequence to a reference sequence with MAFFT v7.309 (Katoh and Standley, 2013) as implemented in Geneious v9.1.8 (Biomatters Ltd., Auckland, New Zealand) and using the 'Extract Regions' option in Geneious. Most sequences corresponded to sequences used in previous phylogenetic inferences of *Ludwigia* (Bedoya and Madriñán, 2015; Liu et al., 2017, 2020) and a comprehensive sampling of *Ludwigia* species from South America.

Phylogenetic inference, divergence dating and diversification–Sequences for each locus were aligned with MAFFT v7.309 as implemented in Geneious v9.1.8 using default settings followed by manual inspection. A time-calibrated tree was inferred in RevBayes (Höhna et al., 2016) using an uncorrelated lognormal relaxed clock and a birth-death tree model. We specified the GTR+G substitution model for each partition. Node ages were calibrated using five fossil calibrations and one secondary calibration following Freyman and Höhna, 2019. (1) A fossil assigned to *Fuchsia* section Skinnera (23 Ma; lognormal distribution with μ=10, SD 1.0 offset by age of fossil for stem age of the group; Lee et al., 2013); (2) The oldest fossils in Lythraceae from the Late Cretaceous (*Lythrum* and *Peplis*, 81.5 Ma; lognormal distribution with μ=20, SD=2.0 offset by age of fossil for stem age of the group; Grímsson et al., 2011); (3) *Ludwigia* fossil pollen from the Paleocene (57.6 Ma; lognormal distribution with μ=10, SD=1.5 offset by age of fossil; Song et al., 2004); (4) *Epilobium* fossil pollen from the Miocene (12 Ma; lognormal distribution with μ=10, SD=2.0 offset by age of fossil; Grímsson et al., 2012); (5) Circaeae fossil pollen from the Miocene (12 Ma; lognormal distribution with μ=10, SD=2.0 offset by age of fossil; Grímsson et al., 2012); secondary calibration for the Late Cretaceous split of Lythraceae-Onagraceae (ca 93 Ma; normal distribution with μ=93, SD=5.0; Sytsma et al., 2004).

We ran two chains for 1 million generations, sampling every 10 generations, and assessed convergence of chains and stationarity and mixing of all parameters with Tracer v1.7.1 (Rambaut et al., 2018). Trees from

the two independent runs were combined with LogCombiner v2.6.3 after removing a 25% burnin fraction from each run. A maximum clade credibility tree was inferred with RevBayes (trace= readTreeTrace("Rev_combined.trees") and then mccTree(trace, file="Rev_mcc.tre"), annotating node ages using the mean age.

The rate of lineage accumulation through time in the *Ludwigia* clade with most species in Andean foreland basins and basins in central South America were assessed with a lineage-through-time (LTT) plot on the maximum clade credibility time-calibrated tree (MCC) using the *ltt* function of the *phytools* package (Revell, 2012) in R v4.0.2. The outgroups and North America taxa were dropped from the MCC using the function *drop.tip* of the *ape* package (Paradis and Schliep, 2019) in R. To test if lineage accumulation varied through time, we compared the observed LTT to one thousand simulated LTTs assuming a pure-birth process over the same time span and total number of resulting species. Simulated trees used were generated with the *pbtree* function of *phytools*. We implemented the Monte Carlo constant-rates (MCCR) test as implemented in the function *mccr* of *phytools* and estimated the Pybus and Harvey's gamma (γ) statistic (Pybus and Harvey, 2000) from the incompletely sampled phylogeny. A sampling fraction of 0.57 was specified (ca. 59 species of *Ludwigia* in the clade sister to the North American clade, 34 of which are included in this study).

***Biogeography of Ludwigia and its link to landscape change in nSA*–**The maximum clade credibility time-calibrated phylogeny inferred in RevBayes was used for ancestral range reconstruction in BioGeoBEARS (Matzke, 2013). We specified five broad geographic areas modified from Buerki et al., 2011; Berger et al., 2016, to fit the various ranges occupied by *Ludwigia* species (Wagner et al., 2007): (1) North America (N); (2) South America (S; including Southern Mexico, Central America, and the West Indies); (3) Eurasia (A; from western Europe to Asia); (4) Africa and Madagascar (F); (5) Oceania (O). Species were assigned to areas following the current monograph of *Ludwigia* (Wagner et al., 2007), the Onagraceae family website of the Smithsonian National Museum of Natural History (http://botany.si.edu/onagraceae/index.cfm), and the distribution of publicly available records in GBIF (https://www.gbif.org; see Appendix S2–S10 for geographic data of *Ludwigia* species created with the package *rgbif* (Chamberlain and Boettiger, 2017) in R v4.0.2 . Input file in FigShare; available upon publication). To reduce the state space (allowed states), the list of geographic ranges was manually modified to eliminate disjunct geographic ranges that are not displayed by any *Ludwigia* species (South America-Oceania, North America-Asia, Africa-Oceania, North America-Oceania, and North America-Africa). Biogeographic inference was conducted under the dispersal-extinction-cladogenesis (DEC-like) model (Ree et al., 2005; Ree and Smith, 2008). The analysis was non-stratified (unconstrained) and we did not specify a distance matrix given the well-known ability of *Ludwigia* to disperse long distances and

invade new habitats (Ruaux et al., 2009; Green, 2016; Gillard, Grewell, Futrell, et al., 2017), in addition to the unequal distribution of species across the defined geographic areas. We did not analyze our data under the DEC+J model, because the jump ("J") parameter is conceptually flawed (Ree and Sanmartín, 2018).

RESULTS

***Phylogenetic inference, divergence dating, and diversification*–**The time-calibrated tree inferred with RevBayes (Fig. 2) is in line with previous phylogenetic work on *Ludwigia* (Bedoya and Madriñán, 2015; Liu et al., 2017), recovering *Ludwigia* species endemic to North America as a clade sister to the rest of the species in the genus. *Ludwigia ovalis*, a species distributed in northern China, Taiwan, Japan, and Korea, (Wagner et al., 2007) is nested in this North American clade. The four species endemic to Africa and included in this study are nested in a clade including all Central and South American *Ludwigia*. The latter inhabit Andean foreland basins as well as drainage basins in central and eastern South America (Appendix S2–S10). Strongly supported relationships recovered across species, as well as the timing for the split of the North American clade (ca. 31–19 Ma) from the rest of *Ludwigia* also agree with previous phylogenetic work using both chloroplast and nuclear data (Liu et al., 2017) and with the *ITS* gene tree reported by (Bedoya and Madriñán, 2015). Discordance in the previously published trees exists in the placement of *L. peploides* (sister to *L. stolonifera* and *L. adscendens* in Liu et al., 2017; Fig. 2). Other topological incongruences are not highly supported either in this or in previous studies.

The split between the North American *Ludwigia* clade from the rest of *Ludwigia* took place ca. 25 Ma, before the closure of the isthmus of Panama (Fig. 2). Divergence events in *Ludwigia* inhabiting Andean foreland basins and basins in central South America took place before, during, and after previously proposed rapid pulses of Andean uplift, with most speciation events taking place during and after these previously proposed geologic events (Fig. 2). However, lineage-though-time plots show that lineage accumulation showed no significant changes through time. Indeed, Monte Carlo constant-rates accounting for incomplete sampling suggest that lineage accumulation did not vary through time in the Central and South American clade (negative gamma (γ) statistic; γ=-1.9209, p-value=0.294).

Biogeography of Ludwigia and its link to landscape change in nSA–Ancestral range reconstruction with BioGeoBEARS indicated that after the North American clade of *Ludwigia* diverged from the predominantly Central and South American clade, the ancestor of the latter clade was most likely distributed in South/Central America (Fig. 3). This at a time when a vast wetland, the Pebas lake, dominated a great portion of northern and Central South America. Long-distance dispersal events to Africa, Asia, and Oceania were inferred to have taken place repeatedly in this clade, mostly within the last 10 My.

A clade of five species restricted to the Paraná basin in central-eastern South America (*L. hassleriana*, *L. irwinii*, *L. martii*, and *L. myrtifolia*; (Fig. 3; Appendix S2, S3) is recovered. Three other species included in this study (*L. lagunae, L. neograndiflora, L. major,* and *L. pseudonarcissus*; Fig. 3, Appendix S4, S5, S6) are restricted to that drainage basin but are not nested in this clade. Four species (*L. bonariensis, L. grandiflora, L. hexapetala,* and *L. peploides*) have disjunct distributions in central-eastern South America and North America (although *L. peploides* extends to Central America; Appendix S8, S9). The remaining South American species have wide ranges that span Central America through South America (Fig. 3). Of the 28 species included and present in South America, nine species have distributions outside the Neotropics (Fig. 3).

DISCUSSION

Landscape changes through the Miocene onwards in northern and central South America did not have a marked impact on diversification patterns of *Ludwigia*, a group of aquatic plants living in lentic ecosystems with center of diversity in the Neotropics. Here we show that divergence times in *Ludwigia* species living in Andean foreland basins as well as drainage basins in eastern South America took place before, during, and after suggested times for major recent pulses of the Andes. We further show that the rate of lineage accumulation did not vary significantly through time in this group. Our results imply that extant lineages of *Ludwigia* in South America originated when the Pebas system was a dominant feature in the landscape of Central and northern South America (Wesselingh et al., 2001; Antonelli et al., 2009; Hoorn et al., 2010). The group continued to diversify as further uplift of the Andes resulted in the disappearance of the Pebas Lake and in the assembly of the heterogeneous landscape matrix that exists today. These results contrast with the previously reported strong impact of drainage basin reconfiguration linked to Andean uplift on the evolution of aquatic plants restricted to lotic ecosystems, where divergence times occurred in conjunction with pulses of uplift of the Andes (Bedoya et al., 2021). Our study provides the framework for testing of subsequent hypotheses for the role of landscape and climate change on the evolution of plants in lentic ecosystems across South America. Below we discuss the possible reasons for

the steady diversification of *Ludwigia* in South America through time and suggest future research avenues to continue assessing the impact of landscape change in South America on the evolution of plants in standing-water ecosystems.

Phylogenetic inference, divergence dating, and diversification– Although our data overall show that the diversification of *Ludwigia* in South America remained constant through time, a slight increase in species accumulation is detected around the time when the Pebas lake disappeared (Fig. 2). However, this does not differ significantly from what would be expected under a pure-birth process. These findings imply that 1) factors promoting diversification in the group remained constant at least from the late Oligocene onwards, 2) that changes in climate and landscape change in South America did not exert any influence on diversification patterns in this group of plants, or 3) that different factors taking place at various times exerted a similar impact on species accumulation through time in *Ludwigia* species in South America. These hypotheses are not mutually exclusive. Given the dynamic history of landscape change in South America (Hoorn et al., 2010) and the known global climatic changes throughout the Cenozoic (Zachos, 2001), the first scenario seems unlikely.

