TRACING THE EVOLUTIONARY HISTORY OF COCA (*ERYTHROXYLUM*)

by

MELISSA ISLAM

B.S, Colorado State University, 1997

M.S., Colorado State University, 2005

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Ecology & Evolutionary Biology

2011

This thesis entitled:

Tracing the Evolutionary History of Coca (*Erythroxylum*)

written by Melissa Islam

has been approved for the Department of Ecology & Evolutionary Biology

Tom A. Ranker, Chair

Susan Beatty

Pam K. Diggle

Robert P. Guralnick

Andrew P. Martin

Date: May 2, 2011

The final copy of this thesis has been examined by the signatories, and we

Find that both the content and the form meet acceptable presentation standards

Of scholarly work in the above mentioned discipline.

Islam, Melissa (Ph.D., Ecology & Evolutionary Biology)

Tracing the Evolutionary History of Coca (*Erythroxylum*)

Thesis directed by Professor Tom. A. Ranker

Abstract

The tropical, flowering shrubs in the genus *Erythroxylum* (230 species) are commonly known as coca. For 8,000 years, two *Erythroxylum* species have been exploited for their production of cocaine. Coca plants remain culturally significant and economically important. Yet, their origin and evolutionary history, especially of the cultivated species, have not been explored in a modern phylogenetic framework. Molecular and morphological characters were used to reconstruct the evolutionary history of the genus to determine the intrageneric relationships within *Erythroxylum* and origin of cocaine synthesis, to infer the evolutionary history of the cultivated species and the number of domestication events, and to determine the origin and diversification of the Caribbean *Erythroxylum*.

Distribution of *Erythroxylum* across the tropics resembles a Northern Tropical Gondwanan distribution resulting from the break-up of Gondwana. The age of divergence of *Erythroxylum* along with the hypothesized origin of the South American species supports a boreotropical dispersal rather than a Gondwanan origin. Eighty years ago, the genus was split into 19 sections based on a few morphological characters. Many of these sections are heterogeneous and likely not predictive of evolutionary relationships. Testing the monophyly of these sections revealed only one monophyletic section. Strongly supported clades were given informal names with at least one morphological feature identified that is shared by the members of each clade. Multiple origins of cocaine synthesis were inferred.

For the cultivated species, *E. novogranatense,* was hypothesized to be derived through human selection from *E. coca*. These two species were supported as distinct species indicating at least two domestication events of wild *Erythroxylum* species. Their closest wild relatives are section *Archerythroxylum* species with deciduous leaves.

The Caribbean *Erythroxylum* species arose from multiple colonization events, including three long-distance dispersal events from northern South America and one from Mexico or Central America. Species from the same island were not monophyletic indicating species on the same island originated independently from either other islands or continental species, except for a large Cuban and Hispañiolan clade. This clade indicates a radiation within Cuba with one or two dispersal events to Hispañiola. Updated taxonomy and occurrence of the Caribbean species is reviewed.

Table of Contents

List of Tables

List of Figures

CHAPTER ONE: Phylogeny and systematics of *Erythroxylum* (Erythroxylaceae) and the origin of cocaine synthesis

Abstract *Erythroxylum* is a tropical, flowering plant genus of 230 species that is distributed across the tropics with highest diversity in the Neotropics. Eighty years ago, the genus was split into 19 sections based on a few morphological characters. Many of these sections are heterogeneous and not predictive of evolutionary relationships let alone helpful in classifying the species, as a number of sections are monotypic. I tested the evolutionary validity of 10 sections with 66 species using morphological and molecular characters, *rpl32-trn*L and ITS, in a phylogenetic framework to determine a division of the genus that reflects evolutionary history. Based on the phylogeny, only one of the sections is monophyletic. Strongly supported clades were given informal names with at least one morphological feature identified that is shared by the members of each clade. Taxonomy of certain species complexes is discussed in light of the phylogenies. Presence of cocaine and other alkaloids for 44 species were included in the phylogenetic reconstructions, and multiple origins of cocaine synthesis are inferred from the phylogeny, with reservations.

Introduction

The tropical flowering shrubs and trees in the genus *Erythroxylum* P. Browne (Erythroxylaceae Kunth; 230 species) are ecologically and geographically diverse and occur throughout the world with the highest species diversity in the Neotropics (Rury 1982; Plowman 1991; Heywood 1993). The genus is commonly known as coca. Several species of coca are exploited for their production of a secondary metabolite, the tropane alkaloid cocaine (25 of 51 sampled; Bieri et al. 2006). The recreational drug cocaine is typically the salt form of the chemical present in the leaves of coca, and both chemicals are simply referred to as cocaine. Cocaine is related to other economically important tropane alkaloids such as nicotine and scopolamine (Solanaceae). The synthesis of cocaine is only known to occur in *Erythroxylum*.

