Biogeography, Diversification, and Domestication in the Coca Family (Erythroxylaceae)

BY

Dawson White

B. S., Colorado State University, 2011

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biological Sciences
In the Graduate College of the
University of Illinois at Chicago, 2019

Chicago, Illinois

Defense Committee:

Roberta J. Mason-Gamer, Chair and Advisor
Mary Ashley
Boris Igic
Richard Ree, Field Museum of Natural History
Nyree Zerega, Chicago Botanic Garden

ACKNOWLEDGEMENTS

I would like to thank my mother, Margi, and father, Bill, for the opportunities they have provided.

I would like to thank my wife, Emily, for making the sacrifices she has to allow me to pursue this research.

Thank you to Ramona Gaylord for sending me to the Amazon and Dr. John P. Janovec for his outstanding mentorship, friendship, and introduction to the world of tropical botany.

CONTRIBUTIONS OF AUTHORS

Dawson White is responsible for all of the content of this thesis. Chapter 1 has been reprinted with permission from the Botanical Society of America and includes contributions from other authors: Roberta J. Mason-Gamer and Dawson White developed the project. Melissa B. Islam provided DNA samples and expertise on the study system. Dawson White collected data, performed analyses, and wrote the manuscript.

TABLE OF CONTENTS

CHAPTER		PAGE
I	Phylogenetic inference in section *Archerythroxylum* informs taxonomy, biogeography, and the domestication of coca (*Erythroxylum* species)……	1
	A. Introduction……	1
	B. Materials & Methods……..	5
	C. Results……	10
	D. Discussion……	13
II	Multiple origins of the coca plant in South America……	21
	A. Introduction……	21
	B. Materials & Methods……..	23
	C. Results & Discussion.……	26
III	The Pantropical diversification of the Coca family highlights boreotropical migration and biome shifts out of rainforests……	32
	A. Introduction……	32
	B. Materials & Methods……..	36
	C. Results……	42
	D. Discussion……	53
IV	A new *Erythroxylum* (Erythroxylaceae) variety from the Sierra Escalera of Peru..	63
	A. Introduction……	63
	B. Materials & Methods……..	65
	C. Taxonomic treatment……..….	67
	D. Discussion……	68
	Cited Literature……	71
	Appendix A……..	81
	Appendix B……..	85
	Appendix C……..	108
	Appendix D……..	124
	Permission to reprint Chapter I……	125
	Permission to reprint biome map in Figure 7……..	130
	Vita……	136

LIST OF TABLES

TABLE		PAGE
I.	Sequencing and target assembly results for Chapter 1……..……	9
II.	Erythroxylaceae species sampled in Chapter 3 phylogenetic and biogeographic analyses compared to the total species diversity……..	44
III.	Neotropical biome shifts……...	48
IV.	Paleotropical biome shifts…….	48
V.	Mean pairwise distance (MPD) and mean nearest taxon distance (MNTD) statistics for phylogenetic biome composition..……	48
VI.	Morphological comparison of *E. schunkei, E. ulei, and E. ulei* var. *escalerense*….	65

LIST OF FIGURES

FIGURE		PAGE
1.	Histograms describing the nuclear gene data set for Chapter 1	9
2.	ML-concat phylogram of 68 Erythroxylaceae taxa	11
3.	Three hypotheses of coca domestication	22
4.	Astral II lineage tree and genetic cluster assignment (for 5 and 9 groups) of four coca varieties and two wild relatives	28
5.	ABC tests of coca domestication scenarios	30
6.	Sequencing and target assembly results for Chapter 3	43
7.	The diversification of the Erythroxylaceae in time and space	45
8.	Biome states and shifts in Neotropical and Paleotropical Erythroxylaceae	47
9.	Identification of long-branch taxa and effects of correction pipeline	51
10.	BAMM diversification rate configurations for MLcatLB tree and cohort diversification rate heatmap	52
11.	Photos of habitat and organs of *E. ulei* var.*escalerense*	64
12.	Habit of *E. ulei* var. *escalerense*	66
13.	ASTRAL III lineage tree of *E. ulei* var. *escalerense* clade	68

List of Abbreviations

bp	base-pairs
Ma	Mega-annum

Summary

The field of biological systematics is concerned with understanding how the diversification of lineages through time has shaped our observable biodiversity. This study generated extensive genomic resources and data to explore the diversity and diversification of the Coca family of plants; thus filling in the gap in this branch of the tree of life and using our robust hypothesis of species relationships as a blueprint to understand how the clade has evolved.

The Coca family (Erythroxylaceae) is most infamously known as the natural source of cocaine, isolated from four South American taxa called coca, but it also comprises 283 more species of trees and shrubs distributed in tropical habitats throughout the world. This study has two major foci: First, a population-level analysis of diversity and ancestry of the cultivated cocas and their closest relatives has elucidated the history of domestication of this crop. Coca has been created from the wild species, *Erythroxylum gracilipes,* two or three times, supporting a paradigm that different Holocene peoples were able to breed the same natural resource into domestication to serve their needs; in this case, a mild workaday stimulant and medicine. The second focus is on the patterns of biogeography and diversification as the Coca family evolved and migrated around the tropical regions of the world. The study finds that the Coca family originated in Africa in the late Cretaceous before migrating out of Africa and into the Indo-Pacific region as well as the Americas. The timing of this migration appears to have occurred ~50 Ma and might have been facilitated by a north Atlantic land bridge in combination with the warm climates during the early Eocene climatic optimum. This macroevolutionary chronology also details the ecological evolution of the clade by quantifying the distributions of species in tropical dry forest, rainforest, and savanna/grassland biomes and characterizing the frequencies and directions of biome transitions associated with speciation.

Finally, the thesis includes the description of a new variety of *Erythroxylum* collected by the Field Museum of Natural History's Rapid Inventory team from the Sierra Escalera of Peru.

Chapter I:

Phylogenetic inference in section *Archerythroxylum* informs taxonomy, biogeography, and the domestication of coca (*Erythroxylum* species)

This chapter is a reprint (with reformatting) from an original article published in the American Journal of Botany. The citation is as follows: White, D. M., M. B. Islam, and R. J. Mason-Gamer. 2019. Phylogenetic inference in section *Archerythroxylum* informs taxonomy, biogeography, and the domestication of coca (*Erythroxylum* species). *American Journal of Botany* 106(1): 1–12.

Introduction:

A robust phylogenetic hypothesis of species relationships is a fundamental tool in systematic biology and many other investigations within the fields of ecology and evolution. It is applied here to inform the botanical science of the Coca family, Erythroxylaceae: a pantropical family of ca. 285 species of small trees and shrubs, the majority of which (ca. 272 species) are classified in the genus *Erythroxylum* P. Browne. The remaining Erythroxylaceae species belong to three genera found exclusively in Africa and Madagascar: *Aneulophus* Benth. (1 species), *Pinacopodium* Exell & Mendonça (2 species), and *Nectaropetalum* Engl. (8 species) (Schulz, 1907; Hegnauer, 1981; Rury, 1981; Plowman and Hensold, 2004; pers. obs.). *Erythroxylum* is an ecologically diverse and abundant clade; it is the 22nd most speciose genus in Amazonia (82 species; ter Steege et al., 2016), yet most of the diversity is found in the seasonally dry forests of eastern Brazil and the Venezuelan Guayana (Plowman and Berry, 1999; Daly, 2004). There are also ca. 80 *Erythroxylum* species distributed throughout the African and Indo-Pacific tropical regions, with a concentration in Madagascar (Daly, 2004). The small actinomorphic flowers of all species are primarily pollinated by various Hymenoptera and Diptera, and the small red to purple drupes are readily dispersed by birds (Rury, 1982; Gryj and Domínguez, 1996; Oviedo, 2002). Nearly all Erythroxylaceae species are distylous and several have been utilized in investigations into mating system evolution (Darwin, 1877; Avila-Sakar and Domínguez, 2000; Abarca et al., 2008). Yet the Coca family is most infamously known as the natural source of cocaine, which is commercially isolated from two cultivated South American species called coca (*Erythroxylum coca* Lam. and *E. novogranatense* (D. Morris) Hieron.).

Since its description in 1756 (Browne, 1756), one systematic treatment of the Erythroxylaceae was completed using herbarium material, and is the primary systematic framework for understanding *Erythroxylum* morphological patterns and geographic distribution (Schulz, 1907; keys to sections updated in Schulz, 1931). Schulz described 19 morphological sections within *Erythroxylum* based primarily on calyx, stipule, and style characters. The largest section is *Archerythroxylum* O.E. Schulz (Schulz, 1907; ca. 70 species), which comprises Central and South American and Caribbean species with non-striate stipules and cataphylls, perfect flowers with free styles, and calyces with valvate aestivation and generally triangular lobes (Loiola, 2001). However, several morphological synapomorphies have been revealed to be more variable and subsequent researchers have cast doubt on the monophyly of these sections (Schultes et al., 1976; Rury, 1981; Plowman and Rivier, 1983). Preliminary phylogenetic analyses have also supported this later view (Emche et al., 2011; Islam, 2011). Thus, the first purpose of this analysis is to test the monophyly of Schulz' sections with the goal of eventually improving upon this intrageneric classification scheme.

We sample from nine of Schulz' sections but focus primarily on species in section *Archerythroxylum* for two reasons: First, because it is a large section representing a wide range of ecological, geographic, and morphological variation, possibly containing several large paraphyletic clades whose relationships are key to testing the validity of this and other taxonomic sections (Schultes et al., 1976). Instead of Schulz' classification scheme, we believe that phytogeographic regions could be an important factor shaping *Erythroxylum* diversification, given the difficulty of transitioning between ecological biomes (Crisp et al., 2009), and be featured more prominently in future intrageneric systematic revisions. As such, we analyze the biogeographic patterns of *Archerythroxylum* species for a preliminary observation of how diversification has been geographically structured.

The second reason for focusing on *Archerythroxylum* is because it contains the cultivated cocas (*E. coca* and *E. novogranatense*) and their most morphologically and geographically proximate relatives (Rury, 1982). Thorough sampling of the species belonging to *Archerythroxylum* is essential to the second purpose of this analysis: identifying the closest relatives of the coca taxa.

With macrofossils dating to 8,000 years BP, coca is one of the oldest crop plants in the Americas, and the leaves are still chewed by over 5 million South Americans a nutritive supplement, medicine, and mild stimulant (Plowman, 1984b; Dillehay et al., 2010; Conzelman and White, 2016). Coca is also the natural source of cocaine, first isolated from the leaves by Niemann in 1860 and later praised as one of the most significant contributions to Western medicine coming directly from the Neotropical flora (Mortimer, 1901; Schultes, 1979). The botanical diversity of this tropical crop was not fully described until 1979, even though hypotheses of its wild ancestors and geographic origins began with some of the earliest South American botanical explorers (Schultes et al., 1976; Plowman, 1979). From the colonial period until modern times, the four cultivated varieties of coca have been grown in separate geographic areas in South America (Plowman, 1984b). Huánuco coca (*Erythroxylum coca*) is the most widely cultivated variety; it is found on the eastern slopes of the Andes in Peru and Bolivia in the moist, montane belt known as the *montaña* (Peru) or *Yungas* (Bolivia). Amazonian coca (*E. coca* var. *ipadu* Plowman) is grown in discrete localities throughout the Amazon basin by native tribes, especially near the Rio Napo, Rio Negro, and Madre de Dios tributaries. Until restrictions in the past thirty years (Perl, 1992) began limiting cultivation, Colombian Coca (*E. novogranatense*) was grown in the drier montane valleys of the Sierras in Colombia, it is the only cultigen that is partially self-compatible (Bohm et al., 1982). Lastly, Trujillo coca (*E. novogranatense* var. *truxillense* (Rusby) Plowman) is grown in the dry valleys of northwestern Peru, as well as a disjunct location around the Colombia/Ecuador border (Plowman, 1986).

Confounding our understanding of its ancestry, arguably none of the coca varieties grow in the wild (see Schultes, 1979), and the other ~200 Neotropical *Erythroxylum* congener species are all relatively similar in sexual and vegetative morphology (Schulz, 1907). Five wild species have been seriously considered as ancestors of coca by Neotropical botanists: Poeppig believed coca arose from *E. hondense* HBK from the Colombian Andes (fide Hartwich, 1911), Kunth hypothesized it was the Peruvian *E. mamacoca* Mart. (fide Hartwich, 1911), but MacBride noted he was probably referring to *E. gracilipes* Peyr. (Macbride, 1949). *Erythroxylum cataractarum* Spruce ex Peyr. is frequently used in place of coca and called by the Barasana Indians of the Colombian Amazon “the coca of our fathers”

(Schultes, 1981). This is possibly the species responsible for Gutierrez-Noriega and von Hagen's (1950) hypothesis that coca arose in the Amazon basin (Schultes et al., 1976). Lastly, H. H. Rusby (1933) believed *E. anguifugum* Mart. of southern Brazil to be the ancestor, adding yet another potential region of domestication. However, after circumscription of our current coca taxonomy, Timothy Plowman and colleagues posited a linear evolutionary series hypothesis in which Huánuco coca (*E. coca*) once occurred in the wild and was domesticated on the eastern slopes of the Andes in Peru or Bolivia. This single domestication subsequently gave rise to Trujillo coca, which gave rise to Colombian coca as it was transported further north, and independently, Huánuco coca produced Amazonian coca (Bohm et al., 1982).

As bottlenecks and genetic drift are expected to reduce character diversity during the domestication process, progenitor-derivative relationships can be supported when the suspected derivative taxon exhibits a subset of characters present in the progenitor taxon or population (Gross and Olsen, 2010). Chemotaxonomic studies, however, have found distinct secondary metabolite profiles between the two varieties of *E. coca* and the varieties of *E. novogranatense*, suggesting they shared a common ancestor instead of progenitor-derivative relationship (but Amazonian coca and Colombian coca could still be derived from their respective conspecifics; Johnson et al., 1998, 2002). Phylogenetic and cluster analyses using AFLP's also supported the two species being sister lineages (Johnson et al., 2005; Emche et al., 2011). The only other putative ancestor in the AFLP analysis, *E. hondense,* was not immediately similar to the cocas. Our phylogenetic analysis permits testing of Plowman's linear evolutionary series hypothesis (Bohm et al., 1982) as well as Johnson and Emche's sister species hypothesis (Johnson et al., 2005; Emche et al., 2011) by sampling from all the cultivated coca taxa as well as thorough sampling of possible wild progenitors.

Coca is one of the most potent symbols of the clash of ethnic culture and globalization. In the Andes and Amazon, coca is believed to be a sacred gift from mother earth – an aliment, medicine, and practical stimulant; yet as the source of cocaine, it is an internationally acknowledged poison and the nucleus of a war on drugs (Conzelman and White, 2016). Given the dynamic relationship of this plant and

humans, it is paramount to understand coca as a medicine, a narcotic, and a plant. Establishing a robust hypothesis of phylogenetic relationships and the origins of this plant are essential to its botanical and cultural appreciation.

In this analysis we sequenced 547 nuclear genes in 68 Erythroxylaceae taxa and inferred their relationships using concatenation and gene-tree summary methods. We present the first gene-based phylogeny of the majority of section *Archerythroxylum* as well as representatives from eight other sections to test the current taxonomy in a phylogenetic framework, understand biogeographic patterns, and to identify the relationships and closest relatives of the coca taxa.

Materials & Methods:

***Exon-capture probe design*—**We generated a scaffold-level draft genome for *E. coca* by sequencing silica-dried leaf material from a live plant at the USDA Agricultural Research Service (accession B145; contact USDA ARS for references). Two libraries were prepared, the first a paired-end WGS library (Illumina Truseq DNA PCR-Free LT Library Prep. kit) sequenced on the Illumina HiSeq 2500 platform at the University of Wisconsin-Madison. The second was a mate-pair library prepared using the Nextera Mate Pair Library Prep Kit with insert size 8-10 kbp and was sequenced on the Illumina MiSeq platform, 2 x 300 bp, at the Field Museum of Natural History, Chicago, IL. The genome was assembled *de novo* by Ray (Boisvert et al., 2012) with an optimum kmer length of 41 and an expected genome size of ca. 600 Mbp as determined from flow cytometry following Arumuganathan and Earle (1991). The chloroplast genome was assembled by Velvet with a kmer size of 91 and an expected coverage of 150 (Zerbino and Birney, 2008).

Our probe design for target-capture DNA sequencing utilized the bioinformatic pipeline of Weitemier et al. (2014). Briefly, we first excluded all of the genomic scaffolds that mapped to our *E. coca* chloroplast genome. Of the remaining scaffolds, we retained those with ≥99% identity to whole single transcripts from the oneKP *E. coca* transcriptome (Matasci et al., 2014; Wickett et al., 2014). We then removed genes and, subsequently, exons with ≥90% similarity to prevent cross-enrichment of similar

genes and exons across loci. Lastly, we set a minimum concatenated exon length per locus of 960 bp, mapped the target loci back to the genome assembly, and removed genes with zero or one intron, as well as those with any large introns >600 bp in length. Our final target sequences for probe design consisted of 2,551 single-copy exons that were at least 10% divergent from one another, representing 547 nuclear genes. Target sequences were submitted to Arbor Biosciences (Ann Arbor, MI, USA) for custom 120-bp RNA probe synthesis at 3x tiling density, which yielded 16,988 probes.

***DNA extraction, target capture, and sequencing*—**Our sampling effort focused on representation of nine taxonomic sections of Neotropical *Erythroxylum* fide O.E. Schulz (1907), with an emphasis on the ca. 70 species in the section *Archerythroxylum*, to which the coca taxa belong. In addition to *Archerythroxylum,* we sampled five species from the Paleotropical section *Coleocarpus* O.E. Schulz, four species from section *Erythroxylum* Loiola, four species from section *Rhabdophyllum* O.E. Schulz, and a single species from sections *Macrocalyx* O.E. Schulz*, Megalophyllum* O.E. Schulz*, Microphyllum* O.E. Schulz*, Schistophyllum* O.E. Schulz (Paleotropical), and *Venelia* O.E. Schulz (Paleotropical). We selected as outgroup taxa the Paleotropical Erythroxylaceae species *Nectaropetalum zuluense* (Schönl.) Corbishley*,* and *Pinacopodium congolense* (S.Moore) Exell & Mendonça. The dataset included 66 *Erythroxylum* taxa and ten of these have multiple individuals sampled. Sixty-three of the 81 individuals sequenced were herbarium specimens, 14 were silica-dried leaves collected in the field, and four were coca leaves (*E. coca* and *E. novogranatense*) from the USDA ARS living collection in Beltsville, MD, USA (see Appendix A sample information). Coca leaves were received, and DNA extraction performed at the Field Museum under DEA Controlled Substances Registration Certificate PF0108707. Genomic DNA was extracted using a CTAB protocol with 3% PVP and 2% 2-mercaptoethanol in the extraction buffer (Doyle and Doyle, 1990). Discolored samples were cleaned with the MOBIO Laboratories Inc. DNA Clean-Up Kit.

For library preparation, genomic DNA samples were normalized to one µg in 52.5 µL molecular biology grade water and fragmented to 400 bp (PiP 50 W, duty factor 10%, 200 cycles per burst, 70 seconds, 20ºC) with the Covaris M220 Focused-ultrasonicator in 130 µL snap-cap tubes (Covaris Inc.,

Woburn, MA, USA). We then used 50 µL of fragmented DNA in the Illumina TruSeq HT Library Preparation kit (Illumina Inc., San Diego, CA, USA), following manufacturer's instructions, with the provided 96 dual-indexed adapters and 10-14 cycles of library amplification. Samples were combined into six equimolar pools each containing 12 samples and 500 ng total DNA. We tried to pool closely related species together based on morphological and geographic characters. Target capture was completed following the myBaits User Manual v3.01 with 21 hours of probe hybridization at 65ºC. Hybridized libraries were amplified following the myBaits protocol with KAPA HiFi polymerase for 12 cycles in the first round and 9-14 cycles in a second round in the effort to achieve a ~10 ng/ul final pool concentration. We analyzed the final fragment distributions for each pool on a Bioanalyzer (Agilent Genomics, Santa Clara, CA, USA) before combining the pools together at 5 nM concentration and sequencing at the University of Illinois at Chicago DNA Services Facility on the Illumina NextSeq platform (Mid-output, 2 x 150 bp).

Target sequence assembly, alignment, and phylogenetic reconstruction— There is no evidence of polyploidy in *Erythroxylum* as all nine taxa that have been cytologically analyzed (including the cultivated cocas) are diploid with 2n=24 (Mangenot and Mangenot, 1958, 1962; Löve, 1969, 1987; Plowman et al., 1978; Forni-Martins et al., 1995). Reads were filtered and demultiplexed with Trimmomatic (Bolger et al., 2014) and then merged reads were removed using PEAR (Zhang et al., 2014). The leading and trailing 10 bp of each raw read was hard-trimmed and then terminal bases were removed if below quality 20. We then used aTRAM to *de novo* assemble contigs from reads aligning to each target locus (exon sequences concatenated) for each sample (Allen et al., 2015). We chose to use this assembly program because it iterates the blasting of reads and de novo assembly multiple times to build contigs into "splash zones", in our case, introns. The output of aTRAM is a single consensus sequence per locus per individual with polymorphic sites coded as a single nucleotide based on read frequency. Additional steps to code heterozygous bases or phase alleles have not been proven to significantly improve phylogenetic inference (Kates et al., 2018), so we did not take these additional steps to analyze allelic variability.

The output for each gene was then aligned using the local pairwise option in MAFFT iterated up to 1000 times (Katoh et al., 2002). Short gene sequences (<500 bp) were removed and alignments were dropped from the analysis if they contained <62 sequences (at least 75% tip coverage). We made this taxon filter because, in addition to generating a species tree, we were interested in reconstructing phylogeny from a concatenated dataset with the hope there would be minimal discordance. Maximum-likelihood concatenation analyses perform poorly under conditions of nonrandom missing data in the presence of gene rate heterogeneity, further worsened with high levels of incomplete lineage sorting (Xi et al., 2016). While gene recovery using hybridization-based target-capture is a robust method at deep phylogenetic scales (Li et al., 2013), there is still an effect of divergence on the fidelity of RNA baits within hybridization pools (Johnson et al., 2016). While we tried to control for this during pooling, phylogenetic distance is still a non-random factor contributing to missing data.

For phylogenetic inference, we reconstructed a maximum-likelihood tree from 20 random starting trees under a GTR + GAMMA model of nucleotide substitution (Tavaré, 1986; Yang, 1993) for each gene alignment with RAxML v.8 (Stamatakis, 2014). Gene alignments were then viewed in Geneious v.10 (http://www.geneious.com; Kearse et al. 2012) and removed if we identified evidence of paralogous gene sequences, identified by one or more inconsistent "comb-like" clades separated by a significantly long branch from the rest of the tree. The species tree for the nuclear dataset was inferred from these 427 ML gene trees using ASTRAL-II (Mirarab et al., 2014). We then generated another "ML-concat" species tree by concatenating the 427 gene alignments in Geneious v.10, inferring the maximum-likelihood tree using the same methods as above, and generating 1000 bootstrap trees with the RAxML-HPC2 Tool on XSEDE via the CIPRES Science Gateway webportal (Miller et al., 2010; Stamatakis, 2014; www.phylo.org). Topological filtering of gene trees was completed in PAUP* 4.0a157 (Swofford, 2002) and the topological comparison (Appendix A: Figure A1) was constructed using the R package, "phytools" (Revell, 2012).

***Biogeographic dataset assembly*—**Our primary source for species' geographic data was from Plowman and Hensold (2004), which describes geographic distributions of specimens verified by

Erythroxylum expert Timothy Plowman. We also used our own database of Field Museum herbarium specimens and checked GBIF for any new occurrences, for which we found none that could be reliably verified (GBIF.org, 2017). To make Figure 2, shape files were downloaded from the WWF's Terrestrial Ecosystems of the World, and edited to show Neotropical phytogeographic regions following Fine et al. (2014; Atlantic Forest, Amazonia, Andes, Caribbean, Chocó, Cerrado, Caatinga, Ecuatorial Dry Forests, Guiana Shield, Mesoamerica, Orinoquia) using the R package, "maps" (Olson et al., 2001; R Development Core Team, 2013; Becker et al., 2017).

Dataset	value
# gene trees	427
mean gene length	2,803 bp
concatenated length	1.2 Mbp
total exon length	762 kbp
missing data	37%
# PI characters	180,878

Table I: Sequencing and target assembly results for Chapter I.

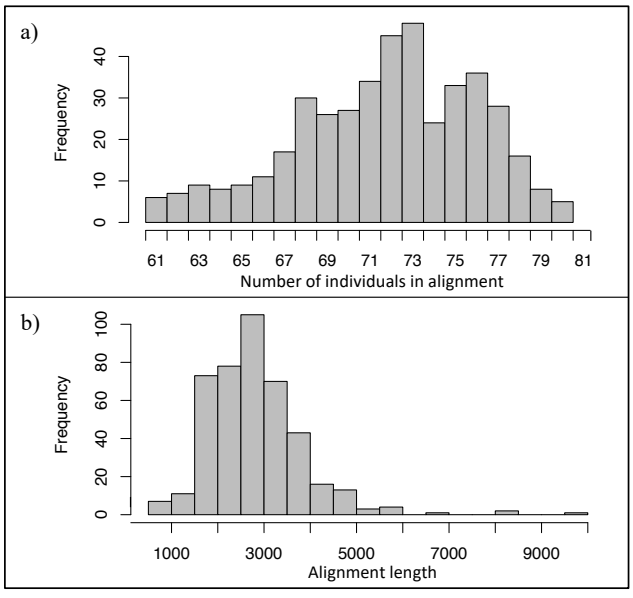


Figure 1: Histograms describing the nuclear gene data set. (A) Number of individuals in the 427 nuclear gene alignments. (B) Lengths of the 427 nuclear gene alignments (bp).

Results:

***Dataset* —** We dropped 29 of the 547 ML gene trees from the analysis that showed evidence of paralogous gene sequences. We dropped 91 additional trees with insufficient (≥75%) taxon coverage; taxon coverage of the final 427 trees is presented in Figure 1a. Thus, our final dataset had a concatenated length 1.2 Mbp and mean nucleotide coverage 80.76% (37% missing data total). Alignments of individual loci ranged from 705 bp to 9,845 bp with a mean of 2803 bp (Figure 1b). The length of the targeted exons in the 427 genes was 762 kbp so we generated ~438 kbp of intron and flanking sequence. A total of 180,878 characters were parsimony-informative (Table I).

***Phylogenetic relationships of* Archerythroxylum *species and relatives* —** The outgroups, *Nectaropetalum zuluense* and *Pinacopodium congolense* were determined from a prior ITS and chloroplast intergenic-spacer study that included samples from the sister family, Rhizophoraceae (Islam 2011; White, unpubl.). These taxa are sister to several clades of Paleotropical *Erythroxylum species* from Africa, Southeast Asia, and Australia. Neotropical *Erythroxylum* form a single monophyletic group sister to these Paleotropical taxa (Figure 2). For the ten species for which we sampled multiple individuals, current species delimitation is corroborated under a species-as-taxa concept (Baum, 2009; with the exception of paraphyletic *E. gracilipes,* but see discussion).

Our analysis showed that section *Archerythroxylum* is paraphyletic and contains at least seven other sections (Figure 2). The largest monophyletic group of *Archerythroxylum* is the clade containing the four cultivated coca taxa as well as ten other species. Every additional section from which we sampled multiple species is also paraphyletic (*Coleocarpus*) or polyphyletic (sect. *Erythroxylum* and *Rhabdophyllum*).

In light of the non-monophyly of the current taxonomic sections, we mapped biogeographic region characters on the phylogeny to evaluate the correlation between geography and phylogenetic structure. These *Erythroxylum* taxa form five monophyletic clades with distinct geographic affinities (Figure 2). Sister to these major clades at the base of the tree, the Caribbean/Mesoamerican species *E. rotundifolium* Lunan is the first sampled species to diverge from the Paleotropical lineages, followed by

the Brazilian Atlantic Forest species *E. splendidum* Plowman. In our concatenated ML analysis, Clades I-IV form a monophyletic group sister to clade IV, all with 100% bootstrap support, whereas in the ASTRAL species tree, clades III and IV are combined in a monophyletic group with *E. roraimae* Klotzsch ex O.E. Schulz (see Appendix A: Figure A2).

Clade I comprises 11 species, all of which occur in the Brazilian Atlantic Forest, Cerrado, and/or Caatinga (Figure 2), making this an entirely Brazilian clade to which many other *Erythroxylum* species from this region probably belong. The species *E. deciduum* A.St.-Hil. and *E. argentinum* O.E. Schulz extend their distributions to the southern Andes while *E. rufum* Cav. occurs in the Caribbean and Guiana

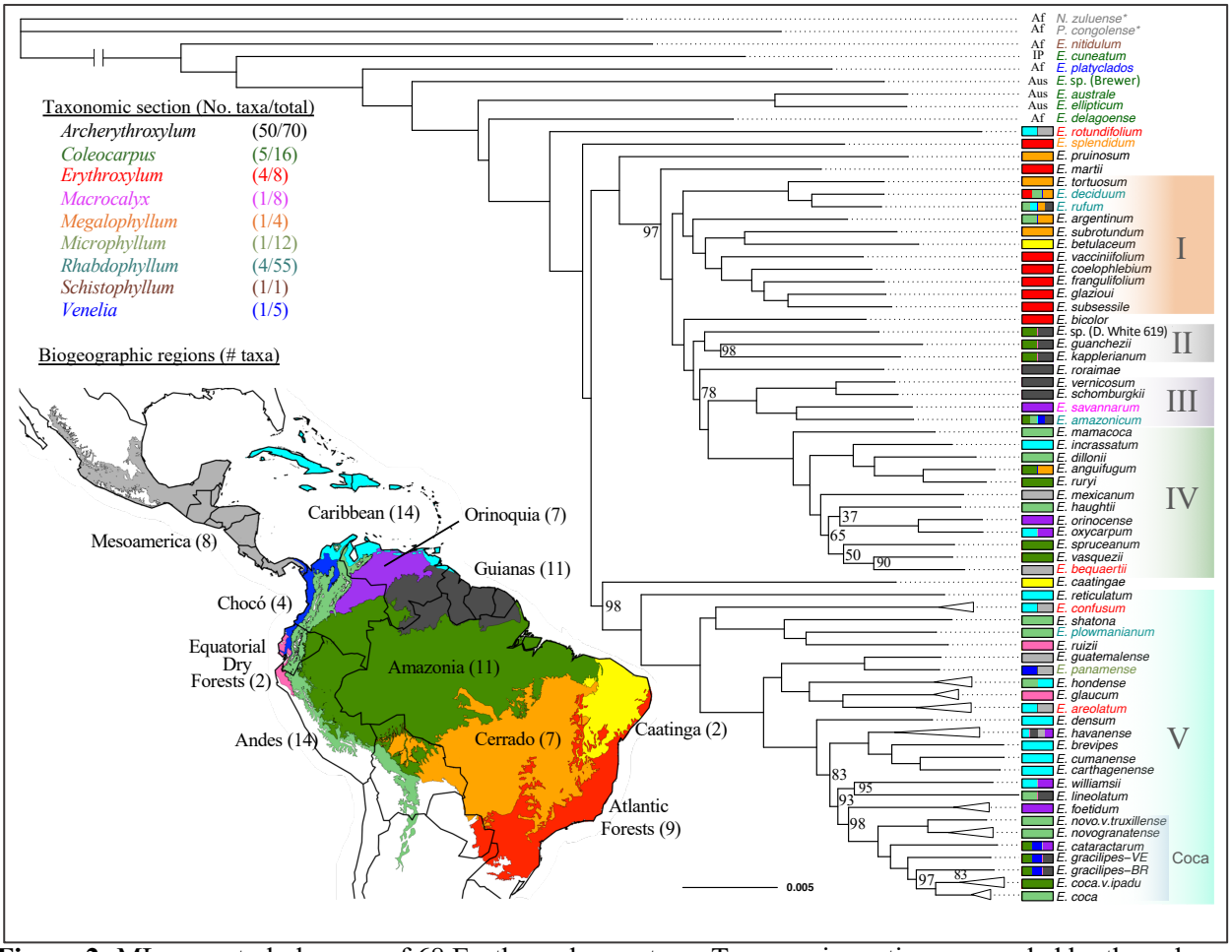


Figure 2: ML-concat phylogram of 68 Erythroxylaceae taxa. Taxonomic sections are coded by the color of the taxon name. Biogeographic regions occupied by each taxon are coded in bars next to taxon names. For Paleotropical species, Af = African, IP = Indo-Pacific, Aus = Australian. Scale bar indicates number of substitutions per site. All bootstrap values are 100% unless otherwise indicated.
regions, and has a disjunct population in the Cuzco department of Peru.

Clade II comprises just three species with distributions restricted to the western Guiana Shield and adjacent Amazonian region in Colombia, Venezuela, and northwestern-most Brazil. Clade III has a nearby distribution, with species from the western Guianas and Orinoquia, as well as *E. amazonicum* Peyr., which has a widespread distribution in northern South America and Amazonia.

Clade IV consists of species from the western and northern part of South America, as well as the Caribbean and Mesoamerica. Species are represented from the Peruvian (*E. mamacoca* Mart., *E. dillonii* Plowman ex Jara) and Colombian (*E. haughtii* W.A. Gentner) Andes, western Amazonia (*E. rury* Plowman*, E. spruceanum* Peyr.*, E. vasquezii* Plowman), and Orinoquia (*E. orinocense* Kunth*, E. oxycarpum* O.E. Schulz). *Erythroxylum oxycarpum* also occurs in the Caribbean along with *E. incrassatum* O.E. Schulz, and there are two Mesoamerican species represented (*E. bequaertii* Standl., *E. mexicanum* Kunth). This clade also contains *E. anguifugum,* which is primarily found in the Pantanal wetlands, and extends to other seasonally inundated areas in the Cerrado and southern Amazonia.

Finally, Clade V comprises 24 taxa occurring in the western and northern regions of the Neotropics. Within this group, there are subclades representing smaller distributions, such as the primarily Caribbean subclade consisting of *E. brevipes* DC.*, E. carthagenense* Jacq.*, E. cumanense* Kunth*,* and *E. havanense* Jacq., and the "Coca" subclade in the Andes/Amazon region. The "Coca" subclade consists of the wild species *E. cataractarum* and *E. gracilipes,* and the cultivated species *E. coca* and *E. novogranatense.*

Phylogenetic concordance among datasets — In addition to the results of the concatenation analysis, we inferred a species tree using the gene-tree reconciliation method implemented in ASTRAL-II. This species tree is almost identical to the ML-concatenation (ML-concat) result except for four single-taxon deviations occurring within clades II, IV, and V, as well as the inferred monophyly of clades II and III in the ASTRAL tree. The phylogenetic comparisons and the ASTRAL tree are shown in Appendix A (Figure A1). In clade II, the relationships are switched within the three-taxon clade with higher support in the ML-concat inference. In clade IV, ML-concat infers *E. bequaertii* to be sister to *E. vasquezii* whereas

ASTRAL infers it to be sister to *E. mexicanum* with improved node support. The differences found in clade V involve the cultivated cocas.

Relationships of cultivated cocas — The ML-concat and ASTRAL analyses both inferred that the cultivated coca species *E. coca* and *E. novogranatense* are polyphyletic. The *E. coca* cultivars are more closely related to *E. gracilipes* and *E. cataractarum* than to the cultivars of *E. novogranatense*; these four species form a clade with 100% bootstrap support (the "Coca" subclade; Figure 2; Appendix A: Figure A1). In order to assess the variation in this topology, we filtered the 427 gene trees to find the number of trees that supported monophyly of the cocas (*E. novogranatense* + *E. novogranatense* var. *truxillense* + *E. coca* + *E. coca* var. *ipadu*). Only seven of the gene trees inferred this topology.

There were two topological differences between the ML-concat and ASTRAL analyses in the "Coca" subclade. First, the two *E. coca* var. *ipadu* samples form a sister relationship in the ML-concat inference versus a paraphyletic relationship in the ASTRAL inference. Second, *E. cataractarum* switches from being sister to *E. gracilipes* and *E. coca* in the ML-concat inference to being sister to *E. novogranatense* in the ASTRAL inference. While there is 100% bootstrap support in the ML-concat tree, the local posterior probability in the ASTRAL tree is only 0.35 in support of this placement, once again suggesting there could be a more complicated history between the cultivated cocas and this wild species requiring more focused analysis. Forty-five of the gene trees inferred the ASTRAL topology, whereas 80 inferred the ML-concat topology. These results refute Plowman's (1979) hypothesis of an ancestor-descendent relationship of coca taxa and elaborate Johnson's (2005) hypothesis that *E. coca* and *E. novogranatense* as distinct evolutionary lineages (see Discussion). It also implicates *E. cataractarum* and *E. gracilipes* in the domestication process, as either progenitors or introgressors, and the possibility of distinct domestication events of *E. coca* and *E. novogranatense.*

Discussion:

***Systematics*—**The coca family, Erythroxylaceae, is a clade of species classified almost entirely as a single genus, *Erythroxylum* (ca. 272 species). A comprehensive intrageneric treatment of this genus was

completed over 100 years ago by O. E. von Schulz, in which he divided *Erythroxylum* into 19 sections primarily based on stipular and floral traits (Schulz, 1907). Since this time, there have been several investigations into the classification of the other Erythroxylaceae genera, *Nectaropetalum* and *Pinacopodium* (see Hegnauer 1981), and an analysis of family anatomy (Rury, 1981), but revisionary systematic treatment of *Erythroxylum* has only been approached regionally (Oviedo, 2002; Barrie and Plowman, 2018) or by taxonomic section (sect. *Rhabdophyllum*; Loiola 2001). This study takes the latter approach to establish the first phylogenomic hypothesis of species relationships in the family, with an emphasis on section *Archerythroxylum*. This section is the largest in the genus and is represented in all biogeographic and ecological regions occupied by the *Erythroxylum* clade (Schulz, 1907; Schultes et al., 1976). These factors make *Archerythroxylum* a logical choice to begin testing the monophyly of Schulz' sections and revising the intrageneric taxonomy of this large genus. Though *Archerythroxylum* is not monophyletic, we don't believe the several clades identified in our analysis will be significantly split apart by other sections as we continue to build the Erythroxylaceae tree.

Our primary application of this phylogeny was to test the monophyly of *Archerythroxylum* with morphologically and geographically proximate relatives from eight other sections, to identify the closest relatives of the cultivated coca taxa, and to analyze the biogeographic patterns of *Erythroxylum* in the Neotropics. We used two methods of species tree inference, maximum likelihood inference from a concatenated dataset and gene tree reconciliation by ASTRAL and found almost complete concordance between the two trees. There were small topological differences, as shown in Appendix A (Figure A1): for four single-taxon deviations occurring within clades II, IV, and V, as well as the inferred monophyly in the ASTRAL tree of clades II and III. Because these specific differences do not affect the intrageneric taxonomy or biogeographic results of this study, we will not rely on one over the other. However, the multispecies coalescent method implemented in ASTRAL is more likely to be accurate in situations with gene rate heterogeneity, high ILS, and nonrandom missing data (Edwards et al., 2016; Xi et al., 2016).

For the sections from which we sampled multiple species, our phylogeny confirms previous suspicions that they do not represent exclusive evolutionary lineages (Rury, 1981; Plowman and Rivier,

1983; Emche et al., 2011; Islam, 2011). Instead, *Archerythroxylum* forms at least five major clades, all but one of which (clade II, with three species) contain members of other sections. In particular, we sampled four out of the eight species in sect. *Erythroxylum,* which are distinguished from *Archerythroxylum* species only by their flower sexuality. Section *Erythroxylum* species occur in the Caribbean and Central America and are characterized by their subdioecious mating systems (functional dioecy and gynodioecy; Avila-Sakar and Domínguez 2000; Abarca et al. 2008). Given the ecological and morphological similarities between the sections, it has been suggested that sect. *Erythroxylum* is polyphyletic within *Archerythroxylum* (Rury, 1981; Islam, 2011). Our analysis supports this relationship and the conclusion that sect. *Erythroxylum* species represent multiple independent origins of dioecy from distylous *Archerythroxylum* ancestors (Avila-Sakar and Domínguez, 2000).

The second largest taxonomic section proposed by Schulz is *Rhabdophyllum* (55 species), from which we sampled four species. This section differs from *Archerythroxylum* by having striated stipules (sclerified vascular bundles) and, occasionally, partially fused styles. Stipule striation is also characteristic of section *Macrocalyx* (1/8 species sampled). Modern researchers have agreed with Schulz that this character is useful in distinguishing subclades of *Erythroxylum* (Loiola, 2001; Plowman and Hensold, 2004). While our analysis does not have sufficient sampling to analyze the taxonomic utility of stipule striations, our phylogeny shows that the five sampled species in *Rhabdophyllum* and *Macrocalyx* form three clades, so there has been at least that many evolutionary gains or losses of this trait. Thus, further sampling of *Rhabdophyllum* of will be an important step towards understanding intrageneric classification and macroevolution of Neotropical *Erythroxylum.*

Considering that Schulz's (1907) systematic treatment resulted from an exhaustive analysis of vegetative and sexual morphology and yet did not produce a stable intrageneric classification, Philip Rury (1981) completed a detailed analysis of the systematic anatomy of the Erythroxylaceae based on samples from 124 species, 45 of which were shared in this analysis. We did not collect any novel anatomical data, but a careful review of Rury's dataset revealed one character that corresponds to our molecular phylogenetic clades. Clades I-IV form a monophyletic group sister to clade V, and these two subdivisions

of the genus differ in their petiole anatomy. Both groups have an adaxially channeled arc of vascular tissue in their petioles, but in clades I-IV, this arc is invaginated and accompanied by dorsal vascular bundles or "plates" (Type 3), whereas this dorsal plate is absent in all species of clade V (Type 2, pg. 163-165). Rury's *E. rotundifolium* samples were heterogeneous, presenting the same vasculature as clades I-IV and also an elliptic trace or flattened arc of tissue (Type 1). Rury's Paleotropical *Erythroxylum*, *Aneulophus, Nectaropetalum,* and *Pinacopodium* samples were mixed between all three of these types. The plant tree of life is full of recalcitrant clades whose morphological evolution has been enlightened only after the addition of phylogenetic data. As the species tree of this family grows, a thorough synthesis of morphological and phylogenetic datasets is warranted to uncover Erythroxylaceae morphological evolution.