The clade composed mainly of Central and South American species (also including some African species) first started to diversify at the time when the Pebas lake was still in place before major recent pulses of Andean uplift took place (Fig. 2). A previous investigation of the biogeographic history of the order Myrtales concluded that the family Onagraceae and *Ludwigia* most likely originated in South America and dispersed to Africa (Berger et al., 2016). This is consistent with our conclusion that the mainly Central American and South American clade was present in South America when a vast wetland covered what is now western Amazonia. The mosaic of aquatic ecosystems comprising the Pebas lake over ca. one million km^2 included permanent and temporal lakes, slow-moving rivers, cut-off meanders, swamps, and gallery forests (Wesselingh et al., 2001, 2006; Latrubesse et al., 2010; Gross et al., 2011) and it is thought to have promoted diversification of fauna adapted to the aquatic conditions characteristic of this formation such as crocodilian species (Muñoz-Torres et al., 2006; Wesselingh, 2006; Salas-Gismondi et al., 2015). This mosaic would have provided ample physical space and diversity of habitats (ponds, lakes, swamps, transition zones, flooded plains, cutoff meanders, etc) to promote diversification in *Ludwigia.* However, whether or not habitat diversity correlates with an increase in ecological strategies represented in *Ludwigia* remains to be investigated.

Divergence events in *Ludwigia* also took place after the draining of the Pebas lake system, concomitant with the onset of recent pulses of Andean uplift (Fig. 2). Major mountain building in the region marked a

transition towards the formation of a complex landscape matrix with wetlands, inter-Andean valleys, and three mountain ranges that created the current lotic environments and caused rain shadow effects and microclimate heterogeneity (Houston and Hartley, 2003; Richardson et al., 2018). Andean uplift also resulted in the establishment of the current drainage basins including the Amazon, the formation of high-elevation tropical ecosystems, and the fragmentation of previously connected areas (Gentry, 1982; Hughes and Eastwood, 2006; Brumfield and Edwards, 2007). Landscape heterogeneity is a major promoter of species diversification (Gentry, 1982, 1988; Fine et al., 2005; Kreft and Jetz, 2007; Antonelli and Sanmartín, 2011b; Smith et al., 2014) and the complex landscape matrix resulting from Andean uplift in South America is believed to have fostered diversification (Smith et al., 2014) with valleys allowing episodic dispersal of lowland animal species (Cadena et al., 2016). The landscape matrix described above could explain why the rate of lineage accumulation through time in *Ludwigia* did not decrease even after the Pebas lake disappeared. As mentioned above, whether or not habitat diversity correlates with an increase in ecological strategies represented in *Ludwigia* remains to be investigated. Additional factors such as climate fluctuations in the Pleistocene could have contributed to speciation in this group of both strictly aquatic and amphibious plants. Pleistocene cooling events have been hypothesized to be one of the drivers of diversification of the Neotropical flora in seasonally-dry forests, mountains, and grasslands (Pennington et al., 2004; Antonelli and Sanmartín, 2011a; Madriñán et al., 2013; Hughes and Atchison, 2015; Flantua et al., 2019; Vasconcelos et al., 2020). At this time, the extent and assembly of lentic ecosystems could have been affected by the expansion of grasslands, perhaps resulting in the fragmentation of lentic habitats. The impact of climate fluctuations in the Pleistocene on lentic ecosystems and its impact of the flora in these aquatic environments remain to be investigated.

The possibility remains however, that when the predominantly Central and South American clade of *Ludwigia* first started to diversify, the lineage was not centered in the area occupied by the Pebas lake, nor where species distributed in areas where mountain uplift took place. The Paraná drainage basin, which include the wet pampas and the Iberá wetlands (second largest wetlands on earth; Fig. 2), was in place during the Cenozoic (Brea and Zucol, 2011). Multiple species of *Ludwigia* currently inhabit this region with a clade <6 Ma old consisting of species centered in the area. The group may have been centered there in the past as well, although marine incursions invaded the region during the Miocene (Ruskin et al., 2011; Appendix S11). Presence of high diversity and endemism in the Paraná drainage basin suggests that mountain uplift was not necessary for diversification of *Ludwigia,* as there has been little topographic change in this region in recent time (Brea and Zucol, 2011).

In either case, Andean uplift did not result in a significant change in diversification of *Ludwigia* species that currently inhabit Andean foreland basins and drainages in northern and central South America. These results contrast with results obtained from several studies on terrestrial plants living in mountains and lowlands showing accelerated rates of diversification linked to Andean uplift (Hughes and Eastwood, 2006; Madriñán et al., 2013; Lagomarsino et al., 2016; Pérez-Escobar et al., 2017; Testo et al., 2019). Furthermore, the results presented in this study contrasts with those of aquatic plants living strictly in rivers, where divergence times occurred in conjunction with recent pulses of Andean uplift (Bedoya et al., 2021). These contrasts highlight that not all plant groups responded in the same way to Andean uplift, and that in the case of *Ludwigia*, their association to standing-water ecosystems may be a significant factor in the steady accumulation of species though time reported here. Particularly, the more effective dispersal ability resulting in wider ranges in plants in wetlands (Paradis et al., 1998; Karrenberg et al., 2002; Hof et al., 2006; Soons, 2006; Soons et al., 2017).

To test the hypotheses laid out above, future studies should focus on inferring phylogenetic relationships of populations within species distributed in regions across northern and central South America, on studying the spatial distribution of genetic variability in those populations, and on investigating ecological strategies in *Ludwigia* in a phylogenetic and diversification framework. These would help determine how populations evolved across the landscape matrix through time, if there are geographic barriers to gene flow, and what the drivers of diversification may be in this group.

***Biogeography of Ludwigia and its link to landscape change in nSA*–** The most recent common ancestor of the Central and South American *Ludwigia* clade was most likely present in the region (Fig.3), as inferred previously for the family Onagraceae (Berger et al., 2016), where *Ludwigia* is sister to the rest of the genera in the family. Multiple inferred long-distance dispersal events are consistent with the invasive potential of *Ludwigia* (Okada et al., 2009; Ruaux et al., 2009; Gillard, Thiébaut, et al., 2017; Skaer Thomason et al., 2018) and the well-documented widespread distribution of some aquatic plants, particularly of those inhabiting wetlands (Paradis et al., 1998; Karrenberg et al., 2002; Hof et al., 2006; Soons, 2006; Soons et al., 2017). Water and wind dispersal have been reported for seeds and vegetative structures in *Ludwigia* (Okada et al., 2009; Ruaux et al., 2009; Gillard, Grewell, Deleu, et al., 2017), with the retention of long-term buoyancy of seeds that allow water dispersal over long distances (Eyde, 1978; Ruaux et al., 2009; Grewell et al., 2019).

Various *Ludwigia* species that were inferred to have originated after the final uplift of the Andes (*L. leptocarpa, L. rigida, L. erecta, L. affinis, L. nervosa, and L. octovalvis*) are currently distributed in

standing-water ecosystems across the Andes (Fig. 3, Appendix S3, S4). This is an indicator that the Andes are not a barrier for dispersal of *Ludwigia*. However, it remains to be determined if populations are genetically structured across space. A trans-Andean fluvial portal is hypothesized to have existed across the Andes in northern South America from ca. 13 to 4 Ma (Montes et al., 2021), and fluvial connections across drainage basins could have remained throughout that interval of time, facilitating the dispersal of *Ludwigia* species across the Andes. Whether or not gene flow is halted by the Andes, it remains to be investigated if isolation by distance takes place in populations of *Ludwigia* species across the Andes and what the modes of speciation in the group are.

The monophyletic group of species restricted to the Paraná drainage basin diverged relatively late compared to other groups in the Central and South American *Ludwigia* clade. This may point to a late migration of *Ludwigia* to southern South America. However, several species of *Ludwigia* with distributions in the Neotropics remain to be included in a complete phylogenetic analysis of the group to make more confident biogeographical inferences within Central and South America *Ludwigia*.

CONCLUSIONS

Higher biodiversity in the Neotropics compared to other tropical areas was explained by Gentry as a consequence of rapid Andean uplift triggering species diversification (Gentry, 1982). This study explores the role of Andean uplift on the evolution of plants in standing-water ecosystems using *Ludwigia* as a model system. It represents a first attempt in elucidating how the dramatic changes to aquatic ecosystems that took place in northern South America from the Miocene onwards shaped the evolution of aquatic plants in standing-water ecosystems. Even though Andean uplift triggered diversification of lowland and high elevation terrestrial taxa and directly shaped the evolution of plants in rivers (Bedoya et al., 2021), it did not cause a change in the rates of accumulation of *Ludwigia* species through time. We provide hypotheses for why diversification in *Ludwigia* remained constant, as well as the probable context in which *Ludwigia* evolved in South America.

The divergence time inference presented here is a means to explore the parameter space of evolutionary scenarios and constitutes the basis for our subsequent analyses. Further testing of these results should include increasing the sampling of species and of genomic regions to conduct species tree inference incorporating gene flow. This given that some speciation events were inferred to be recent in geological time and that natural hybridization is regarded as fairly common in *Ludwigia,* playing a key role in species evolution with some hybrid speciation events reported in the group (Raven and Tai, 1979; Peng, 1989; Zardini et al., 1991; Nesom and Kartesz, 2000; Peng et al., 2005; Hung et al., 2009; Liu et al.,

2020). Therefore, phylogenetic methods that incorporate a strictly bifurcating tree model and ignore incomplete lineage sorting as a cause of gene tree discordance may result in biased estimates of relationships and divergence times (Leaché et al., 2014). In addition, population-level studies would help test if there is gene flow across the Andes or if isolation-by-distance takes place.

This study provides the framework for future research aiming to understand how landscape change shaped the evolution of aquatic plants in standing-water ecosystems in order to gain a better understanding of the processes responsible for the assembly of the Neotropical flora. Furthermore, several species of *Ludwigia* are considered to be invasive and dispersed by propagules and seed dispersal through hydrochory (Okada et al., 2009; Ruaux et al., 2009; Gillard, Thiébaut, et al., 2017; Skaer Thomason et al., 2018). Determining what factors limit species dispersal has important implications in preservation of ecosystems and management of invasive species.

ACKNOWLEDGMENTS

Adam Leaché provided advice and comments on the manuscript. Caroline Strömberg also provided comments to earlier versions of the manuscript. We thank Elena Stiles for insightful discussions. Santiago Madriñán at the Universidad de los Andes and the Guillermo Piñeres Botanical Garden in Colombia provided help with collecting permits and access to tissues of *Ludwigia* stored at the ANDES herbarium for DNA extraction collected as part of an inventory of the aquatic flora of the Orinoco basin of Colombia in 2012. Fernando Bedoya, Mateo Fernández, Maria Paula Contreras, and Adrián Pinzón provided assistance in the field. Samples were collected and transported under the collecting permit XX and export permit XX under the MOU with UNIANDES and BURKE. We thank Mireya Osorio and Yiselle Cano at the Universidad de los Andes for help with permit logistics. Funding was provided by the Botanical Society of America and American Society of Plant Taxonomists Research Awards, Explorers Club Exploration Fund Grant, Giles Award, WRF Hall International Endowed Fellowship, and Denton writing Fellowship from the Biology department at the University of Washington, UW Herbarium Endowment, UW Graduate School Boeing International Fellowship, and the Colciencias fellowship for Graduate studies (Doctorados en el Exterior-679) to A.M.B. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant.