Based on morphology (for overview see Hegnauer 1981; Rury 1982) and molecular analyses, *Erythroxylum* has been placed in the family Erythroxylaceae with three other genera as sister to

Rhizophoraceae (Malpighiales; Setoguchi et al. 1999; Schwarzbach & Ricklefs 2000; Wurdack & Davis 2009). *Erythroxylum* has been classified into 19 sections (Schulz 1907a, 1931). Schulz's sections are based on a few characters primarily relating to the stipules, calyces, and styles, and some sections are believed to represent "artificial assemblages of species based on characters which are now known to be unreliable and highly variable." (p. 652, Plowman and Rivier 1983). Using wood and leaf anatomical features, Rury (1982) investigated Schulz's (1907a, 1931) sections and found that most were heterogeneous or variable morphologically, which is consistent with other studies (Payens 1958; Plowman 1989; Rury 1982; Plowman and Rivier 1983; Loiola 2001; Oviedo 2002; Emche et al. 2011). No one has proposed nomenclatural changes to Schulz's classification and/ or explicitly tested Schulz's sections in a phylogenetic framework. No phylogenetic hypotheses exist for *Erythroxylum* to test Schulz's sections, and, no studies have attempted to reconstruct the ancestral history of the production of cocaine and its precursors in *Erythroxylum*. This study explicitly tests the monophyly of Schulz's sections, constructed a framework for a new division of the genus based on evolutionary lineages, and informs taxonomic questions regarding species delimitation within previously identified species complexes. In addition to a better understanding of the evolutionary relationships in *Erythroxylum,* this study investigates the evolution of cocaine biosynthesis within a phylogenetic framework.

Materials and Methods

Plant material. Five species from the sister family, Rhizophoraceae, were sampled along with three species representing each of the other genera in Erythroxylaceae. Additional species from these three genera in Erythroxylaceae were sampled, but the herbarium material for these species was poor and no amplifiable DNA was extracted after multiple attempts using a variety of extraction protocols.

Erythroxylum is classified into 19 sections (Schulz 1907a, 1931). Schulz's section *Heterogyne* includes the type of the genus, E. areolatum, and the section's name should be Erythroxylum to reflect the inclusion of the type species (Payens 1958; Loiola 2001). To test the monophyly of Schulz's sections, at least two species from 13 of the sections were sampled except for *Megalophyllum, Schistophyllum*, and *Oxystigma* in each of which only one species was sampled (Table 1.1). No species were sampled from the

monotypic sections *Pogonophorum*, *Mastigophorum*, *Melanocladus*, or *Sethia* or the sections *Coelocarpus* and *Venelia* because of the inability to successfully amplify DNA. When possible, multiple accessions of a species were sampled for a total of 119 accessions including the outgroup species for the combined analyses (Appendix 1).

Table 1.1 Total number of species sampled from each of Schulz's (1907a, 1931) sections and the geographical region of each. * indicates a monotypic section. [§]Schulz section *Heterogyne*.

Section #	Schulz Section	Total # species in analyses	Region
I	**Pogonophorum*	0	Americas
II	*Macrocalyx*	3	Americas
III	*Rhabdophyllum*	9	Americas
IV	*Leptogramme*	3	Americas
V	[§]*Erythroxylum*	10	Americas
VI	*Archerythroxylum*	21	Americas
VII	*Megalophyllum*	1	Americas
VIII	**Mastigophorum*	0	Americas
IX	*Microphyllum*	2	Americas
X	**Melanocladus*	0	Africa
XI	*Gonocladus*	2	Madagascar
XII	**Sethia*	0	India
XIII	*Lagynocarpus*	2	Africa/Madagascar
XIV	*Coelocarpus*	0	Australia/South East Asia/South Africa
XV	*Eurysepalum*	2	Madagascar
XVI	*Venelia*	0	Madagascar
XVII	*Pachylobus*	2	Africa
XVIII	**Schistophyllum*	1	Madagascar
XIX	**Oxystigma*	1	South East Asia
	Outgroup	7	
	Total:	**66**	

Specimens for DNA extraction and coding of morphological characters were collected in the field or by acquiring specimens from the following herbaria: Colorado State University (CS), Herbario Nacional Colombiano (COL), The Field Museum (F), New York Botanical Garden Herbarium (NY), Missouri Botanical Garden Herbarium (MO), and Fairchild Tropical Garden Herbarium (FTG). The Drug Enforcement Administration (DEA) classifies coca leaves (*E. coca* var. *coca*) as a Schedule II drug, and the DEA permit necessary for acquiring and keeping coca leaves for DNA extraction (DEA #

RU0357930) was obtained. Several accessions of the cultivated coca were sampled from the USDA ARS, Sustainable Perennial Crops Laboratory, Beltsville, MD.

Genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen Inc., Valencia, CA, USA) or the GenCatch Plant Genomic DNA Purification System (EPOCH Biolabs, Sugarland, TX, USA) with modifications following Johnson et al. (2003) for some species. Because of broad geographic sampling for this study, most species were sampled from herbarium specimens.