***Biogeography of* Erythroxylum—**The lack of correspondence between the current intrageneric taxonomy and *Erythroxylum* evolutionary relationships prompted our investigation of biogeographic structure. Neotropical *Erythroxylum* form a monophyletic clade relative to the Paleotropical species. This has been hypothesized to have resulted from Boreotropical dispersal into the Neotropics because the stem age estimate of the Erythroxylaceae (60-88 Ma) postdates the breakup of Gondwana (Islam, 2011; Magallón et al., 2015). A similar scenario has been proposed for other Malpighiales lineages (Davis et al., 2002).

Our sampled clades do generally represent distinct geographic areas of Latin America and the Caribbean. Clade I is an eastern and southern Brazilian group, species in clades II and III are mostly from the Guiana Shield and adjacent regions, and clades IV and V are from the Andes-Amazon region extending north into the Caribbean and Mesoamerica. Each of these clades contains species from different yet geographically proximal regions. Many of these regions are defined by specific ecologies and precipitation regimes (i.e. Caatinga), but others, such as the Caribbean, contain a full spectrum of ecological gradients from xeric to mesic. The patterns we observe in this biogeographic analysis suggest a prominent role of geography and ecological preference in structuring *Erythroxylum* diversification.

Apart from *E. rotundifolium,* the earliest diverging Neotropical *Erythroxylum* lineages inhabit eastern Brazil. This area is ecologically heterogeneous, where rainforests of the Atlantic Forest, savanna vegetation of the Cerrado, and desert landscapes of the Caatinga converge. It also contains the highest diversity of Neotropical *Erythroxylum*; the Brazilian state of Bahia alone contains >40 *Erythroxylum* species, many from sect. *Rhabdophyllum* (Plowman, 1987; Loiola, 2001). It is possible that *Erythroxylum* lineages in this region have diversified along an ecological axis and have been the source of several dispersals into other areas of Latin America. Our phylogeny reveals the ancestors of species from the Atlantic Forests, Cerrado, and Caatinga are closely related to the ancestors of our earliest diverging clades, I and V. Furthermore, clade I, which is almost strictly Brazilian, has a sister relationship with Atlantic Forest species *E. bicolor* and then clades II-IV. Clades II and III are from the Guianas and adjacent areas, which once again suggests westward dispersal followed by diversification. The biogeographic structure within clades IV and V is not as conducive to interpretation. Mesoamerica, the Caribbean, and the Andes are, of course, ecologically heterogeneous biogeographic zones that, like eastern Brazil, could also be areas of diversification along ecological axes. Unraveling the diversification dynamics within and between Neotropical biogeographic regions could be key to understanding *Erythroxylum* and Neotropical diversification in general, but this type of analysis will require dense taxon sampling (Hughes et al., 2013).

***The domestication of coca*—**There are two standing hypotheses for the domestication of coca: the linear evolutionary series hypothesis, where *E. coca* was the progenitor of the other cultivars (Bohm et al., 1982; Plowman, 1986), and the sister species hypothesis, where the two cultivated species form monophyletic groups sister to one another with a distinct common ancestor (Johnson et al., 2005; Emche et al., 2011). The non-monophyly of the cultivated coca species in our analysis refutes both of these hypotheses and suggests a more complicated history of domestication.

Several prominent botanists have hypothesized the identity of the progenitor of coca and we have sampled from all of these taxa in this analysis: *E. anguifugum* (Rusby, 1933) *, E. cataractarum* (Gutierrez-Noriega and Von Hagen, 1950; Schultes et al., 1976)*, E. gracilipes* (Kunth fide Macbride,

1949), *E. hondense* (Poeppig fide Hartwich, 1911), and *E. mamacoca* (Kunth fide Hartwich, 1911). However, the difficult morphology and lack of taxonomic research in this group confounded any hypothesis until Timothy Plowman conducted extensive fieldwork to finally delimit the cultivated cocas into four taxa (Schultes et al. 1976; Plowman 1979, 1984).

A remarkable feature of these cultivars is that none of the four taxa grow in the wild; only Huánuco coca (*E. coca*) occasionally persists in abandoned fields. After making field observations of many of the close *Archerythroxylum* relatives thought to be allied to coca, Plowman found such feral populations in the upper Huallaga river valley in Peru. He then changed his hypothesis from a distinct wild progenitor to a wild *E. coca* progenitor from the Yungas of Peru or Bolivia, whose populations are now extinct (Schultes et al., 1976). He then added ecological and morphological data and breeding crosses to his theory to establish the *E. coca*(wild) → Huánuco coca (*E. coca*; domestic) → Trujillo coca (*E. novogranatense* var. *truxillense*) → Colombian coca (*E. novogranatense*), and independently Huánuco → Amazonian coca (*E. coca* var. *ipadu*), linear evolutionary series domestication hypothesis (Bohm et al., 1982).

Expecting to corroborate this hypothesis by finding lower genetic diversity in *E. novogranatense*, Johnson et al. (2005), analyzed AFLP data but found equal diversity in *E. coca* and *E. novogranatense*. Expanding this study, Emche et al. (2011) analyzed 36 *Erythroxylum* species (not including *E. gracilipes* or *E. cataractarum*), including 24 accessions of the cultivated coca taxa. They concluded that while the two cultivated species were more closely related to each other than to any sampled wild species, there was a "clear separation of lineage" between them (pg. 130). In yet another approach, flavonoid profiles have also been analyzed to understand the metabolic associations of the cultivars, but they did not reveal any patterns discernable of evolutionary relationships (Johnson et al., 1998, 2002). Thus, while not explicitly refuting Plowman's hypothesis, these researchers concluded that while the cultivated cocas are sister species sharing a recent common ancestor, at least the majority of their domestication was independent.

The phylogenetic placement of the two wild species *E. gracilipes* and *E. cataractarum* identifies these species as the closest wild relatives of the coca taxa and possibly the wild progenitors.

Erythroxylum gracilipes is widespread throughout the Amazon basin and surrounding regions, growing in sympatry with Huánuco and Amazonian cocas. *Erythroxylum cataractarum* grows in the Llanos of Colombia and the upper Rio Negro tributaries, geographically proximal to Colombian coca. Interestingly, *E. gracilipes* does not grow in the region occupied by *E. cataractarum*. Both of these species are morphologically very similar to the cultivated cocas. *Erythroxylum gracilipes* does produce cocaine, but a single analysis of *E. cataractarum* herbarium material did not contain the alkaloid (Holmstedt et al., 1977; Plowman and Rivier, 1983). However, anecdotally, the bitter leaf of *E. cataractarum* does seem to be physiologically active (White, unpubl.) and though they cultivate Amazonian coca, the Barasana Indians of the Rio Piraparana in Colombia refer to *E. cataractarum* as "the coca of our fathers" and use the "strong" leaves by the same method when Amazonian coca is not available (Schultes, 1981; from W. Davis #151).

The taxon sampling effort in this analysis is focused on *Archerythroxylum* and samples from other sections that are morphologically and geographically close to the cultivated cocas. We acknowledge that our intraspecific sampling is limited, and robust support of a history of coca domestication must come from demographic and phylogeographic analyses. However, this phylogenomic result is unequivocal in its support for *E. gracilipes* being the sister taxon of *E. coca* and *E. cataractarum* being closely related to *E. gracilipes* and *E. novogranatense*.

The results of our phylogeny refute Plowman's hypothesis and suggest a more complicated domestication history than presented by Johnson et al. (2005). Since the ASTRAL reconstruction places *E. cataractarum* with *E. novogranatense* (with low support) instead of with the *E. gracilipes* + *E. coca* clade as in the ML-concat reconstruction, this suggests alleles could be shared by both lineages, pointing to *E. cataractarum* as the progenitor of *E. coca* and *E. novogranatense*. It is possible that *E. gracilipes* alleles later introgressed into the *E. coca* lineage either naturally or by directed breeding.

Alternatively, *E. cataractarum* and *E. gracilipes* could be a species complex involved in two independent domestication events of coca: one in which *E. novogranatense* was domesticated from wild populations of *E. cataractarum* in northern South America, and the second in which *E. coca* was

domesticated from *E. gracilipes* populations in the Amazon basin or Andean Yungas regions. Distinguishing among these possibilities will require more focused analyses within the cultivated species and their close relatives, leading to a better understanding of the history of this controversial plant.

Chapter II:

Multiple origins of the coca plant in South America

Introduction:

Called the *Divine Leaf* by the Inca, coca has been cultivated for over 8,000 years and is among the most sacred of medicinal plants for several Andean and Amazonian indigenous cultures; it is also the natural source of cocaine (Plowman, 1986; Dillehay et al., 2010). In parts of Bolivia, Brazil, Colombia, Ecuador, and Peru, the traditional varieties of coca are cultivated by native peoples as they have been since Pre-Columbian times, and over five million South Americans chew the leaves for their mild stimulant and nutritive effects (Conzelman and White, 2016). In order to understand the identity of coca – as a sanctified plant of many South American cultures, as a potential resource for medicinal phytochemicals, and for future identification of the sources of new illicit coca strains – we have conducted the first genomic investigation of the origin, diversity, and evolution of the coca plant and its wild relatives.

The four traditional varieties of coca are grown in separate geographic areas in South America: Huánuco (or Bolivian) coca (*Erythroxylum coca* Lam.) is the most widely cultivated variety; it is found in the moist, montane belt on the eastern slopes of the Andes in Peru and Bolivia, a region called the *montaña* (Peru) or *Yungas* (Bolivia). Amazonian coca (*E. coca* var. *ipadu* Plowman) is grown in discrete localities throughout the Amazon basin and is the only variety believed to be propagated clonally as opposed to by seed (Plowman, 1981). Colombian coca (*E. novogranatense* (D.Morris) Hieron.) was historically grown in the drier montane valleys of the Sierras in Colombia; it is the only variety that is self-compatible (Bohm et al., 1982). Lastly, Trujillo coca (*E. novogranatense* var. *truxillense* (Rusby) Plowman) is grown in the arid valleys of northwestern Peru and on the Colombia/Ecuador border (Plowman, 1986).

Hypotheses of coca's ancestors and geographic origins began with some of the earliest studies of South American flora, including those of E. F. Poeppig and C. S. Kunth (Hartwich, 1911). Our prior

phylogenomic analysis of over 60 wild *Erythroxylum* species, including all previously hypothesized progenitors, identified *E. cataractarum* and *E. gracilipes* as the closest wild relatives, and possible progenitors, of coca (White et al., 2019). With this knowledge, we can test the two standing hypotheses of coca domestication, as well as a new hypothesis derived from our new data, in a statistical phylogeographic framework (Figure 3). First, Plowman's single-origin hypothesis posits that wild (and now extinct) Huánuco coca was domesticated and then taken north where it evolved into two independent lineages: Trujillo coca, which then gave rise to Colombian coca; and a lineage of Huánuco coca that gave rise to Amazonian coca (Bohm et al., 1982). A hypothesis, based on chemical profiles and genetic AFLP analyses, posits that *E. coca* and *E. novogranatense* are distinct lineages, resulting from a single domestication of a common ancestor or two domestications (Johnson et al., 2005; Emche et al., 2011). In this study, we tested this hypothesis from Johnson et al. (2005) as a two-domestication hypothesis after observing our own phylogeographic results. The third hypothesis we test is a three-domestication hypothesis that is a direct interpretation of the phylogeographic analysis we present in this paper.

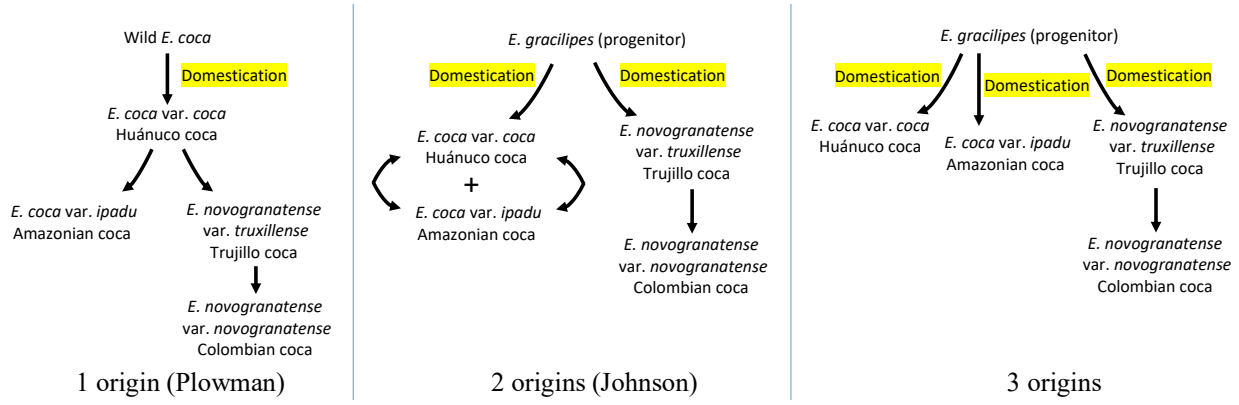


Figure 3: Three hypotheses of coca domestication: Hypothesis 1 posits Colombian and Amazonian coca were independently derived from domesticated Huánuco coca, and Trujillo coca was later derived from Colombian coca *(*Bohm et al., 1982). Hypothesis 2 posits that *Erythroxylum coca* and *E. novogranatense* represent independent lineages derived from two domestication events, with subsequent divergence into four varieties (see Johnson et al., 2005; Emche et al., 2011). Hypothesis three is derived from this phylogeographic analysis (Fig. 2) in which three independent lineages of coca appear to be domesticated independently from *E. gracilipes*.

In order to elucidate the evolutionary relationships and domestication history of coca, we utilized 544 nuclear genes from herbarium tissues to infer phylogeny, calculate population-genetic statistics, and analyze the clustering of individuals into groups. We then tested the two prior domestication hypotheses against novel scenarios we deduce from our results in an approximate Bayesian computation (ABC) framework to evaluate support for these different models (Beaumont et al., 2002; Cornuet et al., 2014).

Materials & Methods:

***DNA extraction, target capture, and sequencing*—**With a focus on maximizing geographic diversity for each taxon, we extracted genomic DNA from 155 *Erythroxylum* herbarium specimens using a 2X CTAB protocol with 3% PVP and 2% 2-mercaptoethanol in the extraction buffer (Doyle and Doyle, 1990) from the following taxa: *E. cataractarum* (12), *E. coca* (44), *E. coca* var. *ipadu* (20), *E. foetidum* (2; outgroup), *E. gracilipes* (40), *E. novogranatense* (25), and *E. novogranatense* var. *truxillense* (18) (Appendix B sample information). Discolored samples were cleaned with the MOBIO Laboratories Inc. DNA Clean-Up Kit. All genomic DNA samples with fragment sizes >500 bp were acoustically sheared to 400 bp (PiP 50 W, duty factor 10%, 200 cycles per burst, 70 seconds, 20ºC) with the Covaris M220 Focused-ultrasonicator in 130 µL snap-cap tubes (Covaris Inc., Woburn, MA, USA). Starting with 10–20 ng/ul^{-1} input DNA, we prepared

samples for target capture using the KAPA Hyper Prep kit with the Illumina TruSeq HT dual-indexed adapters (96 samples) or Adapterama indexed adapters with iTru7_101 and iTru5_101 primer combinations (59 samples; Glenn et al. 2016). The prepared genomic DNA samples were pooled in groups of 10–16 for targeted sequence capture using a custom set of RNA probes designed to capture 544 nuclear genes (White et al., 2019; Arbor Biosciences, Ann Arbor, MI, USA). Target capture preparation and protocol followed the myBaits v3 user manual with 22 hours of hybridization at 65ºC and 12–14 cycles of post-capture amplification. The hybridization pools were checked on an Agilent Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA, USA) and pooled together at 5 nM final concentration for sequencing. All 155 samples were sequenced on the Illumina HiSeq4000 platform (Illumina, Inc., San

Diego, CA, USA) in a single lane (2x100 bp) at the University of Oregon. Sixty of the samples returned suboptimal sequencing results and were rehybridized and amplified in six pools and sequenced on the Illumina NextSeq platform at mid-output (2 x 150 bp) at the University of Illinois at Chicago.

***Target sequence assembly, alignment, and phylogenetic reconstruction*—**Raw reads were trimmed, filtered, and demultiplexed using Trimmomatic (Bolger et al., 2014) by removing bases with a q-score<20 and hard trimming 5–20 bases on the head and/or tail as judged from the base composition graph in FASTQC (Andrews, 2010), and PCR duplicates were removed using SuperDeduper (github.com/dstreett/Super-Deduper.git). Cleaned reads were de novo assembled into contigs and mapped to the concatenated exon sequences for each gene using HybPiper (Johnson et al., 2016). Nine individuals with poor gene recovery (<90 genes) were removed, leaving 146 samples in the analysis including the two outgroup *E. foetidum* samples. Genes with paralog warnings were removed and supercontigs for each of the remaining 424 genes (exon+intron sequence) were aligned using the MAFFT local pairwise algorithm with a maximum of 1000 refinement iterations (Katoh et al., 2002). Poorly aligned regions were removed using the automated algorithm in trimAL (Capella-Gutiérrez et al., 2009). For phylogenetic inference, we reconstructed a maximum-likelihood tree from 20 random starting trees under a GTR + GAMMA model of nucleotide substitution (Tavaré, 1986; Yang, 1993) for each gene alignment with RAxML v.8 (Stamatakis, 2014). The species tree for the nuclear dataset was inferred from these 424 ML gene trees using ASTRAL-II (Mirarab et al., 2014). The final species tree figure was made using the Interactive Tree of Life v.4 webserver (itol.embl.de).

***SNP-calling*—**We generated a consensus sequence for each of the 424 gene alignments by first using Geneious v. 10 to mask all sites with at least 99% missing data in order to exclude insertions unique to single individuals (http://www.geneious.com; Kearse et al. 2012). Second, using a custom script (github.com/airbugs), we generated a majority-rule consensus sequence with the caveat that all sites represented by less than 8 samples were replaced with a "N" in order to mask regions of the alignment with low sample representation. We mapped our demultiplexed reads to the consensus sequences with Bowtie 2 (Langmead et al., 2013) and followed the seqcap_pop pipeline steps 5-10 to convert the .sam to

a clean .bam file, add read groups, mark PCR duplicate reads, and merge our .bam files across individuals (https://github.com/mgharvey/seqcap_pop; Faircloth 2015; Harvey et al. 2016).

We then followed the GATK Best Practices workflow for germline short variant discovery (https://software.broadinstitute.org/gatk/best-practices/workflow?id=11145) to call, annotate, and filter SNPs in each gene with version 4 tools: We first called single-sample variants with HaplotypeCaller in GVCF mode, then consolidated samples for each gene with CombineGVCFs and joint-called all samples with GenotypeGVCFs. To produce the final VCF file, indels, SNPs with >2 alleles, and all SNPs failing any of the following hard filters were removed using SelectVariants: QD < 7.0, FS > 10.0, MQ < 41.0, SOR > 2.5, MQRankSum < -5.0, ReadPosRankSum < -5.0 (Depristo et al., 2011). Lastly, SNPs with >20% missing samples and a minor allele count less than two were removed using VCFtools (Danecek et al., 2011).

***Cluster and population genetic analyses*—**We used our final dataset of 26,655 SNPs for 144 individuals in a genetic cluster analysis using the *snapclust* program from the 'adegenet' version 2.1.1 R package (Jombart and Ahmed, 2011; R Development Core Team, 2013). For up to 14 clusters, five and nine showed significant improvement according to the AIC (Appendix B: Figure B2). The K=5 and K=9 population assignment probabilities were then added to the species tree (Figure 4). Allelic richness, Fst (Weir and Cockerham, 1984), and Jost's D (Jost, 2008) were calculated using the 'diveRsity' R package (Keenan et al., 2013) for only the 10 individuals from each population with the least amount of missing SNPs. This resulted in a SNP dataset of 17,501 sites for 60 individuals. We used the 'diveRsity' R package to calculate several population genetic statistics using our six taxonomic assignments as "populations" after dropping the individuals with "cf." taxonomic identifications (cf.coca_884, cf.trux_824, cf.trux_916). These "cf." determinations were applied to herbarium samples in scenarios where morphological and genetic results were inconsistent with taxonomic names.

We reformatted our VCF file for the fineRADstructure analysis using the *vcf2hapmatrix* python script available from the 'radseq' package (github.com/pimbongaerts/radseq). The two *E. gracilipes* samples from Bolivia (grac_104 & grac_868) were dropped from the final fineRADstructure analysis

because their high pairwise coancestry score of 364 reduced the resolution of the subtler coancestries in the rest of the heatmap. We generated the coancestry matrix with the RADpainter program (Malinsky et al., 2018), and defined our population groups and reconstructed the distance-based tree with the finestructure program following default settings and plotted the heatmap and tree using the provided R codes (Lawson et al. 2012; github.com/millanek/fineRADstructure).

ABC domestication model testing— We completed three runs with DIYabc version 2.0 to test different domestication scenarios as coalescent-based diversification models (Beaumont et al., 2002; Cornuet et al., 2014). For this analysis, we sampled one random SNP from each of our 424 genes. The first run tested three models: Plowman's linear series, two domestications with Amazonian coca derived from Huánuco coca (2D_ci_tn), and three domestications (3D_Gtn). In the second run, we tested 3D_gtn against the two-domestication scenario (2D_ci_tn) and added the second two-domestication scenario in which Huánuco coca was derived from Amazonian coca (2D_ic_cn; see Appendix B: DIYabc run scenarios 1 & 2). In an effort to elucidate the order of events in the second domestication, we sampled two or three SNPs from each gene and reduced the taxa to only *E. gracilipes*, Amazonian and Huánuco coca individuals. We tested five scenarios in this last run: three domestications with Amazonian coca domesticated earlier (3D_ifirst), three domestications with Huánuco coca domesticated earlier (3D_cfirst), simultaneous domestications of Amazonian and Huánuco coca (polytomy), two domestications with Huánuco coca as progenitor of Amazonian coca (2D_ci), and lastly), two domestications with Amazonian coca as progenitor (2D_ic; see Appendix B: DIYabc run scenario 3). Population sizes were given a uniform prior from 10 to $1x10^6$ individuals, except for *E. gracilipes*, which was from 10 to $1x10^7$; time to each coalescence event was given a uniform prior (with some relationship parameters) from 10 to $1x10^6$ generations (see Appendix B: DIYabc parameters).

Results & Discussion:

Our species tree inference shows that *E. gracilipes*, a hypothesized wild progenitor of coca (White et al., 2019), is paraphyletic; individuals form multiple distinct lineages that separate the coca taxa

(Figure 4a). This pattern implies that coca was domesticated multiple times from *E. gracilipes,* a widespread Amazonian species which is also known to produce cocaine (Plowman and Rivier, 1983; Crawford, 2010). However, *E. cataractarum,* from the upper Orinoco and western Rio Negro tributaries, has no apparent role in coca domestication.

Our phylogenetic and genetic clustering analyses suggest three domestications (Figure 4a): Trujillo was the first coca to be domesticated from *E. gracilipes,* presumably in the moist forests of Colombia or Ecuador, and Colombian coca was subsequently derived from Trujillo coca (Dillehay et al., 2010). Given that all coca macrofossils, the earliest dated to 8,000 years BP (Dillehay et al., 2010), are of the Trujillo morphotype and that Colombian coca is the only variety that is self-compatible (a derived trait, unlikely to give rise to self-incompatibility (Goldberg et al., 2010)), we believe that this is the oldest domestication event and the progenitor-derivative relationship (PDR) here was Trujillo→Colombian. More recently, Huánuco coca was domesticated from *E. gracilipes,* possibly in the Andes/Amazon region of southern Peru, before being dispersed throughout the *montaña*. Lastly, the most recent domestication from *E. gracilipes* (recency revealed by the lack of identity in the K=5 cluster analysis) seems to have occurred near the Ecuadorean Amazon, creating Amazonian coca.

To elucidate the PDR of *E. gracilipes* lineages to each domesticated clade, we also calculated population genetic statistics and allele sharing. We found allelic richness of single-nucleotide polymorphisms (SNPs) is highest for coca's wild relatives *E. gracilipes* and *E. cataractarum,* with less genetic diversity in the cultivars: Amazonian and Huánuco coca have equal allelic richness, followed by Trujillo, then Colombian (Figure 4c). This is consistent with the expectation from Plowman's single-domestication hypothesis, though it could also be expected from a three-domestication scenario (Gaut et al., 2015). Weir and Cockerham's (1984) Fst (Weir and Cockerham, 1984) and Jost's D (Jost, 2008) statistics also suggest three domestications (Figure 4b): the cultivars are more similar to *E. gracilipes*

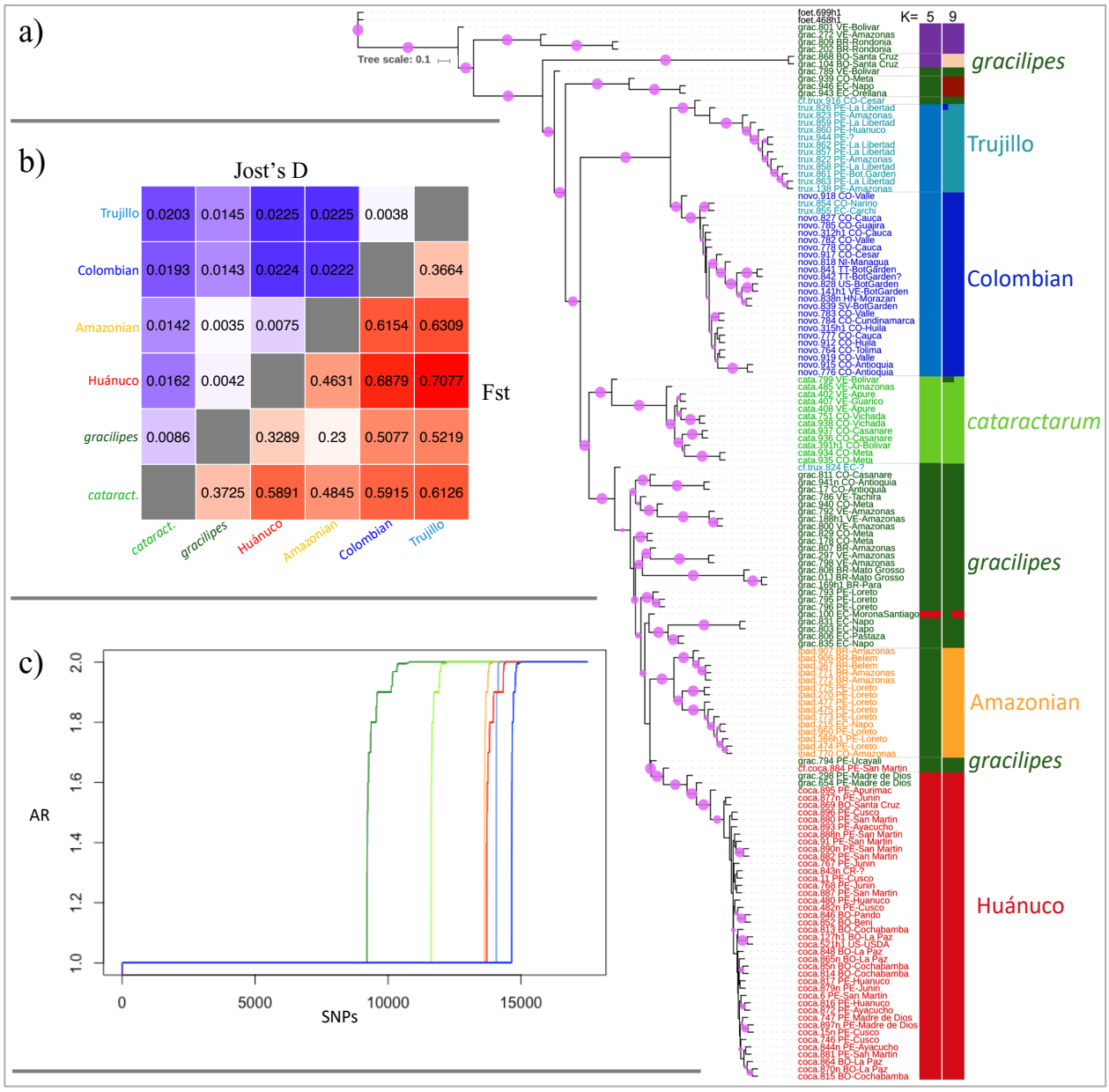


Figure 4: (a) Astral II lineage tree and genetic cluster assignment (for 5 and 9 groups) of four coca varieties and two wild relatives. Branches with local posterior probability 0.5-1 are marked with a circle, the larger sizes reflecting increasing probability. Branch lengths are in coalescent units (terminal branch lengths arbitrary). Tips are color-coded by taxonomic name and cluster colors correspond to the same colors. **(b)** Pairwise Jost's D and Weir and Cockerham's F^{st} statistics of taxa. **(c)** Allelic richness of 17,501 rank-ordered, biallelic SNPs; the rarefaction-corrected number of alleles present for each SNPs for each taxonomic group. Colors of lines correspond to taxa as in (a) and (b). To summarize, *E. gracilipes* populations have the most SNPs with more than one allele present, followed by *E. cataractarum*, Amazonian and Huánuco are roughly equivalent, followed by Trujillo, and Colombian coca has the least, and thus the lowest allelic richness.

populations than *E. cataractarum*, with Trujillo/Colombian being most differentiated from *E. gracilipes,* followed by Huánuco, and lastly Amazonian coca. Interestingly, these statistics support the PDR of Trujillo and Colombian coca but show the that Huánuco and Amazonian coca are less similar to each other than either is to *E. gracilipes*, once again supporting the three-domestication scenario. Our fineRADstructure result suggests Amazonian coca is the only cultigen with significant allele sharing with *E. gracilipes* (see Appendix B: Figure B1). Interestingly, this analysis also shows some signal of ancestral polymorphism or admixture between the two "cf.trux" samples (identified morphologically as Trujillo coca and genetically as *E. gracilipes*), Colombian and Trujillo cocas, as well as some *E. gracilipes* and Amazonian coca individuals. There is a similar situation of widespread allele sharing among the Peruvian *E. gracilipes* samples that are most closely related to Huánuco coca. These individuals could be hinting at a more complicated ancestry involving episodes of introgression from *E. gracilipes* and admixture between cultigens.

The results of ABC model testing showed that Plowman's single-domestication hypothesis, positing that Huánuco coca is the progenitor of the other cocas, was the least supported model of all that we tested. Under our first analysis of all six taxa, we found better support for two-domestication models over the three-domestication model (Figure 5a; Appendix B: DIYabc model comparisons). The analysis supported the first domestication of Trujillo coca from *E. gracilipes*, which then gave rise to Colombian coca. However, it was ambiguous as to the order of events in the second domestication; whether Amazonian coca was domesticated from *E. gracilipes* and Huánuco coca was later derived or vice versa. We then conducted a second analysis of only *E. gracilipes,* Amazonian and Huánuco coca individuals and found support for two models: independent domestications of Amazonian and Huánuco coca (three domestications total) or the domestication of Amazonian coca with Huánuco coca later derived from that lineage (Figure 5b; Appendix B: DIYabc model comparisons).

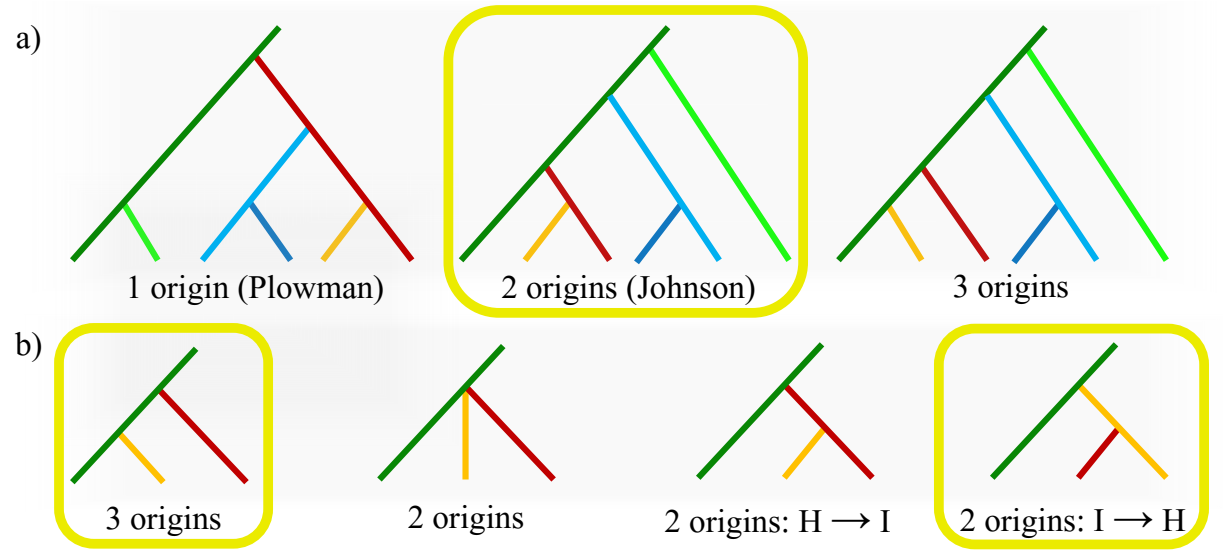


Figure 5: (a) ABC tests of the three models of domestication including all six taxa. The yellow box highlights that the two-origin scenario is most probable of the three models. **(b)** Refined ABC tests of the progenitor-derivative relationship between Amazonian and Huánuco cocas. Two scenarios were supported: independent domestication (3 domestications total) or Huánuco was derived from Amazonian.

To summarize, the taxa and genes sampled in this study provided enough information to clearly distinguish a minimum of two domestication events. First, Trujillo coca was domesticated from wild *E. gracilipes* in northwestern South America over 8,000 years ago and subsequently gave rise to Colombian coca. Amazonian coca was domesticated independently in the western Amazon basin and then taken to the *montaña* of Peru and Bolivia where it evolved into Huánuco coca, but it is also possible that Huánuco coca originated via a third, independent domestication event from *E. gracilipes*. Further geographic and genetic sampling is underway to disentangle these scenarios.

The domestication of coca reveals that peoples in the Amazon basin and the tropical Andes have been continuously sampling and adopting wild *E. gracilipes* into cultivation as a mild stimulant and medicine, resulting in two or three domesticated coca lineages. This contradicts the traditional Vavilovian (Vavilov and Löve, 2009) view of few, distinct centers of origins for crops and instead corroborates widespread and repeated domestication practices (Harlan, 1971). Ethnobotanist R. E. Schultes called coca the most important South American narcotic plant due to its prevalence and significance for indigenous

cultures as well as the revolutionary role of cocaine in Western medicine (Schultes, 1979). In light of the knowledge that *E. gracilipes* forms the large genetic pool from which all coca varieties were derived, investigations into the phytochemical diversity of this widespread Amazonian species could yield useful medicinal compounds (Hegnauer, 1981). Our study has also laid the foundation for understanding the genetic structure of the heirloom coca varieties; important for future identification and classification of new coca strains being cultivated for illicit cocaine production.

Chapter III:

The Pantropical diversification of the Coca family highlights boreotropical migration and biome shifts out of rainforests.

Introduction:

Understanding the pattern and process of organismal distributions is of fundamental importance in ecology and evolutionary biology. Among the most pronounced of these patterns is the latitudinal diversity gradient (Hillebrand, 2004; Mittelbach et al., 2007), yet even within the tropical latitudes plant diversity is also not evenly distributed (Kier et al., 2005). Vegetation plot data indicate around 16,000 tree species occur in Amazonia alone (ter Steege et al., 2016), and while the Neotropics and Indo-Pacific tropics might carry a similar number of tree species, the diversity in tropical Africa is only estimated at 4,500-6,000 species (Slik et al., 2015; Ter Steege et al., 2016). Estimates of species richness for all vascular plants suggest the Neotropics have the highest levels of diversity (80,000-100,000 species), followed by the Indo-Pacific region (40,000-82,000 species), and Africa (30,000-35,000 species) (Gentry, 1982; Antonelli and Sanmartín, 2011).

Phylogenetic studies on the latitudinal diversity gradient support the effect of longer time-integrated area and climatic stability of tropical versus temperate biomes, as these factors enable large population sizes that facilitate speciation and reduce the risk of extinction (Fine and Lohmann, 2018). However, there are very few studies examining broad patterns in tropical plant diversity, and researchers have repeatedly stressed the need for more detailed, clade-specific case-studies (Hughes et al., 2013; Donoghue and Edwards, 2014; Antonelli, Ariza, et al., 2018). This has been met with several historical biogeographic analyses describing the origin, diversification, and ecological evolution of plant lineages.

Several Malpighiales lineages, including the Chrysobalanaceae and Malpighiaceae, as well as the Coca family (Erythroxylaceae), exhibit a pantropical distribution with much higher (70–85%) species richness in the Neotropics compared to the Paleotropics (Davis et al., 2002; Bardon et al., 2013). The Chrysobalanaceae appear to have originated in the Paleotropics in the late Cretaceous (80 Ma) and subsequently dispersed into the Neotropics at least four times starting in the Paleocene-Eocene (60–40

Ma), with the dispersal events associated with diversification rate increases (Bardon et al., 2013). In contrast, the Malpighiaceae appear to have originated in the Neotropics during the early Paleogene (64 Ma) and have migrated into the Paleotropics several times from the Eocene to the Miocene (55–10 Ma). This suggests that their Neotropical diversity can be explained by a longer time for lineage diversification, rather than a change in diversification rate (Davis et al., 2002). These dispersal events between hemispheres are generally accepted to have proceeded via 'boreotropical migration' (Lavin and Luckow, 1993) across northern latitudes during times of pronounced global warming. The early Eocene climatic optimum (52.6-50.3 Ma) enabled warm and humid climates to reach into Greenland, Europe and North America at a time when these continents were geographically proximal, thus creating a north Atlantic land bridge with suitable habitats for migration of tropical plants between the hemispheres (Wolfe, 1975; Tiffney, 1985; Payros et al., 2015).

The tribe Protiae (Burseraceae; Fine et al., 2014) and the Velloziaceae (Alcantara et al., 2018) are two other lineages that seem to have radiated after migration into the Neotropics. The Meliaceae exhibit remarkable diversification rate heterogeneity across Indo-Pacific, African, and Neotropical clades, and have undergone several Oligocene or younger dispersals from Africa to the Neotropics, with two independent radiations in Amazonia (Koenen et al., 2015). While the Meliaceae originated near the Cretaceous-Paleogene boundary (65 Ma), most of the species diversity originated in the mid to late Miocene (15-5 Ma). The multiple cases of recent species diversification in tropical biomes provides a strong hypothesis that species turnover, possibly caused by dynamic speciation and extinction rates, is a pervasive pattern that defines tropical species-level diversity (Hoorn et al., 2010; Hughes et al., 2013).

In addition to documenting how plant lineages have moved and diversified in geographic space, it is important to quantify how these lineages have evolved on a broad ecological axis, since both geographic and ecological barriers define biogeographic patterns (Donoghue, 2008; Fine, 2015; Cavender-Bares et al., 2016; Slik et al., 2018). In order to understand whether one of these processes is more prevalent than the other, it is important to understand the strength of phylogenetic niche conservatism (Wiens, 2004; Donoghue, 2008), or the conservation of organisms' ecological attributes

during speciation. Two outstanding meta-analyses of phylogenetic niche conservatism on the biome scale have been conducted to test the strength of biome conservatism. First, Crisp et al. (2009) estimated that 3.6% of cladogenic events were associated with biome shifts in their phylogenetic analysis of 11,000 Southern-Hemisphere plant taxa, supporting strong biome conservatism. However, 17.7% of their inferred 226 transoceanic dispersal events were associated with biome shifts. In a second meta-analysis of 2,114 angiosperm taxa, Antonelli et al. (2018) analyzed dispersal and biome shifts among Neotropical biogeographical regions, and found biome shifts accompanying 47% of cladogenic events (Table S5 & S6 in Antonelli, Zizka, et al., 2018). The stark difference in biome conservatism versus biome evolutionary lability exhibited by these studies might be explained by the scale, as the first incorporated shifts out of the tropics that require cold-tolerance adaptations while the second only analyzed shifts within tropical biomes.

Another study of the sister families Anacardiaceae and Burseraceae (Weeks et al., 2014) also found remarkable contrasts in the patterns of biome shifting. The common ancestor of these families was inferred to have lived in temperate latitudes about 65 Ma. While extant Burseraceae species diversity mostly originated within the tropics during the Miocene, with a diversification rate increase when the Protiae entered the Neotropics (Fine et al., 2014), only a few biome shifts are inferred between wet and dry tropical regions. Conversely, the Anacardiaceae have steadily accumulated species around the globe and frequently changed climatic niches. These few studies highlight the remarkable heterogeneity of the patterns of global diversity and lineage diversification and emphasize the need for more densely-sampled historical biogeographic analyses.

The Coca family (Erythroxylaceae) is a Pantropical clade of shrubs and small trees in the Malpighiales, with the majority of species diversity (72% of 285 species) in the Neotropics. The family is classified into four genera, with the vast majority of the species belonging to genus *Erythroxylum* (271 species: 202 Neotropical (75%), 33 Malagasy (12%), 16 African (6%), and 20 Indo-Pacific (7%). The other genera are restricted to Africa: *Aneulophus* (1 species), *Nectaropetalum* (8 species), and *Pinacopodium* (2 species). The Erythroxylaceae have been hypothesized to have a fast diversification rate

compared to other Malpighiales families (Xi et al., 2012), and a whole genome duplication has been inferred in the early Eocene (~56 Ma), based on a Ks-based method applied to the *E. coca* transcriptome (Cai et al., 2019).