AUTHOR CONTRIBUTIONS

A.M.B. designed the study advised by R.G.O. A.M.B. collected and analyzed the data; R.G.O. and provided guidance on the methods for data analyses and interpretation of results, as well as reagents, equipment, and laboratory space. A.M.B. wrote the manuscript with significant contributions from R.G.O.

DATA AVAILABILITY STATEMENT

All data will be available upon acceptance in FigShare.

ONLINE SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article and in FigShare. FigShare link will be provided upon acceptance.

Appendix S1. Species of *Ludwigia* present in northern South America.

LITERATURE CITED

Albert, J. S., and W. G. R. Crampton. 2010. The geography and ecology of diversification in Neotropical freshwaters. 3: 13.

Albert, J. S., N. R. Lovejoy, and W. G. R. Crampton. 2006. Miocene tectonism and the separation of cis- and trans-Andean river basins: Evidence from Neotropical fishes. *Journal of South American Earth Sciences* 21: 14–27.

Anderson, V. J., B. K. Horton, J. E. Saylor, A. Mora, E. Tesón, D. O. Breecker, and R. A. Ketcham. 2016. Andean topographic growth and basement uplift in southern Colombia: Implications for the evolution of the Magdalena, Orinoco, and Amazon river systems. *Geosphere* 12: 1235–1256.

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

Antonelli, A., and I. Sanmartín. 2011a. Mass Extinction, Gradual Cooling, or Rapid Radiation? Reconstructing the Spatiotemporal Evolution of the Ancient Angiosperm Genus Hedyosmum (Chloranthaceae) Using Empirical and Simulated Approaches. *Systematic Biology* 60: 596–615.

Antonelli, A., and I. Sanmartín. 2011b. Why are there so many plant species in the Neotropics? *TAXON* 60: 403–414.

Antonelli, A., A. Zizka, F. A. Carvalho, R. Scharn, C. D. Bacon, D. Silvestro, and F. L. Condamine. 2018. Amazonia is the primary source of Neotropical biodiversity. *Proceedings of the National Academy of Sciences* 115: 6034–6039.

Arribas, P., J. Velasco, P. Abellán, D. Sánchez-Fernández, C. Andújar, P. Calosi, A. Millán, et al. 2012. Dispersal ability rather than ecological tolerance drives differences in range size between lentic and lotic water beetles (Coleoptera: Hydrophilidae): Dispersal ability and range size in water beetles. *Journal of Biogeography* 39: 984–994.

Bedoya, A. M., A. D. Leaché, and R. G. Olmstead. 2021. Andean uplift, drainage basin formation, and the evolution of plants living in fast-flowing aquatic ecosystems in northern South America. *New Phytologist*: nph.17649.

Bedoya, A. M., and S. Madriñán. 2015. Evolution of the aquatic habit in Ludwigia (Onagraceae): Morpho-anatomical adaptive strategies in the Neotropics. *Aquatic Botany* 120: 352–362.

Berger, B. A., R. Kriebel, D. Spalink, and K. J. Sytsma. 2016. Divergence times, historical biogeography, and shifts in speciation rates of Myrtales. *Molecular Phylogenetics and Evolution* 95: 116–136.

Boschman, L. M. 2021. Andean mountain building since the Late Cretaceous: A paleoelevation reconstruction. *Earth-Science Reviews*: 103640.

Brea, M., and A. F. Zucol. 2011. The Paraná-Paraguay Basin: Geology and Paleoenvironments. *In* J. Albert [ed.], Historical Biogeography of Neotropical Freshwater Fishes, 68–87. University of California Press.

Brumfield, R. T., and S. V. Edwards. 2007. EVOLUTION INTO AND OUT OF THE ANDES: A BAYESIAN ANALYSIS OF HISTORICAL DIVERSIFICATION IN *THAMNOPHILUS* ANTSHRIKES: EVOLUTION INTO AND OUT OF THE ANDES. *Evolution* 61: 346–367.

Buerki, S., F. Forest, N. Alvarez, J. A. A. Nylander, N. Arrigo, and I. Sanmartín. 2011. An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae: Evaluating parametric versus parsimony-based methods in biogeography. *Journal of Biogeography* 38: 531–550.

Cadena, C. D., C. A. Pedraza, and R. T. Brumfield. 2016. Climate, habitat associations and the potential distributions of Neotropical birds: Implications for diversification across the Andes. *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales* 40: 275.

Cadena, E.-A., T. M. Scheyer, J. D. Carrillo-Briceño, R. Sánchez, O. A. Aguilera-Socorro, A. Vanegas, M. Pardo, et al. 2020. The anatomy, paleobiology, and evolutionary relationships of the largest extinct side-necked turtle. *Science Advances* 6: eaay4593.

Chamberlain, S. A., and C. Boettiger. 2017. R Python, and Ruby clients for GBIF species occurrence data. PeerJ Preprints.

Chambers, P. A., P. Lacoul, K. J. Murphy, and S. M. Thomaz. 2008. Global diversity of aquatic macrophytes in freshwater. *In* E. V. Balian, C. Lévêque, H. Segers, and K. Martens [eds.], Freshwater Animal Diversity Assessment, Developments in Hydrobiology, 9–26. Springer Netherlands, Dordrecht.

Dick, C. W., and R. T. Pennington. 2019. History and Geography of Neotropical Tree Diversity. *Annual Review of Ecology, Evolution, and Systematics* 50: 279–301.

Dobson, M., and C. Frid. 2008. Ecology of aquatic systems. Oxford University Press.

Eyde, R. H. 1978. Reproductive Structures and Evolution in Ludwigia (Onagraceae). II. Fruit and Seed. *Annals of the Missouri Botanical Garden* 65: 656.

Fine, P. V. A., D. C. Daly, G. V. Muñoz, I. Mesones, and K. M. Cameron. 2005. THE CONTRIBUTION OF EDAPHIC HETEROGENEITY TO THE EVOLUTION AND DIVERSITY OF BURSERACEAE TREES IN THE WESTERN AMAZON. *Evolution* 59: 1464.

Flantua, S. G. A., A. O'Dea, R. E. Onstein, C. Giraldo, and H. Hooghiemstra. 2019. The flickering connectivity system of the north Andean páramos. *Journal of Biogeography* 46: 1808–1825.

Freyman, W. A., and S. Höhna. 2019. Stochastic Character Mapping of State-Dependent Diversification Reveals the Tempo of Evolutionary Decline in Self-Compatible Onagraceae Lineages. *Systematic Biology* 68: 505–519.

Garzione, C. N., G. D. Hoke, J. C. Libarkin, S. Withers, B. MacFadden, J. Eiler, P. Ghosh, and A. Mulch. 2008. Rise of the Andes. *Science* 320: 1304–1307.

Gentry, A. H. 1988. Changes in Plant Community Diversity and Floristic Composition on Environmental and Geographical Gradients. *Annals of the Missouri Botanical Garden* 75: 1.

Gentry, A. H. 1982. Neotropical Floristic Diversity: Phytogeographical Connections Between Central and South America, Pleistocene Climatic Fluctuations, or an Accident of the Andean Orogeny? *Annals of the Missouri Botanical Garden* 69: 557.

Gillard, M., B. J. Grewell, C. Deleu, and G. Thiébaut. 2017. Climate warming and water primroses: Germination responses of populations from two invaded ranges. *Aquatic Botany* 136: 155–163.

Gillard, M., B. J. Grewell, C. J. Futrell, C. Deleu, and G. Thiébaut. 2017. Germination and Seedling Growth of Water Primroses: A Cross Experiment between Two Invaded Ranges with Contrasting Climates. *Frontiers in Plant Science* 8: 1677.

Gillard, M., G. Thiébaut, C. Deleu, and B. Leroy. 2017. Present and future distribution of three aquatic plants taxa across the world: decrease in native and increase in invasive ranges. *Biological Invasions* 19: 2159–2170.

Green, A. J. 2016. The importance of waterbirds as an overlooked pathway of invasion for alien species F. Essl [ed.],. *Diversity and Distributions* 22: 239–247.

Gregory-Woodzicki, K. M. 2000. Uplift history of the Central and Northern Andes: A Review. 112: 1091–1105.

Grewell, B. J., M. B. Gillard, C. J. Futrell, and J. M. Castillo. 2019. Seedling Emergence from Seed Banks in Ludwigia hexapetala-Invaded Wetlands: Implications for Restoration. *Plants* 8: 451.

Grímsson, F., R. Zetter, and C.-C. Hofmann. 2011. *Lythrum* and *Peplis* from the Late Cretaceous and Cenozoic of North America and Eurasia: New evidence suggesting early diversification within the Lythraceae. *American Journal of Botany* 98: 1801–1815.

Grímsson, F., R. Zetter, and Q. Leng. 2012. Diverse fossil Onagraceae pollen from a Miocene palynoflora of north-east China: early steps in resolving the phytogeographic history of the family. *Plant Systematics and Evolution* 298: 671–687.

Gross, M., W. E. Piller, M. I. Ramos, and J. Douglas da Silva Paz. 2011. Late Miocene sedimentary environments in south-western Amazonia (Solimões Formation; Brazil). *Journal of South American Earth Sciences* 32: 169–181.

Hof, C., M. Brandle, and R. Brandl. 2006. Lentic odonates have larger and more northern ranges than lotic species. *Journal of Biogeography* 33: 63–70.

Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P. Huelsenbeck, and F. Ronquist. 2016. RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language. *Systematic Biology* 65: 726–736.

Hoorn, C., F. P. Wesselingh, J. Hovikoski, and J. Guerrero. 2009. The development of the Amazonian mega-wetland (Miocene; Brazil, Colombia, Peru, Bolivia). Amazonia: Landscape and species evolution: A look into the past, 123–142. Wiley-Blackwell Publishing Ltd., Oxford, UK.

Hoorn, C., F. P. Wesselingh, H. ter Steege, M. A. Bermudez, A. Mora, J. Sevink, I. Sanmartin, et al. 2010. Amazonia Through Time: Andean Uplift, Climate Change, Landscape Evolution, and Biodiversity. *Science* 330: 927–931.

Houston, J., and A. J. Hartley. 2003. The central Andean west-slope rainshadow and its potential contribution to the origin of hyper-aridity in the Atacama Desert. *International Journal of Climatology* 23: 1453–1464.

Hughes, C. E., and G. W. Atchison. 2015. The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains. *New Phytologist* 207: 275–282.

Hughes, C. E., R. T. Pennington, and A. Antonelli. 2013. Neotropical Plant Evolution: Assembling the Big Picture: Neotropical Plant Evolution. *Botanical Journal of the Linnean Society* 171: 1–18.

Hughes, C., and R. Eastwood. 2006. Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences* 103: 10334–10339.