Morphological characters. Twenty-eight morphological and alkaloid characters were coded and analyzed (Appendix 2). Two characters, habitat and Schulz's sections (Appendix 2), were not included in analyses but mapped on to the Most Parsimonious Trees (MPTs). Nine characters were from Rury (1982), five from Oviedo (2002), and four represent alkaloids. The remaining characters were based on features deemed important for delineation of the sections by Schulz (1907a) and/or by other workers (Rury 1982; Plowman & Hensold 2004). Characters were coded from herbarium specimens, freshly collected specimens, and the literature (Baker 1883; Schulz 1907a; Corbishley 1919; Phillips 1935; Machado 1972; Rury 1982; Plowman 1991; Plowman 1984c; Zappi 1995; Plowman & Berry 1999; Felger et al. 2001; Loiola 2001; Plowman 2001; Heald 2002; Plowman & Hensold; Barrie & Plowman, unpublished).

For the alkaloid characters, two pathways were investigated for the presence or absence of secondary chemicals. Both biosynthetic pathways begin with the amino acid L-ornithine and after a series of reactions the pathway splits into at least two potentially competing pathways derived from the N-methyl pyrrolinium cation. One pathway leads to the production of the drug atropine. The presence of hygrine and tropinone from this pathway were analyzed to investigate if the presence of this pathway effects cocaine occurrence. Hygrine and tropinone are known to occur in *Erythroxylum* (Bieri et al. 2006). The second pathway from the N-methyl pyrrolinium cation leads to cocaine.

Thirty-seven species were sampled from silica-dried material or herbarium specimens and initially tested for the presence of the seven alkaloids, hygrine, tropinone, anhydrous methyl ecognine, methyl ecgonine, anhydrous cis-cinnamoylcocaine, cis-cinnamoylcocaine, and cocaine. Precursors were selected based on cocaine's proposed biosynthetic pathway (Leete et al. 1991). Multiple accessions for some

species were sampled for a total of 69 samples. The alkaloid extraction and gas chromatography/mass spectroscopy (GC/MS) analyses were performed by Dr. Marianne Poxleitner's lab at Gonzaga University according to the protocols below (provided by Dr. Poxleitner).

For herbarium or silica-dried material, alkaloids were extracted from dry *Erythroxylum* leaf tissue using a method provided by Stephen Emche, USDA, Beltsville, MD. When available, one hundred milligrams of dry *Erythroxylum* leaf tissue was ground to a fine powder using a mortar and pestle. The powder was transferred to a 10mL Reactivial (#13225, Pierce, Rockford, IL, 61105), mixed with 7mL 95% ethanol (EtOH) and incubated at room temperature for 10 min. A Reacti-Therm III Heating/Stirring Module (#18935, Pierce) was preheated to 90°C and the vials inserted. The vials were refluxed (not boiled), while stirring for 25 min. The samples were cooled at room temperature and allowed to sit in the vials overnight to optimize extraction. Samples were then Rotovaped the next day to dryness at 45°C (Johnson et al. 1996), and redissolved in 2mL 100% methanol (MeOH, HPLC grade) and filtered through a 0.2um syringe filter. If less than 100mg of leaf tissue was used, the resulting extracted alkaloids were resuspended in an appropriate quantity of MeOH. All of the extracts were green in color except for *E. rufum* samples, which were dark red, most likely due to high quantities of carotenoids, as detected by GC/MS. The samples were stored in airtight vials at 20°C.

For GC/MS analysis, one microliter of extracted alkaloids was injected into a Hewlett-Packard GC/MS: 5890 Series II Plus 6C using a 10ul fine-point needle syringe. Alkaloids were separated using a new Restek RTK-5 (#10223) 30m x 0.25mm dimethyl polysiloxane column with 0.25um bead size following Johnson et al. (1996). In short, the oven was initially set at 70°C for one minute and then increased to 5°C every minute until 285°C was reached, and held for 15 minutes. Alkaloids were detected using an HP 5972 Series Mass Detector. Chromatograph comparison standards were used to confirm the peak analysis of cis-cinnamoylcocaine, methyl ecgonine, and tropinone. A methanol blank was used between each sample to prevent cross contamination of samples and false positive readings. GC/MS library comparisons were used for all other alkaloids considered. Spectra were evaluated for the presence of hygrine, methyl ecgonine, cis-cinnamoylcocaine, and cocaine.

Only qualitative data (presence or absence) of each of alkaloid were used in the phylogenetic analyses since the material sampled was of different quality either by age and/or method of collection. Alkaloid production varies between individual plants (Rivier 1981) and individual leaves (Youssefi, Cooks, and McLaughlin 1979). The alkaloid characters were coded as multiple presence or absence characters.