While the Erythroxylaceae are predominantly known for being the natural source of the tropane alkaloid cocaine, isolated from two cultigen species called coca (*Erythroxylum coca* and *E. novogranatense*), it is also remarkable due to its diversity across many lowland biomes in the Neotropics. *Erythroxylum* is tied as the 22nd most diverse genus in Amazonia with 82 species recorded (ter Steege et al., 2013). Yet its center of diversity is in eastern and northeastern Brazil where the moist Atlantic Forests intersect with the dry forest biome of the Caatinga and the Cerrado (savanna/grassland) biome; the state of Bahia alone harbors over 40 species (Plowman, 1987). While *Erythroxylum* inhabits wet and dry biomes throughout the Neotropics, species in Africa and the Indo-Pacific exhibit more scattered distributions. The majority of Paleotropical species occur in rainforests but dry-broadleaf and savanna/grassland biomes are also inhabited, though less frequently. One notable pattern in Paleotropical diversity is the high concentration of species inhabiting Madagascar (33 species); this comprises 48% of all Paleotropical *Erythroxylum* species.

The reproductive and vegetative morphology of the Erythroxylaceae is remarkably consistent across the family. All species are small trees and shrubs with the majority of species diversity in tropical rainforests. The genera *Erythroxylum, Nectaropetalum,* and *Pinacopodium* have alternate, entire leaves with conspicuous pinnate and camptodromous to brochiodromous venation, subtended by an intrapetiolar stipule that protects axial and terminal buds (Rury, 1982). The axillary flowers are solitary or rarely fascicled, small or medium (1–3 cm) sized, bisexual, pentamerous, and diplostemonous; the petals have a two-lobed ventral ligule that surrounds the androecium (Matthews and Endress, 2011). The flowers are pollinated by various Hymenopter and Diptera (Rury, 1982). The monotypic *Aneulophus* is the most divergent Erythroxylaceae taxon with it has opposite leaves and a septicidal capsule fruit type, presenting a more intermediate morphology with non-mangrove Rhizophoroid relatives. The rest of the

Erythroxylaceae genera produce small (1.5-3 cm) red to purple drupes that are readily dispersed by birds (Gryj and Domínguez, 1996).

The goal of this investigation is to describe the continental dispersal and biome evolution of the Erythroxylaceae during their Pantropical diversification and investigate how the phylogeny informs the tempo of this diversification. We seek to quantify species abundance in the three Neotropical and Pantropical biomes inhabited by this clade (moist-broadleaf forest, dry-broadleaf forests, and savanna/grasslands), and describe the frequency of biome shifts in the Neotropics and Paleotropics to inform the strength of biome conservatism in the diversification process. We also investigate the phylogenetic assemblages within biomes by assessing the mean pairwise distance and mean nearest taxon distance of species within the same biomes (Webb et al., 2002). If species from the same biome are clustered within the phylogenetic tree, it will suggest strong phylogenetic niche conservatism and within-biome diversification, whereas their scattered distribution throughout the tree would suggest dispersal among biomes and ecological evolution.

Given these aspects of Erythroxylaceae distribution and ecology, we hypothesize that the Erythroxylaceae originated in Africa before dispersing into the Indo-Pacific and Neotropical regions. We expect to find an increased diversification rate among one or more lineages in the Neotropics and we believe that biome conservatism is weaker in the Neotropics, as identified by a larger proportion of biome shifts associated with cladogenesis. The patterns of dispersal and ecological evolution are important correlates of the drivers of Erythroxylaceae diversification and its global patterns of species richness.

Materials & Methods:

***DNA extraction, target capture, and sequencing*—**We sampled herbarium and silica-dried leaf tissue from 182 Erythroxylaceae individuals. We did not include any coca individuals (*E. coca* or *E. novogranatense*) because these are believed to be cultigens instead of naturally evolved taxa (White, Huang, et al., 2019). Genomic DNA was extracted using a CTAB protocol with 3% PVP and 2% 2-mercaptoethanol in the extraction buffer (Doyle and Doyle, 1990). Several recently collected samples that

had very mucilaginous first extractions were reextracted with 1/2 volume 5 M NaCl added to the solution before the isopropanol precipitation step (Healey et al., 2014).

Target capture library preparation began with acoustically fragmenting all genomic DNA samples with fragment sizes > 500 bp down to 400 bp on the Covaris M220 Focused-ultrasonicator in 130 µL snap-cap tubes (PiP 50 W, duty factor 10%, 200 cycles per burst, 70 seconds, 20ºC; Covaris Inc., Woburn, MA, USA). All degraded genomic DNA samples with fragment sizes < 500 bp were not acoustically fragmented. Starting with 10-20 ng/µl input DNA, we prepared all samples for target capture using the KAPA Hyper-Prep kit with Adapterama dual-indexed adapters with iTru7_101 or iTru7_102 and iTru5_101 primer combinations (Glenn et al., 2016). The prepared genomic DNA samples were then pooled in groups of 10–16 for targeted sequence capture using a custom set of RNA probes designed to capture 544 nuclear genes (White et al., 2019; Arbor Biosciences, Ann Arbor, MI, USA). Target capture preparation and protocol followed the myBaits v3 user manual with 22 hours of hybridization at 65ºC and 12–14 cycles of post-capture amplification. The hybridization pools were then checked on an Agilent Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA, USA) and pooled together at 5 nM final concentration for sequencing. The majority of samples (177) were sequenced in a single lane (2x150 bp) on the Illumina HiSeq4000 platform (Illumina, Inc., San Diego, CA, USA) at the University of Oregon. Five samples (A_africanus, B_cylindrica, C_brachiata, C_elliptica_444, and R_mangle) were enriched together as a separate pool and sequenced along with a different library (2 x 150 bp) on the Illumina NextSeq platform (Illumina, Inc., San Diego, CA, USA) at the University of Illinois at Chicago.

Target sequence assembly, alignment, and phylogenetic reconstruction—Raw Illumina reads of 70 samples (*N. zuluense, P. congolense,* and *Erythroxylum* spp.) were incorporated from White et al. (2019) and 6 more *Erythroxylum* samples were incorporated from White, Huang, et al. (2019). These reads were obtained following the same protocols described here. Raw reads of our total pool of 258 Erythroxylaceae and Rhizophoraceae individuals were trimmed, filtered, and demultiplexed using the BBDuk program in BBMap (Bushnell, 2015) by removing bases with a q-score <20 and hard trimming 10–30 bases on the head and tail as judged from the base composition graph in FASTQC (Andrews, 2010). We then removed

reads with an average q-score <20 and PCR duplicates were identified and removed using SuperDeduper (github.com/dstreett/Super-Deduper.git).

Cleaned reads were de novo assembled into contigs and mapped to the concatenated exon sequences for each gene using HybPiper with a minimum coverage cutoff of 4 and otherwise default settings (Johnson et al., 2016). Seventeen individuals with poor gene recovery (<100 genes) were removed, including 14 of the 182 samples sequenced in this study, leaving 241 samples in the final analysis. For the 544 genes, supercontigs (exon+intron) were aligned using the MAFFT local pairwise algorithm with a maximum of 1000 refinement iterations (Katoh et al., 2002) and alignments were cleaned using the automated algorithm in trimAL (Capella-Gutiérrez et al., 2009). We removed 26 gene alignments from the analysis that had a total length <500 bp, leaving 519 gene sequence alignments in the final phylogenetic analysis.

For each gene alignment, we reconstructed a maximum-likelihood tree from 20 random starting trees under a GTR + GAMMA model of nucleotide substitution (Tavaré, 1986; Yang, 1993) using RAxML v.8 (Stamatakis, 2014). We then summarized our 519 gene genealogies using the gene tree summary method implemented in ASTRAL III to generate an ASTRAL lineage tree of all 241 samples in the analysis to identify non-monophyletic species. Taxonomic names were verified using the most recent taxonomic resources for the group (Chung, 1996; Loiola, 2001; Oviedo, 2002; Plowman and Hensold, 2004; Loiola et al., 2014; Costa-Lima et al., 2015) but old names were retained if they were not grouped with the presumed synonymous taxon (e.g. *E. tikalense*) in order to help guide future taxonomic revisions. We then reconstructed an ASTRAL III "species" tree with 207 tips that grouped monophyletic species with multiple individuals into a single tip but retained multiple tips for polyphyletic species.

Since the quartet-based coalescent method of ASTRAL does not calculate terminal branch lengths of species represented by a single individual (the majority of the species in our analysis) we also reconstructed a maximum likelihood tree from our concatenated gene alignments of 241 individuals ("MLcat" tree). For this tree, we took the most likely tree from 20 random starting trees inferred under a GTR + GAMMA model of nucleotide substitution with the RAxML-HPC2 Tool on XSEDE via the

CIPRES Science Gateway web portal (Miller et al., 2010; Stamatakis, 2014; www.phylo.org). With the exception of the polyphyletic species identified in the ASTRAL lineage tree, we pruned extra individuals from species with multiple individuals sampled our MLcat as a template to prune our MLcat tree from 241 tips (total individuals) to 207 tips (species). When species were comprised of multiple individuals in the MLcat tree, we pruned individuals with the longest total branch lengths, retaining only a single individual per species. When species were comprised of multiple paraphyletic individuals, we retained the earliest diverging individual from that group. We believe this pruning method of removing extra individuals with the longest branches is justified due to the phenomenon of spuriously long terminal branch lengths of certain samples, as explained in the next section.

***Correcting terminal branch lengths*—**Upon viewing our MLcat tree it was obvious that the terminal branch lengths of several taxa were incredibly long. Inspection of individual gene alignments readily revealed intervals of divergent sequences on the ends of our supercontigs that we believe to be the cause of the long terminal branches. Despite our conservative raw read filtering and trimming steps, we believe these to be spurious nucleotides caused by DNA degradation in herbarium tissues combined with sequencing errors. In order to remove the potentially misleading effects of these sequences on the shape of our MLcat tree in diversification rate and phylogenetic community composition analyses, we removed them from each gene alignment using a heuristic approach.

Using custom R scripts, we rooted each gene tree and identified those species on long branches using a modified version of the long-branch score of Struck (2014) with species' scores calculated by incorporating the distance from root to tip, rather than the mean distance between tips. We identified samples with branch lengths greater than the 60^{th} percentile and removed those sequences from the alignment of the corresponding gene tree. These pruned alignments were then realigned with MAFFT using the same parameters as above, reconcatenated, and input for the final maximum likelihood phylogenetic reconstruction with long-branch correction. We found this method was also effective at identifying and removing possible paralog sequences, where entire clades stuck on long branches.

Outgroup Rhizophoraceae species were excluded as they would be identified as being on long branches due to their phylogenetic position, rather than because of a sequence error. Iterating the analysis with Rhizophoraceae individuals included in the root to tip calculations was not as effective in reducing terminal branch lengths of Erythroxylaceae species. After the automatic sequence-removal method above was applied, we dropped an additional seven species that still appeared to have incredibly long terminal branch lengths, leaving 200 species in the final "MLcatLB" tree. With a lower percentile cut-off value for long branches, we expect these remaining tips would have also been reduced but at the expense of a concatenated gene alignment with even more missing data. Topological comparisons between the ASTRAL tree, MLcat, and MLcatLB trees were made using the R package, 'phytools' (Revell, 2012).

***Estimation of divergence times*—**Two fossil Erythroxylaceae species have been described, both from South America: *Erythroxylum cuneifolioides* from Eocene and Miocene formations and *E. reichei* from Miocene to "early Eocene" strata (Engelhardt, 1905; Berry, 1925; see Graham, 2010). Due to the extensive stratigraphic ranges of these species combined with our uncertainty in placing them within the Neotropical *Erythroxylum* clade, we used this fossil evidence to date the minimum age of the stem node of the Neotropical *Erythroxylum* clade at the end of the Ypresian age (47.8 Ma). We also utilized a secondary calibration at the split of *Carallia gymnorhiza* (Rhizophoraceae) and *Aneulophus africanus* (Erythroxylaceae) with the minimum, 59.3 Ma, and maximum, 87.6 Ma, ages in the 95% highest posterior from the Bayesian dating analysis of Magallón et al. (2015).

We estimated time-calibrated phylogenies using the penalized likelihood dating method implemented in treePL (Smith and O'Meara, 2012). For the MLcat and MLcatLB trees, we first used the priming step to identify the best optimization methods and then estimated divergence times with the leave one out cross-validation option to estimate the smoothing parameter.

***Diversification Analysis*—**We pruned outgroup Rhizophoraceae species from our MLcat and MLcatLB time-calibrated trees and inferred the probability of rates of speciation, extinction, and rate shift configurations on these phylogenetic trees using the Bayesian RJMCMC method implemented in BAMM 2.5 (Rabosky, 2014). We accounted for incomplete taxon sampling using specific sampling fractions of

six clades (*Aneulophus*: 1, *Nectaropetalum*: 0.375, *Pinacopodium* 0.5, Paleotropical: 0.61, Neotropical-A: 0.78, Neotropical-B: 0.77; Appendix C: Table C1, Figure C1). For the MLcat and MLcatLB trees, we estimated diversification priors using the 'BAMMtools', R package (Rabosky et al., 2014) but varied the estimated number of shifts prior between 1 and 10, and the minimum clade size for shift between one, two, and 10, and varied the number of chains between four and eight.

For the MLcat tree, after multiple attempts of MCMC searches for 300 M generations, chains failed to converge using the one expected shift and a minimum clade size of one. We were able to repeatedly achieve chain stationarity, as observed on the likelihood plots and high (>200) ESS values, with priors set to 1 expected shift and a minimum clade size of 2. For the MLcatLB tree, we were able to achieve chain stationarity in multiple runs after 40 million generations with priors set to 1 expected shift and a minimum clade size of 2, or 10 shifts with a minimum clade size of 2. We also evaluated the macroevolutionary cohort analysis in order to identify clades with similar diversification regimes (Shi and Rabosky, 2015).

***Biogeography and biome evolution*—**We coded sampled species by tropical region (Paleotropical vs. Neotropical), world region (Africa, Madagascar, Indo-Pacific, Caribbean, Mesoamerica, South America), and biome (tropical rainforest, tropical dry forest, or tropical savanna/grassland). Our geographic delimitation of biomes follows Pennington et al. (2018) . Neotropical species' characters states were assigned by taxonomic experts Adolfo Jara Munoz and Iracema Bezerra Loiola and verified by me using the same methods for Paleotropical species' character states. Taxonomic expert John L. Clarkson provided data for Australian species. For the other Paleotropical species, we assigned character states using a variety of data sources: GBIF records and live plant images, herbarium ticket information, and plant physiological traits such as drip tips, leaf deciduousness, or suberous bark.

In addition to quantifying the number of species occupying each biome type in each tropical region, we used a maximum parsimony character state reconstruction to estimate the minimum number of biome shifts in each direction that would explain the observed distribution of character states over the phylogeny. Patterns of overdispersion between relatedness and biome will suggest that dispersal and

ecological speciation are drivers of speciation. Alternatively, significant phylogenetic clustering in biomes would suggest that dispersal is rare among regions, supporting a hypothesis of in-situ diversification within these regions. To achieve this, we took our ASTRAL species tree and assigned character states to each species with Mesquite version 3.51 (Maddison and Maddison, 2018). To code multiple biome states for a given species, we added a tip to the tree as a hard polytomy to create two or three tips representing the alternative biome states. Since Neotropical Erythroxylaceae are monophyletic within the paraphyletic Paleotropical lineages, we pruned our Pantropical Erythroxylaceae tree to create a Neotropical tree as well as a Paleotropical tree. We then traced the character history over these trees using an unordered, equal-cost maximum parsimony reconstruction model and utilized the option to summarize state changes over trees on Mesquite to look at the minimum and maximum number of biome state changes over all the most parsimonious character state reconstructions on the ASTRAL species tree topology. Our results on the number of biome shifts is based on the average number of shifts across all most parsimonious trees for the Neotropical clade and the Paleotropical lineages.

We estimated the mean nearest taxon distance (MNTD) and the mean pairwise distance (MPD) of species within each of the three biome types in order to assess the phylogenetic relatedness of these communities; potentially indicating if in-situ diversification or dispersal-based diversification were identifiable patterns. MPD is believed to be more sensitive to phylogeny-wide patterns and MNTD is more sensitive to recent patterns (Webb et al., 2002). To do this, we used our Neotropical and Paleotropical trees derived from the ASTRAL species tree with polytomies added for multiple character states and calculated the MNTD and MPD statistics using the abundance-weighted algorithm in the 'picante' R package (Kembel et al., 2010). We also estimated the standardized effect size of each statistic by randomly shuffling the tips of each phylogeny 10,000 times.

Results:

***Sequencing and phylogenetic inference*—**For the 182 newly-sequenced samples, the median number of genes that had at least 75% target sequence recovery was 520 out of 544 genes. Fourteen samples

recovered <100 genes and these were removed from the analysis. When combined with 76 previously sequenced samples (White, Huang, et al., 2019; White, Islam, et al., 2019), three of which were dropped due to low (<100) gene recovery, we had a total of 241 samples in the final analysis (Figure 6). The total length the 519 gene alignments was 1,397,910 bp. The ASTRAL and MLcat alignments had 13.77% missing data and the MLcatLB alignment had 45.41% missing data.

We sampled 203 of the 283 Erythroxylaceae species (excluding the cultivated cocas; Table II). Of the 29 species from which we sampled multiple individuals, 21 were monophyletic, three were paraphyletic, and five were polyphyletic in the Astral III lineage tree. The two Erythroxylaceae genera from which multiple species were sampled (*Erythroxylum* and *Nectaropetalum*) were inferred to be monophyletic (Appendix C: Figure C2). The earliest diverging lineage in the Erythroxylaceae is *Aneulophus africanus*, followed by the *Nectaropetalum* + *Pinacopodium* clade, followed by a small *Erythroxylum* clade (*E. cambodianum, E. nitidulum, E. seyrigii*). However, we observe very short internodes along the backbone of the tree and some late-diverging lineages did not receive high LPP support. The topological comparison between the ASTRAL tree and MLcat tree (Appendix C: Figure C1), as well as the ASTRAL and MLcatLB tree (Appendix C: Figure C3) showed remarkable congruence of the two tree inference methods.

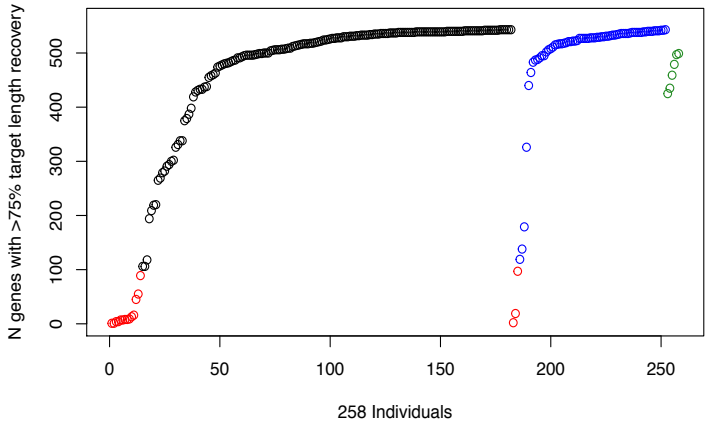


Figure 6: Sequencing and target assembly results showing the number of genes recovered at >75% of the target sequence length for 258 samples. Black circles show new samples from this investigation, blue circles are samples taken from White et al. (2019), and green circles are samples taken from White, Huang, et al., (2019). Samples in red were dropped from the analysis.

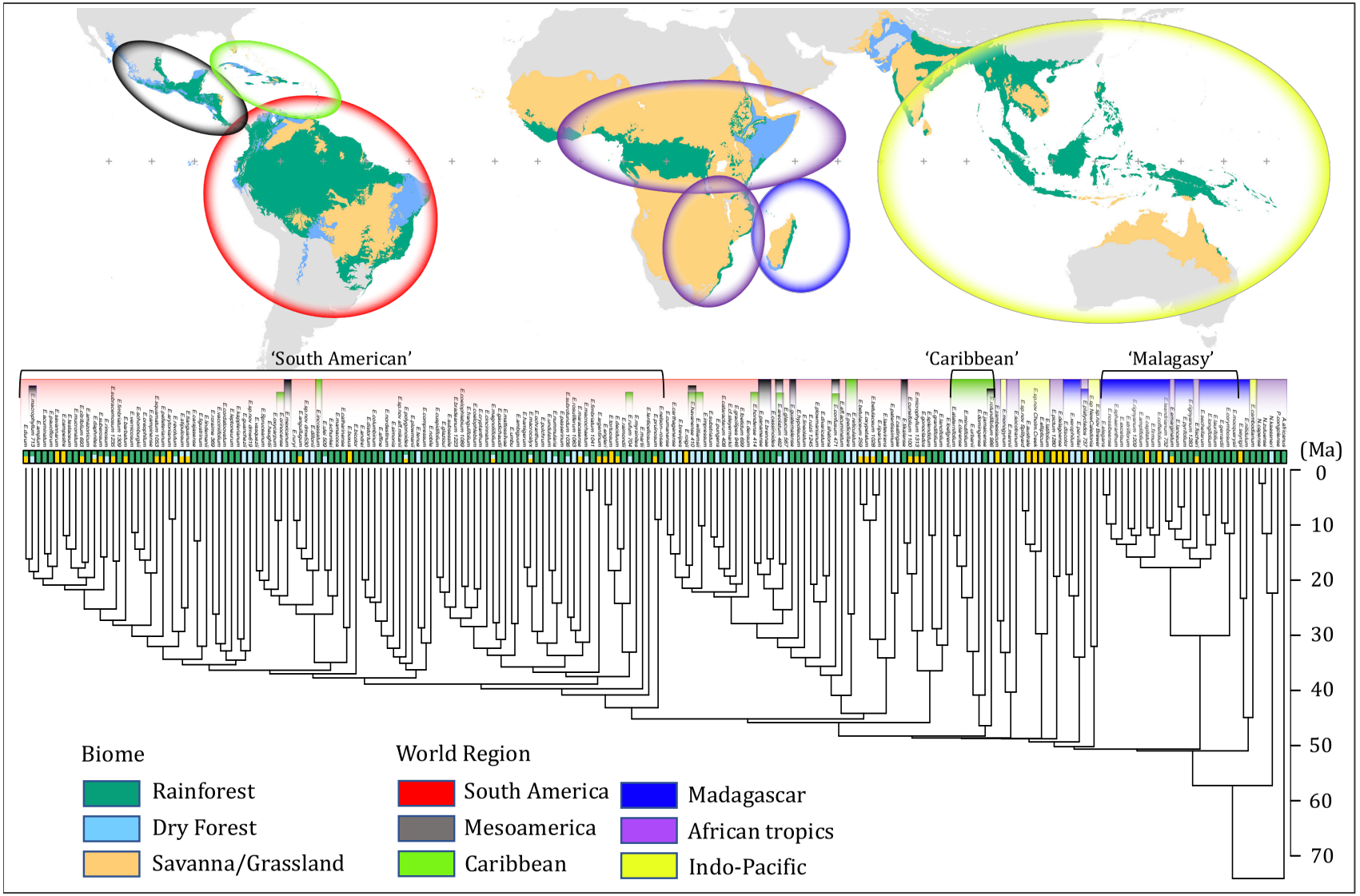
	present	absent	total
NT–Erythroxylum	156	46	202
PT–Erythroxylum	42	28	70
PT–others	5	6	11
total	203	80	283

Table II: Erythroxylaceae species sampled in Chapter 3 phylogenetic and biogeographic analyses compared to the total species diversity. NT=Neotropical, PT=Paleotropical, PT-others= *Aneulophus, Nectaropetalum,* and *Pinacopodium* species. The MLcat tree includes four additional Rhizophoraceae species (total=207 spp.), and the MLcatLB tree for dating and diversification analyses has seven less *Erythroxylum* species (total=200 spp.).

For both comparisons, the top half of the trees have almost perfect congruence, but traversing into the bottom half of the tree, where lower LPP was observed on several branches, there are several clades that share different relationships. Overall, we are confident enough that the degree of congruence of the MLcat and MLcatLB trees with the ASTRAL tree permit using the two trees in our diversification and phylogenetic community composition analyses.

***Biogeography*—**Our biogeographic analysis reveals a paraphyletic relationship of Paleotropical Erythroxylaceae lineages with respect to a single Neotropical clade (Figure 7). *Aneulophus, Nectaropetalum,* and *Pinacopodium* are the earliest-diverging lineages and restricted to rainforests in tropical Africa, with the exception of *Nectaropetalum kaessneri,* which is a dry-forest species. After these older genera diverged, we infer a series of paraphyletic Paleotropical *Erythroxylum* clades are sister to a large, monophyletic Neotropical *Erythroxylum* clade.

Paleotropical *Erythroxylum* biogeography can be broadly described as a series of nested Malagasy clades interspersed with African and Indo-Pacific species and subtended by few other small Indo-Pacific and African clades. The majority of Paleotropical species (21 of 42 sampled Paleotropical species) occur in Madagascar and form three nested clades with the majority (20 species) occurring in a principal ‘Malagasy’ clade that also contains two African species (*E. fischeri* and *E. emarginatum*). Of the



Biome map reprinted with permission from Elsevier.
Original article: Pennington, R.T., C.E.R. Lehmann, and L.M. Rowland. 2018. Tropical savannas and dry forests. Current Biology 28: R527–R548.

Figure 7: The diversification of the Erythroxylaceae in time and space. MLcat chronogram with biome states coded at the tips (rainforest=green, dry-broadleaf forest=light blue, savanna/grassland=gold) corresponding with the global distribution of these three biomes as mapped on the upper half of the figure (biome map from Pennington et al., 2018). Above the species names, world region is color coded and corresponding to the circles on the map (African tropics=purple, Caribbean=chartreuse, Indo-Pacific=yellow, Madagascar=dark blue, Mesoamerica=grey, South America=red).

other two small Malagasy clades, one is younger, and one is older than the main Malagasy clade; the younger comprises four species and contains *E. platyclados*, the only species that occurs in both Africa and Madagascar, and the older clade comprises two species together with a single Indo-Pacific species (*E. cambodianum*). The other eight Indo-Pacific species fall in three other clades, with four Australian species (*E. australe, E. ellipticum, E.sp.nov.Cholm.Creek, E.sp.nov.Splityard*) sister to an Indonesian species (*E. latifolium*). *Erythroxylum monogynum*, from Sri Lanka, is nested within a group of African species. The eight African species are scattered as three singletons and one species pair sister to other geographic regions, and a single three-species clade that is sister to the Neotropical *Erythroxylum*. This clade is represented by dry-forest *E. socotranum* from Socotra (Yemen), *E. mannii* from the central African rainforest, and *E. zambesiacum,* which has a restricted distribution in the savanna of eastern Zimbabwe and southern Zambia.

The earliest-diverging clade of Neotropical *Erythroxylum* is a clade of six endemic Caribbean species plus one species, *E. rotundifolium,* that is distributed in the Caribbean as well as Mesoamerica (Figure 7). Outside of this clade, there are only three other Caribbean endemics distributed between two clades. Two additional species have distributions that span Mesoamerica and the Caribbean, and four species have Caribbean/South American distributions. *Erythroxylum havanense* occurs in dry forests throughout the Caribbean, Mesoamerica, Colombia, and Venezuela. There is also a single record of this plant growing in Florida; to our knowledge, it is the only *Erythroxylum* ever found naturally growing in the United States, resulting from bird or human-mediated dispersal.

Of the 23 Caribbean and Mesoamerican species in our analysis, 18 belong to two of the three oldest Neotropical *Erythroxylum* clades along with 35 South American species. These three clades are paraphyletic with respect to a large ‘South American’ clade with 102 species (Figure 7). The *E. nelson-*

rosae + *E. pruinosum* clade is the earliest diverging lineage in this South American clade and marks the split with the older Caribbean/Mesoamerican/South American section of the tree. Only two of the 102 species in the South American clade also inhabit the Caribbean (*E. rufum* and *E. orinocense*). *Erythroxylum macrophyllum* is a widespread rainforest and dry forest species that also inhabits parts of Mesoamerica. Only two species within the South American clade are endemic to other regions: *E. incrassatum* is from the Caribbean and *E. mexicanum* is from northern Mesoamerica.

***Biome shifts and phylogenetic biome composition*—**The number of species occupying each biome, separated into Neotropical (NT) and Paleotropical (PT) lineages, is presented in Figure 8. In both of the tropical regions, about half of the species are restricted to the rainforest biome (NT=49%, PT=55%). However, 4% of NT species and 23% of PT species are savanna/grassland species, and 24% of NT species and 17% of PT species are dry forest-only species. There are only two PT species in more than one biome; these are *E. fischeri* and *E. emarginatum* from the rainforest and savannas of Africa. There are

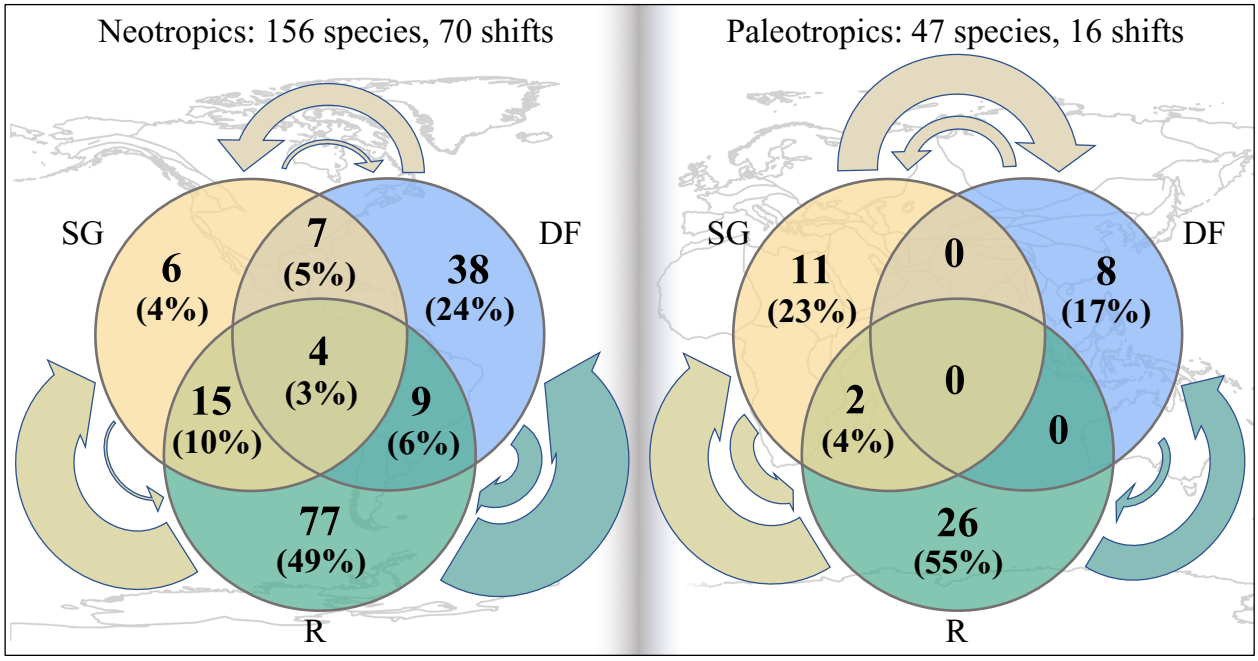


Figure 8: Biome states and shifts in Neotropical (left) and Paleotropical (right) Erythroxylaceae. Numbers inside the Venn diagram show the number and proportion of species inhabiting a biome or multiple biomes. Arrows indicate magnitude and direction of inferred biome shifts during speciation (Tables 1 and 2). Biomes: DF=dry forest, R=rainforest, SG=savanna/grassland.

NT	min	max	avg	pct
R–>DF	24	34	29.08	0.42
R–>SG	22	25	23.5	0.34
DF–>R	4	13	8.08	0.12
DF–>SG	6	9	6.83	0.1
SG–>DF	0	3	1.67	0.02
SG–>R	0	2	0.83	0.01

PT	min	max	avg	pct
R–>SG	5	6	5.5	0.34
SG–>DF	3	5	4	0.25
R–>DF	3	3	3	0.19
SG–>R	1	3	2	0.13
DF–>SG	0	2	1	0.06
DF–>R	0	1	0.5	0.03

Table III (left) showing Neotropical biome shifts and **Table IV** (right) showing Paleotropical biome shifts. Across all most parsimonious trees, the minimum (min), maximum (max), average (avg), and percentage (pct) of directional shifts are provided. Biomes: DF=dry forest, R=rainforest, SG=savanna/grassland.
35 NT species that exist in more than one biome, four of which inhabit all three biome types (*E.*

	MPD - Neotropics				*MPD - Paleotropics*			
	N	**Obs.**	**SES.z**	**SES.p**	**N**	**Obs.**	**SES.z**	**SES.p**
R	103	74.2	–4.02	1e–04	29	76.2	–3.16	0.003
DF	54	78.6	1.97	0.982	13	88.5	1.04	0.863
SG	30	70.9	–2.13	0.023	7	72.9	–0.39	0.295
	MNTD - Neotropics				*MNTD - Paleotropics*			
	N	**Obs.**	**SES.z**	**SES.p**	**N**	**Obs.**	**SES.z**	**SES.p**
R	103	40	0.03	0.512	29	41.5	–0.21	0.411
DF	54	41.6	–1.67	0.051	13	71.8	2.09	0.982
SG	30	43.7	–2.03	0.025	7	48.5	–1.11	0.131

Table V: Mean pairwise distance (MPD) and mean nearest taxon distance (MNTD) statistics for phylogenetic biome composition. Left half shows Neotropical statistics and right half shows Paleotropical statistics, with MPD on the top half and MNTD on the bottom half. Column names are as follows: N=number of species present in biome, Obs.=observed distance, SES.z=standardized effect size z-score, SES.p=standardized effect size p-value. SES values obtained from randomizing tips on phylogenies 10,000 times and recalculating observed MPD and MNTD. Significant SES.p values are highlighted with black boxes. Biomes: DF=dry forest, R=rainforest, SG=savanna/grassland.

daphnites, E. magnoliifolium, E. macrocalyx, and *E. polygonoides*). There are 15 NT species that exist in both rainforest and savanna/grassland, nine in both rainforest and dry forest, and seven in both dry forest and savanna/grassland.

There were 768 most parsimonious state reconstructions explaining the distribution of NT biome states in a total of 70 steps. There were eight most parsimonious trees explaining the distribution of PT biome states in a total of 16 steps. The vast majority of biome shifts originated in the rainforest. In PT,

there were an average of 5.5 shifts (34%) from rainforest to savanna/grassland and 3 shifts (19%) from rainforest to dry forest (Table III). There were 4 shifts (25%) from savanna/grassland to dry forest and 2 shifts (13%) from savanna/grassland to rainforest. Lastly, there was on average 1 shift (6%) inferred from dry forest to savanna/grassland and an average of half a shift (3%) inferred from dry forest to rainforest across all parsimonious trees.

In NT, we inferred an average of 29.1 shifts (42%) from rainforest to dry forest and 23.5 shifts (34%) from rainforest to savanna/grassland. There were 8.1 shifts (12%) from dry forest to rainforest and 6.8 (10%) shifts from dry forest to savanna/grassland. We only inferred 1.67 shifts (2%) from the savanna/grassland to dry forest and 0.8 shifts (1%) from the savanna/grassland to rainforest (Table IV).

With 70 biome shifts inferred across 155 nodes on the NT ASTRAL species tree, this analysis infers that biome shifts were associated with 45% of cladogenic events in Neotropical *Erythroxylum*. Conversely, with 16 biome shifts across 46 nodes on the PT ASTRAL species tree, we infer that biome shifts were associated with 35% of cladogenic events in Paleotropical *Erythroxylaceae*.

We generated two statistics to assess the phylogenetic structure of biome communities in the NT and PT regions: the mean pairwise distance (MPD) statistic is more sensitive to phylogeny-wide patterns in character state distributions and mean nearest taxon distance (MNTD) is more sensitive to patterns near the tips of the tree. Our NT analysis of MPD suggested that rainforests (SES.z=-4.02, p=0.0001) and savanna/grasslands (SES.z=-2.13, p=0.023) are clustered in the phylogeny (alpha=<0.05; Table V). This was corroborated for savanna/grassland by the MNTD score (SES.z=-2.03, p=0.025), but not rainforest (SES.z=0.03, p=0.512). NT dry forests were weakly clustered (SES.z=-1.67, p=0.051) according to the MNTD score but overdispersed according to the MPD score (SES.z=1.97, p=0.982). PT rainforests were significantly clustered by the MPD statistic (SES.z=-3.16, p=0.003), and PT dry forests were significantly overdispersed by the MNTD statistic (SES.z=2.09, p=0.982).

***Tree dating and diversification analysis*—**Our method for correcting terminal branch lengths effectively reduced the number of long branches originally present in the MLcat tree (Figure 9). This also

affected the shape of our time-calibrated trees by reducing the time since divergence for sister species. Both trees, MLcat and MLcatLB, inferred the age of the split between Erythroxylaceae and Rhizophoraceae at 87.6 Ma, the age of the *Aneulophus, Nectaropetalum, Pinacopodium,* and *Erythroxylum* crown group at 73.2 Ma, and age of the *Erythroxylum* crown group at 51 Ma (Appendix C: Figures C4, C5). However, for clades arising after our calibration point at 47.8 Ma, the two trees differ markedly in their age estimates. In the MLcat tree, all major clades originate before the end of the Eocene, whereas in the MLcatLB tree, these clades originate in the Oligocene or Miocene. In the MLcatLB tree, only two cladogenic events happen in the Pliocene, the split of *E. fimbriatum* and *E. macrochaetum*, and the split of *N. zuluense* and *N. capense.*

The BAMM analysis of the MLcatLB tree with priors set to 1 expected rate shift and a minimum clade size of 2 supported a similar diversification rate increase at the *Erythroxylum* crown node followed by a rate slowdown across all lineages except for the Malagasy clade (Figure 10), but less frequently sampled shift configurations showed the *Erythroxylum* rate increase occurring further down the *Erythroxylum* clade (plots 7, 8, & 9, Figure 10). BAMM results from another run with 10 expected shifts provided the same general results but also sampled 3 rate shifts that combined these two previous results, showing rate increases at the crown of *Erythroxylum,* at the stem Malagasy clade, and also at the crown of Neotropical *Erythroxylum* (Appendix C: Figure C7).

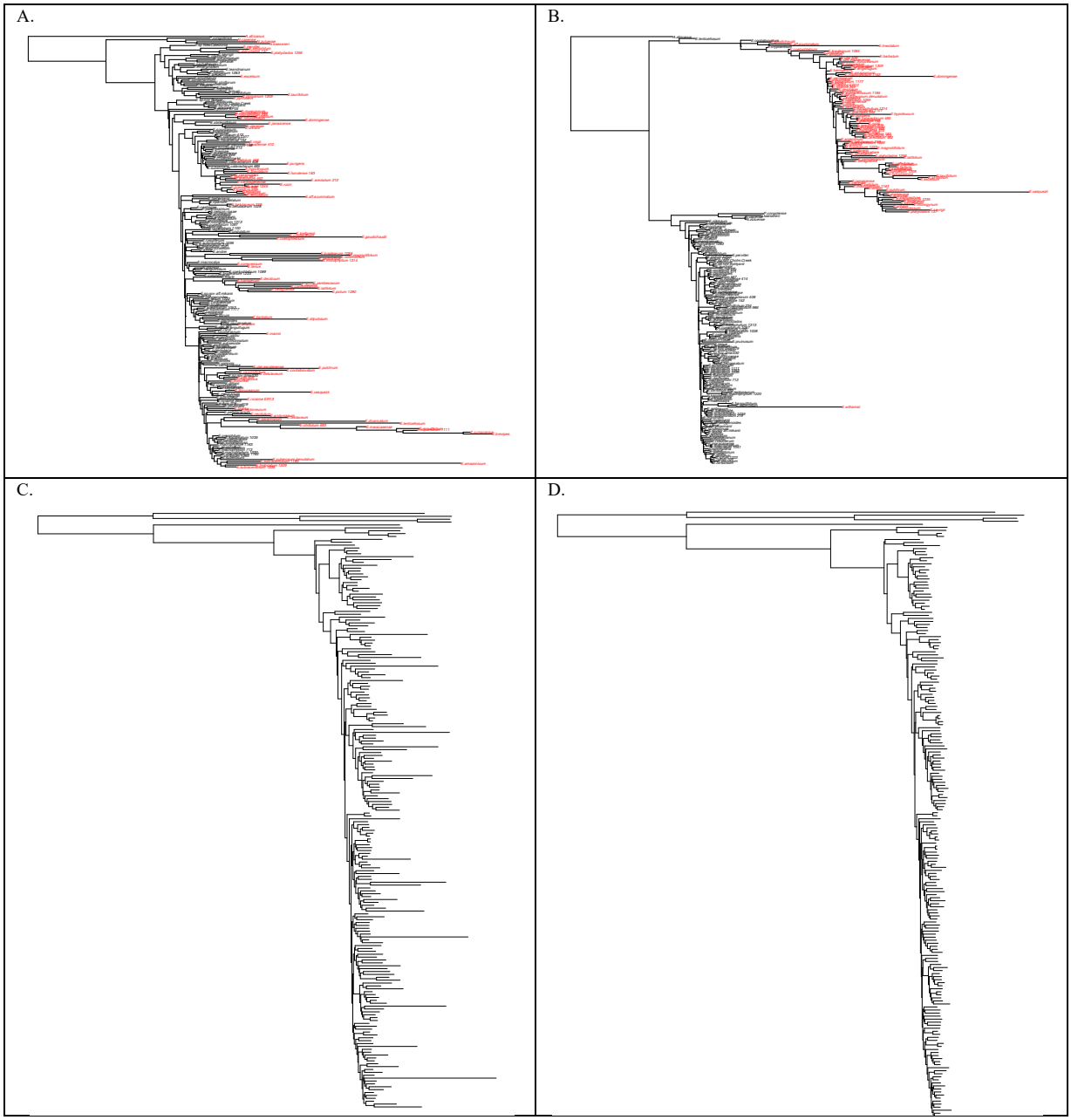


Figure 9: Identification of long-branch taxa and effects of correction pipeline. **(A,B)** Two random gene trees showing the identification of long terminal branch taxa (highlighted in red) that would be subsequently removed from gene alignments via our long-branch correction pipeline. **(A)** Example of gene tree with long terminal branches caused by spurious sequencing and alignment error. **(B)** Example of gene tree with distinct clades possibly caused by paralogs. **(C,D)** Effects of long-branch correction pipeline on concatenated maximum-likelihood tree. **(C)** Maximum-likelihood tree (MLcat) from original concatenated alignment. **(D)** Maximum-likelihood tree from our long-branch-corrected concatenated gene alignment and seven additional long tips removed (MLcatLB). The clade on top with the long branches is the Rhizophoraceae clade.