Hung, K.-H., B. A. Schaal, T.-W. Hsu, Y.-C. Chiang, C.-I. Peng, and T.-Y. Chiang. 2009. Phylogenetic relationships of diploid and polyploid species in *Ludwigia* sect. *Isnardia* (Onagraceae) based on chloroplast and nuclear DNAs. *TAXON* 58: 1216–1226.

Karrenberg, S., P. J. Edwards, and J. Kollmann. 2002. The life history of Salicaceae living in the active zone of floodplains: Salicaceae on flood plains. *Freshwater Biology* 47: 733–748.

Katoh, K., and D. M. Standley. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 30: 772–780.

Kreft, H., and W. Jetz. 2007. Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences* 104: 5925–5930.

Lagomarsino, L. P., F. L. Condamine, A. Antonelli, A. Mulch, and C. C. Davis. 2016. The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytologist* 210: 1430–1442.

Latrubesse, E. M., M. Cozzuol, S. A. F. da Silva-Caminha, C. A. Rigsby, M. L. Absy, and C. Jaramillo. 2010. The Late Miocene paleogeography of the Amazon Basin and the evolution of the Amazon River system. *Earth-Science Reviews* 99: 99–124.

Leaché, A. D., R. B. Harris, B. Rannala, and Z. Yang. 2014. The Influence of Gene Flow on Species Tree Estimation: A Simulation Study. *Systematic Biology* 63: 17–30.

Lee, D. E., J. G. Conran, J. M. Bannister, U. Kaulfuss, and D. C. Mildenhall. 2013. A fossil *Fuchsia* (Onagraceae) flower and an anther mass with in situ pollen from the early Miocene of New Zealand. *American Journal of Botany* 100: 2052–2065.

Liu, S.-H., P. C. Hoch, M. Diazgranados, P. H. Raven, and J. C. Barber. 2017. Multi-locus phylogeny of *Ludwigia* (Onagraceae): Insights on infra-generic relationships and the current classification of the genus. *Taxon* 66: 1112–1127.

Liu, S.-H., H.-A. Yang, Y. Kono, P. C. Hoch, J. C. Barber, C.-I. Peng†, and K.-F. Chung. 2020. Disentangling Reticulate Evolution of North Temperate Haplostemonous Ludwigia (Onagraceae). *Annals of the Missouri Botanical Garden* 105: 163–182.

Madriñán, S., A. J. Cortés, and J. E. Richardson. 2013. Páramo is the world's fastest evolving and coolest biodiversity hotspot. *Frontiers in Genetics* 4.

Matzke, N. J. 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography* 5.

Montes, C., A. F. Rodriguez-Corcho, G. Bayona, N. Hoyos, S. Zapata, and A. Cardona. 2019. Continental margin response to multiple arc-continent collisions: The northern Andes-Caribbean margin. *Earth-Science Reviews* 198: 102903.

Montes, C., C. A. Silva, G. A. Bayona, R. Villamil, E. Stiles, A. F. Rodriguez-Corcho, A. Beltran-Triviño, et al. 2021. A Middle to Late Miocene Trans-Andean Portal: Geologic Record in the Tatacoa Desert. *Frontiers in Earth Science* 8: 587022.

Mora, A., P. Baby, M. Roddaz, M. Parra, S. Brusset, W. Hermoza, and N. Espurt. 2010. Tectonic history of the Andes and sub-Andean zones: implications for the development of the Amazon drainage basin. Amazonia, Landscape and Species Evolution, 38–60. Wiley, Oxford.

Muñoz-Torres, F. A., R. C. Whatley, and D. van Harten. 2006. Miocene ostracod (Crustacea) biostratigraphy of the upper Amazon Basin and evolution of the genus Cyprideis. *Journal of South American Earth Sciences* 21: 75–86.

Nesom, G. L., and J. T. Kartesz. 2000. Observations on the *Ludwigia uruguayensis* complex (Onagraceae) in the Unites States. 65: 123–125.

591 Okada, M., B. J. Grewell, and M. Jasieniuk. 2009. Clonal spread of invasive Ludwigia hexapetala and L.
592 grandiflora in freshwater wetlands of California. *Aquatic Botany* 91: 123–129.

593 Paradis, E., S. R. Baillie, W. J. Sutherland, and R. D. Gregory. 1998. Patterns of natal and breeding
594 dispersal in birds. *Journal of Animal Ecology* 67: 518–536.

595 Paradis, E., and K. Schliep. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary
596 analyses in R R. Schwartz [ed.],. *Bioinformatics* 35: 526–528.

597 Peng, C.-I. 1989. The Systematics and Evolution of Ludwigia Sect. Microcarpium (Onagraceae). *Annals*
598 *of the Missouri Botanical Garden* 76: 221.

599 Peng, C.-I., C. L. Schmidt, P. C. Hoch, and P. H. Raven. 2005. Systematics and evolution of Ludwigia
600 section Dantia (Onagraceae). 92: 307–359.

601 Pennington, R. T., M. Lavin, D. E. Prado, C. A. Pendry, S. K. Pell, and C. A. Butterworth. 2004.
602 Historical climate change and speciation: neotropical seasonally dry forest plants show patterns of
603 both Tertiary and Quaternary diversification Y. Malhi, and O. L. Phillips [eds.],. *Philosophical*
604 *Transactions of the Royal Society of London. Series B: Biological Sciences* 359: 515–538.

605 Pérez-Escobar, O. A., G. Chomicki, F. L. Condamine, A. P. Karremans, D. Bogarín, N. J. Matzke, D.
606 Silvestro, and A. Antonelli. 2017. Recent origin and rapid speciation of Neotropical orchids in the
607 world's richest plant biodiversity hotspot. *New Phytologist* 215: 891–905.

608 Pybus, O. G., and P. H. Harvey. 2000. Testing macro–evolutionary models using incomplete molecular
609 phylogenies. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 267:
610 2267–2272.

611 Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. Posterior Summarization in
612 Bayesian Phylogenetics Using Tracer 1.7 E. Susko [ed.],. *Systematic Biology* 67: 901–904.

613 Raven, P. H., and W. Tai. 1979. Observations of Chromosomes in Ludwigia (Onagraceae). *Annals of the*
614 *Missouri Botanical Garden* 66: 862.

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

617 Ree, R. H., and I. Sanmartín. 2018. Conceptual and statistical problems with the DEC+J model of
618 founder-event speciation and its comparison with DEC via model selection. *Journal of*
619 *Biogeography* 45: 741–749.

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

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

624 Ribera, I., T. G. Barraclough, and A. P. Vogler. 2008. The effect of habitat type on speciation rates and
625 range movements in aquatic beetles: inferences from species-level phylogenies. *Molecular*
626 *Ecology* 10: 721–735.

627 Richardson, J. E., S. Madriñán, M. C. Gómez-Gutiérrez, E. Valderrama, J. Luna, K. Banda-R., J. Serrano,
628 et al. 2018. Using dated molecular phylogenies to help reconstruct geological, climatic, and
629 biological history: Examples from Colombia I. Somerville [ed.],. *Geological Journal* 53: 2935–
630 2943.

631 Ruaux, B., S. Greulich, J. Haury, and J.-P. Berton. 2009. Sexual reproduction of two alien invasive
632 Ludwigia (Onagraceae) on the middle Loire River, France. *Aquatic Botany* 90: 143–148.

633 Ruskin, B. G., F. M. Dávila, G. D. Hoke, T. E. Jordan, R. A. Astini, and R. Alonso. 2011. Stable isotope
634 composition of middle Miocene carbonates of the Frontal Cordillera and Sierras Pampeanas: Did
635 the Paranaense seaway flood western and central Argentina? *Palaeogeography,*
636 *Palaeoclimatology, Palaeoecology* 308: 293–303.

637 Salas-Gismondi, R., J. J. Flynn, P. Baby, J. V. Tejada-Lara, F. P. Wesselingh, and P.-O. Antoine. 2015. A
638 Miocene hyperdiverse crocodylian community reveals peculiar trophic dynamics in proto-
639 Amazonian mega-wetlands. *Proceedings of the Royal Society B: Biological Sciences* 282:
640 20142490.

641 Skaer Thomason, M. J., C. D. McCort, M. D. Netherland, and B. J. Grewell. 2018. Temporal and
642 nonlinear dispersal patterns of Ludwigia hexapetala in a regulated river. *Wetlands Ecology and*
643 *Management* 26: 751–762.

644 Smith, B. T., J. E. McCormack, A. M. Cuervo, Michael. J. Hickerson, A. Aleixo, C. D. Cadena, J. Pérez-
645 Emán, et al. 2014. The drivers of tropical speciation. *Nature* 515: 406–409.

646 Song, Z.-C., W.-M. Wang, and F. Huang. 2004. Fossil pollen records of extant angiosperms in China. 70:
647 425–458.

648 Soons, M. B. 2006. Wind dispersal in freshwater wetlands: knowledge for conservation and restoration. 9:
649 271–278.

650 Soons, M. B., G. A. Groot, M. T. Cuesta Ramirez, R. G. A. Fraaije, J. T. A. Verhoeven, and M. Jager.
651 2017. Directed dispersal by an abiotic vector: wetland plants disperse their seeds selectively to
652 suitable sites along the hydrological gradient via water J. Alahuhta [ed.],. *Functional Ecology* 31:
653 499–508.

654 Sytsma, K. J., A. Litt, M. L. Zjhra, J. Chris Pires, M. Nepokroeff, E. Conti, J. Walker, and P. G. Wilson.
655 2004. Clades, Clocks, and Continents: Historical and Biogeographical Analysis of Myrtaceae,
656 Vochysiaceae, and Relatives in the Southern Hemisphere. *International Journal of Plant Sciences*
657 165: S85–S105.

658 Tagliacollo, V. A., F. F. Roxo, S. M. Duke-Sylvester, C. Oliveira, and J. S. Albert. 2015. Biogeographical
659 signature of river capture events in Amazonian lowlands. *Journal of Biogeography* 42: 2349–
660 2362.

661 Testo, W. L., E. Sessa, and D. S. Barrington. 2019. The rise of the Andes promoted rapid diversification
662 in Neotropical *Phlegmariurus* (Lycopodiaceae). *New Phytologist* 222: 604–613.

663 Vasconcelos, T. N. C., S. Alcantara, C. O. Andrino, F. Forest, M. Reginato, M. F. Simon, and J. R. Pirani.
664 2020. Fast diversification through a mosaic of evolutionary histories characterizes the endemic

flora of ancient Neotropical mountains. *Proceedings of the Royal Society B: Biological Sciences* 287: 20192933.

Wagner, W., P. C. Hoch, and P. H. Raven. 2007. Revised classification of the Onagraceae. 83: 1–240.

Waters, J. M., L. H. Dijkstra, and G. P. Wallis. 2000. Biogeography of a southern hemisphere freshwater fish: how important is marine dispersal? *Molecular Ecology* 9: 1815–1821.

Wesselingh, F. P. 2006. Miocene long-lived lake Pebas as a stage of mollusc radiations, with implications for landscape evolution in western Amazonia. 133: 1–17.