Additional species were recorded from the literature for the presence or absence of cocaine (Aynilian et al. 1974, Holmstedt et al. 1977, Plowman and Rivier 1983, Bieri et al. 2006, review Oliveira et al. 2010). Although additional studies have tested for the presence of cocaine in *Erythroxylum* (review Novák et al. 1984), only results from the studies listed above were used. *Erythroxylum* species are difficult to identify from herbarium material and before the 1970s the taxonomy of *Erythroxylum* was less well known or understood than after the 1970s. Therefore, only those species that had vouchered samples identified by an *Erythroxylum* authority and for which multiple studies obtained a consistent result were recorded. Results listed in Oliveira et al. (2010) were not used directly but as a source for primary literature. A two-tailed Fisher's exact test compared the presence or absence of cocaine depending if the material was freshly-dried or herbarium material for the samples tested by Gonzaga University.

Molecular characters. Total genomic DNA was extracted from herbarium and silica-dried material. Two loci were amplified and sequenced, one from the nuclear genome (nuclear ribosomal internal transcribed spacer (ITS)) and one locus from the chloroplast (plastid) genome (*rpl32-trn*L intergenic spacer (*rpl32-trn*L)). An additional twenty-three loci from the plastid, nuclear, and mitochondrial genomes (Taberlet et al. 1991; Strand et al. 1997; Sang et al. 1997; Kuzoff et al. 1998; Mathews & Donoghue 1999; Davis et al. 2001; Howarth & Baum 2002; Davis & Wurdack 2004; Fan et al. 2004; Samuel et al. 2005; Shaw et al. 2005; Weese & Johnson 2005; Davis et al. 2007; Shaw et al. 2007; Kulji et al. 2007; M. Li et al. 2008) were tested for their ability to resolve the phylogeny based on the number of parsimony informative (PI) characters for five to ten distantly related species based on the preliminary ITS phylogeny (Table 1.2). Although the loci chosen were informative for similar intrageneric studies, for *Erythroxylum* these loci had few or no PI characters except for *idhB,* a region from the nuclear *idh* gene

(Weese & Johnson 2005). This locus was amplified for 44 accessions to test the monophyly of the cultivated species.

Table 1.2. Loci successfully amplified with few to no parsimony informative characters for five to ten *Erythroxylum* species. cpDNA=chloroplast genome; mt=mitochondrial genome.

Locus	**Source**	**Region**
3'rps15-5'trnK	Shaw et al. 2007	cpDNA
atpI-atpH	Shaw et al. 2007	cpDNA
ndhF	Davis et al. 2001	cpDNA
*ndhJ-trn*L (tabE)	Shaw et al. 2007	cpDNA
psbA-trnH intergenic spacer	Sang et al. 1997	cpDNA
psbD-trnT	Shaw et al. 2007	cpDNA
psbJ-petA	Shaw et al. 2007	cpDNA
ndhF-rpl32	Shaw et al. 2007	cpDNA
trnLF intergenic spacer	Taberlet et al. 1991	cpDNA
trnS-trnG-trnG	Shaw et al. 2007	cpDNA
trnV-ndhC	Shaw et al. 2007	cpDNA
ccmB	Davis et al. 2007	Mt
matR	Davis & Wurdack 2004	Mt
26S4F-950R	Kuzoff et al. 1998	nuclear
At103	M. Li et al. 2008	nuclear
G3pdh	Strand et al. 1997	nuclear
idhB	Weese and Johnson 2005	nuclear
myc-like	Fan et al. 2004	nuclear
ncpGS	Kulji et al. 2007	nuclear
NIA-i3	Howarth and Baum 2002	nuclear
PhyC	Mathews & Donoghue 1999	nuclear
PHYC exon 1	Samuel et al. 2005	nuclear

Target loci were amplified using standard PCR procedures. For the ITS locus, the profile started with an initial denaturation of 95°C (3 min), followed by 25 cycles of 95°C denaturation (30 sec), 50°C annealing (20 sec), and 72°C extension (1 min). The PCR profile for *rpl32-trn*L was modified from Shaw et al. 2007 with an initial denaturation of 80°C (5 min) followed by 40 cycles of 95°C denaturation (1 min), 50°C annealing (1 min), and 65°C extension (4 min) and a final extension at 65°C (5 min). *rpl32-trn*L was amplified with a combination of primers, *rpL32F-trn*L$^{(UAG)}$ (whole gene), *1327R-trn*L (whole gene), or *1327R-467F* (5' end ~600 bp) and *467FR-trn*L (3' end ~500 bp) and. Initially, primers *rpL32F* and *trn*L$^{(UAG)}$ (Shaw et al. 2007) were used to amplify the entire region, which varies in *Erythroxylum* from 700-1200 bp. Internal primers and a new 5' primer were developed using *Erythroxylum* sequences and NCBI Blast-Primer3 (S. Rozen & H.J. Skaletsky 2000; Table 1.3). ITS was amplified using either ITSA-ITSB (Blattner et al. 1999) or ITS5-ITS4 (White et al. 1990) primers.