The BAMM analysis of the MLcat tree estimated the best shift configuration to have two shifts: a diversification rate increase at the *Erythroxylum* crown node and another rate increase (following the slowdown) at the stem of the Malagasy clade (Appendix C: Figure C6, plot 1). The rate increase was inferred along the entire backbone of the *Erythroxylum* clade and followed by a universal rate slowdown for all lineages during the mid-Eocene. However, the Malagasy clade was supported to have either maintained a constant yet lower diversification rate or slowed less quickly. This configuration was corroborated by several other probable shift configurations (Appendix C: Figure C6).

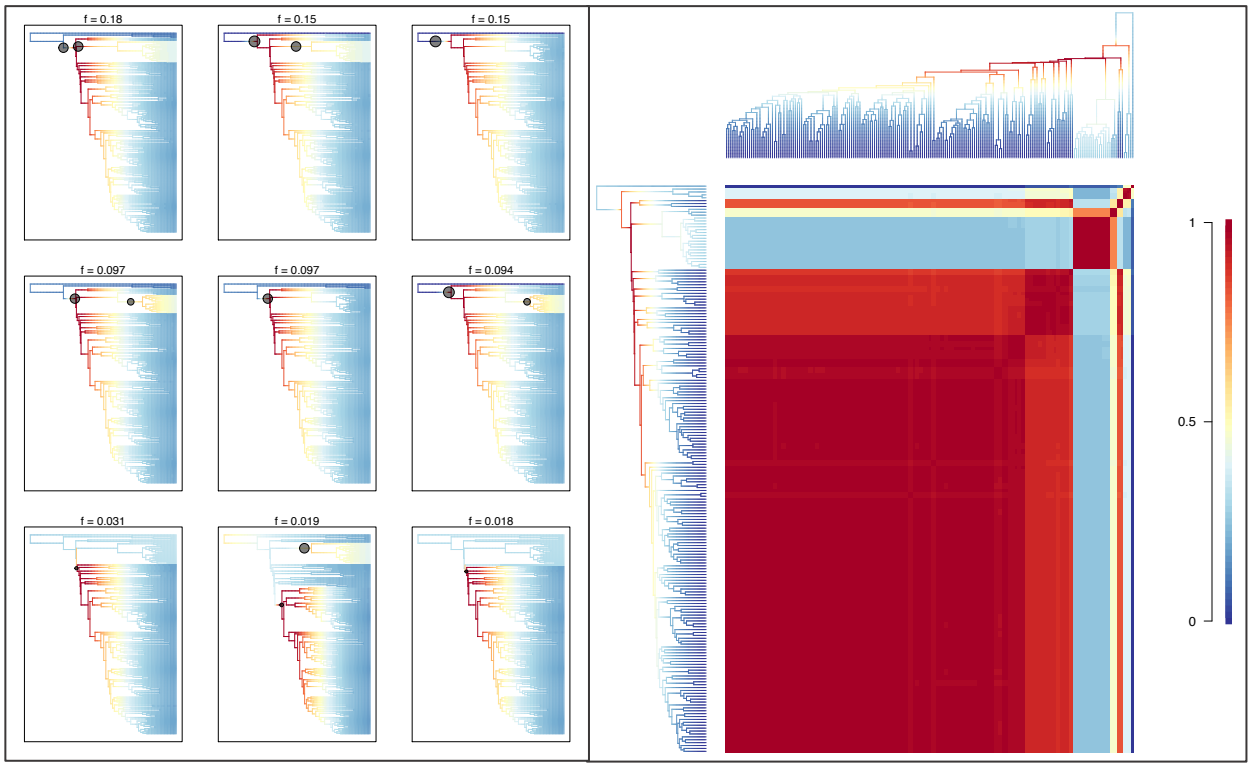


Figure 10: Left – The nine most probable BAMM diversification rate configurations for MLcatLB tree with rate shift locations and magnitudes indicated by circles. F-values indicate the frequency at which the shift configuration was sampled from the posterior. Branch colors indicate the estimated net diversification rate from fast=red to slow=dark blue. **Right –** Cohort diversification rate heatmap showing pairwise comparison of clades to identify the probability from 0 to 1, corresponding to the color on the spectrum on the right side of the heatmap, that any two species share a common macroevolutionary rate regime. The BAMM phylorate plot is presented on the top and left of the heatmap; depicting colors associated with branch diversification rate from fast=red to slow=dark blue.

Discussion:

***Historical Biogeography*—**Our representation of 156 species of Neotropical *Erythroxylum* (77% of total) and 42 Paleotropical *Erythroxylum* (61% of total) provides a fairly randomized sample of the majority (73%) of *Erythroxylum* diversity (Table II). Without performing any ancestral area estimation analysis, the overall shape of the tree, as a series of nested clades, is indicative of probable ancestral geographic areas. The fact that the earliest diverging Erythroxylaceae lineages are the depauperate African genera *Aneulophus, Nectaropetalum,* and *Pinacopodium* suggests that extant Erythroxylaceae lineages share a common African ancestor that existed in the late Cretaceous.

The crown age of *Erythroxylum* is estimated at ~51 Ma and the oldest lineages are represented by species occurring throughout the Paleotropics, but nearly 50% of species diversity is in Madagascar. Our sampling of Paleotropical *Erythroxylum* is least representative of the Indo-Pacific diversity but approximates a fairly random draw, with 12 of 16 (75%) African species, 9 of 20 (45%) Indo-Pacific species, and 21 of 33 (64%) Malagasy species sampled. Based on the distribution of clades and the richness of Malagasy diversity, it is probable that extant *Erythroxylum* species share a common ancestor that existed in Africa or Madagascar, and that this lineage has migrated into the Indo-Pacific region multiple times.

Neotropical species are a single monophyletic group and account for 75% of *Erythroxylum* global diversity. With an estimated Neotropical crown age of 47.8 Ma, our analysis suggests that after *Erythroxylum* originated in Africa or Madagascar, it quickly migrated into the Neotropics. The oldest extant Neotropical *Erythroxylum* lineage comprises a monophyletic group of seven Caribbean species ('Caribbean' clade, Figure 7). Diversifying after this Caribbean clade are several lineages of mixed Caribbean, Mesoamerican, and South American distributions before the diversification of the speciose 'South American' clade. Our divergence time estimation places this migration into the Neotropics between 51 and 47.8 Ma (Appendix C: Figure C5).

These biogeographic patterns, together with the timing of migration, provide compelling support for boreotropical migration wherein *Erythroxylum* dispersed out of Africa, through European and North

American land masses, and into the American tropics (Wolfe, 1975; Tiffney, 1985, as has been hypothesized for numerous other plant lineages (Davis et al., 2002; Bardon et al., 2013; Wei et al., 2015)). While not corresponding exactly with the boundaries of the early Eocene climatic optimum (52.6-50.3 Ma; Payros et al., 2015), this divergence time estimate does correspond to warm temperatures and extensive tropical climates of the early Eocene.

Our results strongly support a diversification rate increase at the *Erythroxylum* crown that was maintained as the lineage migrated into the Neotropics, followed by a global diversification slowdown. While there is not strong evidence of a diversification rate shift in Neotropical *Erythroxylum* versus the Paleotropical lineages, our survey of credible shift configurations and the diversification cohort analysis (Figure 10) shows evidence that the diversification regime in Neotropical *Erythroxylum* is the product of the rate increase across the *Erythroxylum* backbone and is decoupled, though similar to, the diversification regime in Paleotropical *Erythroxylum.*

We also observe a dynamic diversification configuration in the Malagasy clade. This is inferred to be one of the youngest *Erythroxylum* clades (crown age ~29 Ma) and is the only major clade to have diversified primarily in the Miocene (23-5 Ma). When diversification rates in the rest of the Erythroxylaceae lineages appear to be declining, the origin and diversification of this clade in Madagascar represents another pulse of speciation resulting in a significant proportion of Paleotropical species diversity (22 of 42 species, 52%). There is evidence of a subsequent diversification decline in the 'Malagasy' clade as well.

The coupling of migration into a new biogeographic area with a diversification rate increase, followed by a decrease in species-level diversification rates through time is the hallmark of a density-dependent evolutionary radiation (Rabosky and Lovette, 2008). Pulses of diversification, or evolutionary turnover, are believed to be a pervasive pattern in tropical plant diversification (Hoorn et al., 2010; Hughes et al., 2013). This diversification regime is hypothesized to occur when a dispersal event or evolutionary innovation leads to nascent ecological opportunities that accelerate speciation via natural selection or allopatry (Schluter, 2000); yet the confirmation and process of the Erythroxylaceae

diversification as a density-dependent process will require additional investigation. In the case of the Malagasy species, we observe three clades with the majority of species (20/26) occurring in the Malagasy clade. This also suggests a possible evolutionary innovation in the stem of the Malagasy clade due to its higher species diversity compared to the other (younger and older) lineages occupying Madagascar. Future investigations under our phylogenetic framework are warranted to address a central question in the process of ecological radiations: if dispersal and filling of geographic space mediated evolution on ecological axes, or if local ecological diversification enabled geographic colonization.

***Biome evolution*—**Our investigation of biome shifting suggests a combination of dispersal and ecological evolution in the diversification of the Erythroxylaceae. While the primary diversification rate configuration inferred by BAMM was a global diversification rate increase in the backbone of Neotropical + Paleotropical *Erythroxylum*, observation of all credible shifts configurations indicated rate shifts at different places along the backbone, including at the crown of the Neotropical clade (Figure 10). The cohort analysis also showed a difference in diversification regimes among the Neotropical and Paleotropical clades (Figure 10). In addition to describing global patterns in biome shifts during diversification, these results clarify biome diversity and shifting within Neotropical (NT) and Paleotropical (PT) Erythroxylaceae lineages.

The quantification of biome shifts requires ancestral states to be reconstructed under a defined model of character evolution. We chose to trace the history of biome states along our ASTRAL III species tree using maximum parsimony because we were interested in quantifying the minimum number of shifts that would explain observed character states. We believe this method is justified as compared to a more informed model of geographic movement (i.e. DEC, Ree and Smith, 2008) because, instead of trying to model the likely pattern of biome evolution, we are more interested in generating a conservative estimate (i.e. most parsimonious) that is easily repeatable across different datasets to describe the frequency at which lineages evolve and transition between biomes during diversification. The ultimate goal would be to have an index that reflects the taxonomic richness (i.e. the number of species or cladogenic events) with respect to the distribution and phylogenetic diversity of biome character states.

In both hemispheres, about half of all *Erythroxylum* species live in rainforest biomes exclusively. The diversification of the group was accompanied by a large proportion of biome shifts out of the rainforest and into the other biomes, including 42% of shifts in NT from rainforest to dry forest and 34% of shifts from rainforest to savanna/grassland. In PT, most shifts were from rainforest to savanna/grassland (34%), but the next most frequent shift is from savanna/grassland to dry forest (25%), followed by rainforest to dry forest (19%; Tables 1 and 2). These patterns are consistent with our understanding of the latitudinal diversity gradient; a longer time-integrated area and climatic stability of biomes will promote species richness in that region (Fine, 2015). Thus the age and stability of the rainforest biome in both hemispheres has generated and maintained more species than other biomes (Fine and Ree, 2006), and Amazonia has been proven to be a significant source of species diversity dispersing into other biomes (Antonelli, Zizka, et al., 2018). This leads to the expectation that shifts should be more frequent between biomes with longer time-integrated shared perimeters (Donoghue and Edwards, 2014). Across the globe, the savanna/grassland biome is very young (2.5-10 Ma; Pennington and Hughes, 2014), but in Africa and Australia this biome is very large compared to dry forest and rainforest biome areas, possibly explaining the higher frequency of shifts out of the savanna/grasslands in the Paleotropics. Thus, we interpret patterns of biome shifting as follows: In NT and PT, the age and size of the rainforest biome has allowed for more species accumulation, and shared perimeters with savanna/grasslands in both areas has facilitated shifts, more so with the large savanna ecosystems in the Paleotropics (see Pokorny et al., 2015 for a brief summary of African climatic history). The shifts from rainforest to dry forest in the NT could be explained by the shared perimeter between Amazonia and the Caatinga dry forest in South America, whereas the dry forests on the horn of Africa could have had less shared perimeter with the central African rainforests through time. Also, Pliocene and Pleistocene aridification in Africa has resulted in overall instability in climate regimes and reduction of the size of the rainforest, possibly resulting in higher extinction rates in this region.

Another notable observation in biome states is that there are very few species in PT that occur in multiple biome types, only two species (4%) as opposed to 35 (23%) in the NT. Future investigations into

the careful assignment of biome characters, along with the dynamics of biome time-integrated areas and shared perimeters, are warranted in order to explain these patterns.

However, our original hypothesis to explain diversification rate differences between NT and PT was that dispersal-mediated speciation, accompanied by more frequent biome shifts and by differences in phylogenetic biome composition, is the source of higher species richness in the NT. Our analysis inferred that a remarkable 45% of speciation events were associated with biome shifts in the NT, whereas 35% of speciation events were associated with biome shifts in the PT. Though we infer higher rate shifts in the NT, the lower diversity and smaller sample size of PT taxa in our study caution against drawing any conclusions from these results.

Both of these rates are very high, especially when compared to the 'benchmark' 3.6% inferred by Crisp et al. (2009), though their study included shifts out of the tropics, which could require more difficult physiological transitions. Though not presented in the paper itself, our interpretation of the data from Antonelli, Zizka, et al. (2018) is that there was a 47% biome shift rate during speciation events across wet and dry South American forests, which is comparable to our results. Clearly, this field of study will require rigorous comparative analyses before we can understand what the realm of normal is when plant lineages are diversifying and evolving into new biomes.

We tested whether the species occurring within biomes are significantly more related to each other than they are to species selected at random with respect to biome using mean pairwise distance (MPD, more sensitive to phylogeny-wide patterns) and mean nearest taxon distance (MNTD, more sensitive to recent patterns) between all species in each biome in NT and PT regions (Webb et al., 2002). These statistics suggested several remarkable patterns in phylogenetic biome composition. First, rainforest species in NT and PT were significantly clustered by the MPD statistic but randomly distributed by the MNTD statistic. Given about half of Erythroxylaceae species inhabit rainforests (Figure 8), this could be explained by the presence of several large rainforest clades within which several lineages near the tips (causing randomized MNTD statistics) have evolved into different biomes. The one significant and consistent difference between hemispheres is that savanna/grassland species are clustered

in NT but randomly distributed in PT. In the Neotropics, there are only six exclusive savanna/grassland species and 26 species that occupy the savanna/grassland as well as another biome; whereas in the Paleotropics 11 species are savanna/grassland only and two occupy the savannah/grassland and rainforest. In the Neotropics, the clustered phylogenetic community composition as well as high frequency of species occupying more than one biome suggests that a few specific clades might have geographic proximity or be 'preadapted' to evolve into this biome and species within these clades have frequently done so. Although the grassland biome is believed to be very young in both hemispheres (5-10 Ma; Pennington and Hughes, 2014), the relatively larger size of the savanna/grassland biome compared to rainforest in Africa, Australia, and Madagascar might have important effects in generating the 11 Paleotropical savannah/grassland-only species.

Another significant difference between NT and PT is that dry forest species are overdispersed (distributed evenly) in NT according to the MPD statistic and in PT according to the MNTD statistic. However, NT dry forest species are clustered at the tips of the phylogeny, as indicated by the significance of the MNTD statistic. Thus, in NT, nearly all lineages have evolved into the dry forest biome, but among a few lineages, has there been in-situ diversification resulting in small clusters of dry forest species. In the case of PT, however, dry forest species have evolved in a few randomly distributed lineages but dispersal from lineages from other biomes is the dominant pattern in the origin of dry forest species.

Interpretation of the MPD and MNTD statistics are not straightforward, but nonetheless they do reveal an interesting pattern suggesting that savanna/grassland species in the NT are more clustered together in the phylogeny than PT species. Also, the MNTD overdispersion of PT dry forest species in addition to the few shifts from this biome to others suggests that the dry forest biome is a possible evolutionary dead end for Paleotropical species.

***Caveats in the timing and tempo of diversification*—**While we are confident in the biogeographic inference of our analysis, the chronology and tempo of diversification depends on several critical assumptions and characteristics of our sequence data. Upon viewing the MLcat tree, it seemed likely that some of the very long terminal branch lengths were artifactual. Inspection of gene sequence alignments

revealed that sequences from these samples, especially at the gaps and ends of the alignments, contained regions composed of apparently arbitrary bases. Our R code identifies these long branches on individual gene trees, and they were removed from their respective gene alignments before they were reconcatenated for ML inference. Within the Erythroxylaceae, there is an observable reduction in branch lengths from the MLcat tree to the MLcatLB tree for over 60 tips on the phylogeny (Figure 9). However, when we included Rhizophoraceae samples in the long-branch identification procedure, they seemed to swamp the signal of long branches among ingroup taxa. We are pleased with the resulting MLcatLB tree but the masking of Rhizophoraceae taxa in the procedure has one obvious drawback: that the branch lengths in this clade are still spuriously long and this will bias our estimation of divergence dates based on our secondary calibration of the split of the Erythroxylaceae and Rhizophoraceae by pushing the age of the Erythroxylaceae forward in time.

The result of our dating procedure with treePL (Smith and O'Meara, 2012), however, consistently placed the MRCA of *A. africanus* and *C. elliptica* at its maximum allowed date of 87.6 Ma. This behavior can be explained by the distribution of branch lengths on the MLcatLB tree before dating combined with the inclusion of our macrofossil calibration. Two fossil *Erythroxylum* species are described from leaf macrofossils from Miocene formations in Patagonia (Engelhardt, 1905; Berry, 1925). *Erythroxylum* exhibit characteristic venation and vernation patterns that are readily identifiable and we can be confident enough that these fossils are represent valid *Erythroxylum* taxa, yet the reports from other authors (reviewed in Graham, 2010) that extend the age of these taxa into the early Eocene and even Paleocene that must be viewed with skepticism.

The shape of the MLcatLB tree is remarkable based on the short internodes along the backbone (Appendix C: Figure C3) compared to the long branches leading to the *Aneulophus, Nectaropetalum,* and *Pinacopodium* clade (these branches are truncated in Appendix C: Figure C3). Thus, our calibrations appear to be constricting the relatively long branch lengths between the *Erythroxylum* backbone and the split of Erythroxylaceae and Rhizophoraceae such that our calibration points are pushed to their minimum and maximum possible ages, respectively. Another effect of pushing the *Erythroxylum* backbone back to

the early Eocene is that it is stretching the subtending *Erythroxylum* clades back in time and causing a signal of diversification slowdown in the BAMM analysis. This almost universal slowdown is more prominent in the MLcat tree than the MLcatLB tree, as an effect of the even longer terminal branch lengths creating a series of very short internodes along the back of the *Erythroxylum* clade. If we remove the Neotropical *Erythroxylum* crown calibration, then the diversification of Neotropical and Paleotropical lineages is inferred by treePL to have occurred mostly in the Miocene (not shown).

Based on the present literature, this analysis presents our most informed estimate of the timing of diversification. However, the issues with long terminal branches, though rectified to some extent, compounded with the contrasting signals of the tree shape and the calibration points, have produced a very specific chronology of *Erythroxylum* diversification that must be regarded as a hypothesis. Just based on the shape of the tree and a hesitant assumption of a molecular-clock, either the age of the Rhizophoraceae-Erythroxylaceae split is considerably older than we predict here, or the early Eocene *Erythroxylum* macrofossils do not belong to the Neotropical *Erythroxylum* clade.

Explaining global patterns in Erythroxylaceae diversity— This study provides a robust hypothesis of the biogeographic history of the Erythroxylaceae that highlights its boreotropical dispersal from the Paleotropics into the Neotropics. While the causes of diversification rate dynamics are not within the scope of this study, dispersal-mediated ecological evolution seems to be a prominent characteristic of Erythroxylaceae evolution, with a significant number of speciation events being associated with shifts from the rainforests into the grasslands and dry forests. In light of the phylogenetic hypothesis presented here, new investigations into the morphology and evolutionary ecology of the Erythroxylaceae are warranted to understand how these organismal traits have shaped the family's diversification history.

Neotropical *Erythroxylum* comprises 75% of Erythroxylaceae diversity but the clade is younger than all Paleotropical lineages. Despite diversifying into several other genera and giving rise to other *Erythroxylum* lineages that (eventually) dispersed between Africa, Australia, Madagascar, Melanesia, Micronesia, southeast Asia, and Sri Lanka between the crown age of 73 Ma and the origin of Neotropical *Erythroxylum* 51 Ma, these Paleotropical lineages have never accumulated many species and thus have

been marked by low diversification rates or significant recent extinction across many lineages. One exception to this Paleotropical pattern is an *Erythroxylum* lineage that started to diversify in Madagascar and now accounts for about half of the Paleotropical species richness.

The short backbone internodes on our Erythroxylaceae phylogeny suggest several periods of fast diversification early in the family's history (Appendix C: Figure C3). These areas are near the crown of Neotropical *Erythroxylum* and along the backbone of the South American clade.

Around the time of the early Eocene climatic optimum, *Erythroxylum* migrated out of Africa and across a north Atlantic land bridge before descending latitudes into the Neotropics. Our inference of diversification rates suggests that this period of migration was marked by high diversification rates that slowed down as lineages colonized and filled ecological space in the western hemisphere. Though we only infer two shifts towards increased diversification rate (in the Malagasy clade and at the crown of *Erythroxylum*) the fast diversification rate persisted slightly longer in the Neotropical lineage as identified by the difference in diversification regimes in our cohort analysis.

The patterns of biome evolution in the Erythroxylaceae reveal that just less than half of the speciation events are associated with a biome shift. Our analysis suggests that phylogenetic niche conservatism and even biome conservatism are weak in the Erythroxylaceae in both Neotropical and Paleotropical lineages. Instead, dispersal is a key factor in species diversification and lineages are either retaining a broad set of physiological adaptations through speciation events or rapidly adapting to abiotically heterogeneous environments. About half of the species in the Neotropical and Paleotropical regions are exclusively distributed in rainforests, and the majority of biome shifts occur as lineages move from rainforests to dry forests and savanna/grasslands. The Paleotropics harbor a larger proportion of savanna/grassland species and this also contributes to a higher frequency of shifts out of this biome when compared to the Neotropics. Dry forests in the Neotropics are the source of some dispersal-based speciation events into rainforest and savanna/grassland biomes, but while Paleotropical Erythroxylaceae will disperse into the dry forest and speciate, it is very rare that Paleotropical species diversify within this biome or disperse out of it.

Dispersal-mediated speciation is acknowledged as an important driver of tropical plant diversification (Fine and Lohmann, 2018). Boreotropical migration into the Neotropics and dispersal from rainforests into other biomes are prominent characteristics of Erythroxylaceae diversification. However, the scale at which we study dispersal, within or between phytogeographic regions, biomes, or continents, will provide different patterns that must be interpreted as a whole in order to draw major conclusions on the patterns of global biodiversity.

Chapter IV:

A new *Erythroxylum* (Erythroxylaceae) variety from the Sierra Escalera of Peru

Introduction:

The Cordillera Escalera is an Andean "tepui" running parallel to the Andes in northern Peru. Uplift and erosion of the mostly sedimentary deposits that comprise this mountain range has created a tremendous diversity of landscapes that have in turn shaped a variety of vegetation types and plant communities. An estimated 30 new species to have been collected from this region during the Field Museum of Natural History's Rapid Inventory of the Cordillera Escalera in 2013, with several suspected endemics (Pitman et al., 2014).

The Erythroxylaceae Kunth is a pantropical family with most of the diversity belonging to the genus *Erythroxylum* P. Browne (ca. 274 of ca. 285 species) (Daly, 2004; pers. obs.). Although pantropical, the majority of species in this genus are found in the Neotropics, with centers of diversity in eastern and northeastern Brazil and secondarily in the Venezuelan Guyana, and Andes/Amazon region (Plowman and Berry 1999; Plowman and Hensold 2004; Loiola and Costa-Lima 2015). *Erythroxylum* are easily identified in the field as glabrous shrubs or treelets which have alternate, reticulate-veined leaves with intrapetiolar stipules, and small, axillary flowers which have free, appendaged petals and a 3-loculed, superior ovary surrounded by a filament tube (Woodson Jr. et al., 1975; Plowman and Berry, 1999). In this region, the stipule morphology, leaf shape, ovoid fruits, and consistent presence of foliicolous lichens is diagnostic of the closely related *E. ulei* O.E. Schulz, an understory shrub found in wet and deciduous forests in the Andes/Amazon region from Colombia to Bolivia, and *E. schunkei* Plowman a species known from few collections in the lowland forests of Ucayali Department, Peru (Schulz, 1907; Plowman, 1984a).

The current intrageneric classification system in *Erythroxylum* follows Schulz (1907), which defines 19 sections on the basis of morphological and geographic characters. These sections are useful enough for grouping species based on general morphology but several authors have expressed doubt that

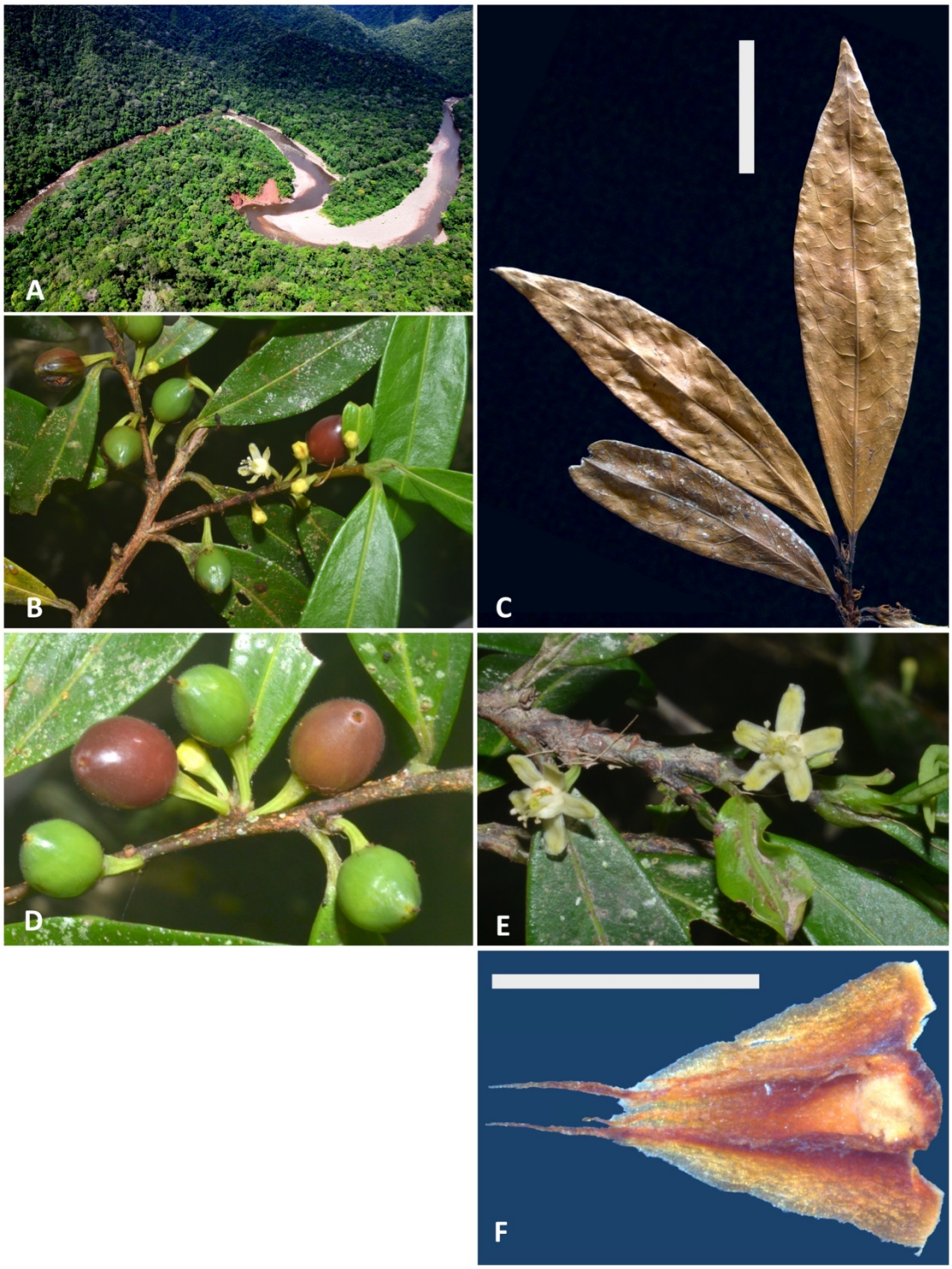


Figure 11. A) Habitat of *E. ulei* var. *escalerense*, Rio Cachiyacu at base of the Cordillera Escalera range, Loreto, Peru. B) Flowering and fruiting branchlet. C) Leaves. Bar = 10 mm. D) Fruit detail, notice apparent pubescence. E) Flower detail. F) Young foliar stipule and setae, mature stipules are larger but setae are fully formed in this picture. Bar = 10 mm. Photo credits: A=Alvaro del Campo; B,D,E = David Neill; C,F=Dawson White.

these sections represent distinct evolutionary lineages, especially the largest sections *Archerythroxylum* and *Rhabdophyllum* (Rury, 1982; Plowman and Rivier, 1983; Emche et al., 2011; Islam, 2011). Thus, phylogenetic inference is key to understanding the interspecific relationships and macroevolutionary patterns in *Erythroxylum*. My dissertation research has focused on phylogenetic reconstruction at multiple taxonomic scales – culminating in a phylogenetic hypothesis that includes the majority of Erythroxylaceae species from across the globe. I describe the morphological differences with *E. ulei, E. ulei* var. *escalerense*, and *E. schunkei* and present a molecular phylogenetic hypothesis of this *Erythroxylum* subclade and discuss its application to Schulz' sections.

Materials & Methods:

***Morphological data*—**As a continuation of the author's dissertation research on the systematics and biogeography of the Erythroxylaceae, *Erythroxylum* specimens have been examined from ANDES, F, LLANOS, MO, MOL, and USM. Morphological measurements were made (Table VI) with a digital caliper, ruler, or eyepiece micrometer.

	E. schunkei	*E. ulei*	*E. ulei v. escalerense*
Leaf shape	lanceolate to oblong elliptic	ovate to broadly elliptic, rarely lanceolate	lanceolate to narrowly elliptic
Leaf size	30-55 mm long, 10-21 mm wide	34-82 mm long, 16.5-39 mm wide	26-53 mm long, 8-14 mm wide
Leaf lamina color	bifacial; dark gray-green and shiny above and ochreous green and dull below	bifacial; gray-green to dark green above and pallid below	green above and light green below
Adaxial midrib	sharply acute in cross section	slightly raised, blunt, or broadly acute in cross section	slightly raised, broadly acute in cross section
Leaf margin	undulate, not revolute	plane, somewhat revolute	plane, revolute
Lateral stipular setae	stout, terete, persistent, 2-5 mm long	slender, flattened, evanescent, 0.5-0.8 mm long	slender, flattened, evanescent, 0.5 mm long
Margin of staminal cup	irregularly 10-crenate	truncate (entire)	truncate (entire)

Table VI: Morphological comparison of *E. schunkei, E. ulei, and E. ulei* var. *escalerense.*

***DNA extraction and phylogenetic analysis*—**Genomic DNA from 4 Rhizophoraceae and 251 Erythroxylaceae samples from herbarium or silica-dried leaf material was extracted using a 2X CTAB protocol with 3% PVP and 2% 2-mercaptoethanol in the extraction buffer (Doyle and Doyle, 1990). Discolored samples were cleaned with the MOBIO Laboratories Inc. DNA Clean-Up Kit. I prepared DNA samples for target capture and sequenced them following the protocols described in Chapter III and utilized the phylogenetic result from Chapter III, which included *E. ulei* var. *escalerense*, in order to draw the conclusions for this new taxon description.

Figure 12: Habit of *E. ulei* var. *escalerense.* Bar = 50 mm.

Taxonomic treatment:

Erythroxylum ulei* var. *escalerense D. M. White, **var. nov.** (Fig. 1,2)

E. ulei var. *escalerense* is similar to *E. ulei* and *E. schunkei* in having stipules two elongated lateral setae, evergreen leaves with foliicolous lichens, and ovoid fruits, but differs by its smaller (26-53 mm long x 8-14 mm wide), narrowly elliptic to lanceolate leaves which are not strongly bifacial and revolute margins (Table VI).

Type:—PERU. **Loreto**: Cordillera Escalera, Provincia Alto Amazonas, Distrito Balsapuerto: Bosque a orillas del Río Cachiyacu, 267 m, 76º36'15.7" W, 5º53'22" S, 18 Sept. 2013 (fl, fr), *Marcos Ríos, Tony Mori, David Neill, Luis Torres, Corine Vriesendorp 3080* (holotype: USM!, isotypes: F!, AMAZ!).

Shrub 1.5 m, *Branchlets* compressed, 1 mm by 1-1.3 mm wide, greenish-brown in young branchlets, drying-reddish brown, lenticels elliptic. *Cataphylls* persistent, occasionally forming ramenta near ends of shoots, similar to stipules, chartaceous. *Foliar stipules* distichous, erect, triangular, 1.2-1.8 mm long, membranous, nonstriate, green, drying light brown, apex obtuse, with 2 erect setae 0.5 mm long extending from two dorsal ridges, third apical seta shorter and usually absent margin entire, scarious in young stipules. *Leaves* persistent, distichous, short-petiolate, laminas narrowly elliptic, 26-53 mm long, 8-14 mm wide, base cuneate, apically acute to obtuse, mucronulate with mucron 0.2-0.5 mm, the upper surface smooth and frequently with lichens, dark green, adaxial midrib light green, sulcate to slightly raised, the lower surface light green, not bilineate or with areole, mixed eucamptodromous venation with 20-40% of secondary nerves forming brochiodromous arches, the secondary nerves 6-9, more distinct on upper surface than lower surface. *Petiole* 2.5-3.2 mm long, adaxially canaliculate, drying dark brown. *Flowers* few, axillary, sequentially one per node. *Bracteoles* 0.9-1.4 mm long, triangular ovate, 1-keeled, apex acute, 1-setulose, seta 0.2-0.4 mm. *Pedicel* 2.8-3.1 mm long in flower, 5 mm long in fruit, diameter increasing in size from 0.4 mm at base to 0.6 mm at tip, 5-ribbed. *Calyx* 1.5 mm long, divided to 2/3 its length, the lobes 1 mm long, triangular to lanceolate, slightly acuminate at apex. *Petal* lamina oblong, slightly concave, ca. 2 mm long (excl. claw), 1 mm wide, the claw ca. 1 mm long, the ligule bilobed, 1.1-1.7 mm long, forming tube around stamens, with labiate anterior auricle at base of lobes. *Staminal tube*

shorter than calyx, ca. 1 mm tall, margin smooth. *Brachystylous flowers:* filaments 2.5-2.8 mm long, the anthers ovate to elliptic, ca. 035 mm long; styles 1-1.2 mm long, free; stigmas capitellate, 0.2-0.3 mm long. *Dolichostylous flowers:* unknown. *Ovary* oblongoid, 1-1.1 mm long, equal or slightly longer than staminal tube. *Drupe* elliptic or slightly ovate, tapering to obtuse apex, purple, the endocarp ovoid-elliptic, semiterete, apex acute and trigonous, 6-7 mm long and 4 mm in diameter, 3-locular, two of the locules empty and almost obsolete, the fertile locule elliptic in cross-section, the endosperm occupying >80% of area.

Distribution:—Cordillera Escalera, Loreto, Peru (Only one locality). The shrub was found along the banks of the Rio Cachiyacu under dense canopy (Fig. 1a).

Etymology: —The varietal epithet is derived from the type locality in the hope that it will draw attention to this understudied region.

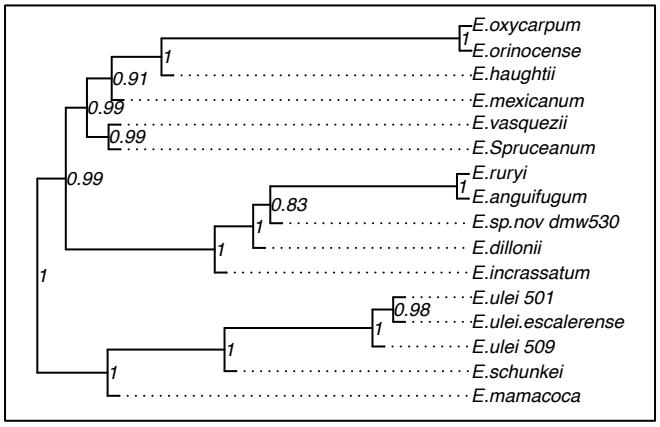


Figure 13: ASTRAL III lineage tree (full 241 samples) from Chapter III, pruned to show only the *E. ulei* clade. Node labels show local posterior probability.

Discussion:

While field biologists and ecologists can frequently identify *Erythroxylum* to the genus level, specific determination in this group is significantly more challenging due to the importance of a handful of vegetative characteristics. Upon viewing photos of *E. ulei* var. *escalerense*, it was hypothesized to be an undescribed species by myself and taxonomic experts Adolfo Jara Muñoz and James da Costa Lima

due to the color, size, and shape of leaves; quite distinct from *E. ulei* or any other Andes/Amazon region *Erythroxylum* (pers. Comm.). However, other vegetative parts, including the highly variable stipules, and flower and fruit morphology are within the variation exhibited by *E. ulei*. The other similar species, *E. schunkei,* a nearby regional endemic from the lowland forests of the Ucayali Department of Peru, is easily distinguished from *E. ulei* by the stout, persistent, and long (2-5 mm) lateral stipular setae. The leaves of *E. schunkei* are more similar to *E. ulei* var. *escalerense* in their size and shape, but still show the prominent bifacial coloring not seen in *E. ulei* var. *escalerense* (Table VI). If more *E. ulei* var. *escalerense* individuals were examined and found to have close morphology to the type specimen, this taxon could warrant elevation to the level of species.

The phylogeny also supports the close relationship of *E. ulei* to *E. ulei* var. *escalerense* with strong support, and *E. schunkei* is sister to this species (Figure 13; Appendix D sample information; Plowman 1984). In his intrageneric classification system, Schulz placed *E. ulei* and four other species into the small section *Leptogramme* O. E. Schulz based on the presence of indistinctly striated stipules (Schulz, 1907; Rury, 1982). Considering the presence or absence of these striations is a very prominent character for several of his other sections, sect. *Leptogramme* may have been created to accommodate the residuals. The phylogeny reveals the other four species in the section (*E. pulchrum, E. passerinum, E. ovalifolium, E. substriatum*) are all distant relatives, distributed throughout the tree (see Appendix C: Figure C2). However, overlooking the striations, Plowman (1984a) places *E. ulei* and *E. schunkei* in sect. *Archerythroxylum* O. E. Schulz, a large section of ca. 65 species characterized by unstriated stipules, perfect flowers, and free styles (Schulz, 1907; Plowman, 1984a). Indeed, the close relatives to the *E. ulei* clade are all *Archerythroxylum* species from the western Amazon and Andean region. These species include *E. anguifugum, E. dillonii, E. sp. nov. dmw530, E. haughtii*, *E. incrassatum* (from Jamaica), *E. mamacoca*, *E. mexicanum, E. orinocense, E. oxycarpum, E. ruryi, E. spruceanum,* and *E. williamsii.* The addition of the molecular phylogeny will be most useful in *Erythroxylum* systematics, particularly toward revision of Schulz' sections.

Lastly, it must be noted that the photograph from the field presented in Figure 11d appears to show the presence of short, velutinous hairs on the fruits. As the Erythroxylaceae have only ever been described as glabrous, this would be a remarkable discovery. I could not find this feature during examination of the exsiccate, so it remains a mysterious observation and one that should be noted if observed on this or any other live *Erythroxylum*.

CITED LITERATURE

Abarca, C.A., A. Martinez-Bauer, F. Molina-Freaner, and C.A. Dominguez. 2008. The genetic consequences of evolving two sexes: the genetic structure of distylous and dioecious species of Erythroxylum. *Evolutionary Ecology Research* 10: 281–293.

Alcantara, S., R.H. Ree, and R. Mello-Silva. 2018. Accelerated diversification and functional trait evolution in Velloziaceae reveal new insights into the origins of the campos rupestres' exceptional floristic richness. *Annals of Botany* 122: 165–180.

Allen, J.M., D.I. Huang, Q.C. Cronk, and K.P. Johnson. 2015. aTRAM - automated target restricted assembly method: a fast method for assembling loci across divergent taxa from next-generation sequencing data. *BMC Bioinformatics* 16: 98.

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. Available at: www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Antonelli, A., M. Ariza, J. Albert, T. Andermann, J. Azevedo, C. Bacon, S. Faurby, et al. 2018. Conceptual and empirical advances in Neotropical biodiversity research. *PeerJ* 6: e5644.

Antonelli, A., and I. Sanmartín. 2011. 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 of the United States of America* 115: 6034–6039.