Wesselingh, F. P., A. Ranzi, and M. E. Räsänen. 2006. Miocene freshwater Mollusca from western Brazilian Amazonia. 133: 419–437.

Wesselingh, F. P., M. E. Räsänen, G. Irion, H. B. Vonhof, R. Kaandorp, W. Renema, L. Romero Pittman, and M. Gingras. 2001. Lake Pebas: a palaeoecological reconstruction of a Miocene, long-lived lake complex in western Amazonia. 1: 35–81.

Zachos, J. 2001. Trends, Rhythms, and Aberrations in Global Climate 65 Ma to Present. *Science* 292: 686–693.

Zardini, E. M., C. Peng, and P. C. Hoch. 1991. Chromosome numbers in *Ludwigia* sect. *Oligospermum* and sect. *Oocarpon (Onagraceae)*. *TAXON* 40: 221–230.

683 Table 1. *Ludwigia* species included in this study and NCBI accession numbers for the taxa included in this study.

Species	*ITS*	*rpl16*	*rps16*	*trnG*	*trnL*
L. abyssinica	KX168306	KX154596	KX154371	KX154280	KX168264, KU941992
L. adscendens	KX168307	KX154597	KX154372	KX154281	KX168274, KU941993
L. affinis	KP026969				
L. alternifolia	KX168309	KX154600	KX154375	KX154284	KX168220, KX154261
L. bonariensis	KX168311	KX154602	KX154377	KX154286	KX168241, KX154263
L. decurrens	KX168317	KX154607	KX154382	KX154291	KX168278, KX154267
L. elegans	KX168318	KX154608	KX154383	KX154292	KX168283, KX154269
L. erecta	KX168320	KX154610	KX154385	KX154294	KX168250, KX154271
L. grandiflora	KX168323	KX154613	KX154388	KX154297	KX168284
L. hassleriana	KX168325	KX154615	KX154390	KX154299	KX168287, KX154276
L. helminthorrhiza	KP026983				
L. hexapetala		KX154616		KX154300	KX168215, KX168143
L. hyssopifolia	KX168328	KX154618	KX154393	KX154302	KX168238, KX168145
L. inclinata	KP026973		KX154457	KX168211	KX168308, KX168210
L. irwinii	KX168331	KX154621	KX154396	KX154305	KX168294, KX168148
L. jussiaeoides	KX168332	KX154622	KX154397	KX154306	KX168239, KX168149
L. lagunae	KX168333	KX154624	KX154399	KX154308	KX168259, KX168151
L. leptocarpa	KX168336	KX154627	KX154402	KX154311	KX168282, KX168154
L. linearis	KX168341	KX154632	KX154407	KX154316	KX168225, KX168159
L. major	KX168343	KX154634	KX154409	KX154318	KX168279, KX168161
L. martii	KX168345	KX154636	KX154411	KX154320	KX168293, KX168163
L. myrtifolia	KX168348	KX154640	KX154415	KX154324	KX168295, KX168167
L. neograndiflora	KX168349	KX154641	KX154416	KX154325	KX168285, KX168168
L. nervosa	KX168351	KX154643	KX154418	KX154327	KX168286, KX168170
L. octovalvis	KX168356	KX154648	KX154423	KX154332	KX168258, KX168175
L. ovalis	KX168359	KX154651	KX154426	KX154335	KX168216, KX168178
L. peploides	KX168363	KX154655	KX154430	KX154339	KX168218, KX168182
L. perennis	KX168364	KX154656	KX154431	KX154340	KX168296, KX168183
L. peruviana	KX168366	KX154659	KX154434	KX154343	KX168271, KX168186
L. pseudonarcisssus	KX168369	KX154662	KX154437	KX154346	KX168291, KX168189
L. quadrangularis	KX168371	KX154664	KX154439	KX154348	KX168191
L. ravenii	HE585708	KX154665	KX154440	KX154349	KX168240, KX168192
L. rigida	KP026971	KX154668	KX154443	KX154365	KX168300, KX168195
L. sedioides	KX168375	KX154669	KX154444	KX154352	KX168217, KX168196
L. sericea	KX168377	KX154671	KX154446	KX154354	KX168281, KX168198
L. stenorraphe	KX168381	KX154675	KX154450	KX154358	KX168254, KX168202
L. stolonifera	KX168383	KX154677	KX154452	KX154360	KX168265, KX168204
L. suffruticosa	KX168384	KX154678	KX154453	KX154361	KX168229, KX168205
L. torulosa	KP026984	KX154682	KX154458	KX154366	KX168301, KX168209
Chamerion angustifolium	JF976295	NC_052848.1	AY267389	NC_052848.1	AY264505
Chamerion latifolium	L28023				DQ860545
Circaea alpina	AY357769	AY357851			AY264498
Circaea lutetiana	GQ232526	GQ232649			GQ232710
Epilobium ciliatum	L28015				AY264508
Epilobium cylindricum	JF976300				
Fuchsia boliviana	GQ232536	GQ232660			GQ232722
Fuchsia cyrtandroides	AY357779	AY357861			AY905460
Fuchsia excorticata	AY357781	AY357863			GQ232724
Fuchsia procumbens	GQ232543	AY357882			GQ232729
Lythrum salicaria	AY035750	NC_042891.1	NC_042891.1	NC_042891.1	NC_042891.1
Trapa natans	MH712715	NC_042895.1	NC_042895.1	NC_042895.1	MT316134

684

FIGURE LEGENDS

Figure 1. Examples of hydrophytic and helophytic *Ludwigia* species. a) *L. nervosa* and b) *L. rigida* rooted-emergent helophytes inhabiting flooded savannas and shallow ponds. c) *L. sedioides* and d) *L. inclinata* are rooted-floating and rooted-submerged hydrophytes that live in ponds and lakes.

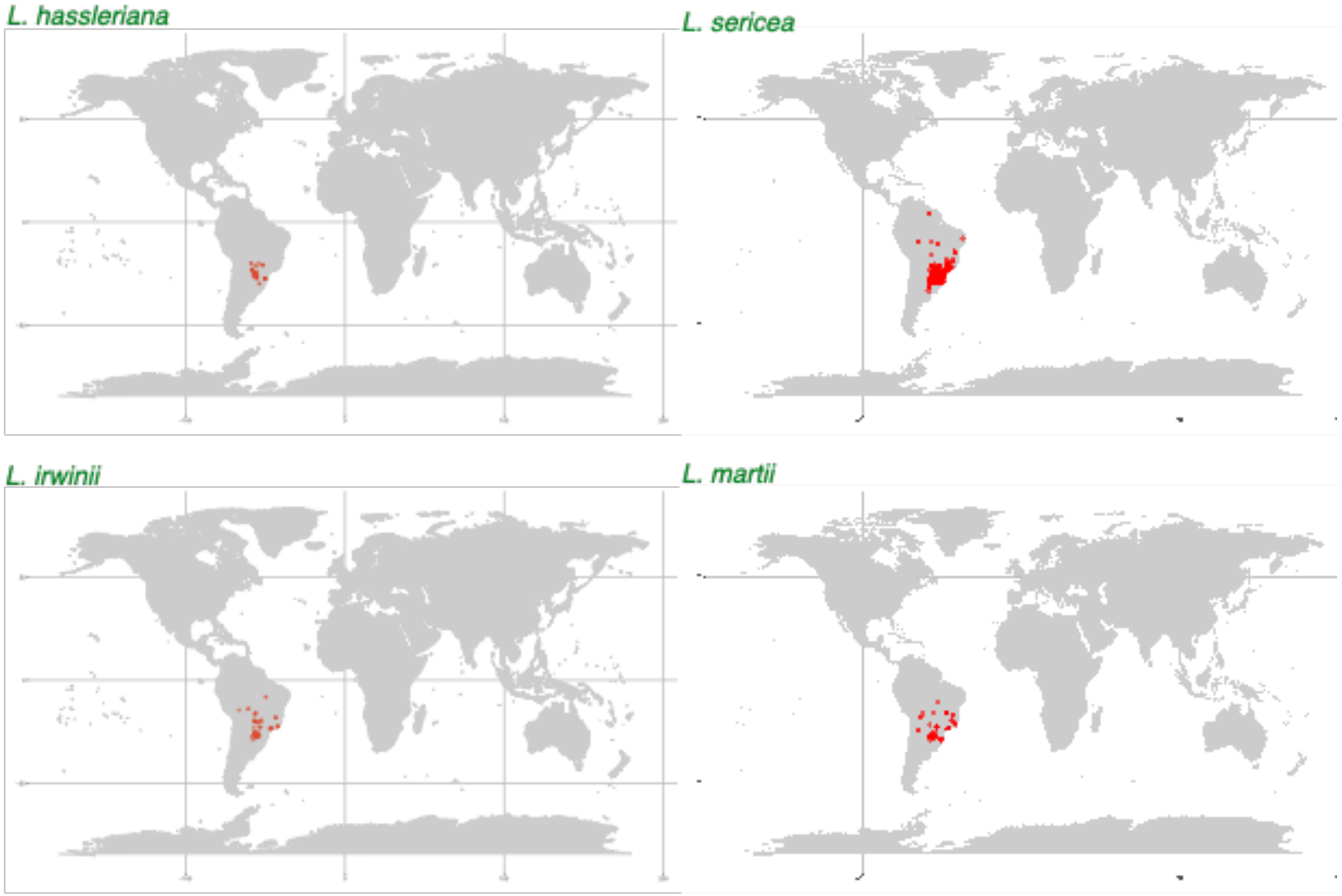
Figure 2. Time-calibrated tree of *Ludwigia* and lineage-through-time plots (LTT). Blue vertical lines delimit the approximate time frame when the Pebas lake was in place (23–10 Ma). Posterior probabilities >0.7 are shown above branches. Brown vertical lines indicate the approximate times when two major recent pulses of Andean uplift took place in northern South America (12–6 Ma and 4.5 Ma). Orange: North American clade. Green: predominantly Central and South American clade. Underlined text: *Ludwigia* species whose range currently include northern South America. *Ludwigia* species endemic to Africa are indicated with asterisks (*). LTT: semi-logarithmic plot where log-transformed values for number of lineages through time are shown (bold black line). Results from simulations under a pure-birth model are in grey. Prediction under a pure-birth process is shown with a red-dashed line.

Figure 3. Results from the biogeographic analysis with BioGeoBEARS. Pie charts on nodes indicate the relative probabilities of ancestral ranges. Ranges reconstructed at ancestral nodes are shown. The hypothesized time frame when the Pebas system existed (23–10 Ma) and the recent pulses of Andean uplift took place (12–6 Ma and ~4.5 Ma) are indicated with dashed lines in blue and brown, respectively. The current distribution of species is shown at the tips. Bold: *Ludwigia* species whose ranges currently include northern South America.