Table 1.3. New *rpl32-trn*L primers.

Primer Name	**Sequence**
1327R	TGG TCA GCG TTG AAA GCT TTT TCG T
467FR	GCG ATC AAT ACC TAA TTG GGT TGC CA
467F	TGG CAA CCC AAT TAG GTA TTG ATC GC

The PCR products were then purified using ExoSapIT® (Amersham, Piscataway, NJ, USA), and DNA sequences were generated by Macrogen (Seoul, South Korea) or the University of Chicago Cancer Research Center, DNA Sequencing Facility (Chicago, IL). The same primers used to amplify the sequences were used for sequencing.

Data analyses. The nucleotide sequences obtained for each locus were aligned separately using the online version of MAFFT (Katoh et al. 2005) after the removal of primer sequences. For *rpl32-trn*L, the E-INS-i option was selected as recommended by MAFFT for fewer than 200 sequences "with multiple conserved domains and long gaps" (Katoh 2011). The scoring matrix 1PAM/K2 was selected because the taxa are closely related along with the default gap opening penalty 1.53 and an offset value 0.1 recommended for long gaps. For ITS, the G-INS-i option was selected because of the number of

sequences and no long gaps were expected based on preliminary alignments. The scoring matrix 1PAM/K=2 was selected with the default gap opening penalty 1.53 and an offset gap penalty 0.0 because no large gaps were expected.

Manual adjustments to the alignments followed the procedure outlined by Simmons (2004), following Zurawski & Clegg (1987). Outgroup taxa were often highly divergent, especially for the chloroplast region, resulting in large areas that were ambiguously aligned between the ingroup and the outgroup; these unalignable areas were coded as ambiguous for the outgroup only. Ambiguously aligned areas for all sequences were removed prior to the analyses (*rpl32-trn*L 534/1209bp; ITS 14/723bp). Gap characters were coded for the parsimony analyses. If a locus contained parsimony-informative gap characters, these characters were coded in unambiguously aligned regions using modified complex indel coding (Simmons & Ochoterena 2000; Müller 2006). Most gaps coded were simple and did not require a step matrix except for two gaps in ITS. Twenty gap characters were coded for each locus.

Phylogenies were estimated using morphological and molecular characters that were analyzed separately and in combination (Kluge 1989; Nixon & Carpenter 1996) using maximum parsimony in PAUP* 4.0b10 (Swofford 2002) and maximum likelihood in RAxML 7.2.6 (Stamatakis 2005, 2006). Most of the likelihood bootstrap analyses were run either using RAxMLGui 0.9 beta 1 (Silvestro & Michalak 2010) or RAxML in CIPRES (Portal 3, Miller et al. 2009). Four data matrices were analyzed: each locus separately, loci together, and a combined analysis of morphological and molecular characters. In the combined analyses of all characters, two accessions of a putative hybrid *E. rotundifolium*X*flavicans* were dropped because no herbarium specimens or literature was available for coding the morphological characters for this taxon. Morphological characters and Schulz's classification were mapped onto the tree using Mesquite 2.72 (build 527, Maddison & Maddison 2009) and parsimony ancestral state reconstructions. Schulz's classification was not used in the phylogenetic analyses, but mapped on for illustrative purposes. Equally weighted parsimony heuristic searches were conducted using 2,000 tree-bisection-reconnection (TBR) searches with a maximum of 10 trees held per search. Parsimony jackknife analyses (JK; Farris et al. 1996) were conducted with the removal probability set to approximately e^{-1}

(36.7879), and "jac" resampling emulated. Two-thousand replicates were performed with 100 TBR searches per replicate and a maximum of 10 trees held per TBR search.

Maximum likelihood analyses (Felsenstein 1973) were conducted for molecular characters only excluding gap characters using the only model available in RAxML, GTR, and four discrete rate categories. One-thousand separate heuristic searches were performed, using randomized stepwise parsimony trees generated by RAxML with the model above to find the best tree. A thorough bootstrap search (Felsenstein 1985) was performed until a sufficient number of bootstrap (BS) replicates was found using the majority rule criterion, autoMR (Pattengale et al. 2010). For the combined molecular analyses, the data were partitioned for the two loci to allow for different estimates of model parameters. Results of the bootstrap analyses were drawn on to the best tree for that dataset in RAxML. All tree figures were drawn using TreeGraph2 (Müller & Müller 2004; Stöver & Müller 2008).

Two clades were of particular interest based on the results from the analyses above and their possible monophyly was tested using Wilcoxon rank test (Templeton 1983) in PAUP. First, the *Macrocalyx* species were supported within a clade of *Rhabdophyllum* species, and therefore, the monophyly of *Rhabdophyllum* was tested. Second, the placement of *E. sauve* as sister to the majority of Caribbean endemic species forming a monophyletic clade, erythroxylum2, was tested. A combined heuristic search of all characters constraining either the monophyly of *Macrocalyx* or *E. rotundifolium* with erythroxylum2 was conducted as above, with the results compared to the unconstrained heuristic search for the same dataset.