Arumuganathan, K., and E.D. Earle. 1991. Estimation of nuclear DNA content of plants by flow cytometry. *Plant Molecular Biology Reporter* 9: 229–241.

Avila-Sakar, G., and C.A. Domínguez. 2000. Parental effects and gender specialization in a tropical heterostylous shrub. *Evolution* 54: 866–877.

Bardon, L., J. Chamagne, K.G. Dexter, C.A. Sothers, G.T. Prance, and J. Chave. 2013. Origin and evolution of Chrysobalanaceae: insights into the evolution of plants in the Neotropics. *Botanical Journal of the Linnean Society* 171: 19–37.

Barrie, F.R., and T. Plowman. 2018. Erythroxylaceae. *In* G. Davidse, M. Sousa Sánchez, S. Knapp, and F. Chiang Cabrera [eds.], Flora Mesoamericana, Missouri Botanical Garden, St. Louis.

Baum, D.A. 2009. Species as ranked taxa. *Systematic Biology* 58: 74–86.

Beaumont, M.A., W. Zhang, and D.J. Balding. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162: 2025–2035.

Becker, R.A., R. Brownrigg, T.P. Minka, A.R. Wilks, and A. Deckmyn. 2017. maps: Draw Geographical Maps. Available at: https://cran.r-project.org/package=maps.

Berry, E.W. 1925. A Miocene flora from Patagonia. *Johns Hopkins University Studies in Geology* 6: 6.

Bohm, B.A., F.R. Ganders, and T. Plowman. 1982. Biosystematics and evolution of cultivated coca

(Erythroxylaceae). *Systematic Botany* 7: 121–133.

Boisvert, S., F. Raymond, E. Godzaridis, F. Laviolette, and J. Corbeil. 2012. Ray Meta: scalable de novo metagenome assembly and profiling. *Genome biology* 13: R122.

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

Browne, P. 1756. The civil and natural history of Jamaica. London.

Bushnell, B. 2015. BBMap (version 37.75). *Available at https://sourceforge.net/projects/bbmap/*.

Cai, L., Z. Xi, A.M. Amorim, M. Sugumaran, J.S. Rest, L. Liu, and C.C. Davis. 2019. Widespread ancient whole-genome duplications in Malpighiales coincide with Eocene global climatic upheaval. *New Phytologist* 221: 565–576.

Capella-Gutiérrez, S., J.M. Silla-Martínez, and T. Gabaldón. 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973.

Cavender-Bares, J., D.D. Ackerly, S.E. Hobbie, and P.A. Townsend. 2016. Evolutionary legacy effects on ecosystems: biogeographic origins, plant traits, and implications for management in the era of global change. *Annual Review of Ecology, Evolution, and Systematics* 47: 433–462.

Chung, R.C.K. 1996. Taxonomic notes on Erythroxylum (Erythroxylaceae) in Malesia. *Sandakania* 7: 67–80.

Conzelman, C.S., and D.M. White. 2016. The Botanical Science and Cultural Value of Coca Leaf in South America. *In* P. Gootenberg [ed.], Roadmaps to Regulation: Coca, Cocaine, and Derivatives, The Beckley Foundation, Oxford. Available at: http://beckleyfoundation.org/policy/roadmaps-to-regulation-coca-cocaine-derivatives/the-botanical-science-and-cultural-value-of-coca-leaf-in-south-america/.

Cornuet, J.M., P. Pudlo, J. Veyssier, A. Dehne-Garcia, M. Gautier, R. Leblois, J.M. Marin, and A. Estoup. 2014. DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* 30: 187–1189.

Costa-Lima, J.L., M.I.B. Loiola, and M. Alves. 2015. New synonyms and lectotypifications in Brazilian Erythroxylum (Erythroxylaceae). *Phytotaxa* 1: 100–102.

Crawford, D.J. 2010. Progenitor-derivative species pairs and plant speciation. *Taxon* 59: 1413–1423.

Crisp, M.D., M.T.K. Arroyo, L.G. Cook, M. a Gandolfo, G.J. Jordan, M.S. McGlone, P.H. Weston, et al. 2009. Phylogenetic biome conservatism on a global scale. *Nature* 458: 754–756.

Daly, D.C. 2004. Erythroxylaceae (Coca Family). *In* Flowering Plants of the Neotropics, 143–145. Princeton University Press, Princeton.

Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, et al. 2011. The variant call format and VCFtools. *Bioinformatics*.

Darwin, C. 1877. The different forms of flowers on plants of the same species. Cambridge University

Press, Cambridge, UK.

Davis, C.C., C.D. Bell, S. Mathews, and M.J. Donoghue. 2002. Laurasian migration explains Gondwanan disjunctions: evidence from Malpighiaceae. *Proceedings of the National Academy of Sciences of the United States of America* 99: 6833–6837.

Depristo, M.A., E. Banks, R. Poplin, K. V. Garimella, J.R. Maguire, C. Hartl, A.A. Philippakis, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43: 491–501.

Dillehay, T.D., J. Rossen, D. Ugent, A. Karathanasis, V. Vásquez, and C.P.J. Netherly. 2010. Early Holocene coca chewing in northern Peru. *Antiquity* 84: 939–953.

Donoghue, M.J. 2008. Phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences of the United States of America* 105: 11549–55.

Donoghue, M.J., and E.J. Edwards. 2014. Biome shifts and niche evolution in plants. *Annual Review of Ecology, Evolution, and Systematics* 45: 547–572.

Doyle, J.J., and J.L. Doyle. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12: 13–15.

Edwards, S. V., Z. Xi, A. Janke, B.C. Faircloth, J.E. McCormack, T.C. Glenn, B. Zhong, et al. 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution* 94: 447–462.

Emche, S.D., D. Zhang, M.B. Islam, B. a. Bailey, and L.W. Meinhardt. 2011. AFLP phylogeny of 36 Erythroxylum species. *Tropical Plant Biology* 4: 126–133.

Engelhardt, H. 1905. Bemerkungen zu chilenischen Tertiärpflanzen. *In* Sitzungsberichte und Abhandlungen der Naturwissenschaftlichen Gesellschaft Isis in Dresden, 69–72.

Faircloth, B.C. 2015. PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics* 32: 786–788.

Fine, P.V.A. 2015. Ecological and evolutionary drivers of geographic variation in species diversity. *Annual Review of Ecology, Evolution, and Systematics* 46: 369–392.

Fine, P.V.A., and L.G. Lohmann. 2018. Importance of dispersal in the assembly of the Neotropical biota. *Proceedings of the National Academy of Sciences of the United States of America* 115: 5829–5831.

Fine, P.V.A., and R.H. Ree. 2006. Evidence for a time-integrated species-area effect on the latitudinal gradient in tree diversity. *The American Naturalist* 168: 796–804.

Fine, P.V.A., F. Zapata, and D.C. Daly. 2014. Investigating processes of neotropical rain forest tree diversification by examining the evolution and historical biogeography of the protieae (Burseraceae). *Evolution* 68: 1988–2004.

Forni-Martins, E.R., C.A. Pinto-Maglio, and N.D. da Cruz. 1995. Chromosome numbers in Brazilian cerrado plants. *Revista Brasileira de Genética* 18: 281–288.

Gaut, B.S., C.M. Díez, and P.L. Morrell. 2015. Genomics and the contrasting dynamics of annual and perennial domestication. *Trends in Genetics* 31: 709–719.

GBIF.org. 2017. GBIF Home Page. Available at: http://gbif.org.

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: 557–593.

Glenn, T.C., R. Nilsen, T.J. Kieran, J.W. Finger, T.W. Pierson, K.E. Bentley, S. Hoffberg, et al. 2016. Adapterama I: universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext). Available at: http://biorxiv.org/lookup/doi/10.1101/049114.

Goldberg, E.E., J.R. Kohn, R. Lande, K. a Robertson, S. a Smith, and B. Igić. 2010. Species selection maintains self-incompatibility. *Science* 330: 493–495.

Graham, A. 2010. Late Cretaceous and Cenozoic history of Latin American vegetation and terrestrial environments. Missouri Botanical Garden Press, St. Louis.

Gross, B.L., and K.M. Olsen. 2010. Genetic perspectives on crop domestication. *Trends in Plant Science* 15: 529–537.

Gryj, E.O., and C.A. Domínguez. 1996. Fruit removal and postdispersal survivorship in the tropical dry forest shrub Erythroxylum havanense: ecological and evolutionary implications. *Oecologia* 108: 368–374.

Gutierrez-Noriega, C., and V.W. Von Hagen. 1950. The strange case of the coca leaf. *The Scientific Monthly* 70: 81–89.

Harlan, J.R. 1971. Agricultural origins: centers and noncenters. *Science* 174: 468–474.

Hartwich, C. 1911. Die Menschlichen Genussmittel. Chr. Herm. Tauchnitz, Leipzig.

Harvey, M.G., B.T. Smith, T.C. Glenn, B.C. Faircloth, and R.T. Brumfield. 2016. Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic Biology* 65: 910–924.

Healey, A., A. Furtado, T. Cooper, and R.J. Henry. 2014. Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* 10: 1–8.

Hegnauer, R. 1981. Chemotaxonomy of Erythroxylaceae (including some ethnobotanical notes on old world species). *Journal of ethnopharmacology* 3: 279–292.

Hillebrand, H. 2004. On the generality of the latitudinal diversity gradient. *The American Naturalist* 163: 192–211.

Holmstedt, B., E. Jäätmaa, K. Leander, and T. Plowman. 1977. Determination of cocaine in some South American species of Erythroxylum using mass fragmentography. *Phytochemistry* 16: 1753–1755.

Hoorn, C., F.P. Wesselingh, H. ter Steege, M. a Bermudez, A. Mora, J. Sevink, I. Sanmartín, et al. 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science (New York, N.Y.)* 330: 927–31.

Hughes, C.E., R.T. Pennington, and A. Antonelli. 2013. Neotropical plant evolution: assembling the big picture. *Botanical Journal of the Linnean Society* 171: 1–18.

Islam, M.B. 2011. Tracing the evolutionary history of coca (Erythroxylum). University of Colorado.

Johnson, E.L., W.F. Schmidt, and D. Cooper. 2002. Flavonoids as chemotaxonomic markers for cultivated Amazonian coca. *Plant Physiology and Biochemistry* 40: 89–95.

Johnson, E.L., W.F. Schmidt, and H.A. Norman. 1998. Flavonoids as markers for Erythroxylum taxa: E. coca var. ipadu and E. novogranatense var. truxillense. *Biochemical Systematics and Ecology* 26: 743–759.

Johnson, E.L., D. Zhang, and S.D. Emche. 2005. Inter- and intra-specific variation among five Erythroxylum taxa assessed by AFLP. *Annals of Botany* 95: 601–8.

Johnson, M.G., E.M. Gardner, Y. Liu, R. Medina, B. Goffinet, A.J. Shaw, N.J.C. Zerega, and N.J. Wickett. 2016. HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* 4: 1600016.

Jombart, T., and I. Ahmed. 2011. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* 27: 3070–3071.

Jost, L. 2008. GST and its relatives do not measure differentiation. *Molecular Ecology* 17: 4015–4026.

Kates, H.R., M.G. Johnson, E.M. Gardner, N.J.C. Zerega, and N.J. Wickett. 2018. Allele phasing has minimal impact on phylogenetic reconstruction from targeted nuclear gene sequences in a case study of Artocarpus. *American Journal of Botany* 105: 404–416.

Katoh, K., K. Misawa, K. Kuma, and T. Miyata. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30: 3059–3066.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.

Keenan, K., P. McGinnity, T.F. Cross, W.W. Crozier, and P.A. Prodöhl. 2013. {diveRsity}: An {R} package for the estimation of population genetics parameters and their associated errors. *Methods in Ecology and Evolution* 4: 782–788.

Kembel, S.W., P.D. Cowan, M.R. Helmus, W.K. Cornwell, H. Morlon, D.D. Ackerly, S.P. Blomberg, and C.O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26: 1463–1464.

Kier, G., J. Mutke, E. Dinerstein, T.H. Ricketts, W. Küper, H. Kreft, and W. Barthlott. 2005. Global patterns of plant diversity and floristic knowledge. *Journal of Biogeography* 32: 1107–1116.

Koenen, E.J.M., J.J. Clarkson, T.D. Pennington, and L.W. Chatrou. 2015. Recently evolved diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on the origins of rainforest hyperdiversity. *New Phytologist* 207: 327–339.

Langmead, B., S.L. Salzberg, and Langmead. 2013. Bowtie2. *Nature methods* 9: 357–359.

Lavin, M., and M. Luckow. 1993. Origins and relationships of tropical North America in the context of the boreotropics hypothesis. *American Journal of Botany* 80: 1–14.

Lawson, D.J., G. Hellenthal, S. Myers, and D. Falush. 2012. Inference of population structure using dense haplotype data. *PLoS Genetics* 8: e1002453.

Li, C., M. Hofreiter, N. Straube, S. Corrigan, and G.J.P. Naylor. 2013. Capturing protein-coding genes across highly divergent species. *BioTechniques* 54: 321–326.

Loiola, M.I.B. 2001. Revisão taxonômica de Erythroxylum sect. Rhabdophyllum O.E. Schulz (Erythroxylaceae Kunth). Universidade Federal Rural de Pernambuco.

Loiola, M.I.B., and J.L. Costa-Lima. 2015. Erythroxylaceae. *Lista de Espécies da Flora do Brasil*. Available at: http://floradobrasil.jbrj.gov.br/jabot/FichaPublicaTaxonUC/FichaPublicaTaxonUC.do?id=FB7680.

Loiola, M.I.B., J.L. Costa-Lima, S.J. Mayo, and M. Ferreira De Sales. 2014. Nomenclatural notes on Erythroxylum sect. Rhabdophyllum (Erythroxylaceae): new synonyms and lectotypifications. *Nordic Journal of Botany* 33: .

Löve, A. 1987. IOPB Chromosome Number Reports XCV. *Taxon* 36: 493–498.

Löve, A. 1969. IOPB Chromosome Number Reports XX. *Taxon* 18: 215.

Macbride, J.F. 1949. Erythroxylaceae. *In* Flora of Peru, 632–647. Field Museum of Natural History, Chicago, IL, USA.

Maddison, W.P., and D.R. Maddison. 2018. Mesquite: a modular system for evolutionary analysis.

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

Malinsky, M., E. Trucchi, D.J. Lawson, and D. Falush. 2018. RADpainter and fineRADstructure: Population Inference from RADseq Data. *Molecular Biology and Evolution* 35: 1284–1290.

Mangenot, S., and G. Mangenot. 1958. Deuxieme liste de nombres chromosomiques nouveaux chez diverses Dicotyledones et Monocotyledones d'Afrique occidentale. *Bulletin du Jardin botanique de l'État a Bruxelles* 28: .

Mangenot, S., and G. Mangenot. 1962. Enquete sur les nombres chromosomiques dans une collection d'especes tropicales. *Bulletin de la Société Botanique de France* 25: 411–447.

Matasci, N., L.-H. Hung, Z. Yan, E.J. Carpenter, N.J. Wickett, S. Mirarab, N. Nguyen, et al. 2014. Data access for the 1,000 Plants (1KP) project. *GigaScience* 3: 17.

Matthews, M.L., and P.K. Endress. 2011. Comparative floral structure and systematics in Rhizophoraceae, Erythroxylaceae and the potentially related Ctenolophonaceae, Linaceae, Irvingiaceae and Caryocaraceae (Malpighiales). *Botanical Journal of the Linnean Society* 166: 331–416.

Miller, M.A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *In* 2010 Gateway Computing Environments Workshop, GCE 2010,

Mirarab, S., R. Reaz, M.S. Bayzid, T. Zimmermann, M.S. Swenson, and T. Warnow. 2014. ASTRAL:

genome-scale coalescent-based species tree estimation. *Bioinformatics (Oxford, England)* 30: i541-8.

Mittelbach, G.G., D.W. Schemske, H. V Cornell, A.P. Allen, J.M. Brown, M.B. Bush, S.P. Harrison, et al. 2007. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters* 10: 315–31.

Mortimer, W.G. 1901. History of coca: the divine plant of the Incas. Vail, New York.

Olson, D.M., E. Dinerstein, E.D. Wikramanayake, N.D. Burgess, G.V.N. Powell, E.C. Underwood, J.A. D'amico, et al. 2001. Terrestrial ecoregions of the world: a new map of life on Earth. *BioScience* 51: 933.

Oviedo, R.P. 2002. Contribución al conocimiento de Erythroxylaceae Kunth en Cuba y las Antillas. Universidad de la Habana.

Payros, A., S. Ortiz, I. Millán, J. Arostegi, X. Orue-Etxebarria, and E. Apellaniz. 2015. Early Eocene climatic optimum: Environmental impact on the North Iberian continental margin. *Bulletin of the Geological Society of America* 127: 1632–1644.

Pennington, R.T., and C.E. Hughes. 2014. The remarkable congruence of New and Old World savanna origins. *New Phytologist* 204: 4–6.

Pennington, R.T., C.E.R. Lehmann, and L.M. Rowland. 2018. Tropical savannas and dry forests. *Current Biology* 28: R527–R548.

Perl, R.F. 1992. United States Andean drug policy: background and issues for decisionmakers. *Journal of Interamerican Studies and World Affairs* 34: 13–35.

Pitman, N., C. Vriesendorp, D. Alvira, J.A. Markel, M. Johnston, E. Ruelas inzunza, A. Lancha Pizango, et al. eds. . 2014. Perú: Cordillera Escalera-Loreto. *In* Rapid Biological and Social Inventories Report 26, The Field Museum, Chicago.

Plowman, T. 1981. Amazonian coca. *Journal of Ethnopharmacology* 3: 195–225.

Plowman, T. 1979. Botanical perspectives on coca. *Journal of Psychedelic Drugs* 11: 103–117.

Plowman, T. 1986. Coca chewing and the botanical origins of coca (Erythroxylum spp.) in Latin America. *In* D. Pacini, and C. Franquemont [eds.], Coca and Cocaine: Effects on People and Policy in Latin America, 5–34. Cultural Survival, Inc./LASP, Cornell University.

Plowman, T. 1984a. New taxa of Erythroxylum (Erythroxylaceae) from the Amazon basin. *Acta Amazonica* 14: 117–143.

Plowman, T. 1987. Ten new species of Erythroxylum (Erythroxylaceae) from Bahia, Brazil. *Fieldiana* Bot. no. 1: 1–41.

Plowman, T. 1984b. The ethnobotany of coca (Erythroxylum spp., Erythroxylaceae). *Advances in Economic Botany* 1: 62–111.

Plowman, T., and P.E. Berry. 1999. Erythroxylaceae. *In* J. A. Steyermark, P. E. Berry, B. Holst, and K. Yatskievych [eds.], Flora of the Venezuelan Guyana, 59–71. Missouri Botanical Garden Press, St.

Louis.

Plowman, T., and N. Hensold. 2004. Names, types, and distribution of neotropical species of Erythroxylum (Erythroxylaceae). *Brittonia* 56: 1–53.

Plowman, T., and L. Rivier. 1983. Cocaine and cinnamoylcocaine content of Erythroxylum species. *Annals of Botany* 51: 641–659.

Plowman, T., L. Rudenberg, and C.W. Greene. 1978. Chromosome Numbers in Neotropical Erythroxylum (Erythroxylaceae). *Botanical Museum Leaflets, Harvard University* 26: 203–209.

Pokorny, L., R. Riina, M. Mairal, A.S. Meseguer, V. Culshaw, J. Cendoya, M. Serrano, et al. 2015. Living on the edge: timing of Rand Flora disjunctions congruent with ongoing aridification in Africa. *Frontiers in Genetics* 6: 1–15.

R Development Core Team. 2013. R: A language and environment for statistical computing.

Rabosky, D.L. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PloS one* 9: e89543.

Rabosky, D.L., M. Grundler, C. Anderson, P. Title, J.J. Shi, J.W. Brown, H. Huang, and J.G. Larson. 2014. BAMMtools: An R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution* 5: 701–707.

Rabosky, D.L., and I.J. Lovette. 2008. Density-dependent diversification in North American wood warblers. *Proceedings of the Royal Society B* 275: 2363–2371.

Ree, R.H., and S.A. Smith. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* 57: 4–14.

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

Rury, P.M. 1981. Systematic anatomy of Erythroxylum P. Browne: practical and evolutionary implications for the cultivated cocas. *Journal of Ethnopharmacology* 3: 229–263.

Rury, P.M. 1982. Systematic anatomy of the Erythroxylaceae. University of North Carolina.

Rusby, H.H. 1933. Jungle memories. McGraw-Hill, New York.

Schluter. 2000. The ecology of adaptive radiation. Oxford University Press, Oxford.

Schultes, R.E. 1981. Coca in the northwest Amazon. *Journal of Ethnopharmacology* 3: 173–194.

Schultes, R.E. 1979. Evolution of the identification of the major South American narcotic plants. *Journal of Psychedelic Drugs* 11: 119–134.

Schultes, R.E., T. Plowman, and C. Sheviak. 1976. Semi-annual reports under USDA Contract No. 12-14-1001-230.

Schulz, O.E. 1907. Erythroxylaceae. *In* A. Engler [ed.], Das Pflanzenreich, 1–164.

Schulz, O.E. 1931. Erythroxylaceae. *In* A. Engler, and K. Prantl [eds.], Die natürlichen Pflanzenfamilien, 227–658. Engelmann, Leipzig.

Shi, J.J., and D.L. Rabosky. 2015. Speciation dynamics during the global radiation of extant bats. *Evolution* 69: 1528–1545.

Slik, J.W.F., V. Arroyo-Rodríguez, S.-I. Aiba, P. Alvarez-Loayza, L.F. Alves, P. Ashton, P. Balvanera, et al. 2015. An estimate of the number of tropical tree species. *Proceedings of the National Academy of Sciences of the United States of America* 112: 7472–7477.

Slik, J.W.F., J. Franklin, V. Arroyo-Rodríguez, R. Field, S. Aguilar, N. Aguirre, J. Ahumada, et al. 2018. Phylogenetic classification of the world's tropical forests. *Proceedings of the National Academy of Sciences of the United States of America* 115: 1837–1842.

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

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

ter Steege, H., N.C.A. Pitman, D. Sabatier, C. Baraloto, R.P. Salomão, J.E. Guevara, O.L. Phillips, et al. 2013. Hyperdominance in the Amazonian tree flora. *Science* 342: 1243092.

Ter Steege, H., R.W. Vaessen, D. Cárdenas-López, D. Sabatier, A. Antonelli, S.M. De Oliveira, N.C.A. Pitman, et al. 2016. The discovery of the Amazonian tree flora with an updated checklist of all known tree taxa. *Scientific Reports* 6: 29549.

Struck, T.H. 2014. Trespex-detection of misleading signal in phylogenetic reconstructions based on tree information. *Evolutionary Bioinformatics* 31: 51–67.

Swofford, D.L. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4.

Tavaré, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences* 7: 57–86.

Tiffney, B.H. 1985. The Eocene North Atlantic Land Bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. *Journal of the Arnold Arboretum* 66: 243–273.

Vavilov, N.I., and D. Löve. 2009. Origin and geography of cultivated plants. Cambridge University Press.

Webb, C.O., D.D. Ackerly, M.A. McPeek, and M.J. Donoghue. 2002. Phylogenies and community ecology. *Annual Review of Ecology and Systematics* 33: 475–505.

Weeks, A., F. Zapata, S.K. Pell, D.C. Daly, J.D. Mitchell, and P.V.A. Fine. 2014. To move or to evolve: contrasting patterns of intercontinental connectivity and climatic niche evolution in "Terebinthaceae" (Anacardiaceae and Burseraceae). *Frontiers in Genetics* 5: .

Wei, R., Q. Xiang, H. Schneider, M.A. Sundue, M. Kessler, P.W. Kamau, A. Hidayat, and X. Zhang. 2015. Eurasian origin, boreotropical migration and transoceanic dispersal in the pantropical fern genus Diplazium (Athyriaceae). *Journal of Biogeography* 42: 1809–1819.

Weir, B.S., and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure.

Evolution 38: 1358–1370.

Weitemier, K., S.C.K. Straub, R.C. Cronn, M. Fishbein, A. McDonnell, and A. Liston. 2014. Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 2: 3–9.

White, D.M., J.P. Huang, and R.J. Mason-Gamer. 2019. Multiple origins of the coca plant in South America. *In prep.*

White, D.M., M.B. Islam, and R.J. Mason-Gamer. 2019. Phylogenetic inference in section Archerythroxylum informs taxonomy, biogeography, and the domestication of coca (Erythroxylum species). *American Journal of Botany* 106: 1–12.

Wickett, N.J., S. Mirarab, N. Nguyen, T. Warnow, E. Carpenter, N. Matasci, S. Ayyampalayam, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences of the United States of America* 111: E4859–E4868.

Wiens, J.J. 2004. Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. *Evolution* 58: 193–197. Available at: http://doi.wiley.com/10.1111/j.0014-3820.2004.tb01586.x.

Wolfe, J.A. 1975. Some aspects of plant geography of the northern hemisphere during the late Cretaceous and Tertiary. *Annals of the Missouri Botanical Garden* 62: 264–279.

Woodson Jr., R.E., R.W. Shery, W.G. D'Arcy, and N. Schanen. 1975. Flora of Panama. Part IV. Family 87. Erythroxylaceae. *Annals of the Missouri Botanical Garden* 62: 21–33.

Xi, Z., L. Liu, and C.C. Davis. 2016. The impact of missing data on species tree estimation. *Molecular Biology and Evolution*.

Xi, Z., B.R. Ruhfel, H. Schaefer, A.M. Amorim, M. Sugumaran, K.J. Wurdack, P.K. Endress, et al. 2012. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proceedings of the National Academy of Sciences of the United States of America* 109: 17519–24.

Yang, Z. 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Molecular Biology and Evolution* 10: 1396–401.

Zerbino, D.R., and E. Birney. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* 18: 821–829.

Zhang, J., K. Kobert, T. Flouri, and A. Stamatakis. 2014. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30: 614–620.

Appendix A: Additional figures and sample information for Chapter I

Figures:

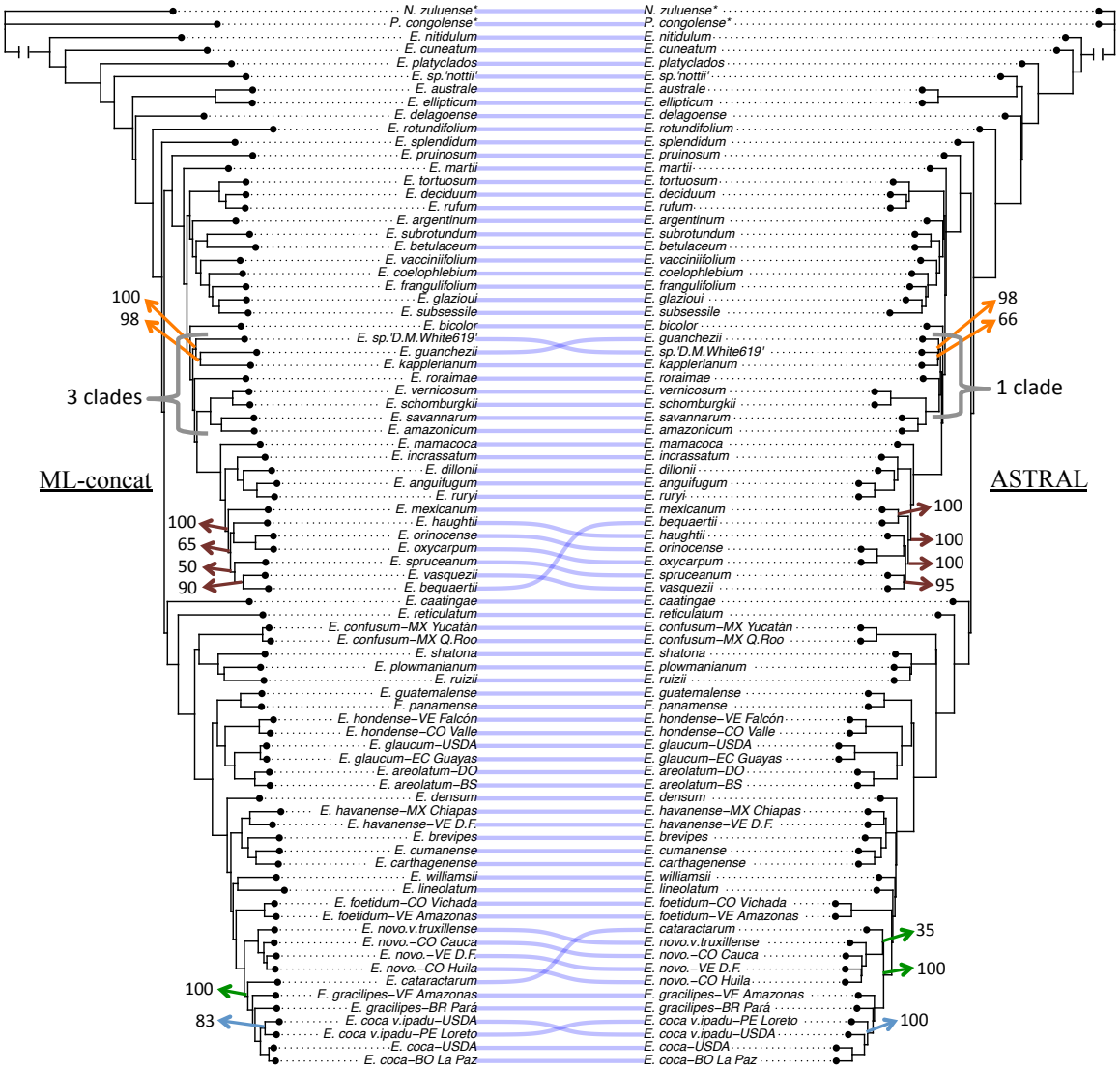


Figure A1: Comparative topologies of ML-Concat phylogeny (left) and ASTRAL II species tree (right). Incongruent clades have bootstrap support values (ML-Concat) or local posterior probability (ASTRAL) labeled.

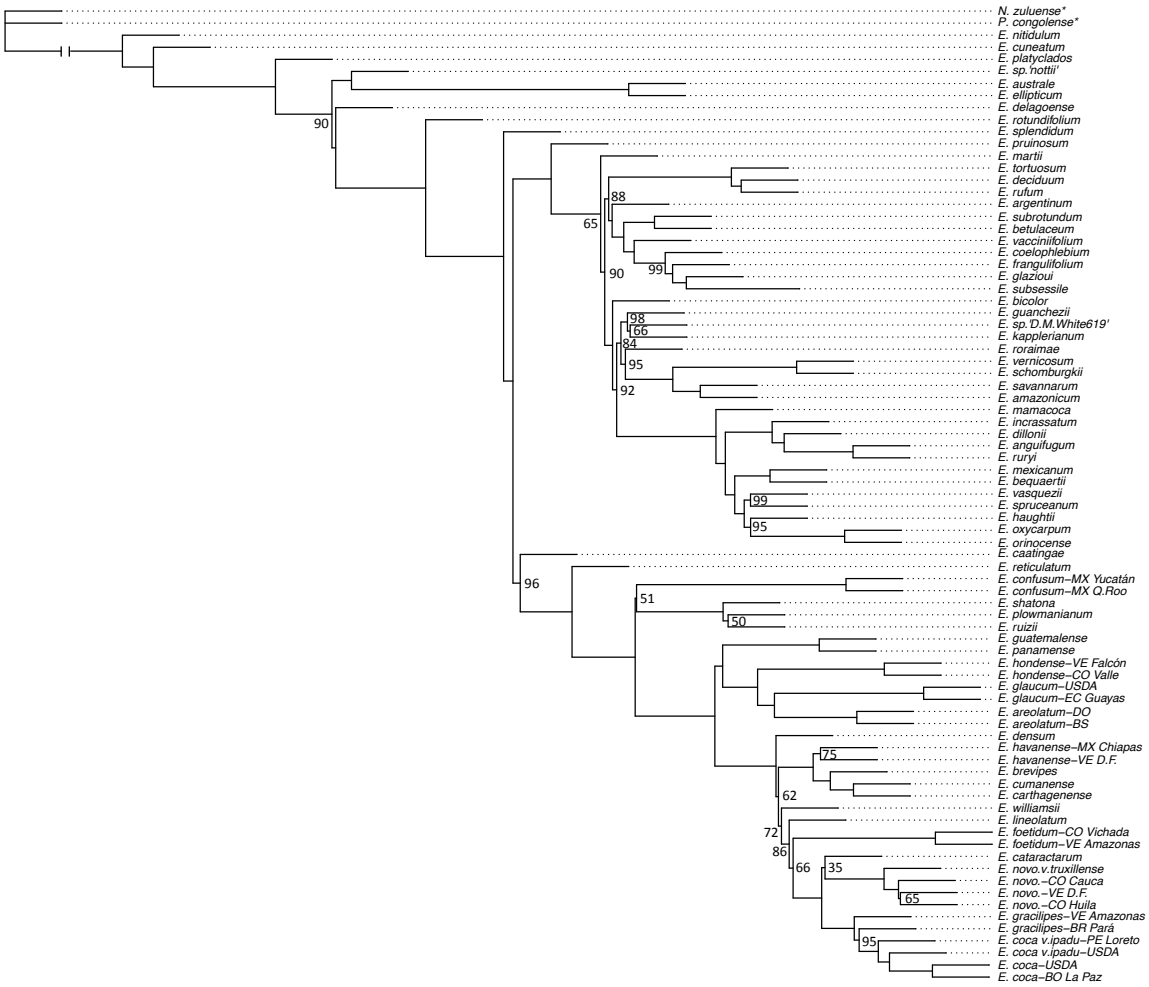


Figure A2: ASTRAL II lineage tree. Local posterior probability values <100 are labeled.

Chapter I sample information:

sample	Schulz section	Region (AF=Atlantic Forest, AM=Amazonia, AN=Andes, Ca=Caribbean, Ch=Chocó, Cr=Cerrado, Ct=Caatinga, EC= Ecuatorial Dry Forests, GS= Guiana Shield, MA= Mesoamerica, Or=Orinoquia)	Herbarium: number	collector:number
E. amazonicum	Rhabdophyllum	Am,An,Ch,GS	F:2198813	Vasquez:24660
E. anguifugum	Archerythroxylum	Am,Cr	F:2179162	Schinini:31708
E. areolatum-BS	Erythroxylum	MA,Ca	F:1758721	Correll:45197
E. areolatum-DO	Erythroxylum	MA,Ca	F:2171213	Jimenez:2004
E. argentinum	Archerythroxylum	An,Cr	F:1757134	Amaral Jr.:969
E. australe	Coleocarpus	Australia	BRI:*	Clarkson:11785
E. bequaertii	Erythroxylum	MA	F:2135738	Vincent:6118
E. betulaceum	Archerythroxylum	Ct	F:1916628	Plowman:12715
E. bicolor	Archerythroxylum	AF	F:1955248	Furlan:6441
E. brevipes	Archerythroxylum	Ca	F:2171167	Garcia:6024
E. caatingae	Archerythroxylum	Ct	F:2116327	Carvalho:3837
E. carthagenense	Archerythroxylum	Ca	F:1938825	Gentry47468
E. cataractarum	Archerythroxylum	Am,Ch,Or	F:2198726	Callejas:4371
E. coca-BO La Paz	Archerythroxylum	An	F:2138572	Marko Lewis:36915
E. coca-USDA	Archerythroxylum	An	USDA**:B145	-:-
E. coca v. ipadu-PE Loreto	Archerythroxylum	Am	F:1823932	Plowman:6923
E. coca v. ipadu-USDA	Archerythroxylum	Am	USDA**:B503	-:-
E. coelophlebium	Archerythroxylum	AF	F:1922037	Plowman:12900
E. confusum-MX Yucatán	Erythroxylum	MA,Ca	MO:2285708	Tapia:1891
E. confusum-MX Q. Roo	Erythroxylum	MA,Ca	F:1951069	Olmsted:s.n.
E. cumanense	Archerythroxylum	Ca	F:1853471	Plowman:7654
E. cuneatum	Coleocarpus	IndoPacific	F:2087947	Phillipson:2878
E. deciduum	Rhabdophyllum	AF,An,Cr	F:1744854	Amaral Jr.:11481
E. delagoense	Coleocarpus	Africa	MO:2449460	Kemp:532
E. densum	Archerythroxylum	Ca	F:2324373	White:543
E. dillonii	Archerythroxylum	An	F:2320287	White:522
E. ellipticum	Coleocarpus	Australia	DNA: D0185653	Westaway:2439
E. foetidum-VE Amazonas	Archerythroxylum	Or	F:1987304	Guanchez & Urbina:1772
E. foetidum-CO Vichada	Archerythroxylum	Or	F:2324334	White:605
E. frangulifolium	Archerythroxylum	AF	F:1916625	Plowman:12860
E. glaucum-EC Guayas	Archerythroxylum	EC	F:1900484	Dodson:11369
E. glaucum-USDA	Archerythroxylum	EC	USDA**: FOX221	-:-
E. glazioui	Archerythroxylum	AF	F:1976059	Martinelli:11658
E. gracilipes-BR Pará	Archerythroxylum	Am,Ch,GS	F:2198679	Beck:427
E. gracilipes-VE Amazonas	Archerythroxylum	Am,Ch,GS	F:1982323	Holst & Liesner:3146
E. guanchezii	Archerythroxylum	Or,GS	F:2324335	White:618
E. guatemalense	Archerythroxylum	MA	F:1967917	Lundell:19312
E. haughtii	Archerythroxylum	An	F:1769140	Plowman:5360
E. havanense-MX Chiapas	Archerythroxylum	MA,Ca,GS,Or	F:1921838	Breedlove:50501
E. havanense-VE D.F.	Archerythroxylum	MA,Ca,GS,Or	F:1930162	Berry:3651
E. hondense-VE Falcón	Archerythroxylum	An,Ca	F:1902194	Berry:3921

E. hondense-CO Valle	Archerythroxylum	An,Ca	F:1986037	Silverstone-Sopkin:3097
E. incrassatum	Archerythroxylum	Ca	KHD:63027	Islam:09-02
E. kapplerianum	Archerythroxylum	Am,GS	F:1924745	Strudwick:4084
E. lineolatum	Archerythroxylum	An,GS	F:2241987	Evans:2714
E. mamacoca	Archerythroxylum	An	F:1897281	Plowman:11740
E. martii	Archerythroxylum	AF	F:1960559	dos Santos:4012
E. mexicanum	Archerythroxylum	MA	F:1992215	Plowman:14546
E. nitidulum	Schistophyllum	Africa	F:2116948	Zarucchi:7425
E. novo. v. truxillense	Archerythroxylum	An	F:1873746	Plowman:5600
E. novo.-VE D.F.	Archerythroxylum	An	F:1853454	Plowman:7670
E. novo.-CO Cauca	Archerythroxylum	An	F:1898296	Plowman:10980
E. novo.-CO Huila	Archerythroxylum	An	F:1746941	Plowman:4152
E. orinocense	Archerythroxylum	Or	ANDES:8903	White:613
E. oxycarpum	Archerythroxylum	Ca,Or	F:2324360	White:563
E. panamense	Microphyllum	MA,Ch	F:1962829	Schatz:1166
E. platyclados	Venelia	Africa	MO:4249313	Robertson:5409
E. plowmanianum	Rhabdophyllum	An	ANDES:8873	White:582
E. pruinosum	Archerythroxylum	Cr	F:2199294	Santos:62
E. reticulatum	Archerythroxylum	Ca	F:1774549	Correll:49694
E. roraime	Archerythroxylum	GS	MO:5072362	Chacon:681
E. rotundifolium	Erythroxylum	MA,Ca	F:1935681	Lott:1738
E. rufum	Rhabdophyllum	Am,An,Ca,GS	USDA**: B453SS	-:-
E. ruizii	Archerythroxylum	EC	F:1973406	Plowman:14342
E. ruryi	Archerythroxylum	Am	F:2320256	White:479
E. savannarum	Macrocalyx	Or	F:2324371	White:555
E. schomburgkii	Archerythroxylum	GS	F:2076646	Maas:7181
E. shatona	Archerythroxylum	An	F:1774523	Martin:1851
E. sp. 3248	Coleocarpus	Australia	BRI:*	Clarkson:11733
E. sp. 'D.M.White619'	Archerythroxylum	Or	F:2324377	White:619
E. splendidum	Megalophyllum	AF	F:1910746	Carvalho:1125
E. spruceanum	Archerythroxylum	Am	F:1944256	Zarucchi:3096
E. subrotundum	Archerythroxylum	Cr	F:2283655	Paredes:91
E. subsessile	Archerythroxylum	AF	F:1982136	Fialho:3
E. tortuosum	Archerythroxylum	Cr	F:2302271	Mendes:292
E. vacciniifolium	Archerythroxylum	AF	F:2188284	Mendoça:50
E. vasquezii	Archerythroxylum	Am	F:1994766	Spichiger:1983
E. vernicosum	Archerythroxylum	GS	F:1958896	Jensen-Jacobs:30480
E. williamsii	Archerythroxylum	Ca,Or	F:1933195	Plowman:13495
N. zuluense	Erythroxylaceae	Africa	F:1977936	Ansell:s.n.
P. congolense	Erythroxylaceae	Africa	MO: 4238697	McPherson:15533

Appendix B: Additional figures and sample information for Chapter II

Figures:

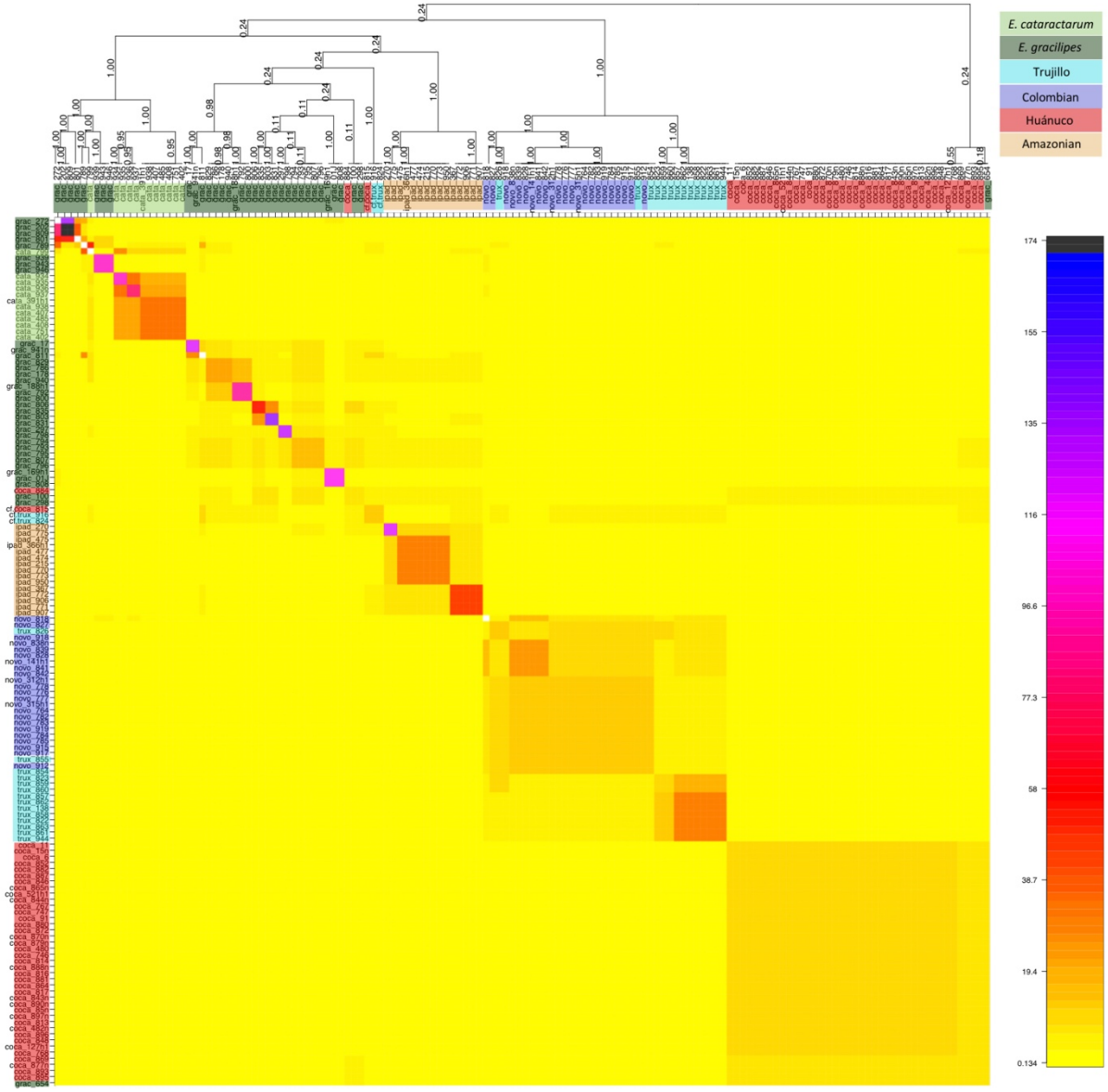


Figure B1: Heatmap of RADpainter coancestry index values. Results of RADpainter analysis showing population-averaged haplotype (SNP) sharing due to coancestry. Sample names are color-coded according to taxonomic variety, as identified in the legend on the top right corner. The top-right diagonal of the heatmap shows the coancestry index values from the unlinked SNPs and the bottom-left diagonal shows coancestry index values with linkage.