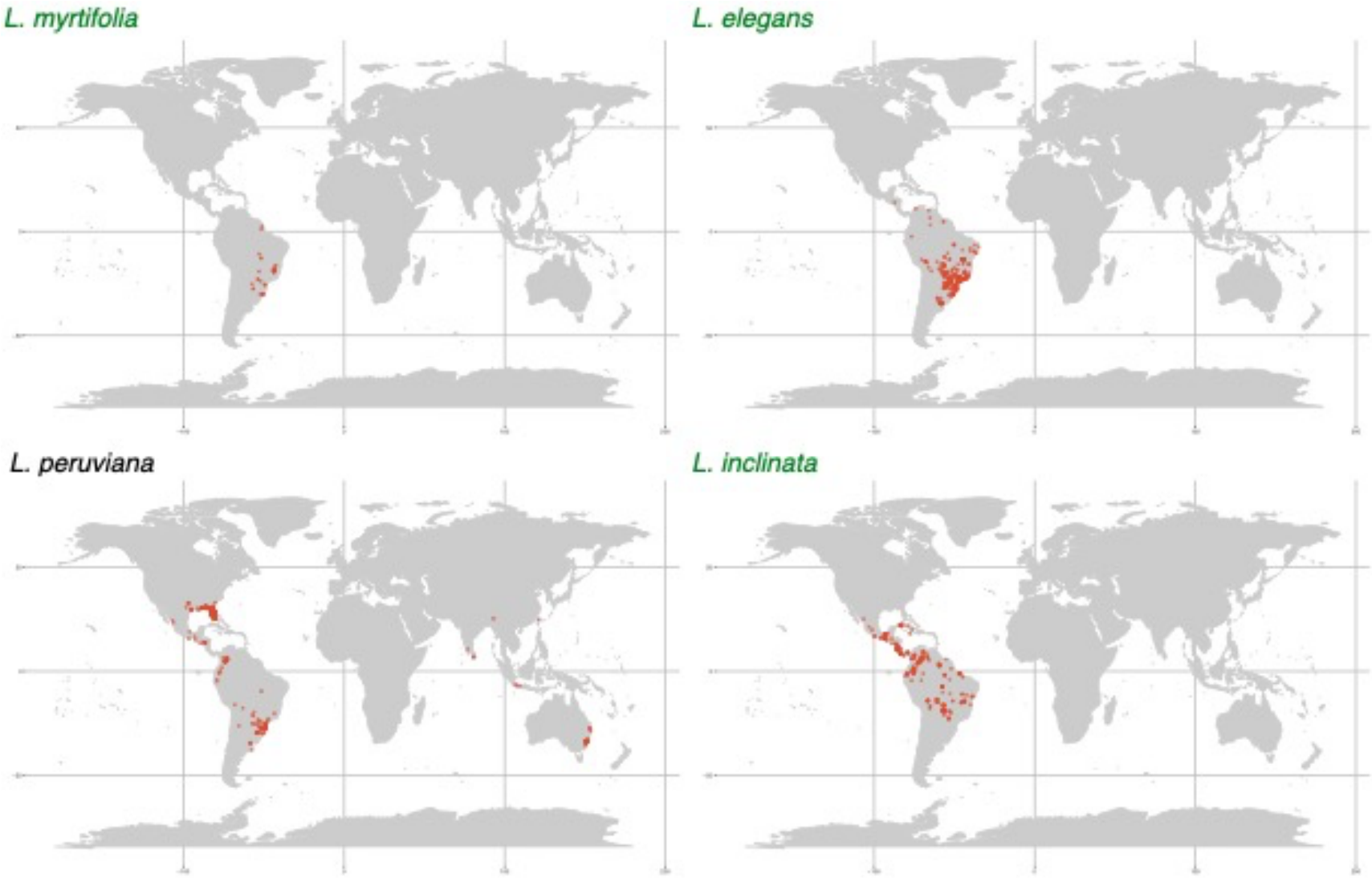
Appendix S1. Currently accepted *Ludwigia* species. Species included in this study (Table 1) are indicated with (*). Distribution in the five geographic areas defined is shown; South America (S), North America (N), Eurasia (A), Africa (F), Oceania (O)

Species	Distribution
*L. abyssinica**	F
*L. adscendens**	A/F/O
*L. affinis**	S/F
L. africana	F
L. alata	N
L. albiflora	S
*L. alternifolia**	N
L. anastomosans	S
L. arcuata	N
*L. bonariensis**	S/N
L. brachyphylla	S
L. brenanii	F
L. brevipes	N
L. bullata	S
L. burchellii	S
L. caparosa	S
L. curtisii	N
*L. decurrens**	S/N/A/F
L. densiflora	S
L. dodecandra	S
*L. elegans**	S
L. epilobioides	A
*L. erecta**	S/N/F
L. filiformis	S
L. foliobracteolata	S
L. glandulosa	N
*L. grandiflora**	S/N
*L. hassleriana**	S
*L. helminthorrhiza**	S
*L. hexapetala**	S/N/A
L. hirtella	N
L. hookeri	S
*L. hyssopifolia**	S/N/A/F/O
*L. inclinata**	S
*L. irwinii**	S
*L. jussiaeoides**	F
*L. lagunae**	S
L. lanceolata	N
L. laurotteana	S
L. latifolia	S
*L. leptocarpa**	S/N/F
*L. linearis**	N
L. linifolia	N
L. longifolia	S
*L. major**	S
L. maritima	N
*L. martii**	S
L. mexiae	S
L. microcarpa	N
L. multinervia	S
*L. myrtifolia**	S
*L. neograndiflora**	S
*L. nervosa**	S
*L. octovalvis**	S/N/A/F/O
*L. ovalis**	A
L. palustris	S/N/A/F/O
L. peduncularis	S
*L. peploides**	S/N/A/O
*L. perennis**	A/F/O
*L. peruviana**	S/N/A/O
L. pilosa	N

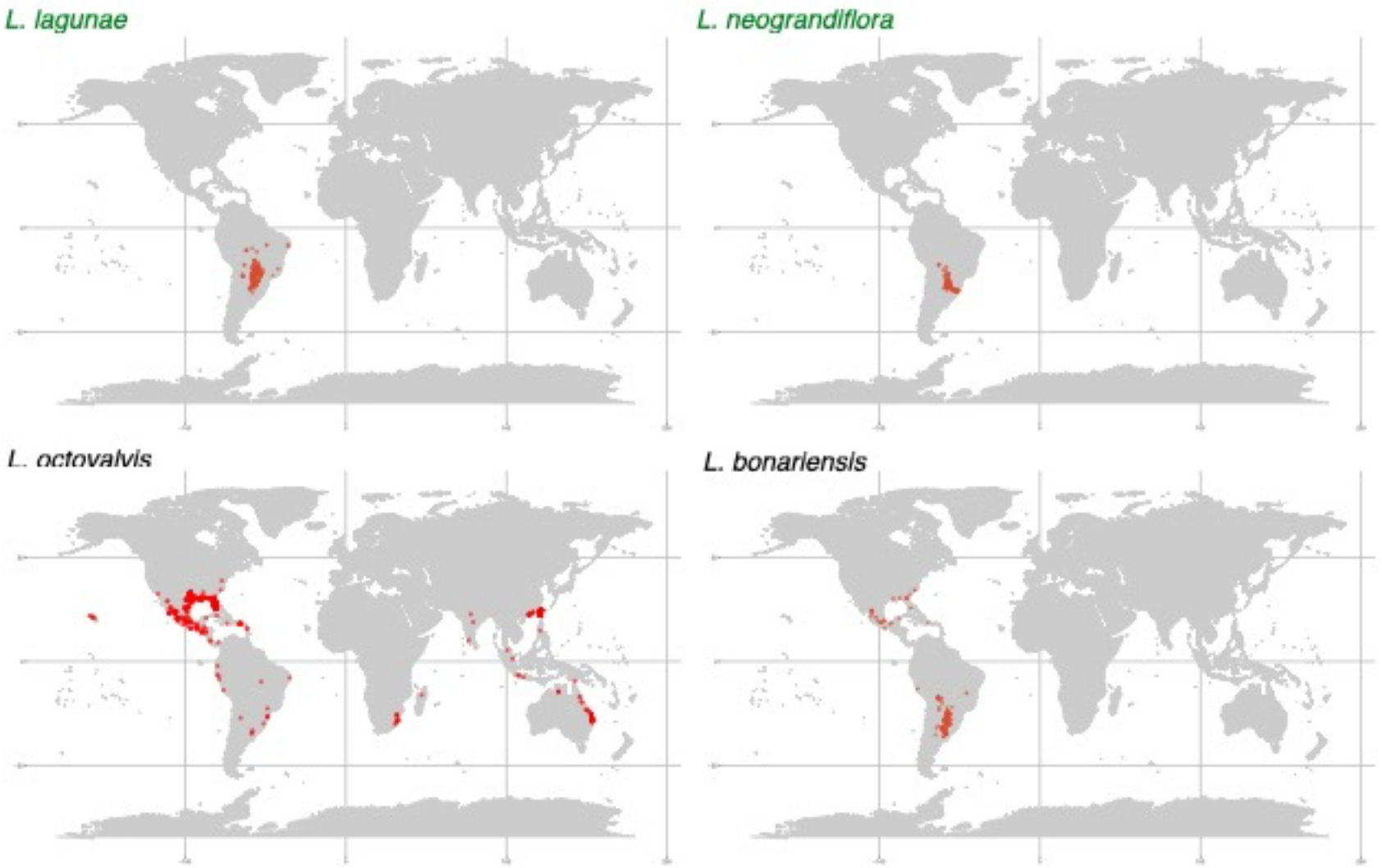
L. polycarpa	N
L. prostrata	A
*L. pseudonarcisssus**	S
*L. quadrangularis**	S
*L. ravenii**	N
L. repens	N
*L. rigida**	S
*L. sedioides**	S
L. senegalensis	F
*L. sericea**	S
L. simpsonii	N
L. spathulata	N
L. speciosa	F
L. sphaerocapra	N
*L. stenorraphe**	F
*L. stolonifera**	F
L. stricta	N
*L. suffruticosa**	N
L. tomentosa	S
*L. torulosa**	S
L. virgata	N