Results

Of the 37 species sampled for presence of alkaloids, eight species were positive for the presence of at least one alkaloid (Appendix 1). The majority-rule parsimony JK tree for the combined analysis of all molecular characters (Fig. 1.1) and the combined analysis of all characters is presented (Fig. 1.2). The parsimony JK support values are above the branches, and the likelihood BS support values are below each branch. If a contradictory clade is present in the likelihood analysis compared to the parsimony analysis

and supported more than 50%, then the support value is in square brackets below the node containing all the taxa in the contradictory clade. When two support values are listed in the text, the first value is for JK

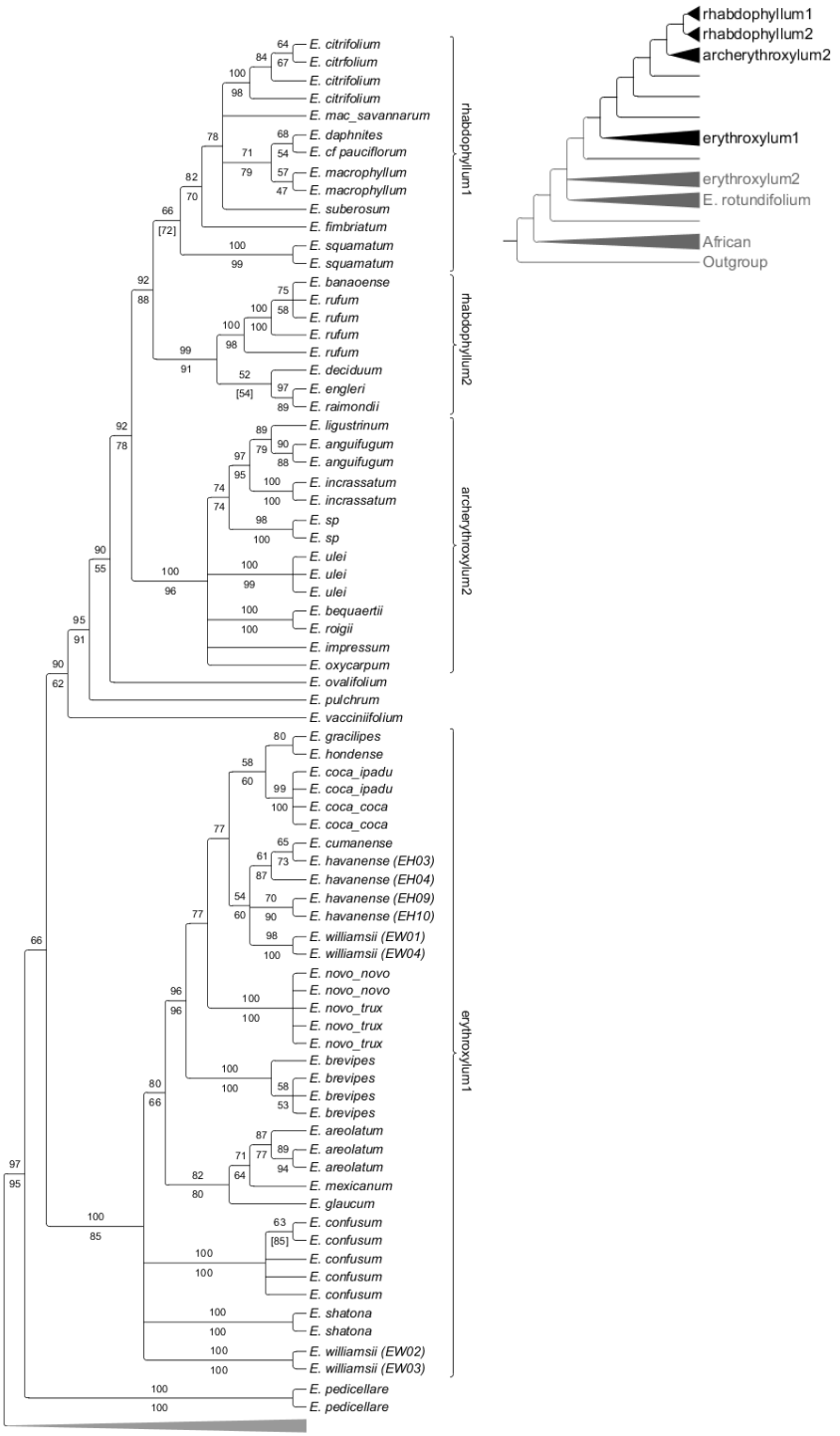


Fig. 1.1a. All molecular character majority-rule parsimony jackknife tree with jackknife values above each branch branch and likelihood values below supported clades. Support for incongruent clades in the best likelihood tree were placed in []. Clades annotated with informal clade names. *E. mac_savannarum* (*E. macrophyllum* var. *savannarum*), *E. novo_novo* (*E. novogranatense* var. *novogranatense*), *E. novo_trux* (*E. novogranatense* var. *truxillense*), *E. coca_coca* (*E. coca* var. *coca*), *E. coca_ipadu* (*E. coca* var. *ipadu*).