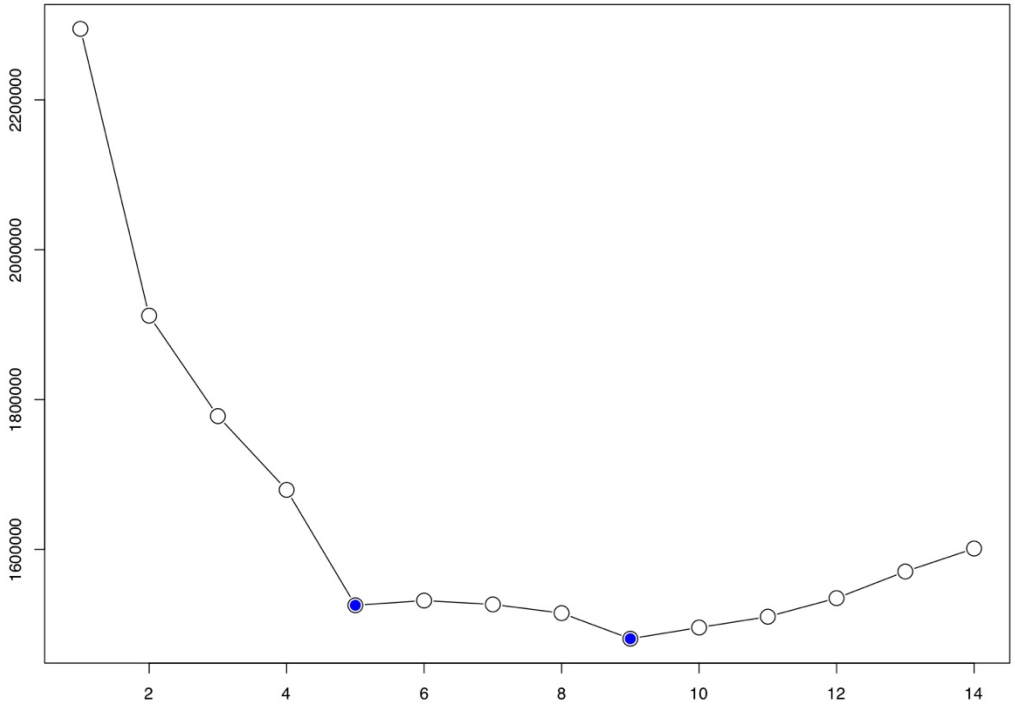


Figure B2: Snapclust AIC values for 1-14 genetic clusters. We present the samples' cluster assignment based on five or nine genetic groups.

DIYABC parameters:

################ DIYabc run 1 parameters
Scenario 1.1:
N1 N2 N3 N4 N5 N6
0 sample 1
0 sample 2
0 sample 3
0 sample 4
0 sample 5
0 sample 6
t35 merge 3 5
t26 merge 2 6
t32 merge 3 2
t41 merge 4 1
ta merge 4 3

Scenario 1.2:
N1 N2 N3 N4 N5 N6
0 sample 1
0 sample 2
0 sample 3
0 sample 4
0 sample 5
0 sample 6
t35 merge 3 5
t26 merge 2 6
t43 merge 4 3
t42 merge 4 2
ta merge 4 1

Scenario 1.3:
N1 N2 N3 N4 N5 N6
0 sample 1
0 sample 2
0 sample 3
0 sample 4
0 sample 5
0 sample 6
t45 merge 4 5
t26 merge 2 6
t43 merge 4 3
t42 merge 4 2
ta merge 4 1

Run 1 Parameters:
Uniform
N1 min 10 max 1000000
N2 min 10 max 1000000

N3 min 10 max 1000000
N4 min 10 max 10000000
N5 min 10 max 1000000
N6 min 10 max 1000000
t35 min 10 max 1000000
t26 min 10 max 1000000
t32 min 10 max 1000000
t41 min 10 max 1000000
ta min 10 max 1000000
t43 min 10 max 1000000
t42 min 10 max 1000000
t45 min 10 max 1000000
t42>=t26
ta>t32
ta>t35
t32>t26
t43>=t35

############### DIYabc run 2 parameters
Scenario 2.1:
N1 N2 N3 N4 N5 N6
0 sample 1
0 sample 2
0 sample 3
0 sample 4
0 sample 5
0 sample 6
t35 merge 3 5
t26 merge 2 6
t43 merge 4 3
t42 merge 4 2
ta merge 4 1

Scenario 2.2:
N1 N2 N3 N4 N5 N6
0 sample 1
0 sample 2
0 sample 3
0 sample 4
0 sample 5
0 sample 6
t35 merge 3 5
t26 merge 2 6
t43 merge 4 3
t42 merge 4 2
ta merge 4 1

Scenario 2.3:
N1 N2 N3 N4 N5 N6
0 sample 1
0 sample 2

0 sample 3
0 sample 4
0 sample 5
0 sample 6
t43 merge 4 3
t26 merge 2 6
t45 merge 4 5
t42 merge 4 2
ta merge 4 1

Run 2 Parameters:
Uniform
N1 min 10 max 1000000
N2 min 10 max 1000000
N3 min 10 max 1000000
N4 min 10 max 10000000
N5 min 10 max 1000000
N6 min 10 max 1000000
t35 min 10 max 1000000
t26 min 10 max 1000000
t43 min 10 max 1000000
t42 min 10 max 1000000
ta min 10 max 1000000
t53 min 10 max 1000000
t45 min 10 max 1000000
t42>=t26
t53<=t45
t43>=t35

############### DIYabc run 3 parameters
Scenario 3.1:
N1 N2 N3
0 sample 1
0 sample 2
0 sample 3
t21 merge 2 1
t23 merge 2 3

Scenario 3.2:
N1 N2 N3
0 sample 1
0 sample 2
0 sample 3
t23 merge 2 3
t21 merge 2 1

Scenario 3.3:
N1 N2 N3
0 sample 1
0 sample 2
0 sample 3

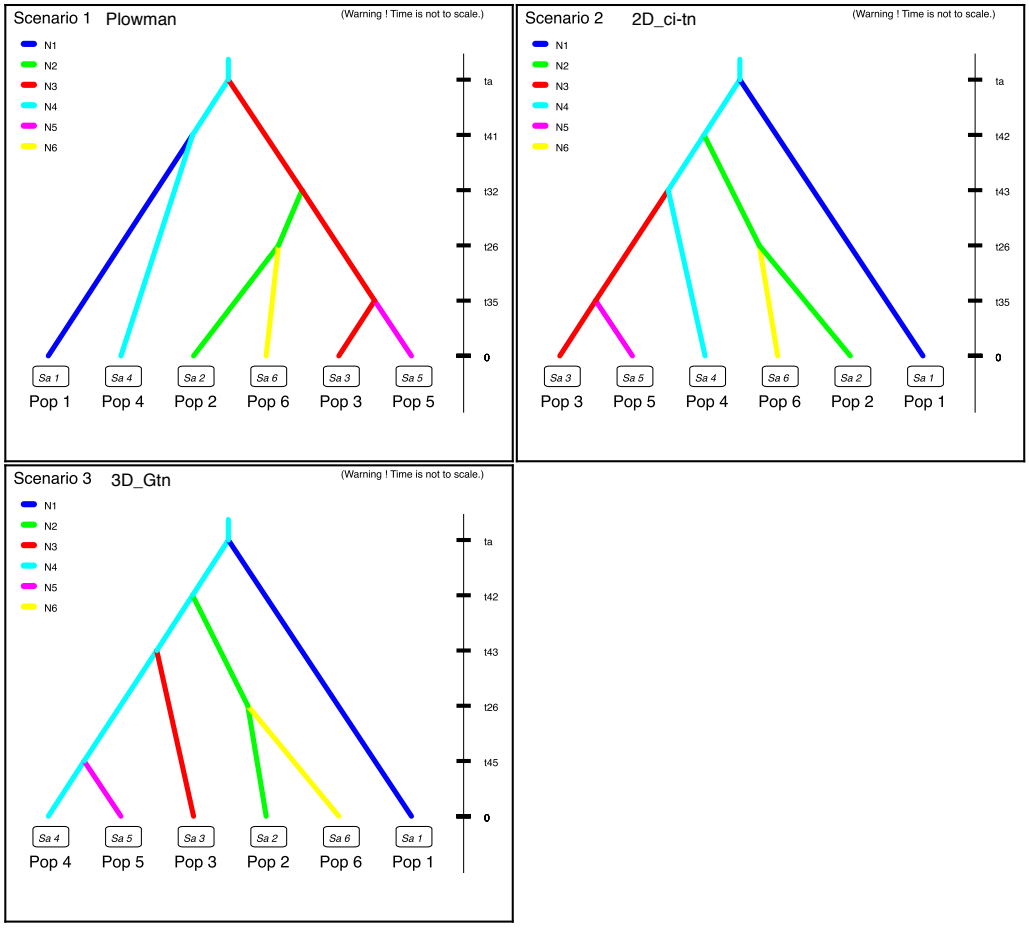
t231 merge 2 1
t231 merge 2 3

Scenario 3.4:
N1 N2 N3
0 sample 1
0 sample 2
0 sample 3
t13 merge 1 3
t21 merge 2 1

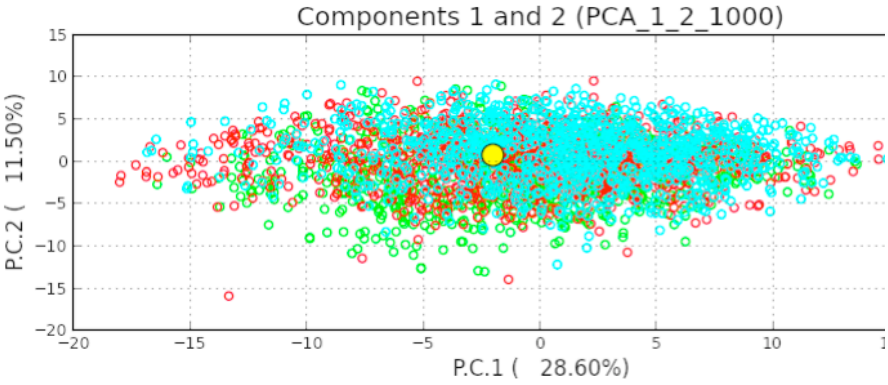
Scenario 3.5:
N1 N2 N3
0 sample 1
0 sample 2
0 sample 3
t31 merge 3 1
t23 merge 2 3

Run 3 parameters:
N1 min 10 max 1000000
N2 min 10 max 10000000
N3 min 10 max 1000000
t21 min 10 max 1000000
t23 min 10 max 1000000
t231 min 10 max 1000000
t13 min 10 max 1000000
t31 min 10 max 1000000
t23>=t31
t13<=t21

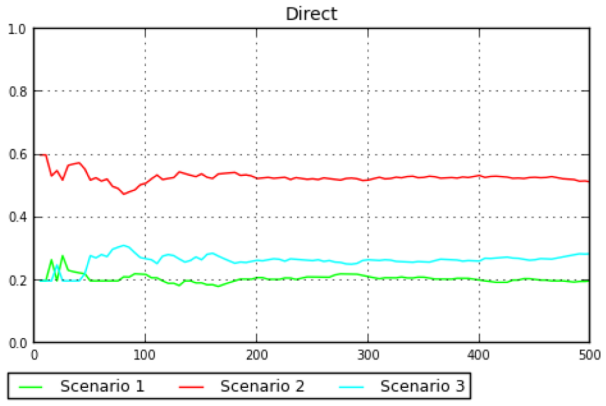
DIYabc run 1 scenarios

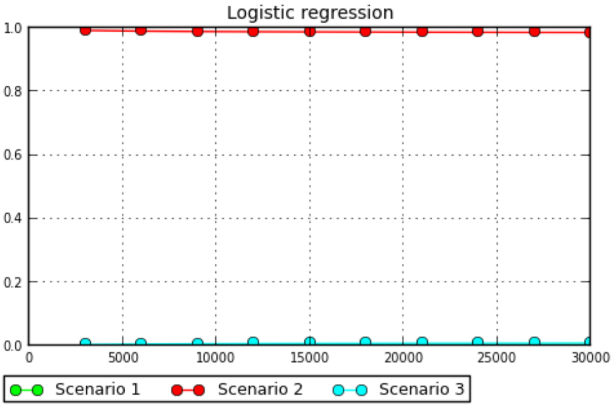


DIYabc run 1 PCA result

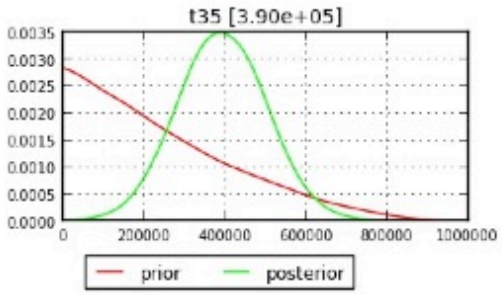


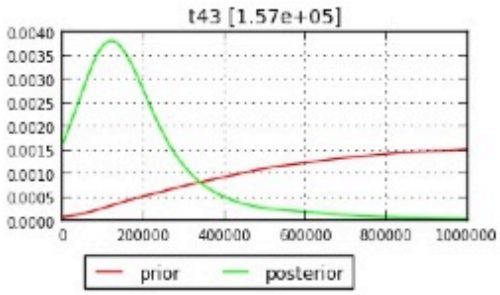
DIYabc run 1 model comparisons

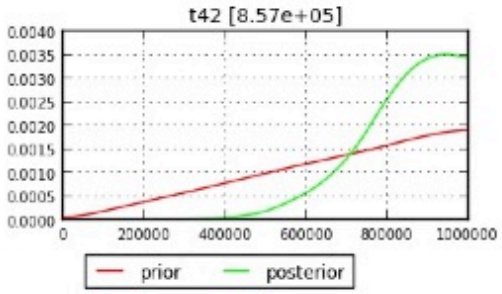


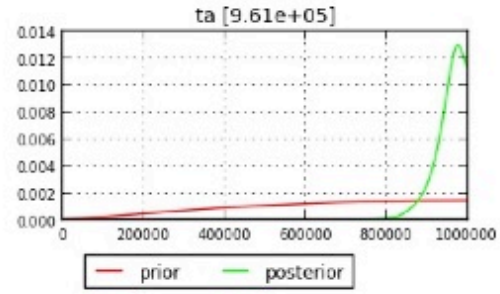


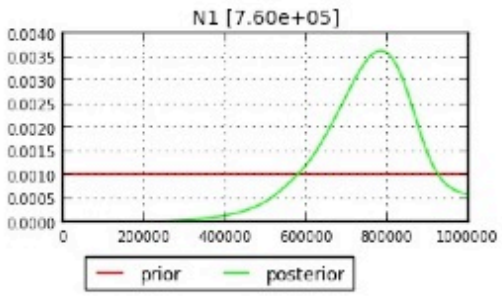
DIYabc run 1, model 2 parameter estimations

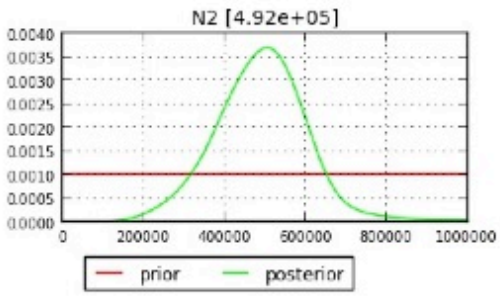




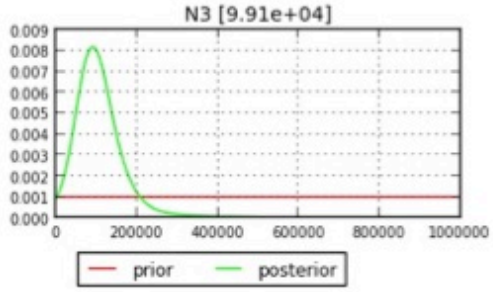


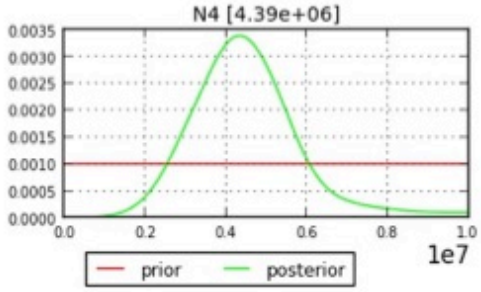


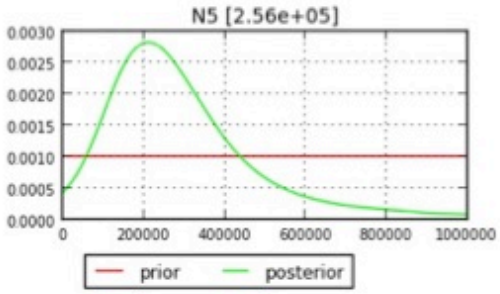


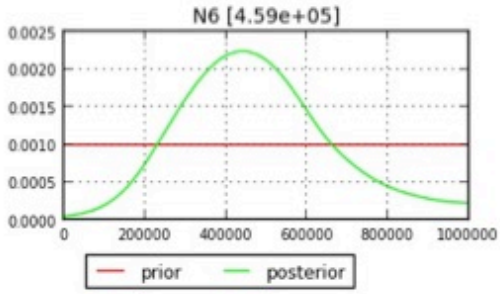


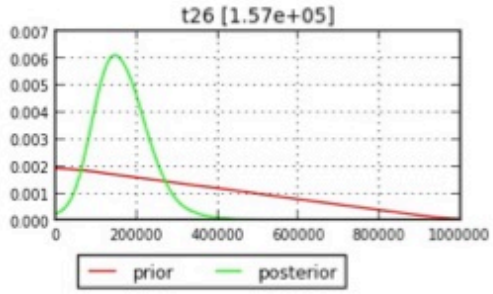
DIYabc run 1, model 2 parameter estimations



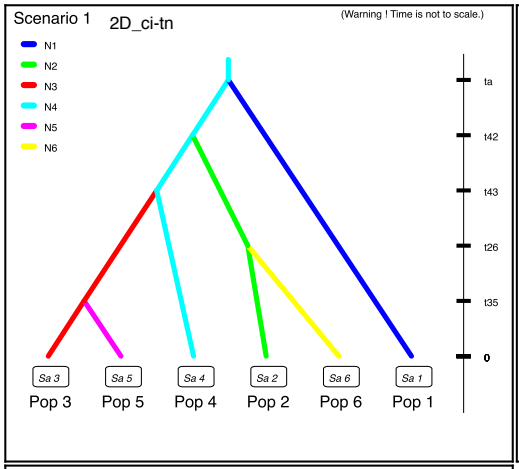


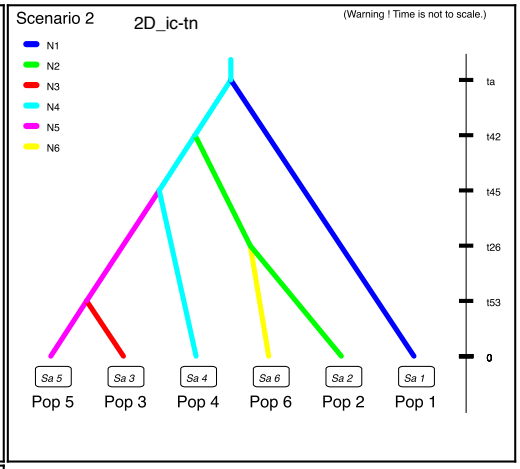


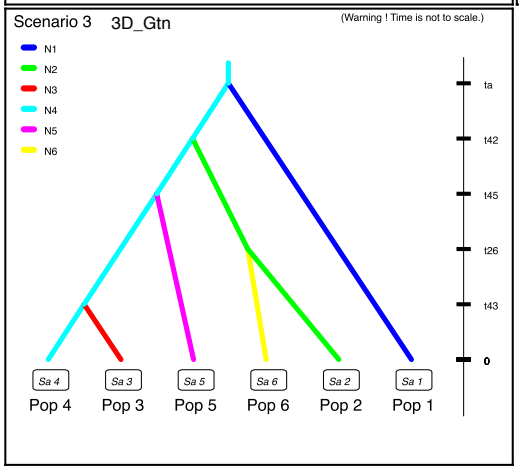




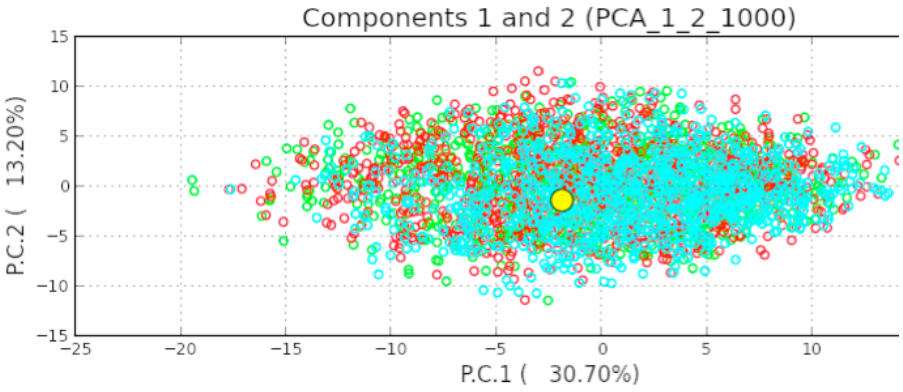
DIYabc run 2 scenarios



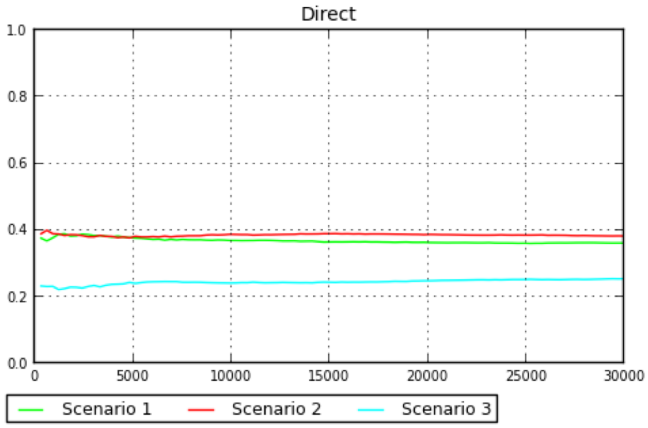


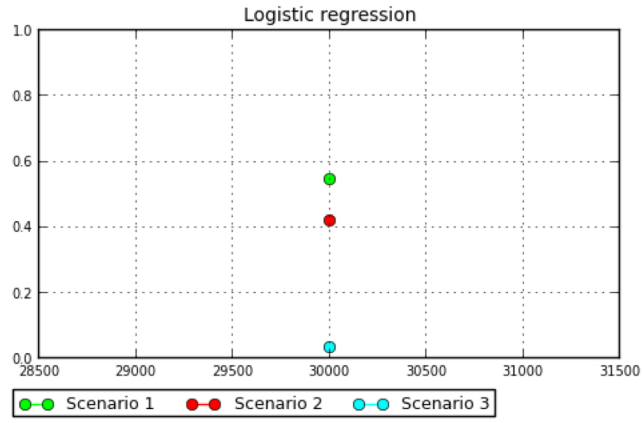


DIYabc run 2 PCA result

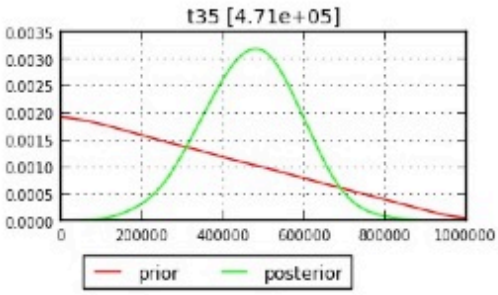


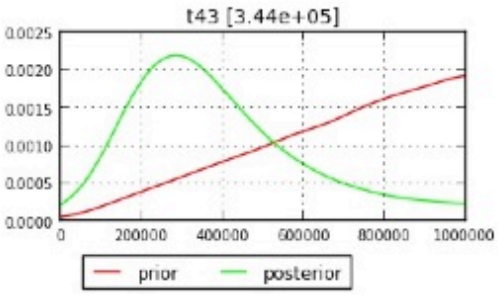
DIYabc run 2 model comparisons

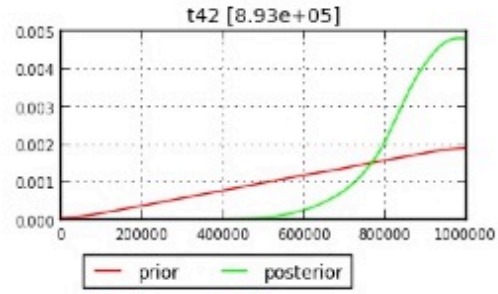


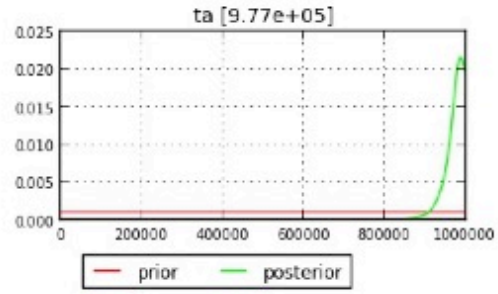


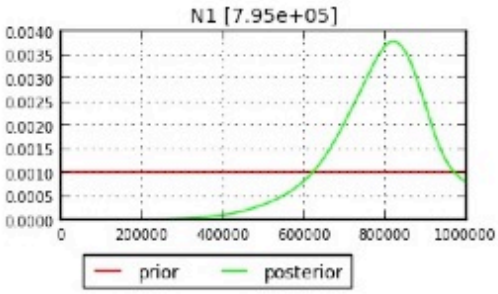
DIYabc run 2, model 2 parameter estimations

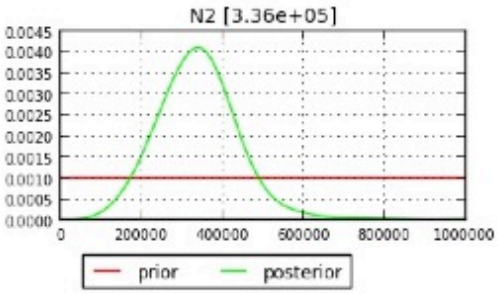




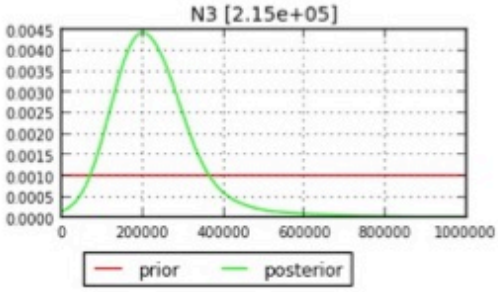


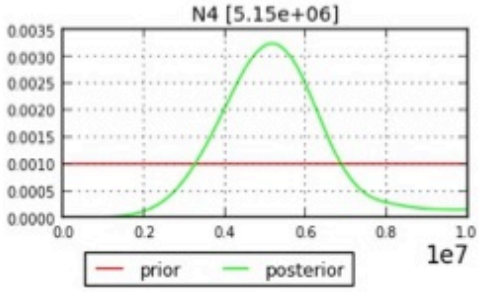


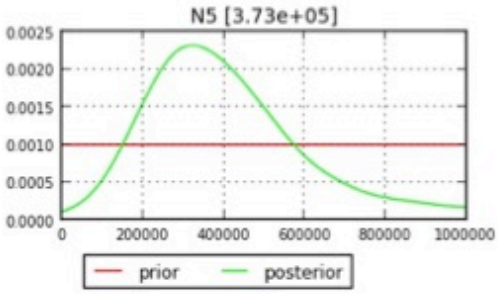


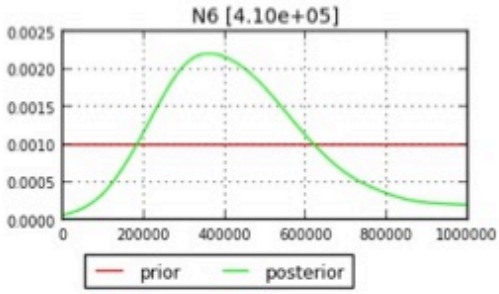


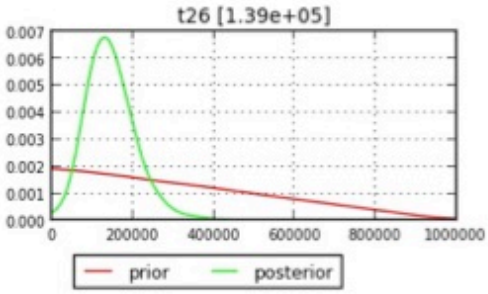
DIYabc run 2, model 2 parameter estimations



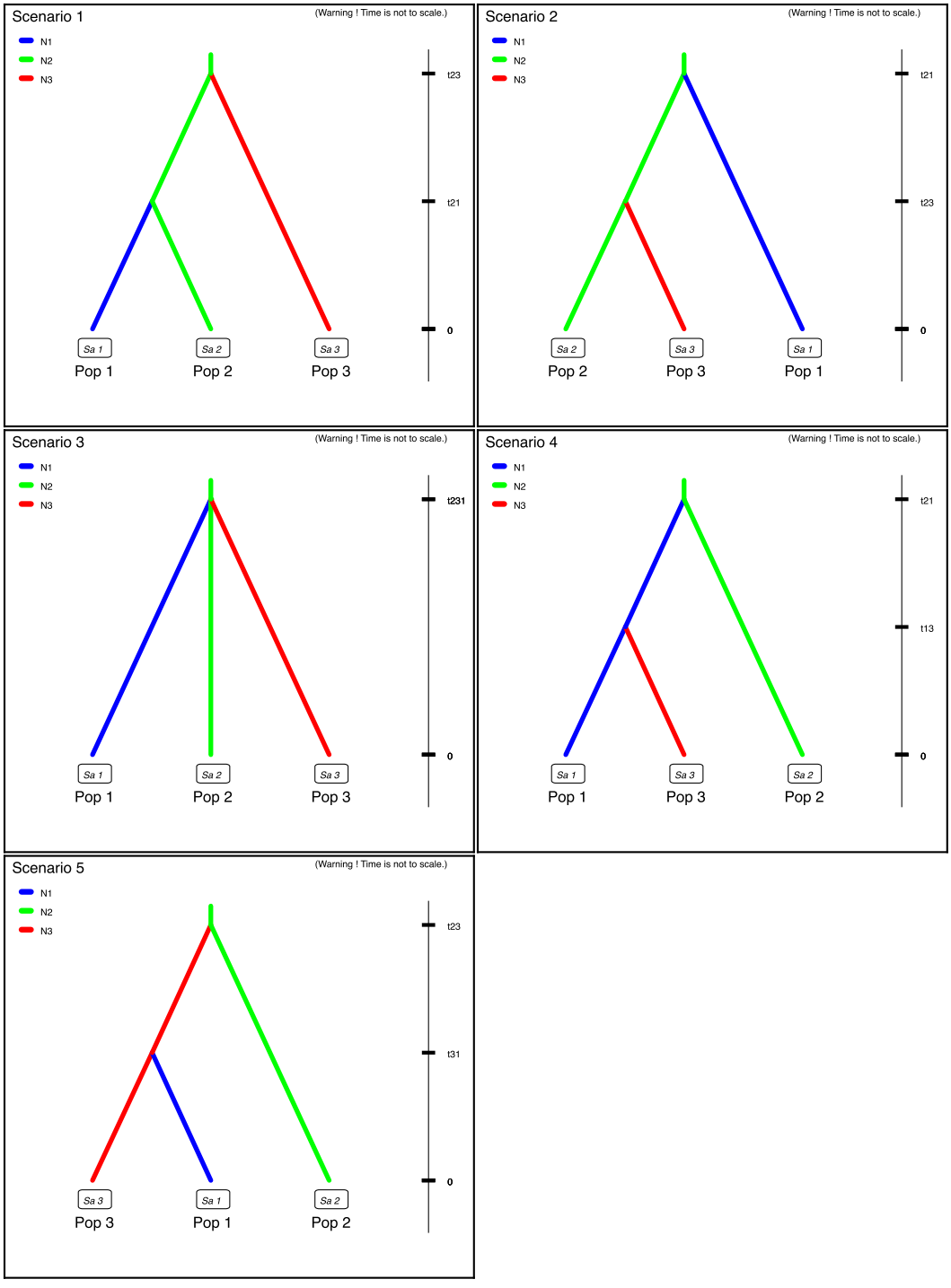




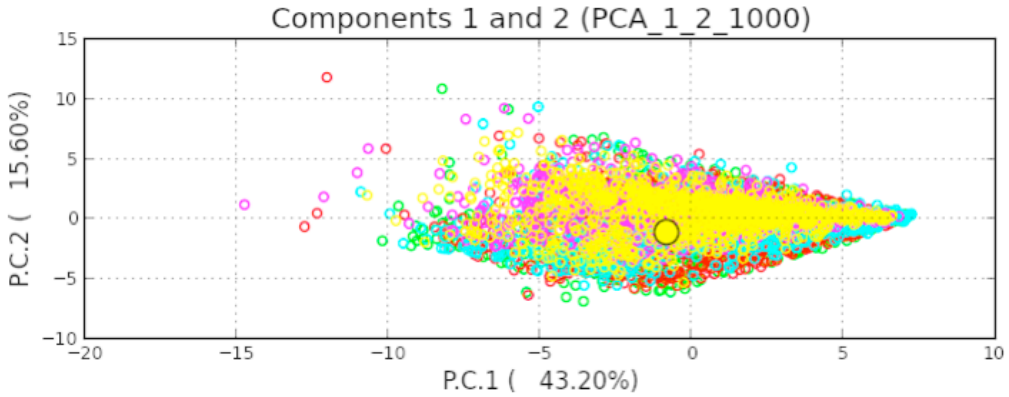




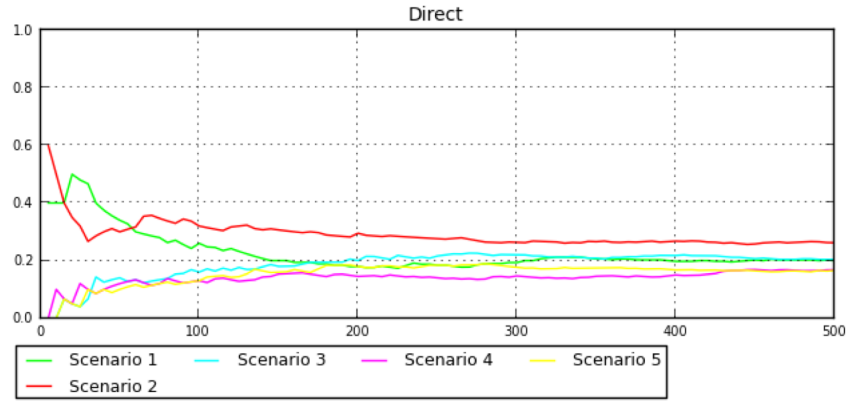
DIYabc run 3 scenarios

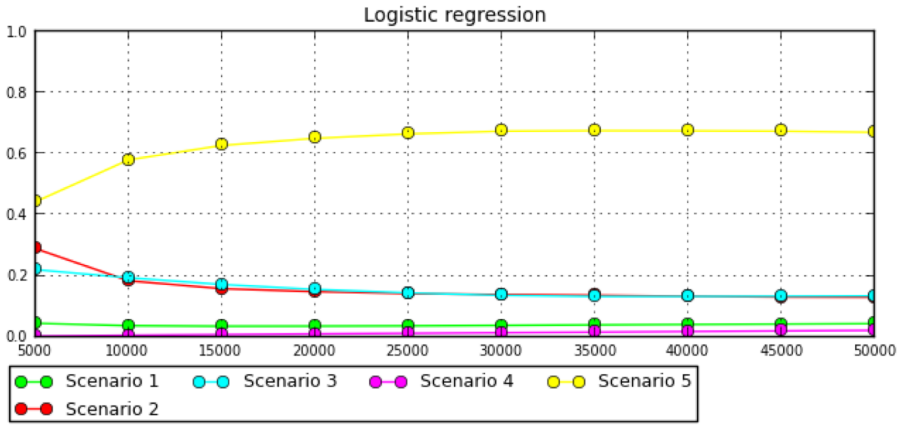


DIYabc run 3 PCA result

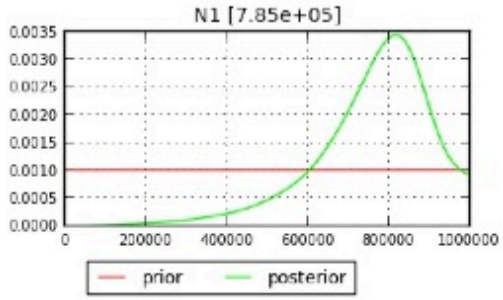


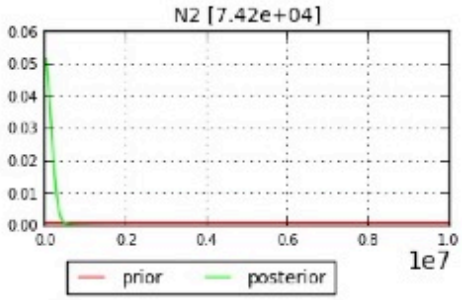
DIYabc run 3 model comparisons

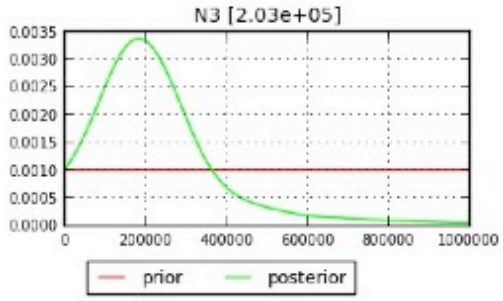


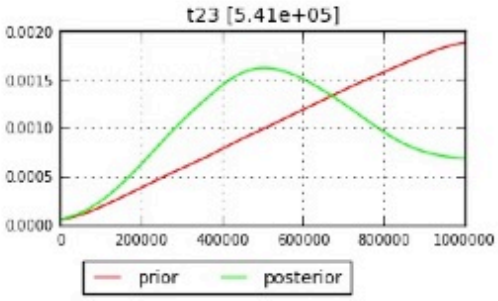


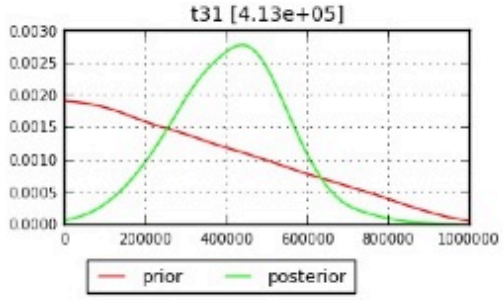
DIYabc run 3, model 5 parameter estimations