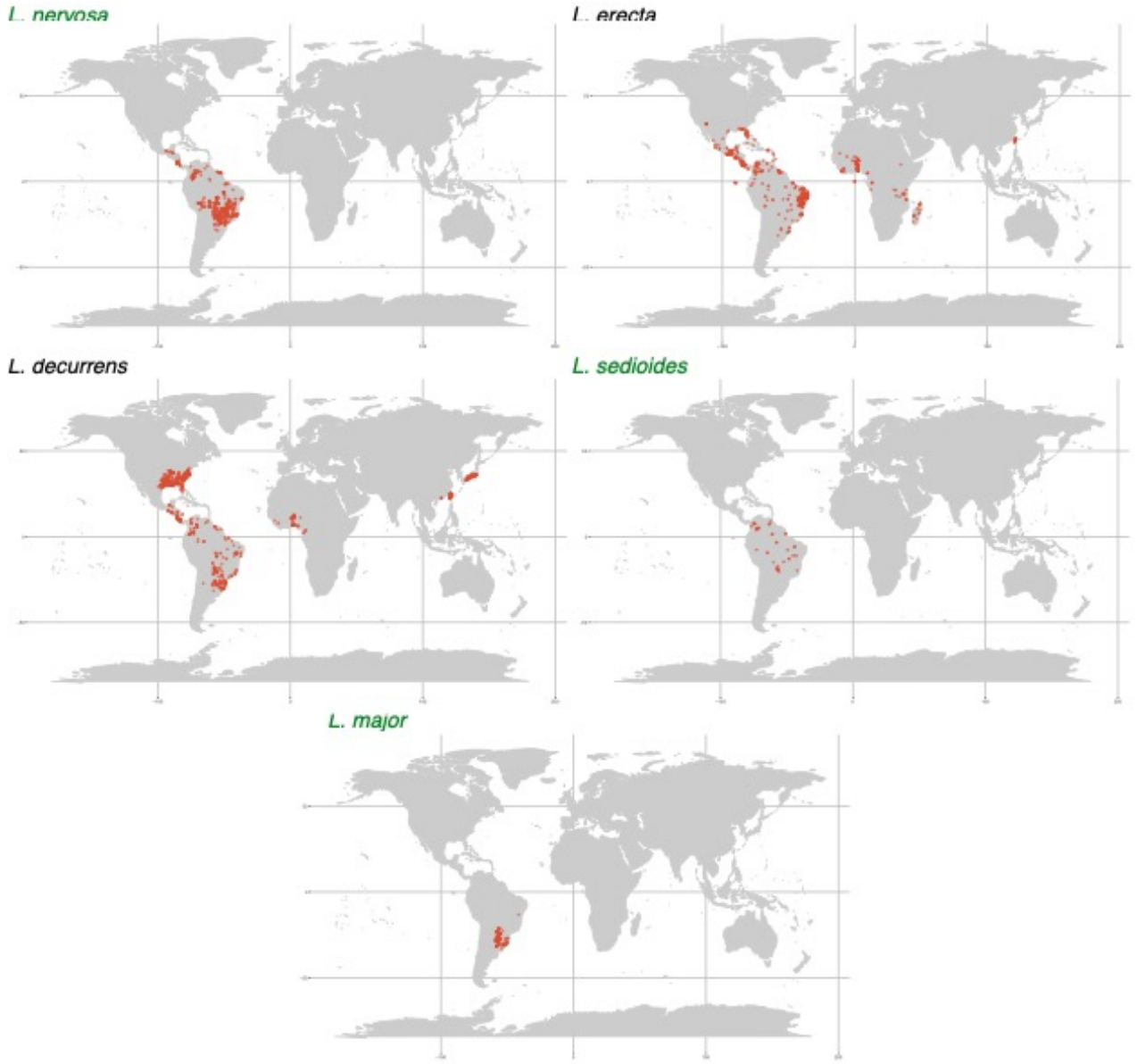
Appendix S2. Distribution of *Ludwigia* species. Species color indicates the region they are endemic to. Green: South/Central America. Black: Broad distribution.



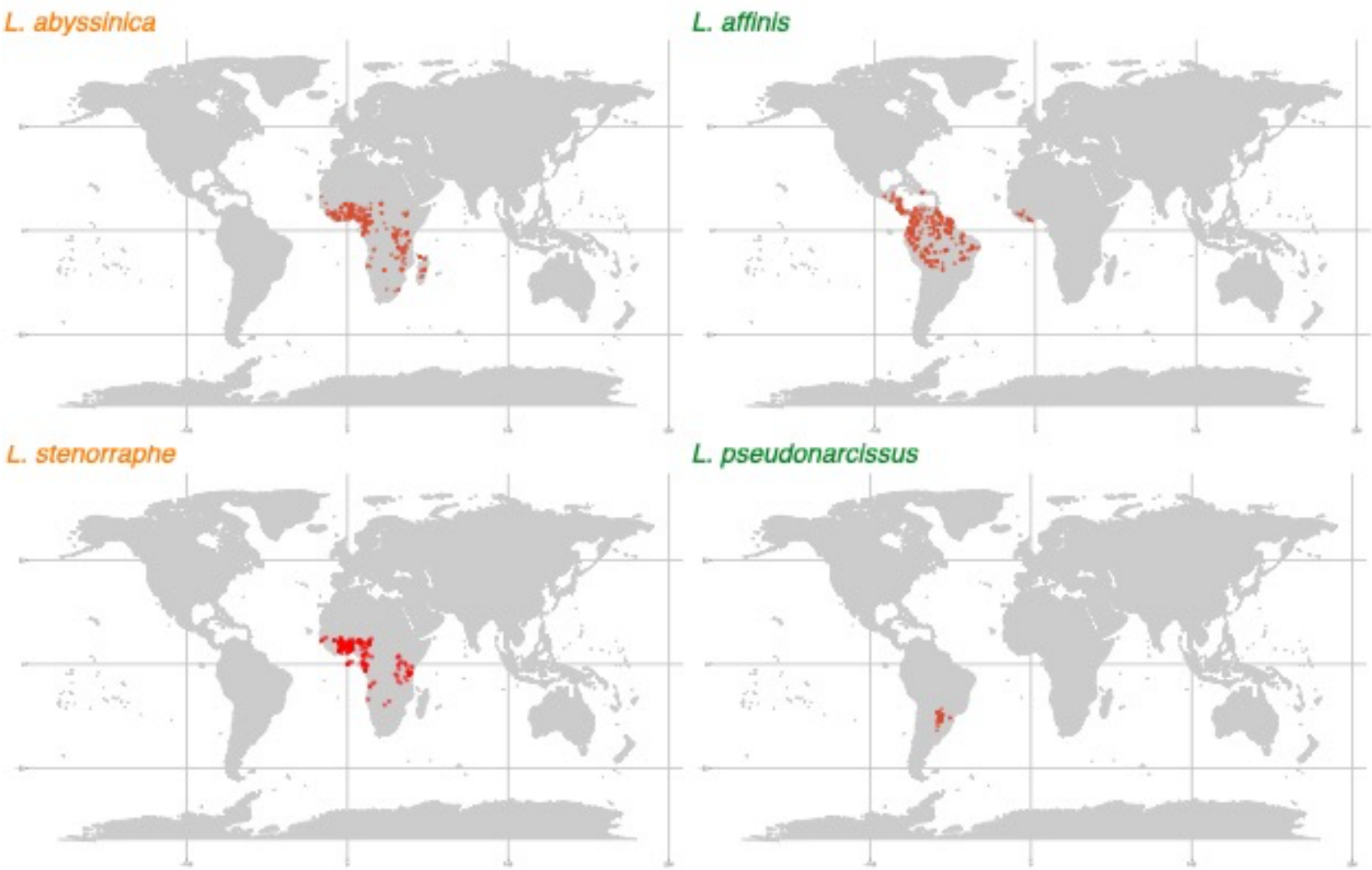
Appendix S3. Distribution of *Ludwigia* species. Species color indicates the region they are endemic to. Green: South/Central America. Black: Broad distribution.



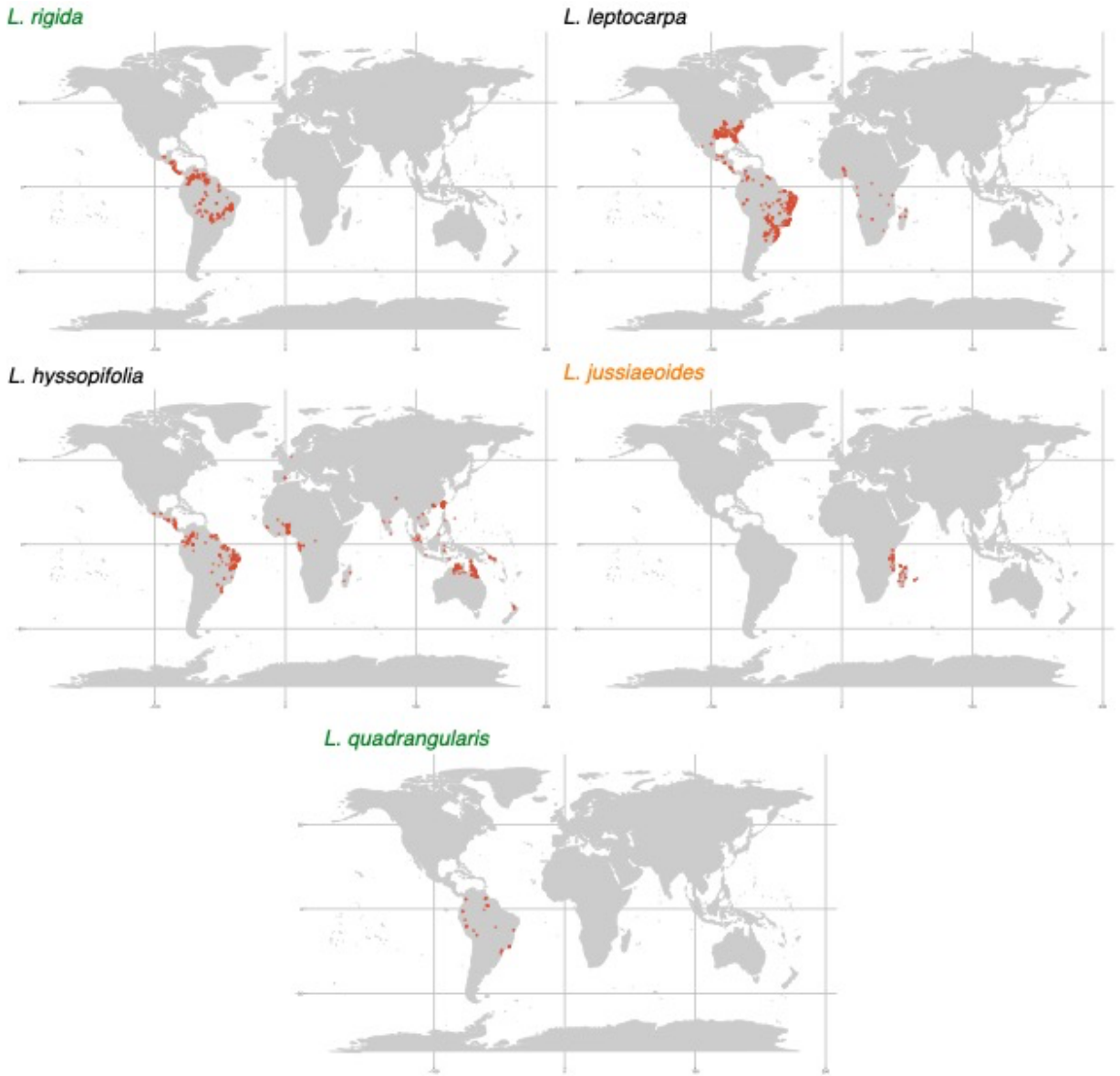
Appendix S4. Distribution of *Ludwigia* species. Species color indicates the region they are endemic to. Green: South/Central America. Black: Broad distribution.