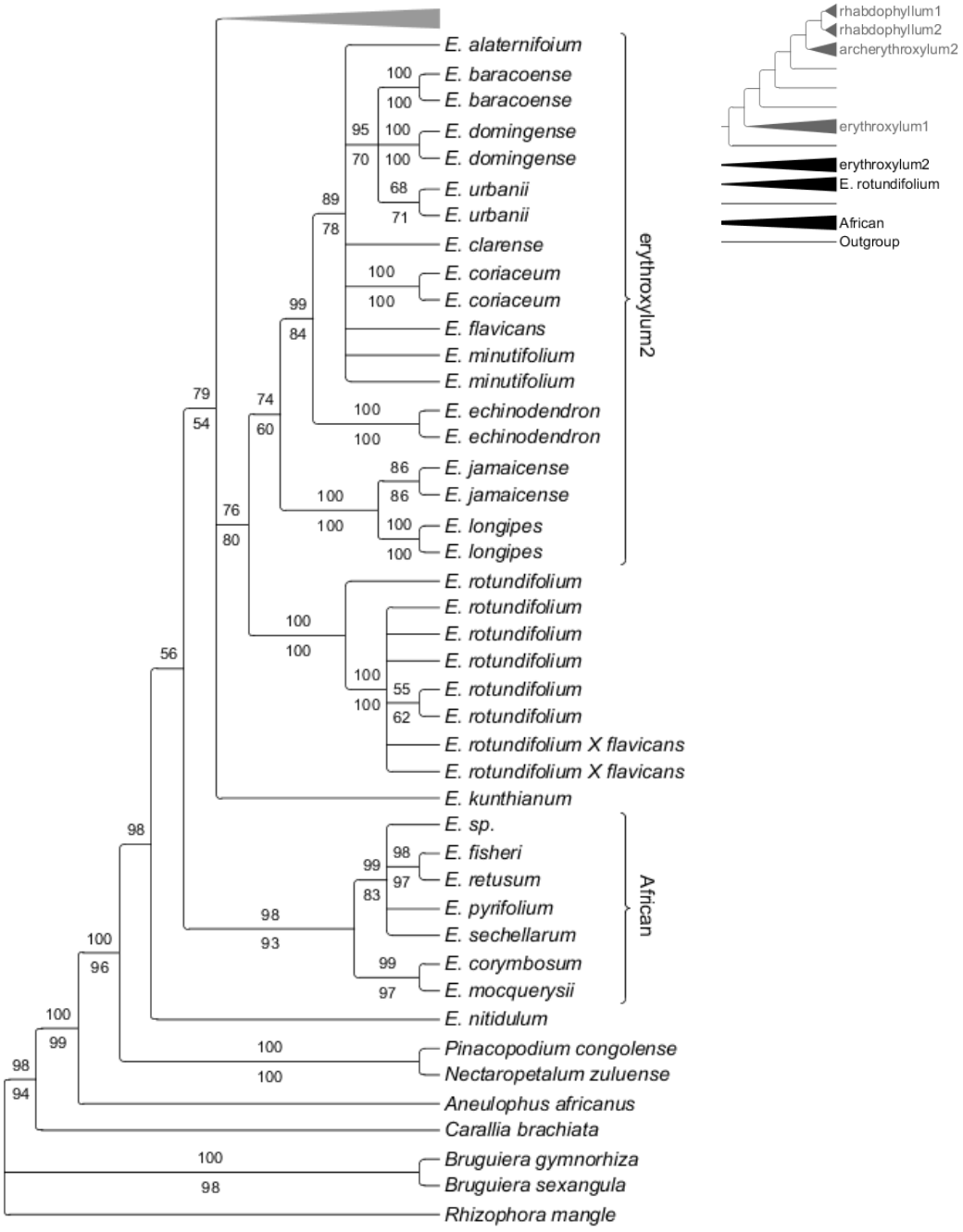


Fig. 1.1b. All molecular character majority-rule parsimony jackknife tree with jackknife values above each branch and likelihood values below supported clades. Support for incongruent clades in the best likelihood tree were placed in []. Clades annotate with informal clade names. E. mac_savannarum (E. macrophyllum var. savannarum), E. novo_novo (E. novogranatense var. novogranatense), E. novo_trux (E. novogranatense var. truxillense), E. coca_coca (E. coca var. coca), E. coca_ipadu (E. coca var. ipadu).