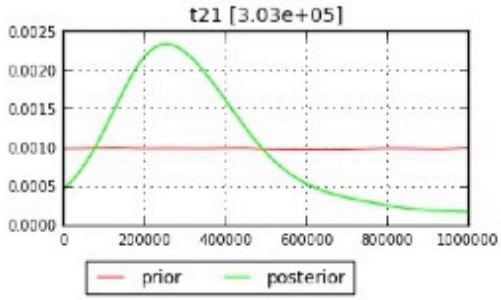


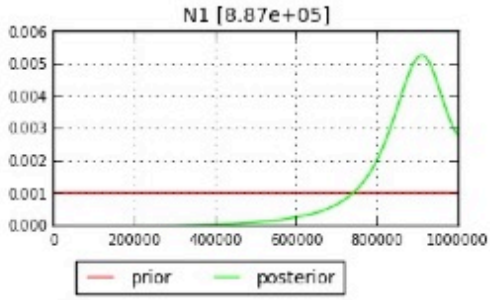


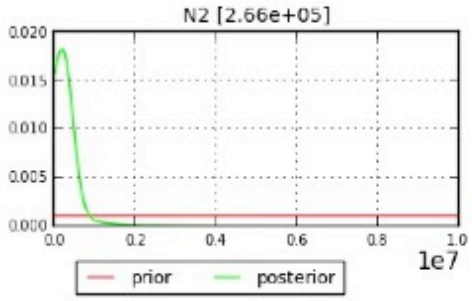


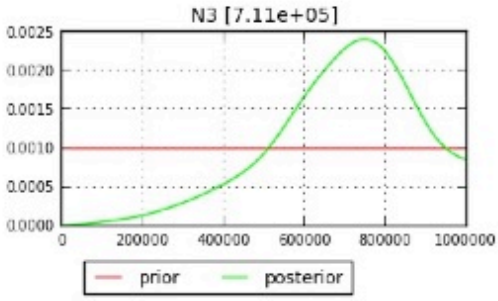


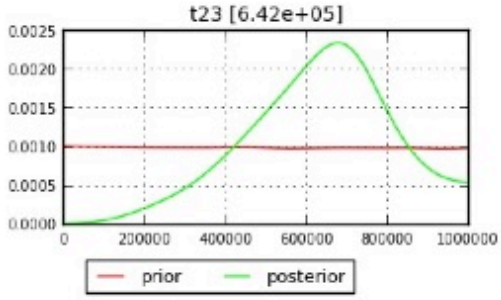
DIYabc run 3, model 2 parameter estimations











Chapter II sample information:

sample_name	species	Collector_No	Location	year	Herbarium:No
cata_402	E. cataractarum	Davidse 12489	Venezuela: Apure	1977	F:1841906
cata_407	E. cataractarum	Delascio 11178	Venezuela: Guarico	1981	F:1910770
cata_408	E. cataractarum	Davidse 15975	Venezuela: Apure	1979	F:1862224
cata_485	E. cataractarum	Guanchez 2938	Venezuela: Amazonas	1984	F:1952199
cata_751	E. cataractarum	White 615	Colombia: Vichada	2015	F:2324366
cata_799	E. cataractarum	Liesner 11001	Venezuela: Bolivar	1981	F:1895309
cata_934	E. cataractarum	White 551	Colombia: Meta	2015	F:2324365
cata_935	E. cataractarum	White 553	Colombia: Meta	2015	F:2324364
cata_936	E. cataractarum	White 557	Colombia: Casanare	2015	F:2324363
cata_937	E. cataractarum	White 558	Colombia: Casanare	2015	F:2324362
cata_938	E. cataractarum	White 612	Colombia: Vichada	2015	F:2324367
coca_11	E. coca	Plowman 4819	Peru: Cusco	1975	F:1774902
coca_15n	E. coca	Galiano 6827	Peru: Cusco	2004	F:2287081
coca_480	E. coca	Plowman 11272	Peru: Huanuco	1981	F:1899084
coca_482n	E. coca	Nunez 6837	Peru: Cusco	1986	F:1984280
coca_6	E. coca	Quipuscoa 2302	Peru: San Martin	2000	F:2229135
coca_746	E. coca	White 469	Peru: Cusco	2014	MOL:40676
coca_747	E. coca	White 490	Peru: Madre de Dios	2014	F:2320313
coca_766	E. coca	White 502	Peru: Huanuco	2014	MOL
coca_767	E. coca	White 524	Peru: Junin	2014	MOL:40677
coca_768	E. coca	White 525	Peru: Junin	2014	F:2320285
coca_813	E. coca	McMullen_694	Bolivia: Cochabamba	1988	F:2314960
coca_814	E. coca	McMullen 695	Bolivia: Cochabamba	1988	F:2314959
coca_815	E. coca	McMullen 697	Bolivia: Cochabamba	1988	F:2314957
coca_816	E. coca	McMullen 693	Peru: Huanuco	1988	F:2314952
coca_817	E. coca	McMullen 691	Peru: Huanuco	1988	F:2313953
coca_843n	E. coca	Chacon 760	Costa Rica	1983	F:1929662
coca_844n	E. coca	Plowman 4713	Peru: Ayacucho	1975	F:1774895
coca_846	E. coca	Nee 31594	Bolivia: Pando	1985	F:1979830
coca_848	E. coca	Plowman 5181	Bolivia: La Paz	1975	F:1774928
coca_852	E. coca	Davis 1205	Bolivia: Beni	1981	F:1891565
coca_85n	E. coca	Michel 3339	Bolivia: Cochabamba	2003	F:2303521
coca_864	E. coca	Herman 308	Bolivia: La Paz	1989	F:2216318
coca_865n	E. coca	Mello-Silva 2130	Bolivia: La Paz	2002	F:2245684
coca_866	E. coca	Plowman 5208	Bolivia: La Paz	1975	F:1873755
coca_869	E. coca	Steinbach 7110	Bolivia: Santa Cruz	1925	F:563988
coca_87	E. coca	Davis 1101	Bolivia: Beni	1981	F:1891570
coca_870n	E. coca	Plowman 5179	Bolivia: La Paz	1975	F:1744807
coca_872	E. coca	Plowman 4648	Peru: Ayacucho	1975	F:1873725
coca_877n	E. coca	Killip 26609	Peru: Junin	1929	F:616154
coca_879n	E. coca	Plowman 4628	Peru: Junin	1974	F:1873723
coca_880	E. coca	Plowman 11425	Peru: San Martin	1981	F:1915011
coca_881	E. coca	Plowman 5793	Peru: San Martin	1976	F:1763757
coca_882	E. coca	Plowman 5987	Peru: San Martin	1976	F:1774899
coca_887	E. coca	Belshaw 3579	Peru: San Martin	1937	F:1680333
coca_888n	E. coca	Plowman 5795	Peru: San Martin	1976	F:1763743
coca_890n	E. coca	Schunke 10018	Peru: San Martin	1978	F:1851142
coca_893	E. coca	Plowman 4712	Peru: Ayacucho	1975	F:1744808
coca_895	E. coca	Plowman 6095	Peru: Apurimac	1976	F:1873753
coca_896	E. coca	Plowman 5203	Peru: Cusco	1975	F:1873751
coca_897n	E. coca	Fernandez Distel s.n.	Peru: Madre de Dios	1979	F:1863187
coca_91	E. coca	Plowman 7540	Peru: San Martin	1978	F:1856758
cf.coca_884	E. cf. coca	Wesshausen 1053	Peru: San Martin	1979	F:1900127
grac_01J	E. gracilipes	Engels 4544	Brazil: Mato Grosso	2016	HUEFS:299520
grac_100	E. gracilipes	Baker 6841	Ecuador: Morona Santiago	1986	F:1992274
grac_104	E. gracilipes	Pena 231	Bolivia: Santa Cruz	1991	F:2074357
grac_17	E. gracilipes	Velasquez 245	Colombia: Antioquia	1989	F:2198680
grac_178	E. gracilipes	Croat 55511	Colombia: Meta	1983	F:1921743
grac_202	E. gracilipes	Nee 34658	Brazil: Rondonia	1987	F:2000201
grac_272	E. gracilipes	Davidse 27790	Venezuela: Amazonas	1984	F:1962371
grac_297	E. gracilipes	Liesner 8977	Venezuela: Amazonas	1980	F:1895285

grac_298	E. gracilipes	Foster 11483	Peru: Madre de dios	1986	F:1980401
grac_654	E. gracilipes	White 476	Peru: Madre de Dios	2014	F:2320258
grac_745	E. gracilipes	Janovec 2666	Peru: Madre de Dios	2005	MOL:14760
grac_786	E. gracilipes	Liesner 10224	Venezuela: Tachira	1981	F:1895308
grac_789	E. gracilipes	Fernandez 835	Venezuela: Bolivar	1984	F:1964691
grac_790	E. gracilipes	Lopez Figueiras 30788	Venezuela: Barinas	1983	F:1918742
grac_792	E. gracilipes	Liesner 10967	Venezuela: Amazonas	1981	F:1895286
grac_793	E. gracilipes	McDaniel 29917	Peru: Loreto	1988	F:2155567
grac_794	E. gracilipes	Schunke 1218	Peru: Ucayali	1966	F:1688199
grac_795	E. gracilipes	Vasquez 7142	Peru: Loreto	1986	F:1984271
grac_796	E. gracilipes	Vasquez 2647	Peru: Loreto	1981	F:1984265
grac_798	E. gracilipes	Liesner 7575	Venezuela: Amazonas	1979	F:1875168
grac_800	E. gracilipes	Steyermark 113958	Venezuela: Amazonas	1977	F:1848454
grac_801	E. gracilipes	Steyermark 88080	Venezuela: Bolivar	1960	F:1621159
grac_803	E. gracilipes	Marles 139	Ecuador: Napo	1985	F:1960019
grac_806	E. gracilipes	Shemluck 236	Ecuador: Pastaza	1979	F:1863368
grac_807	E. gracilipes	Maas 6746	Brazil: Amazonas	1987	F:2198678
grac_808	E. gracilipes	Silva 4683	Brazil: Mato Grosso	1979	F:2042208
grac_809	E. gracilipes	Nee 34664	Brazil: Rondonia	1987	F:2000200
grac_811	E. gracilipes	Alvira 40	Colombia: Casanare	1996	F:2214855
grac_829	E. gracilipes	Plowman 4265	Colombia: Meta	1974	F:1746942
grac_831	E. gracilipes	Neill 7016	Ecuador: Napo	1985	F:1959875
grac_835	E. gracilipes	Brandbyge 31689	Ecuador: Pastaza	1980	F:1982373
grac_868	E. gracilipes	Vargas 4072	Bolivia: Santa Cruz	1995	F:2171782
grac_939	E. gracilipes	White 566	Colombia: Meta	2015	F:2324355
grac_940	E. gracilipes	White 570	Colombia: Meta	2015	F:2324350
grac_941n	E. gracilipes	White 587	Colombia: Antioquia	2015	F:2324356
grac_943	E. gracilipes	Villa 499	Ecuador: Orellana	2000	F:2227303
grac_946	E. gracilipes	Romoleroux 1963	Ecuador: Napo	1995	F:2163874
grac_98n	E. gracilipes	Cabrera 3637	Colombia: Caqueta	1975	F:1842755
ipad_215	E. coca var. Ipadu	Madison 6775	Ecuador: Napo	1979	F:1853639
ipad_270	E. coca var. Ipadu	Ruiz 1254	Peru: Loreto	1988	F:2198727
ipad_367	E. coca var. Ipadu	Carreira 203	Brazil: Belem	1981	F:1923911
ipad_474	E. coca var. Ipadu	Plowman 6922	Peru: Loreto	1977	F:1823930
ipad_475	E. coca var. Ipadu	Salvador Flores P 15(P2-07)	Peru: Loreto	1982	F:1968326
ipad_477	E. coca var. Ipadu	Ayala 2806	Peru: Loreto	1980	F:1993557
ipad_770	E. coca var. Ipadu	Plowman 6360	Colombia: Amazonas	1977	F:1824622
ipad_771	E. coca var. Ipadu	Plowman 12117	Brazil: Amazonas	1982	F:1925059
ipad_772	E. coca var. Ipadu	Cid Ferreira 3165	Brazil: Amazonas	1982	F:1937122
ipad_773	E. coca var. Ipadu	Plowman 6924	Peru: Loreto	1977	F:1823933
ipad_775	E. coca var. Ipadu	Plowman 7136	Peru: Loreto	1977	F:1824532
ipad_900	E. coca var. Ipadu	Zarucchi 1146	Colombia: Vaupes	1975	F:1774922
ipad_903	E. coca var. Ipadu	Zarucchi 1145	Colombia: Vaupes	1975	F:1774924
ipad_905	E. coca var. Ipadu	Jangoux 1221	Brazil: Belem	1980	F:1935034
ipad_906	E. coca var. Ipadu	Nelson 422	Brazil: Belem	1980	F:1874378
ipad_907	E. coca var. Ipadu	Cid Ferreira 3439	Brazil: Amazonas	1982	F:1937121
ipad_908	E. coca var. Ipadu	Silva 1310	Brazil: Amazonas	1973	F:1858322
ipad_950	E. coca var. Ipadu	Plowman 10224	Peru: Loreto	1979	F:1872708
novo_764	E. novogranatense	White 576	Colombia: Tolima	2015	F:2324281
novo_776	E. novogranatense	Plowman 3734	Colombia: Antioquia	1974	F:1774506
novo_777	E. novogranatense	Plowman 5385	Colombia: Cauca	1976	F:1823473
novo_778	E. novogranatense	Plowman 5381	Colombia: Cauca	1976	F:1823476
novo_782	E. novogranatense	Maas 1836	Colombia: Valle	1974	F:1823807
novo_783	E. novogranatense	Plowman 5272	Colombia: Valle	1976	F:1823467
novo_784	E. novogranatense	Garcia-Barriga 20950	Colombia: Cundinamarca	1976	F:1820738
novo_785	E. novogranatense	Moser s.n.	Colombia: Guajira	1980	F:1881984
novo_818	E. novogranatense	Garnier s.n.	Nicaragua: Managua	1935	F:1248358
novo_827	E. novogranatense	Plowman 5373	Colombia: Cauca	1976	F:1993590
novo_828	E. novogranatense	Plowman 3500	USA: Florida	1974	F:1993581
novo_838n	E. novogranatense	Molina 21913	Honduras: Morazan	1984	F:1675440
novo_839	E. novogranatense	Carlson 271	El Salvador: La Libertad	1946	F:1186707
novo_840	E. novogranatense	Broadway 5094	Trinidad & Tobago: Botanic Garden	1923	F:549513

novo_841	E. novogranatense	Dahlgren s.n.	Trinidad & Tobago: Botanic Garden	1932	F:654806
novo_842	E. novogranatense	Broadway s.n.	Trinidad & Tobago: Botanic Garden	1919	F:492735
novo_912	E. novogranatense	Plowman 4176	Colombia: Huila	1974	F:1746942A
novo_913	E. novogranatense	Plowman 4151-A	Colombia: Huila	1974	F:1873735
novo_915	E. novogranatense	Hodge 6710	Colombia: Antioquia	1945	F:1338563
novo_917	E. novogranatense	Moser s.n.	Colombia: Cesar	1980	F:1881985
novo_918	E. novogranatense	Silverstone-Sopkin 2284	Colombia: Valle	1986	F:1672686
novo_919	E. novogranatense	Maas 1886	Colombia: Valle	1974	F:1757563
trux_138	E. novo. var. truxillense	Plowman 6099	Peru: Amazonas	1976	F:1873737
trux_556h1	E. novo. var. truxillense	- -	USA: Maryland	2013	USDA:B304
trux_822	E. novo. var. truxillense	Plowman 5583	Peru: Amazonas	1976	F:1993569
trux_823	E. novo. var. truxillense	Plowman 5590A	Peru: Amazonas	1976	F:1993578
trux_826	E. novo. var. truxillense	Plowman 5620	Peru: La Libertad	1976	F:1993531
trux_854	E. novo. var. truxillense	Boeke 854	Colombia: Narino	1977	F:1813247
trux_855	E. novo. var. truxillense	Madison 4447	Ecuador: Carchi	1978	F:1854084
trux_856	E. novo. var. truxillense	Gentry 26380	Ecuador: Carchi	1979	F:1870648
trux_857	E. novo. var. truxillense	Young 1204	Peru: La Libertad	1985	F:1977944
trux_858	E. novo. var. truxillense	Plowman 5600	Peru: La Libertad	1976	F:1823791
trux_859	E. novo. var. truxillense	Plowman 5614	Peru: La Libertad	1976	F:1825164
trux_860	E. novo. var. truxillense	Plowman 5828	Peru: Huanuco	1976	F:1825167
trux_861	E. novo. var. truxillense	Plowman 5209	Peru: Lima	1975	F:1757565
trux_862	E. novo. var. truxillense	Ramon Ferreyra s.n.	Peru: La Libertad	1967	F:1774929
trux_863	E. novo. var. truxillense	Plowman 5612	Peru: La Libertad	1976	F:1818086
trux_944	E. novo. var. truxillense	Rusby 2684	Peru	1900	F:1863090
cf.trux_824	E. novo. cf. var. truxillense	Duke s.n.	Ecuador: From seed cultivated at USDA	1988	F:1982139
cf.trux_916	E. novo. cf. var. truxillense	Plowman 3628-d	Colombia: Cesar	1974	F:1873740

Appendix C: Additional tables, figures, and sample information for Chapter III

Tables:

species	WorldRegion: Africa (Af), Caribbean (Ca), Indo-Pacific (IP), Madagascar (Md), Neotropics (NT)	CladeAssignment:
E.dekindtii	Af	Paleotropical
E.moonii	Af	Paleotropical
E.schliebenii	Af	Paleotropical
E.zeylanicum	Af	Paleotropical
N.acuminatum	Af	Nectaropetalum
N.carvalhoi	Af	Nectaropetalum
N.eligulatum	Af	Nectaropetalum
N.evrardii	Af	Nectaropetalum
N.lebrunii	Af	Nectaropetalum
P.gabonense	Af	Pinacopodium
E.armatum	Ca	Neotropical-A
E.banaoense	Ca	Neotropical-A
E.baracoense	Ca	Neotropical-A
E.barahonense	Ca	Neotropical-A
E.coriaceum	Ca	Neotropical-B
E.dumosum	Ca	Neotropical-A
E.echinodendron	Ca	Neotropical-A
E.flavicans	Ca	Neotropical-A
E.longipes	Ca	Neotropical-A
E.mogotense	Ca	Neotropical-A
E.annamense	IP	Paleotropical
E.calyptranum	IP	Paleotropical
E.couveleense	IP	Paleotropical
E.cuneatum	IP	Paleotropical
E.gracile	IP	Paleotropical
E.iwahigense	IP	Paleotropical
E.kochummenii	IP	Paleotropical
E.lanceolatum	IP	Paleotropical
E.novocaledonicum	IP	Paleotropical
E.obtusifolium	IP	Paleotropical
E.sarawakanum	IP	Paleotropical
E.sinense	IP	Paleotropical
E.acranthum	Md	Paleotropical
E.boinense	Md	Paleotropical
E.boivinianum	Md	Paleotropical
E.buxifolium	Md	Paleotropical
E.capitatum	Md	Paleotropical
E.ferrugineum	Md	Paleotropical
E.hypericifolium	Md	Paleotropical
E.lamprocarpum	Md	Paleotropical
E.macrocarpum	Md	Paleotropical
E.mangorense	Md	Paleotropical
E.retusum	Md	Paleotropical
E.sideroxyloides	Md	Paleotropical
E.acrobeles	NT	Neotropical-A
E.apiculatum	NT	Neotropical-B
E.bangii	NT	Neotropical-B
E.bequaertii	NT	Neotropical-B
E.bezerrae	NT	Neotropical-B
E.ectinocalyx	NT	Neotropical-A
E.gentryi	NT	Neotropical-A
E.jaimei	NT	Neotropical-B
E.leal-costae	NT	Neotropical-B
E.longisetulosum	NT	Neotropical-B

E.loretense	NT	Neotropical-A
E.lygoides	NT	Neotropical-B
E.membranaceum	NT	Neotropical-B
E.mikanii	NT	Neotropical-B
E.occultum	NT	Neotropical-B
E.opacum	NT	Neotropical-B
E.ovalifolium	NT	Neotropical-B
E.pachyneurum	NT	Neotropical-B
E.pacificum	NT	Neotropical-B
E.paraguariense	NT	Neotropical-B
E.parvistipulatum	NT	Neotropical-B
E.pauferrense	NT	Neotropical-B
E.petraecaballi	NT	Neotropical-B
E.popayanense	NT	Neotropical-B
E.santosii	NT	Neotropical-A
E.simonis	NT	Neotropical-B
E.spinescens	NT	Neotropical-A
E.strobilaceum	NT	Neotropical-B
E.subglaucescens	NT	Neotropical-B
E.timothei	NT	Neotropical-B
E.tucuruiense	NT	Neotropical-B
E.vaginatum	NT	Neotropical-B
E. sp.nov. (aff. macrophyllum)	NT	Neotropical-B
E.cogolloi	NT	Neotropical-B
E.riverae	NT	Neotropical-B
E.umbrosum	NT	Neotropical-B

Table C1: List of Erythroxylaceae species not sampled with geographic range (world region) and clade assignment for determining missing data in clades defined in Figure C1 in the BAMM diversification analysis.

Figures:

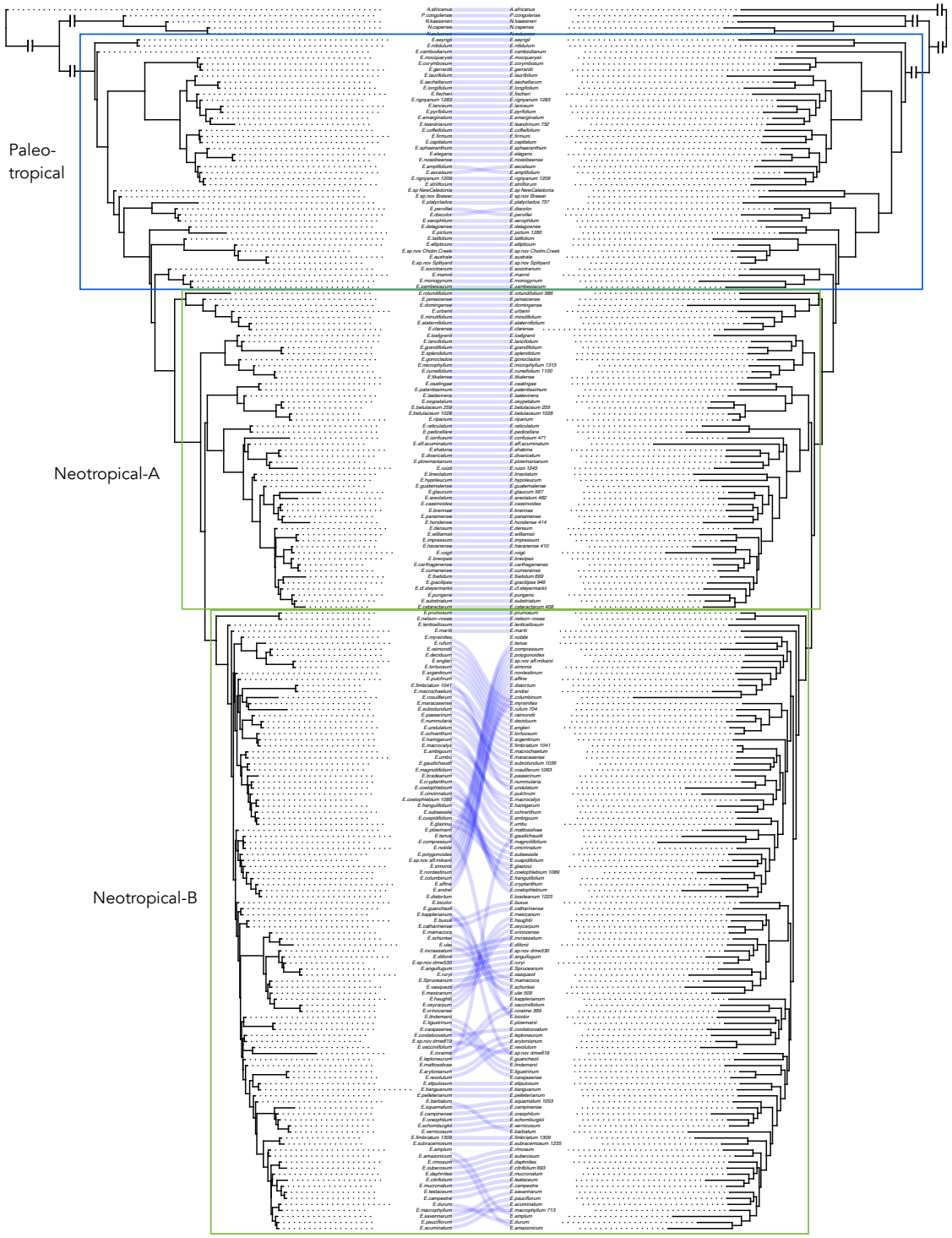


Figure C1: Delimitation of clades for assigning missing taxa proportions in BAMM analysis labeled on right side of cophylo plot. Topological comparison between ASTRAL III species tree (left) and MLcat tree (right) are shown by blue lines connecting tips. Short terminal branches on ASTRAL III tree are arbitrary as they represent a single individual for that given species.

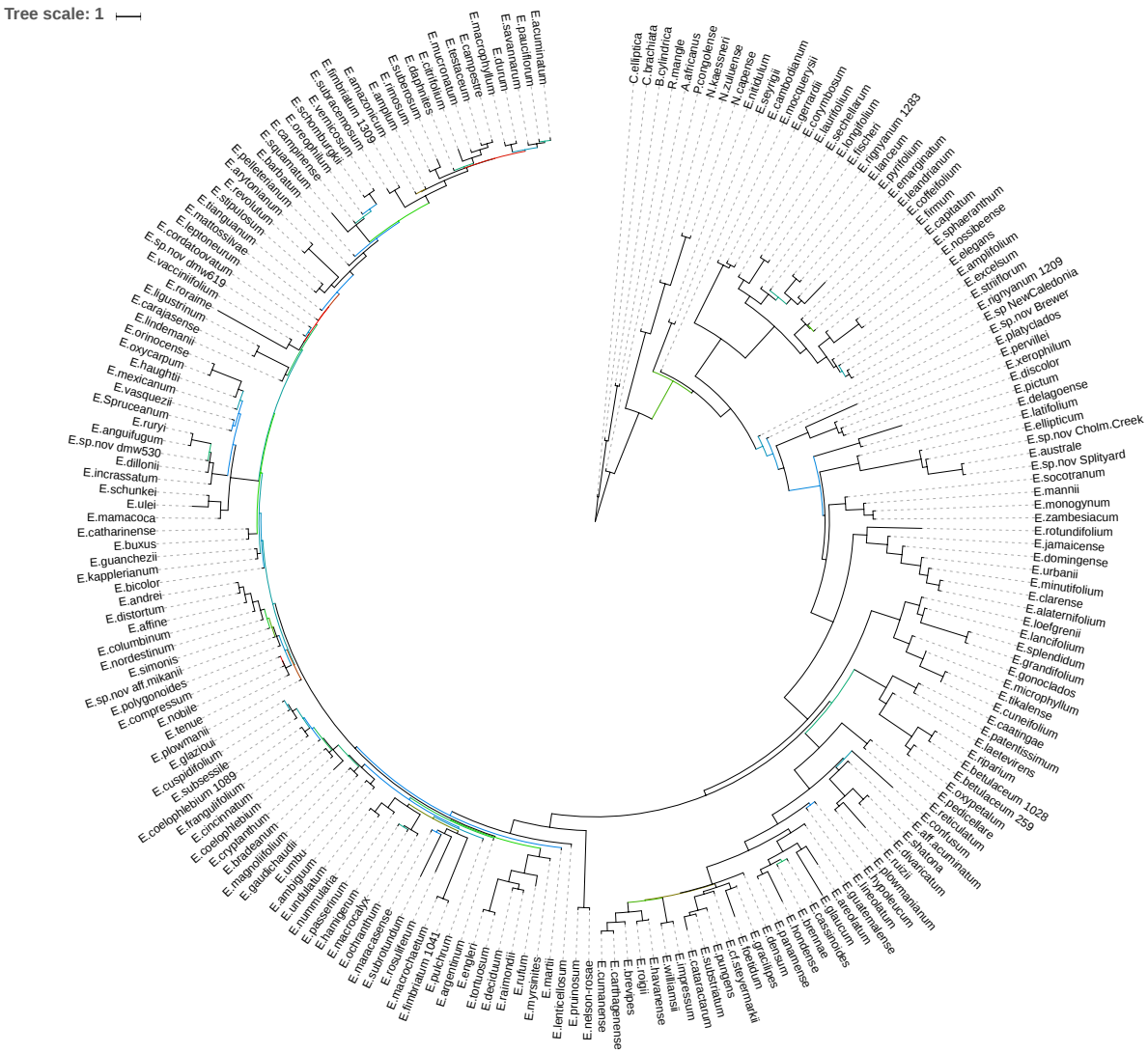


Figure C2: Astral III Species tree for 207 species including 4 Rhizophoraceae outgroups (*Cassipourea elliptica*, *Carralia brachiata, Bruguiera cylindrica, Rhizophora mangle*). Genera abbreviations for the Erythroxylaceae taxa are as follows: *A.=Aneulophus, E.=Erythroxylum, N.=Nectaropetalum, P.=Pinacopodium.* Branches are colored by local posterior probability (LPP) with lowest support in red (>0.37), transitioning to green then blue (<1). Branches with LPP of 1 are black.

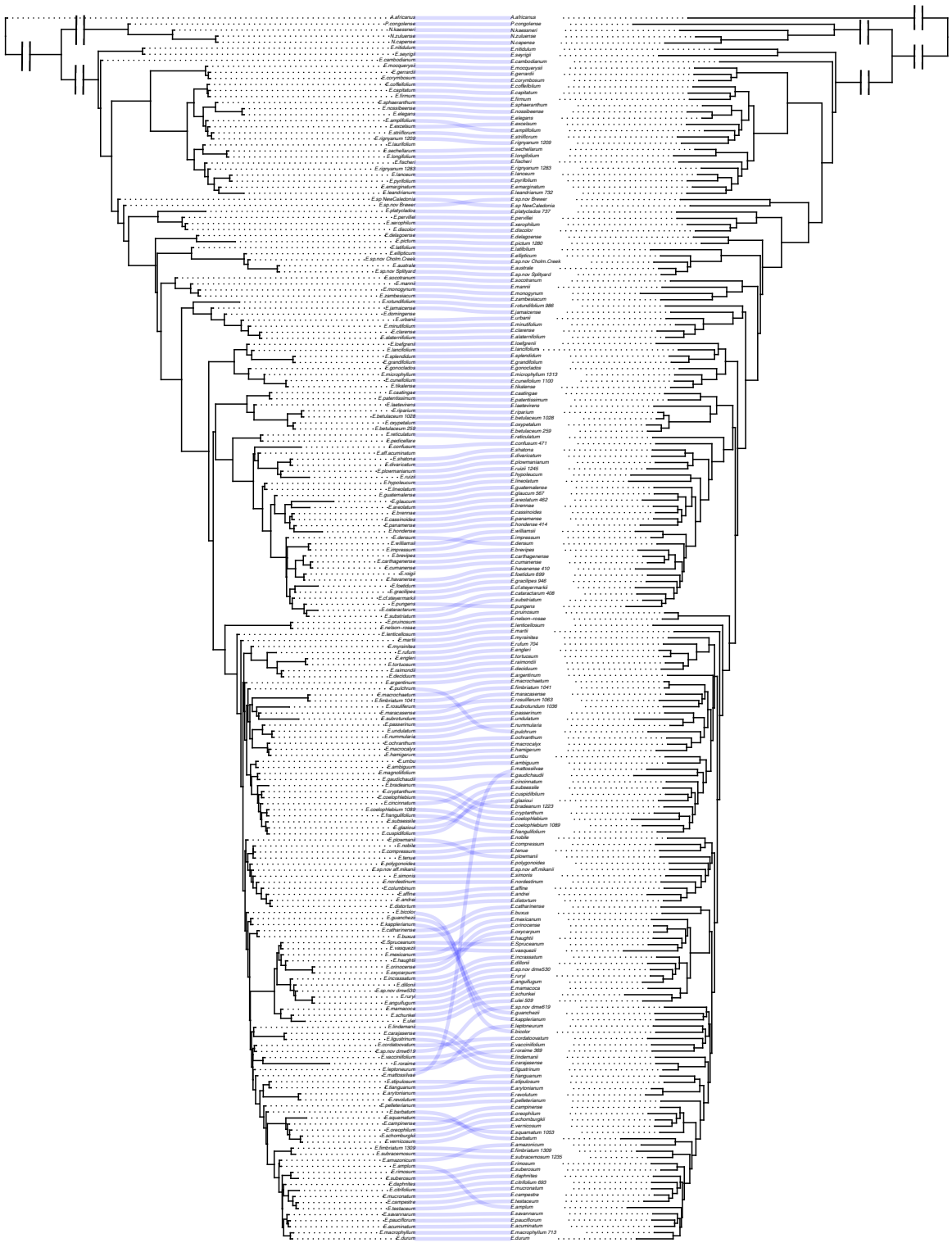

A.africanus
P.congolense
N.kassneri
N.zuluense
N.capense
E.nitidulum
E.eayrigii
E.cambodianum
E.mocquerysii
E.gerrardii
E.corymbosum
E.coffeifolium
E.capitatum
E.firmum
E.sphaeranthum
E.nossibeense
E.elegans
E.ampliflolium
E.excelsum
E.atrolflorum
E.rignarum 1209
E.laurifolium
E.aechellarum
E.longifolium
E.fischeri
E.rignarum 1283
E.lanceum
E.pyrifolium
E.emarginatum
E.leandrianum 732
E.sp.nov Brewer
E.sp.NewCaledonia
E.platyclados 737
E.pervillei
E.xerophilum
E.discolor
E.delagoense
E.pictum 1280
E.latifolium
E.ellipticum
E.sp.nov Cholm.Creek
E.australe
E.sp.nov Spillyard
E.scotranum
E.mannii
E.monogynum
E.zambesiacum
E.rotundifolium 986
E.jamaicense
E.urbanii
E.minutifolium
E.clarense
E.alaternifolium
E.loeflgranii
E.lancifolium
E.splendidum
E.grandifolium
E.goncloclados
E.microphyllum 1313
E.cuneifolium 1100
E.likalense
E.caatingae
E.patentissimum
E.laetevirens
E.riparium
E.betulaceum 1028
E.oxypetalum
E.betulaceum 259
E.reticulatum
E.confusum 471
E.xhatona
E.divaricatum
E.plowmanianum
E.ruizii 1245
E.hypoleucum
E.linealatum
E.guatemalense
E.glaucum 587
E.arealatum 462
E.brennae
E.casasnoides
E.panamense
E.hondense 414
E.williamsii
E.impressum
E.densum
E.brevipes
E.carthagenense
E.cumanense
E.haenense 410
E.foetidum 699
E.gracilipes 946
E.cf.ateyermarkii
E.cataractarum 408
E.substriatum
E.pungens
E.pruinosum
E.nelson--rosae
E.lentiscifolium
E.martii
E.myrsinites
E.rufum 704
E.englerl
E.tortuosum
E.raimondii
E.decidUum
E.argentinum
E.macrochaetum
E.fimbriatum 1041
E.maracaiense
E.rosulifolium 1063
E.subrotundum 1036
E.passerinum
E.undulatum
E.nummularia
E.pulchrum
E.ochranthum
E.macrocalyx
E.hamiguron
E.umbu
E.ambiguum
E.mattossilvae
E.gaudichaudii
E.cincinnatum
E.subsessile
E.cuspidifolium
E.glaziovi
E.bradleanum 1223
E.cryptanthum
E.coelophlebium
E.coelophlebium 1089
E.hanguifolium
E.nobile
E.compressum
E.tenue
E.plowmanii
E.polygonoides
E.sp.nov aff.mikanii
E.almonia
E.nordestinum
E.affine
E.andrei
E.dlatortum
E.catharinense
E.bucus
E.mexicanum
E.crinocense
E.oxycarpum
E.haughii
E.Spruceanum
E.vasquezii
E.incrassatum
E.dilloni
E.sp.nov dmw530
E.ruryi
E.anguifugum
E.mamacoca
E.schunkei
E.ulei 509
E.sp.nov dmw619
E.guanchezii
E.kappleriarum
E.leptoneurum
E.bicolor
E.cordatoovatum
E.vaccinifolium
E.coriaceum 369
E.lindemanii
E.canajasense
E.ligustrinum
E.tianguanum
E.stipulosum
E.arytonianum
E.revolutum
E.pelleterianum
E.campinense
E.oneophilum
E.schomburgkii
E.verricosum
E.aquamatum 1053
E.barbatum
E.amazonicum
E.fimbriatum 1309
E.subracemosum 1255
E.rimosum
E.suberosum
E.daphnites
E.citrifolium 693
E.mucronatum
E.campestre
E.testaceum
E.asperum
E.asamarum
E.pauciflorum
E.acuminatum
E.macrophyllum 713
E.durum

Figure C3: Topological comparison between ASTRAL III species tree (left) and MLcatLB tree (right) are shown by blue lines connecting tips. Short terminal branches on ASTRAL III tree are arbitrary as they represent a single individual for that given species.

Figure C4: ML-cat time-calibrated tree. X-axis is millions of years and shows geologic epochs. Red lines delimit the early Eocene climatic optimum. The gold star is the minimum 47.8 Ma calibration point for Neotropical *Erythroxylum.*

Figure C5: ML-catLB time-calibrated tree. X-axis is millions of years and shows geologic epochs. Red lines delimit the early Eocene climatic optimum. The gold star is the minimum 47.8 Ma calibration point for Neotropical *Erythroxylum.*

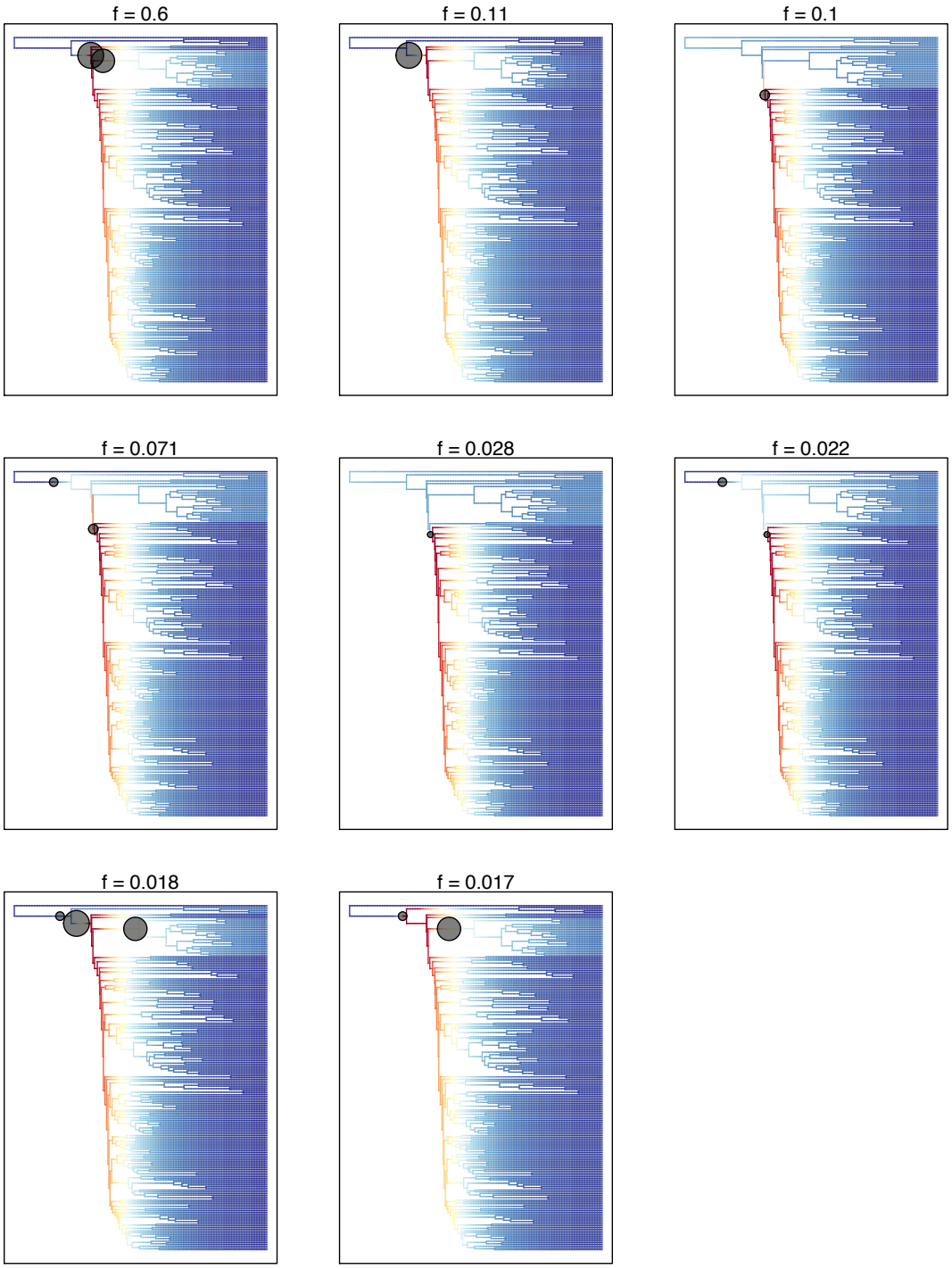


Figure C6: MLcat BAMM credible shift set. 8 shift configurations with rate shift locations and magnitudes indicated by circles. F-values indicate the frequency at which the shift configuration was sampled in the posterior.