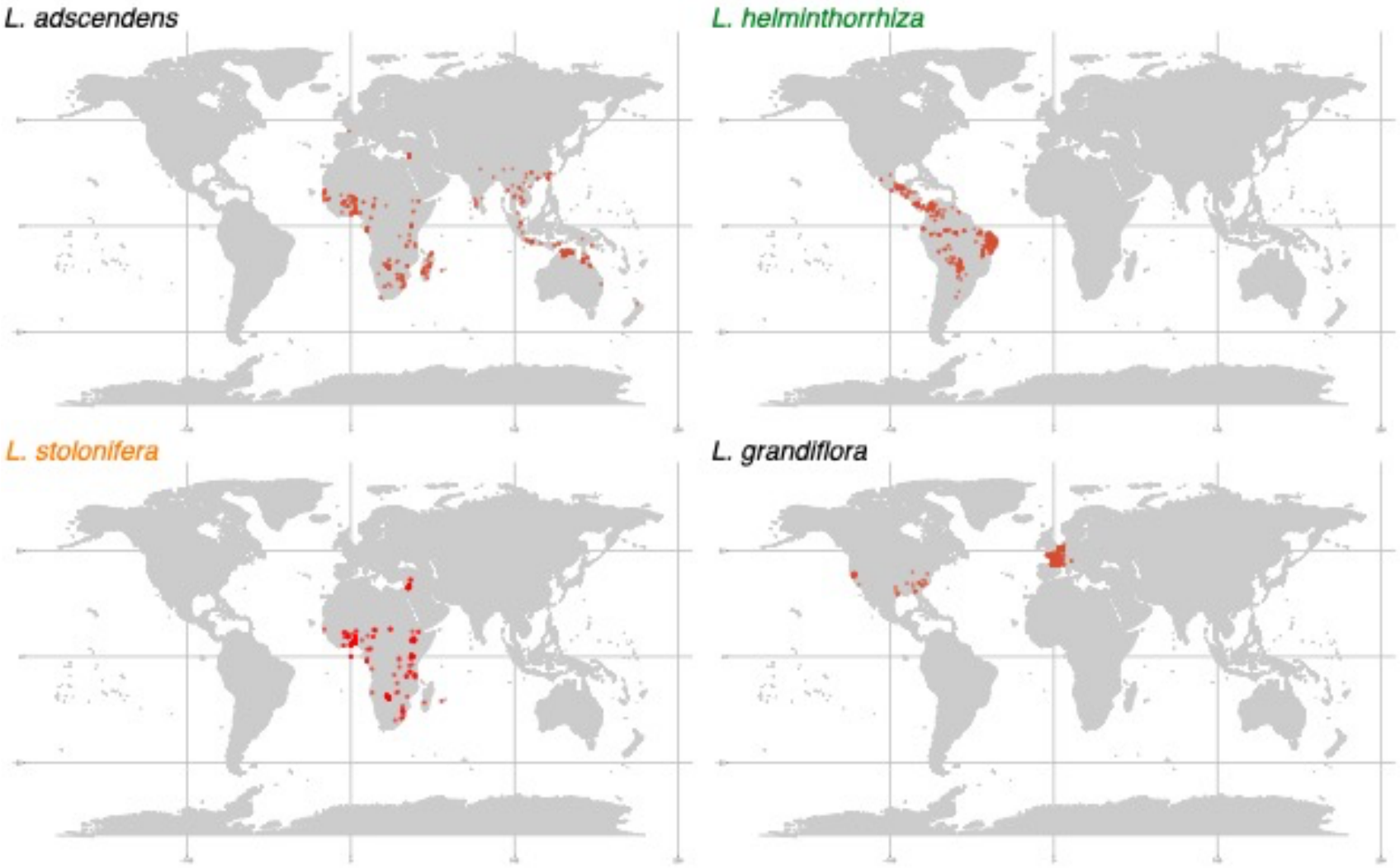
Appendix S5. Distribution of *Ludwigia* species. Species color indicates the region they are endemic to. Green: South/Central America. Black: Broad distribution.



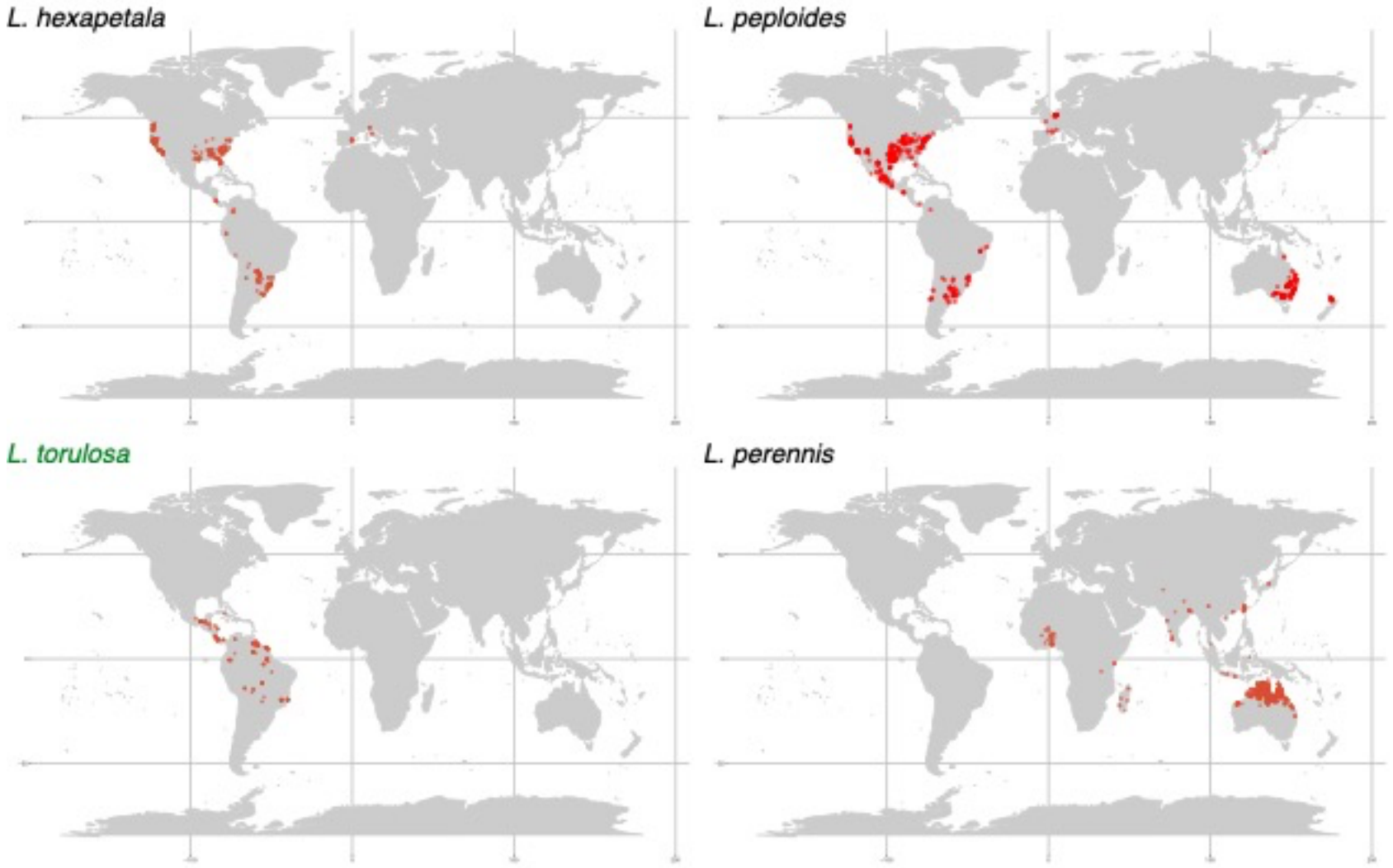
Appendix S6. Distribution of *Ludwigia* species. Species color indicates the region they are endemic to. Green: South/Central America. Orange: Africa. Black: Broad distribution.



Appendix S7. Distribution of *Ludwigia* species. Species color indicates the region they are endemic to. Green: South/Central America. Orange: Africa. Black: Broad distribution.

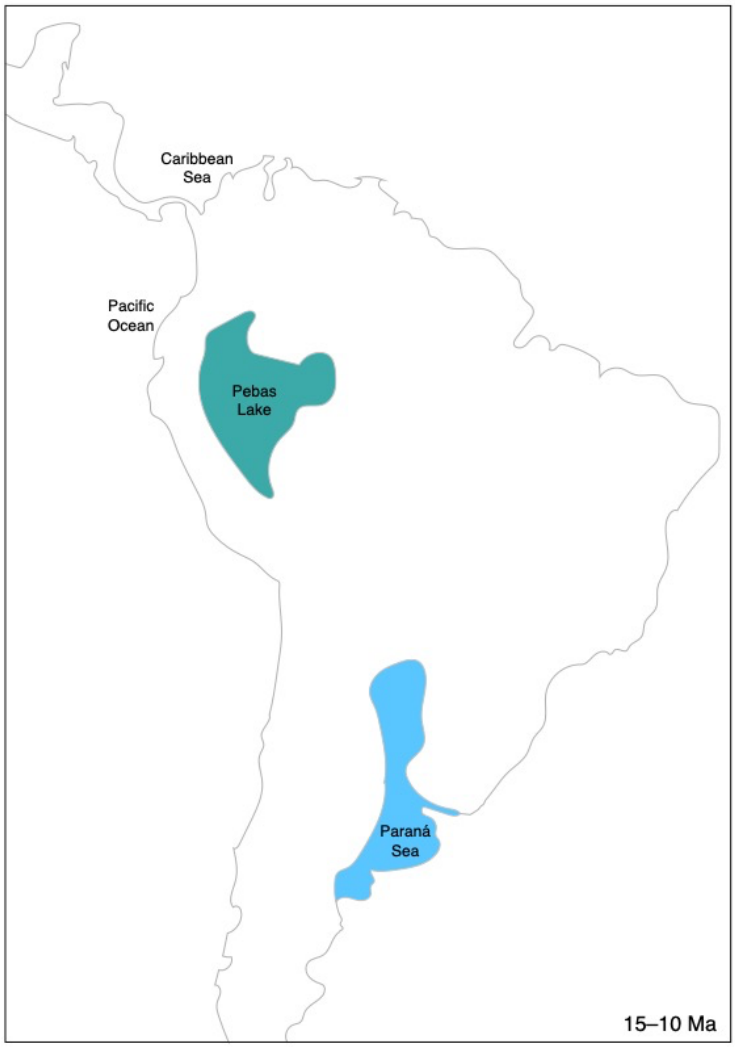


Appendix S8. Distribution of *Ludwigia* species. Species color indicates the region they are endemic to. Green: South/Central America. Orange: Africa. Black: Broad distribution.



Appendix S9. Distribution of *Ludwigia* species. Species color indicates the region they are endemic to. Green: South/Central America.

Appendix S10. Distribution of *Ludwigia* species. Species color indicates the region they are endemic to. Yellow: North America. Red: Eurasia.



Appendix S11. Map of South America in the middle Miocene. Modified from Gross *et al*. 2016. The Pebas lake and the Paraná Sea were in place.