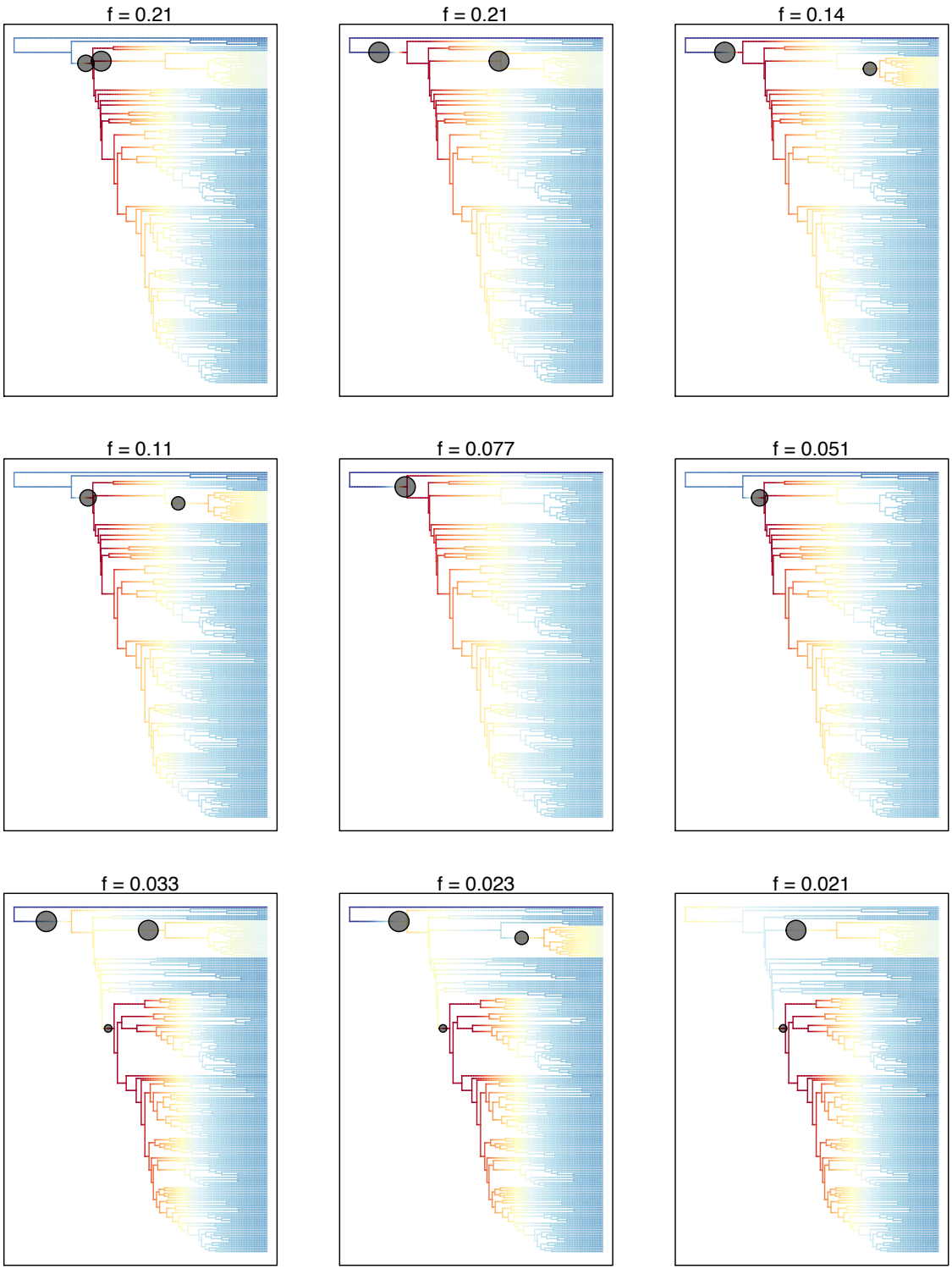


Figure C7: MLcatLB BAMM credible shift set with priors set to 10 expected shifts and a minimum clade size of 2. 8 shift configurations with rate shift locations and magnitudes indicated by circles. F-values indicate the frequency at which the shift configuration was sampled in the posterior.

Chapter III sample information:

sample code	species	Region-SouthAmerica (SA), Caribbean (Ca), Mesoamerica (Ma), Africa (Af), Madagascar (Md), Indo-Pacific (IP)	Biome: Rainforest (R), DryForest (DF), Savanna/ Grassland (SG)	Source: This study=1, White et al. 2019=2, White, Huang, et al. 2019=3	Herb:number	Collector:number	Country:Location
A.africanus	Aneulophus africanus	Af	R	1	MO	McPherson:16920	Gabon: Ogooue-Maritime
B.cylindrica	Bruguiera cylindrica	-	-	1	F:2054703	Frodin:s.n.	Malaysia:Lundu
C.brachiata	Carallia brachiata	-	-	1	MO	White:ECOF26	Unknown
C.elliptica	Cassipourea guianensis	-	-	1	F:2104817	Haber:6541	Costa Rica
E.acuminatum	E. acuminatum	SA	R	1	F:2320297	White:509	Peru
E.aff.acuminatum	E. aff. acuminatum	SA	R	1	F:2198812	Silveira:1459	Brazil:Acre
E.affine	E. affine	SA	R	1	HUEFS: 242658	Costa-Lima:2485	Brazil:Espiritu Santo
E.alaternifolium	E. alaternifolium	Ca	DF	1	F:186175	Britton:611	Cuba:Madruga
E.amazonicum	E. amazonicum	SA	R	2	F:2198813	Vasquez:24660	Brazil:Mato Grosso
E.ambiguum	E. ambiguum	SA	R	1	F:2072865	Ribas:181	Brazil:Parana
E.amplifolium	E. amplifolium	Md	R	1	TROPID: 100451815	Rasoazanany:56	Madagascar
E.amplum	E. amplum	SA	R	1	F:1854023	Gentry:22350	Peru:Loreto
E.andrei	E. andrei	SA	R	1	F:2034891	Plowman:1990	Brazil:Bahia
E.anguifugum	E. anguifugum	SA	R, SG	2	F:2179162	Schinini:31708	Paraguay: Amambay
E.areolatum_212	E. areolatum	Ca,MA	DF, R	2	F:1758721	Correll:45197	Bahamas
E.areolatum_462	E. areolatum	Ca,MA	DF, R	2	F:2171213	Jimenez:2004	Dominican Republic
E.argentinum	E. argentinum	SA	R, SG	2	F:1757134	Amaral Jr.:969	Brazil:São Paulo
E.australe	E. australe	IP	SG	2	BRI	Westaway:2439	Australia: Queensland
E.ayrtonianum	E. ayrtonianum	SA	SG	1	F:1995219	Plowman:9208	Brazil:Goias
E.barbatum	E. barbatum	SA	R, SG	1	F:1962364	Thomas:4729	Brazil:Mato Grosso
E_bequaertii	E. bequaertii	dropped	dropped	2	F:2135738	Vincent:6118	Belize
E.betulaceum_1028	E. betulaceum	SA	DF, SG	1	HUEFS:224038	Costa-Lima:1991	Brazil:Manha
E.betulaceum_259	E. betulaceum	SA	DF, SG	1	F:1955218	Stannard:36271	Brazil
E.bicolor	E. bicolor	SA	R	2	F:1955248	Furlan:6441	Brazil:Minas Gerais
E.bradeanum_1084	E. bradeanum	SA	R	1	F:1916492	Plowman:12950	Brazil:Rio de Janeiro
E.bradeanum_1223	E. bradeanum	SA	R	1	MO:5950028	Plowman:12950	Brazil
E.brennae	E. brennae	MA	R	1	F:2296150	McPherson:20896	Panama:Colon
E.brevipes	E. brevipes	Ca	DF	2	F:2171167	Garcia:6024	Dominican Republic

E.buxus	E. buxus	SA	SG	1	HUEFS:224059	Costa-Lima:2012	Brazil
E.caatingae	E. caatingae	SA	DF	2	F:2116327	Carvalho:3837	Brazil:Bahia
E.cambodianum	E. cambodianum	IP	R	1	F:2290165	Soejarto:14378	Vietnam
E.campestre	E. campestre	SA	SG	1	HUEFS:242739	Costa-Lima:2566	Brazil
E.campinense	E. campinense	SA	R	1	HUEFS:242950	Costa-Lima:2846	Brazil:Amazonas
E.capitatum	E. capitatum	Md	DF	1	MO:6489577	Phillipson:6065	Madagascar
E.carajasense	E. carajasense	SA	R	1	HUEFS:242583	Costa-Lima:2410	Brazil:Para
E.carthagenense	E. carthagenense	SA	DF	2	F:1938825	Gentry:47468	Colombia:Bolivar
E.cassinoides	E. cassinoides	SA	R	1	F:2324465	White:578	Colombia
E.cataractarum_408	E. cataractarum	SA	R	3	F:1862224	Davidse:15975	Venezuela:Apure
E.cataractarum_485	E. cataractarum	SA	R	3	F:1952199	Guanchez:2938	Venezuela: Amazonas
E.catharinense	E. catharinense	SA	R	1	F:1926265	Reitz:3253	Brazil:Santa Catarina
E.cf.steyermarkii	E. cf. steyermarkii	SA	DF	1	F:1853596	Plowman:7795	Venezuela:Sucre
E.cincinnatum	E. cincinnatum	SA	R	1	MO:2816859	Plowman:10174	Brazil
E.citrifolium_693	E. citrifolium	SA	R, SG	1	F:2324346	White:573	Colombia:Meta
E.citrifolium_1065	E. citrifolium	SA	R, SG	1	HUEFS:242887	Costa-Lima:2715	Brazil:Para
E.clarense	E. clarense	Ca	DF	1	F:719548	Jack:7912	Cuba
E.coelophlebium	E. coelophlebium	SA	R	2	F:1922037	Plowman:12900	Brazil:Rio de Janeiro
E.coelophlebium_1089	E. coelophlebium	SA	R	1	F:1872255	Martinelli:3182	Brazil:Rio de Janeiro
E.coffeifolium	E. coffeifolium	Md	R	1	MO:6185583	Lowry II:6684	Madagascar
E.columbinum	E. columbinum	SA	R	1	F:1884886	Carvalho:525	Brazil:Bahia
E.compressum	E. compressum	SA	R	1	F:1916868	Mattos Silva:1666	Brazil:Bahia
E.confusum_472	E. confusum	Ca,MA	R	2	F:1951069	Olmsted:s.n.	Mexico:Quintana Roo
E.confusum_471	E. confusum	Ca,MA	R	2	MO:2285708	Tapia:1891	Mexico:Yucatán
E.confusum_1107	E. confusum	Ca,MA	R	1	F:1967983	Contreras:8535	Guatemala :El Peten
E.cordatoovatum	E. cordato-ovatum	SA	R	1	F:1895058	Cid Ferreira:277	Brazil:Amazonas
Ecoriaceum_993	E. coriaceum	dropped	dropped	1	F:1477614	Ekman:2728	Cuba
E.corymbosum	E. corymbosum	Md	R	1	F:2253673	Rabenantoandro:250	Madagascar:Toliara
E.cryptanthum	E. cryptanthum	SA	R	1	F:1976058	Martinelli:11798	Brazil:Rio de Janeiro
E.cumanense	E. cumanense	SA	DF	2	F:1853471	Plowman:7654	Venezuela:D.F.
E.cuneifolium_1100	E. cuneifolium	SA	R, SG	1	F:1756843	Amaral Jr.:1532	Brazil:Sao Paulo
E.cuneifolium_1097	E. cuneifolium	SA	R, SG	1	F:2084018	Willian:136	Bolivia: Chuquisaca
E.cuspidifolium	E. cuspidifolium	SA	R	1	MO:5162689	Berry:4497	Brazil
E.daphnites	E. daphnites	SA	DF, R, SG	1	HUEFS:242697	Costa-Lima:2524	Brazil:Tocantins
E.deciduum	E. deciduum	SA	R, SG	2	F:1744854	Amaral Jr.:11481	Brazil:São Paulo
E.delagoense	E. delagoense	Af	SG	2	MO:2449460	Kemp:532	Swaziland

E.densum	E. densum	SA	R	2	F:2324373	White:543	Colombia: Magdalena
E.dillonii	E. dillonii	SA	DF	2	F:2320287	White:522	Peru:Cajamarca
E.discolor	E. discolor	Md	DF	1	MO:6227188	Randrianaivo:1390	Madagascar
E.distortum	E. distortum	SA	R	1	HUEFS:242653	Costa-Lima:2480	Brazil
E.divaricatum	E. divaricatum	SA	R	1	F:2324480	White:610	Colombia
E.domingense	E. domingense	Ca	DF	1	F:1953753	Zanoni:15346	Dominican Republic
E.durum	E. durum	SA	R, SG	1	INPA:137882	Cid Ferreira:6247	Brazil:Mato Grosso
E.elegans	E. elegans	Md	R	1	MO:5741761	Barthelat:560	Madagascar
E.ellipticum	E. ellipticum	IP	SG	2	BRI	-:-	Australia
E.emarginatum	E. emarginatum	Af	R,SG	1	MO:6444782	Mwangoka:6307	Tanzania
E.engleri	E. engleri	SA	R, SG	1	F:2199298	Walter:3223	Brazil:Brazilia.
E.excelsum	E. excelsum	Md	R	1	MO:6563911	Rakotoarivelo:573	Madagascar
E.fimbriatum_1041	E. fimbriatum	SA	R	1	F:2320249	White:486	Peru
E.fimbriatum_1309	E. fimbriatum	SA	R	1	F:1898612	Plowman:11250	Peru:Huanuco
Efimbriatum_1312	E. fimbriatum	dropped	dropped	1	F:1992282	Cid Ferreira:5112	Brazil:Amazonas
E.firmum	E. firmum	Md	R	1	MO:6609434	Ravelonarivo:4133	Madagascar
E.fischeri	E. fischeri	Af	R,SG	1	MO:5291781	Festo:1000	Tanzania
E.foetidum_699	E. foetidum	SA	DF	3	F:2324334	White:605	Colombia:Vichada
E.foetidum_468	E. foetidum	SA	DF	3	F:1987304	Guanchez:1772	Venezuela: Amazonas
E.frangulifolium	E. frangulifolium	SA	R	2	F:1916625	Plowman:12860	Brazil:Rio de Janeiro
E.gaudichaudii	E. gaudichaudii	SA	R	1	F:1950266	Plowman:13929	Brazil:Rio de Janeiro
E.gerrardii	E. gerrardii	Md	R	1	MO:5968474	Randriatafika:403	Madagascar
E.glaucum_567	E. glaucum	SA	DF	2	USDA:FOX221	-:-	Ecuador:P.N. Churute
E.glaucum_148	E. glaucum	SA	DF	2	F:1900484	Dodson:11369	Ecuador:Guayas
E.glazioui	E. glazioui	SA	R	2	F:1976059	Martinelli:11658	Brazil:Rio de Janeiro
E.gonoclados	E. gonocladum	SA	R, SG	1	F:2240089	Groppo Jr.:644	Brazil:Minas Gerais
E.gracilipes_809	E. gracilipes	SA	R	3	F:2000200	Nee:34664	Brazil:Rondonia
E.gracilipes_946	E. gracilipes	SA	R	3	F:2163874	Romoleroux:1963	Ecuador:Napo
E.grandifolium	E. grandifolium	SA	R	1	F:1873538	Plowman:10068	Brazil:Bahia
E.guanchezii	E. guanchezii	SA	DF	2	F:2324335	White:618	Colombia:Guainía
E.guatemalense	E. guatemalense	Ca	R	2	F:1967917	Lundell:19312	Guatemala:Izabal
E.hamigerum	E. hamigerum	SA	R	1	F:1910738	Carvalho:1365	Brazil:Bahia
E.haughtii	E. haughtii	SA	DF	2	F:1769140	Plowman:5360	Colombia:Cauca
E.havanense_152	E. havanense	Ca,MA,SA	DF	2	F:1921838	Breedlove:50501	Mexico:Chiapas
E.havanense_410	E. havanense	SA	DF	2	F:1930162	Berry :3651	Venezuela:D.F.
E.hondense_414	E. hondense	SA	DF	2	F:1986037	Silverstone-Sopkin:3097	Colombia:Valle
E.hondense_183	E. hondense	SA	DF	2	F:1902194	Berry:3921	Venezuela:Falcón

E.hypoleucum	E. hypoleucum	SA	R	1	INPA:206421	Clarke:8127	Brazil
E.impressum	E. impressum	SA	DF	1	F:1962326	Prance:29756	Brazil:Amazonas
E.incrassatum	E. incrassatum	Ca	R	2	KHD:63027	Islam:2-Sep	Jamaica
E.jamaicense	E. jamaicense	Ca	R	1	F:1471048	Proctor:11398	Jamaica
E.kapplerianum	E. Kapplerianum	SA	R	2	F:1924745	Strudwick:4084	Brazil:Pará
E.laetevirens	E. laetevirens	SA	DF, SG	1	HUEFS:209840	Costa-Lima:2260	Brazil
Elanceolatum_1247	E. lanceolatum	dropped	dropped	1	MO:2295415	Nicolson:288	Sri Lanka
E.lanceum	E. lanceum	Md	R	1	TROPID:100698734	Martial:288	Madagascar
E.lancifolium	E. lancifolium	SA	R	1	F:1871895	Martinelli:149	Brazil:Rio de Janeiro
E.latifolium	E. latifolium	IP	R	1	MO:6180848	Suzana:100569	Indonesia
E.laurifolium	E. laurifolium	Md	R	1	MO:2369901	Lorence:1490	Mauritius
E.leandrianum_732	E. leandrianum	Md	DF	2	F:2087947	Phillipson:2878	Madagascar: Toliara
E.leandrianum_1287	E. leandrianum	Md	DF	1	MO:6408053	Razakamalala:5935	Madagascar
E.lenticellosum	E. lenticellosum	SA	R	1	F:1871919	Ducke:11320	Brazil:Rio de Janeiro
E.leptoneurum	E. leptoneurum	SA	R	1	MO:3244929	de Granville:6480	Guyana
E.ligustrinum	E. ligustrinum	SA	R	1	HUEFS:242930	Costa-Lima:2448	Brazil:Amapa
E_ligustrinum	E. ligustrinum	dropped	dropped	2	F:1882658	Davidse:17961	Brazil:Para
E.lindemanii	E. lindemanii	SA	R	1	F:2077046	Lindeman:793	Suriname
E.lineolatum	E. lineolatum	SA, Ca	R	1	F:1993880	Pipoly:7908	Guyana
E.loefgrenii	E. loefgrenii	SA	DF	1	F:1910065	Fernandes:S.n.	Brazil
E.longifolium	E. longifolium	Md	R	1	F:1877226	Robertson:3713	Seychelles
Eloretense_1129	E. loretense	dropped	dropped	1	F:1949368	Encarnacion:1075	Peru:Loreto
E.macrocalyx	E. macrocalyx	SA	DF, R, SG	1	HUEFS:224052	Costa-Lima:2005	Brazil
E.macrochaetum	E. macrochaetum	SA	DF, R	1	HUEFS:243256	Costa-Lima:2183	Brazil:Bahia
E.macrophyllum_713	E. macrophyllum	SA	DF, R, SG	1	F:2324347	White:575	Colombia
E.macrophyllum_1220	E. macrophyllum	MA,SA	R	1	MO:6356781	Velasco-Sinaca:644	Mexico
E.macrophyllum_1163	E. macrophyllum	SA	DF, R, SG	1	F:1903555	Lindeman:277	Suriname
E.macrophyllum_1162	E. macrophyllum	SA	DF, R, SG	1	F:2033387	Mori:21153	French Guiana
E.magnoliifolium	E. magnoliifolium	SA	DF, R, SG	1	MO:2816852	Plowman:10101	Brazil
E.mamacoca	E. mamacoca	SA	R	2	F:1897281	Plowman:11740	Peru: Huánuco
E.mannii	E. mannii	Af	R	1	MO:4664633	Jongkind:1582	Ghana
E.maracasense	E. maracasense	SA	DF, SG	1	HUEFS:224029	Costa-Lima:1982	Brazil
E.martii	E. martii	SA	R	2	F:1960559	dos Santos:4012	Brazil:Bahia
E.mattossilvae	E. mattos-silvae	SA	R	1	F:1944045	Plowman:13958	Brazil:Bahia
E.mexicanum	E. mexicanum	MA	DF	2	F:1992215	Plowman:14546	Mexico: Jalisco
E.microphyllum_1313	E. microphyllum	SA	R, SG	1	F:2179166	Schinini:29948	Paraguay
E.microphyllum_1314	E. microphyllum	SA	R, SG	1	F:1910314	Oliveira:358	Brazil:Brazil

E.minutifolium	E. minutifolium	Ca	DF	1	LE	Howard:139	Cuba
E.mocquerysii	E. mocquerysii	Md	R	1	MO:5843370	Rabenantoandro: 1440	Madagascar
E.monogynum	E. monogynum	IP	DF	1	LE	Duthie:9328	India?
E.mucronatum	E. mucronatum	SA	R	1	F:2320309	White:494	Peru
E.myrsinites	E. myrsinites	SA	R	1	FUEL:28710	Francisco:s.n.	Brazil
E.nelson-rosae	E. nelson-rosae	SA	R	1	F:2072859	Silva:1363	Brazil:Para
E.nitidulum	E. nitidulum	Md	R	2	F:2116948	Zarucchi:7425	Madagascar: Toamasina
E.nobile	E. nobile	SA	R	1	HUEFS:242684	Costa-Lima:2511	Brazil:Espiritu Santo
E.nordestinum	E. nordestinum	SA	DF	1	HUEFS:242852	Costa-Lima:2680	Brazil:Algoas
E.nossibeense	E. nossibeense	Md	R	1	MO:6196803	Labat:3212	Comores
E.nummularia	E. nummularia	SA	DF, R	1	HUEFS:242802	Costa-Lima:2630	Brazil:Bahia
Eobtusifolium_1250	E. obtusifolium	dropped	dropped	1	MO:4379766	Jayasuriya:2968	Sri Lanka
E.ochranthum	E. ochranthum	SA	R	1	MO:3252385	Plowman:13962	Brazil
E.oreophilum	E. oreophilum	SA	R	1	F:1955105	Pipo:7204	Venezuela: Bolivar
E.orinocense	E. orinocense	SA	DF	2	F:2324385	White:613	Colombia: Casanaré
Eovalifolium_1173	E. ovalifolium	dropped	dropped	1	F:1916622	Plowman:12842	Brazil:Rio de Janeiro
E.oxycarpum	E. oxycarpum	SA	DF	2	F:2324360	White:563	Colombia: Vichada
E.oxypetalum	E. oxypetalum	SA	DF, SG	1	F:1842073	Irwin:23883	Brazil:Minas Gerais
E.panamense	E. panamense	SA	R	1	F:2324474	White:594	Colombia
E.passerinum	E. passerinum	SA	R	1	UPF:78184	Costa-Lima:644	Brazil
E.patentissimum	E. patentissimum	SA	DF	1	F:1871925	Silva Costa:160	Brazil:Mato Grosso
E.pauciflorum	E. pauciflorum	SA	R	1	MO:2909699	CayoPérez:2129	Bolivia
E.pedicellare	E. pedicellare	Ca	DF	1	F:1477615	Ekman:2310	Cuba:Oriente
E.pelleterianum	E. pelleterianum	SA	R, SG	1	HUEFS:242800	Costa-Lima:2628	Brazil:Bahia
Eflexuosum_1226	E. pelleterianum	dropped	dropped	1	MO:2695878	Irwin:29653	Brazil
E.pervillei	E. pervillei	Md	DF	1	F:2284951	Randrianaivo:923	Madagascar: Toliara
Epetraecaballi_1238	E. petrae-caballi	dropped	dropped	1	MO:3396650	Noblick:3361	Brazil
E.pictum_1280	E. pictum	Af	R,DF	1	MO:3285073	Hutchings:1050	South Africa
E.pictum_EP05	E. pictum	Af	R,DF	1	F	Pirani:5495	South Africa
E.platyclados_737	E. platyclados	Af, Md	DF	2	MO:4249313	Robertson:5409	Kenya
E.platyclados_1256	E. platyclados	Af, Md	DF	1	MO:6459708	Festo:2300	Kenya
Eacranthum_973	E. platycladum	dropped	dropped	1	LE	Frazier:58	Seychelles
E.plowmanianum	E. plowmanianum	SA	R	2	F:2324469	White:582	Colombia: Antioquia
E.plowmanii	E. plowmanii	SA	R	1	RB:509622	Silva:121	Brazil
E.polygonoides	E. polygonoides	SA	DF, R, SG	1	HUEFS:226983	Costa-Lima:1315	Brazil:Bahia
Epopayanense_1310	E. popayanense	dropped	dropped	1	F:2065655	Callejas:5471	Colombia: Antioquia
E.pruinosum	E. pruinosum	SA	R, SG	2	F:2199294	Santos:62	Brazil:Goias

E.pulchrum	E. pulchrum	SA	R	1	MO:6462731	Braga:3553	Brazil
E.pungens	E. pungens	SA	D	1	F:1895060	Daly:716	Brazil:Maranhão
E.pyrifolium	E. pyrifolium	Md	R	1	F:1982513	Dorr:4620	Madagascar: Tamatave
E.raimondii	E. raimondii	SA	R	2	USDA:B453SS	-:-	Peru
E.reticulatum	E. reticulatum	Ca	DF	2	F:1774549	Correll:49694	Bahamas
E.revolutum	E. revolutum	SA	DF, R	1	F:1954826	Pirani:886	Brazil:Mato Grosso
E.rignyanum_1283	E. rignyanum	Md	R	1	MO:6214797	Ratovoson:1317	Madagascar
E.rignyanum_1209	E. rignyanum	Md	R	1	F:2276381	McPherson:18775	Madagascar: Antsiranana
E.rimosum	E. rimosum	SA	DF, R	1	UPF:78185	Costa-Lima:646	Brazil:Sergipe
E.riparium	E. riparium	SA	R	1	CEPEC:144117	Araujo:211	Brazil
E.roigii	E. roigii	Ca	TC	1	F:460063	Shafer:13374	Cuba
E.roraime_ER13	E. roraimae	SA	R	2	MO:5072362	Chacon:681	Venezuela: Bolivar
E.roraime_369	E. roraimae	SA	R	2	MO:5072362	Chacon:681	Venezuela: Bolivar
E.rosuliferum_257	E. rosuliferum	SA	D	2	F:1916628	Plowman:12715	Brazil:Ceará
E.rosuliferum_1063	E. rosuliferum	SA	D	1	HUEFS:242816	Costa-Lima:2644	Brazil:Bahia
E.rotundifolium_234	E. rotundifolium	Ca,MA	DF	2	F:2149713	SalinTovar:6452	Mexico:Oaxaca
E.rotundifolium	E. rotundifolium	Ca,MA	DF	2	F:1935681	Lott:1738	Mexico:Jalisco
E.rotundifolium_1219	E. rotundifolium	Ca,MA	DF	1	MO:6425076	Elorsa:2210	Mexico
E.rotundifolium_986	E. rotundifolium	Ca,MA	DF	1	F:2278718	Veliz:14573	Guatemala
E.rufum_704	E. rufum	SA	DF	1	F:2324479	White:606	Colombia
E.rufum_1222	E. rufum	Ca	R, DF	1	MO:5942794	Garcia:6609	Dominican Republic
Erufum_965	E. rufum	dropped	dropped	1	IAN:156984	Rodriguez:791	Brazil:Bahia
E.ruizii_174	E. Ruizii	SA	DF	2	F:1973406	Plowman:14342	Ecuador: Manabí
E.ruizii_1245	E. ruizii	SA	DF	1	MO:6069936	Neill:116617	Ecuador
E.ruryi	E. ruryi	SA	R	2	F:2320256	White:479	Peru:Madre de Dios
Esantosii_1124	E. santosii	dropped	dropped	1	F:1839581	dos Santos:1618	Brazil
E.savannarum	E. savannarum	SA	SG	2	F:2324371	White:555	Colombia:Meta
E.schomburgkii	E. Schomburgkii	SA	R	2	F:2076646	Maas:7181	Guyana: Essequibo
E.schunkei	E. schunkei	SA	R	1	F:1688195	Schunke:2188	Peru:Ucayali
E.sechellarum	E. sechellarum	Md	R	1	F	Robertson:2704	Seychelles
E.seyrigii	E. seyrigii	Md	DF	1	MO:6648513	Luino:13	Madagascar
E.shatona	E. shatona	SA	DF	2	F:1774523	Martin:1851	Peru:San Martín
E.simonis	E. simonis	SA	R	1	HUEFS:224030	Costa-Lima:1983	Brazil:Bahia
E.socotranum	E. socotranum	Af	D	1	TROPID:100430854	Kilian:YP4163	Yemen:Socotra
E_sp.nov_Brewer	E. sp. (Brewer LA B.Hyland 13373)	IP	R	2	BRI	-:-	Australia: Queensland
E.sp.nov_dmw530	E. sp. (D.M. White 530)	SA	DF	1	F:2320281	White:530	Peru

E.sp.nov_dmw619	E. sp. (D.M. White 619)	SA	R	2	F:2324377	White:619	Colombia: Guainía
E.sp.nov_aff.mikanii	E. sp. aff. mikanii	SA	R	1	HUEFS:242654	Costa-Lima:2481	Brazil:Bahia
E.sp.nov_Cholm.Creek	E. sp. Cholmondely Creek (J.R.Clarkson 9367)	IP	SG	1	BRI	-:-	Australia
E.sp_NewCaledonia	E. sp. indet.	IP	DF	1	TROPID: 100880302	Lowry II:7413	New Caledonia
E.sp.nov_Splityard	E. sp. Splityard Creek (L.Pedley 5360)	IP	DF	1	BRI	-:-	Australia
E.sphaeranthum	E. sphaeranthum	Md	R	1	F:2284922	Martin Callmander: 560	Madagascar: Antsiranana
Espinescens_1000	E. spinescens	dropped	dropped	1	F:1477611	Ekman:6124	Cuba
E.splendidum	E. splendidum	SA	R	2	F:1910746	Carvalho:1125	Brazil:Bahia
E.Spruceanum	E. Spruceanum	SA	R	2	F:1944256	Zarucchi:3096	Brazil:Amazonas
E.squamatum_1035	E. squamatum	SA, Ca	R, SG	1	F:2320264	White:470	Peru
E.squamatum_1053	E. squamatum	SA	R, SG	1	F:2324475	White:595	Colombia
E.squamatum_1111	E. squamatum	SA, Ca	R, SG	1	HUEFS:242886	Costa-Lima:2714	Brazil:Para
E.squamatum_1177	E. squamatum	SA, Ca	R, SG	1	F:2251158	Forzza:2620	Brazil:Tocantins
E.stipulosum	E. stipulosum	SA	DF, SG	1	HUEFS:224049	Costa-Lima:2002	Brazil
E.striiflorum	E. striiflorum	Md	R	1	MO:6616691	Gautier:5694	Madagascar
E.suberosum	E. suberosum	SA	R, SG	1	HUEFS:242885	Costa-Lima:2713	Brazil:Para
E.suberosum. denudatum	E. suberosum var. denudatum	SA	R, SG	1	F:1842063	Anderson:36306	Brazil:Gerais
E.subracemosum_1235	E. subracemosum	SA	R, SG	1	MO:5932157	Nee:35024	Brazil:Rondonia
E.subracemosum_1185	E. subracemosum	SA	R, SG	1	F:2199452	Pereira da Silva:3519	Brazil:Goias
E.subrotundum_219	E. subrotundum	SA	R, SG	2	F:2283655	Paredes:91	Bolivia:La Paz
E.subrotundum_1036	E. subrotundum	SA	R, DF	1	F:2320284	White:526	Peru
E.subsessile	E. subsessile	SA	R	2	F:1982136	Fialho:3	Brazil:Rio de Janeiro
E.substriatum	E. substriatum	SA	R	1	F:1915443	Sobral:14006	Brazil:Rio Grande do Sul
E.tenue	E. tenue	SA	DF	1	F:1916664	Plowman:12812	Brazil:Bahia
E.testaceum	E. testaceum	SA	R	1	MO:3481417	Irwin:18705	Brazil
E.tianguanum	E. tianguanum	SA	DF, SG	1	HUEFS:223459	Costa-Lima:2267	Brazil
E.tikalense	E. tikalense	MA	R	1	F:1967981	Contreras:1051	Guatemala:El Peten
E.tortuosum	E. tortuosum	SA	SG	2	F:2302271	Mendes:292	Brazil:Minas Gerais
E.ulei_501	E. ulei	SA	DF, R	1	F:2258350	Araujo:1353	Bolivia
E.ulei_509	E. ulei	SA	DF, R	1	F:1926523	Plowman:4184	Colombia
E_ulei	E. ulei	dropped	dropped	2	F:2297680	Rojas:722	Peru:Amazonas
E.ulei.escalerense	E. ulei var. escalerense	SA	DF, R	1	F	Marcos-Rios:3080	Peru:San Martín
E.umbu	E. umbu	SA	R	2	F:2188284	Mendoça:50	Brazil:São Paulo
E.undulatum	E. undulatum	SA	DF	1	F:1859530	Paul Berry :3479	Venezuela: Caracas.

E.urbanii	E. urbanii	Ca	DF	1	F:2235714	Garcia:3917	Dominican Republic
E.vacciniifolium	E. vacciniifolium	SA	DF	1	F:1916650	Plowman:12717	Brazil:Ceara
E.vasquezii	E. vasquezii	SA	R	2	F:1994766	Spichiger:1983	Peru:Loreto
E.vernicosum	E. vernicosum	SA	R	2	F:1958896	Jensen-Jacobs:30480	Guyana
E.williamsii	E. williamsii	SA	DF	2	F:1933195	Plowman:13495	Colombia: Vichada
E.xerophilum	E. xerophilum	Md	D	1	F:1982518	Dorr:4539	Madagascar: Antananarivo
E.zambesiacum	E. zambesiacum	Af	SG	1	MO:5725244	Smith:4236	Botswana
Ezeylanicum_1251	E. zeylanicum	dropped	dropped	1	MO:2351290	Waas:1355	Sri Lanka
N.capense	N. capense	Af	R	1	MO:3295085	Wyk:B2529	South Africa
N.kaessneri	N. kaessneri	Af	DF	1	MO:6455800	Festo:2744	Kenya
N.zuluense	Necaropetalum zuluense	Af	R	2	F:1977936	Ansell:s.n.	South Africa: Zululand
P.congolense	Pinacopodium congolense	Af	R	2	MO:4238697	McPherson:15533	Gabon
R.mangle	Rhizophora mangle	-	-	1	COLO:	Islam:6	Unknown

Appendix D: Sample information for Chapter IV

sample code	species	Herb:number	Collector:number	Country:Location
E.anguifugum	E. anguifugum	F:2179162	Schinini:31708	Paraguay:Amambay
E.dillonii	E. dillonii	F:2320287	White:522	Peru:Cajamarca
E.haughtii	E. haughtii	F:1769140	Plowman:5360	Colombia:Cauca
E.incrassatum	E. incrassatum	KHD:63027	Islam:2-Sep	Jamaica
E.mamacoca	E. mamacoca	F:1897281	Plowman:11740	Peru:Huánuco
E.mexicanum	E. mexicanum	F:1992215	Plowman:14546	Mexico:Jalisco
E.orinocense	E. orinocense	F:2324385	White:613	Colombia:Casanaré
E.oxycarpum	E. oxycarpum	F:2324360	White:563	Colombia:Vichada
E.ruryi	E. ruryi	F:2320256	White:479	Peru:Madre de Dios
E.schunkei	E. schunkei	F:1688195	Schunke:2188	Peru:Ucayali
E.sp.nov_dmw530	E. sp. (D.M. White 530)	F:2320281	White:530	Peru
E.Spruceanum	E. Spruceanum	F:1944256	Zarucchi:3096	Brazil:Amazonas
E.ulei_501	E. ulei	F:2258350	Araujo:1353	Bolivia
E.ulei_509	E. ulei	F:1926523	Plowman:4184	Colombia
E.ulei.escalerense	E. ulei var. escalerense	F	Marcos-Rios:3080	Peru:San Martín
E.vasquezii	E. vasquezii	F:1994766	Spichiger:1983	Peru:Loreto

COPYRITE PERMISSION TO REPRINT CHAPTER I

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Mar 19, 2019

This Agreement between University of Illinois at Chicago -- Dawson White ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4552760409141
License date	Mar 19, 2019
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	AMERICAN JOURNAL OF BOTANY
Licensed Content Title	Phylogenetic inference in section Archerythroxylum informs taxonomy, biogeography, and the domestication of coca (Erythroxylum species)
Licensed Content Author	Dawson M. White, Melissa B. Islam, Roberta J. Mason-Gamer
Licensed Content Date	Jan 10, 2019
Licensed Content Volume	106
Licensed Content Issue	1
Licensed Content Pages	12
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Title of your thesis / dissertation	Biogeography, Diversification, and Domestication in the Coca Family (Erythroxylaceae)
Expected completion date	May 2019
Expected size (number of pages)	130
Requestor Location	University of Illinois at Chicago 845 W. Taylor St. #3272 Biological Sciences MC 066 CHICAGO, IL 60608 United States Attn: Robie Mason-Gamer
Publisher Tax ID	EU826007151
Total	0.00 USD
Terms and Conditions	

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction

(along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.

- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order,** is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner.**For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts,** You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY,

EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.

- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.

- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and

WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.

- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The Creative Commons Attribution Non-Commercial (CC-BY-NC)License permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library

http://olabout.wiley.com/WileyCDA/Section/id-410895.html

Other Terms and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

COPYRITE PERMISSION TO REPRINT BIOME MAP IN FIGURE 7

ELSEVIER LICENSE
TERMS AND CONDITIONS

Mar 19, 2019

This Agreement between University of Illinois at Chicago -- Dawson White ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4552620657666
License date	Mar 19, 2019
Licensed Content Publisher	Elsevier
Licensed Content Publication	Current Biology
Licensed Content Title	Tropical savannas and dry forests
Licensed Content Author	R. Toby Pennington,Caroline E.R. Lehmann,Lucy M. Rowland
Licensed Content Date	7 May 2018
Licensed Content Volume	28
Licensed Content Issue	9
Licensed Content Pages	5
Start Page	R541
End Page	R545
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Figure 1
Title of your thesis/dissertation	Biogeography, Diversification, and Domestication in the Coca Family (Erythroxylaceae)
Expected completion date	May 2019
Estimated size (number of pages)	130
Requestor Location	University of Illinois at Chicago 845 W. Taylor St. #3272 Biological Sciences MC 066 CHICAGO, IL 60608 United States Attn: Robie Mason-Gamer
Publisher Tax ID	98-0397604
Total	0.00 USD
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com). No modifications can be made to any Lancet figures/tables and they must be reproduced in full.
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:
15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.
16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.
17. **For journal authors:** the following clauses are applicable in addition to the above:
Preprints:
A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).
Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).
If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available

version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.
Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.
Authors can share their accepted author manuscript:

- immediately
  - via their non-commercial person homepage or blog
  - by updating a preprint in arXiv or RePEc with the accepted manuscript
  - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
  - directly by providing copies to their students or to research collaborators for their personal use
  - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- After the embargo period
  - via non-commercial hosting platforms such as their institutional repository
  - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.
Policies for sharing publishing journal articles differ for subscription and gold open access articles:
Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.
Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.
If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.
Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.
Please refer to Elsevier's posting policy for further information.
18. **For book authors** the following clauses are applicable in addition to the above:
Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.

Terms & Conditions applicable to all Open Access articles published with Elsevier:

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.

The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

Additional Terms & Conditions applicable to each Creative Commons user license:

CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by/4.0.

CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at http://creativecommons.org/licenses/by-nc-sa/4.0.

CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by-nc-nd/4.0.

Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. **Other Conditions**:

v1.9

VITA

NAME:	Dawson Martin White
EDUCATION:	B.S., Colorado State University, Fort Collins, CO, 2011
	Ph.D., University of Illinois at Chicago, Chicago, IL, 2019
TEACHING:	Department of Biological Sciences, University of Illinois at Chicago; Biology of Populations and Communities. 2012-2013.
	Department of Biological Sciences, University of Illinois at Chicago; Fundamentals of Genetics Laboratory. 2017-2018.
HONORS:	Graduate Research Fellowship, National Science Foundation, 2012-2017.
PRESENTATIONS:	White, D.M. & Roberta J. Mason-Gamer: Old specimens, new results: Preliminary analysis of Coca domestication from intrageneric relationships of Neotropical *Erythroxylum.* Botany 2013, New Orleans, LA. Poster.
	White, D.M. & Roberta J. Mason-Gamer: Genetic analysis of the domestication of coca. Department Seminar, University of Illinois at Chicago. 2015.
	White, D.M. & Roberta J. Mason-Gamer: Diversification and diversity gradients in the Coca family of flowering plants. Department Seminar, University of Illinois at Chicago. 2016.
	Mason-Gamer, Roberta J. & D.M. White: Phylogenetic relationships of Triticeae estimated from the chloroplast genome and 123 nuclear loci. Botany 2016, Savannah, GA.
	White, D.M. & Roberta J. Mason-Gamer: The origins of Coca: identifying progenitors of a South American shrub (*Erythroxylum* spp.). Evolution 2017, Portland, OR.
	White, D.M. & Roberta J. Mason-Gamer: Phylogeographic analysis of the domestication of coca (*Erythroxylum* spp.). Botany 2018, Rochester, MN.
PUBLICATIONS:	White, D.M., M.B. Islam, and R.J. Mason-Gamer. 2019. Phylogenetic inference in section Archerythroxylum informs taxonomy, biogeography, and the domestication of coca (*Erythroxylum* species). *American Journal of Botany* 106: 1–12.
PROFESSIONAL MEMBERSHIP:	Botanical Society of America Colorado Native Plant Society Society of Systematic Biologists