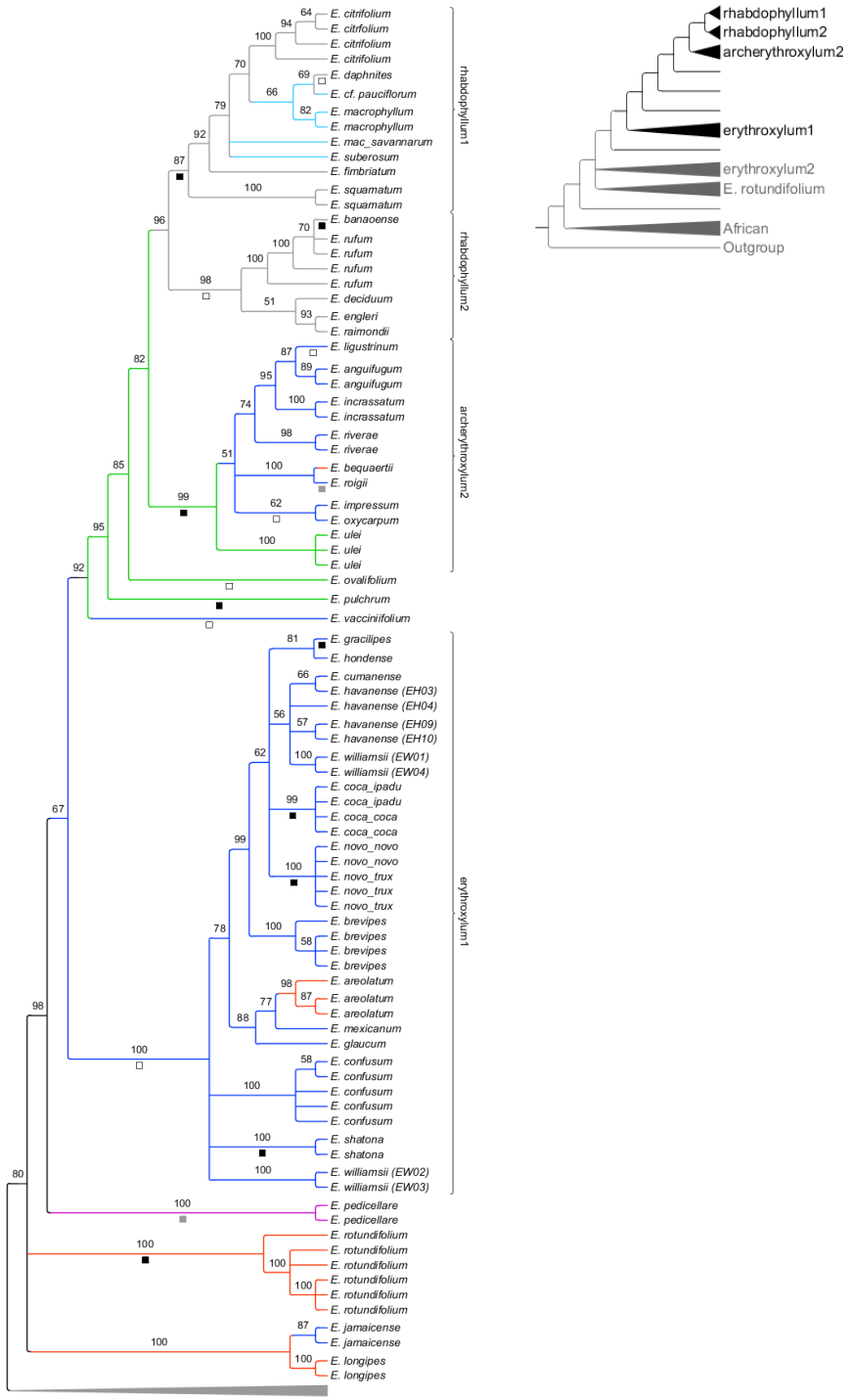


Fig. 1.2a. All character majority-rule parsimony jackknife tree with jackknife values above each branch. Clades annotated with informal clade names. Leaf duration for Neotropical clade marked as follows below the branch, evergreen = dark box, semideciduous = grey box, and deciduous = white box. Branches colored based on Schulz's sections, *Archerythroxylum* (dark blue), *Erythroxylum* (orange), *Eurysepalum* (brown), *Gonocladus* (light green), *Lagynocarpus* (light orange) *Leptogramme* (green), *Macrocalyx* (light blue), *Megalophyllum* (light grey), *Microphyllum* (purple), *Rhabdophyllum* (grey), except for *Pachylobus*, *Schistophyllum*, *Oxystigma* which with just one representative were left black. *E. mac_savannarum*, *E. macrophyllum var. savannarum*,*E. novo_novo* (*E. novogranatense* var. *novogranatense*), *E. novo_trux* (*E. novogranatense* var. *truxillense*), *E. coca_coca* (*E. coca* var. *coca*), *E. coca_ipadu* (*E. coca* var. *ipadu*).

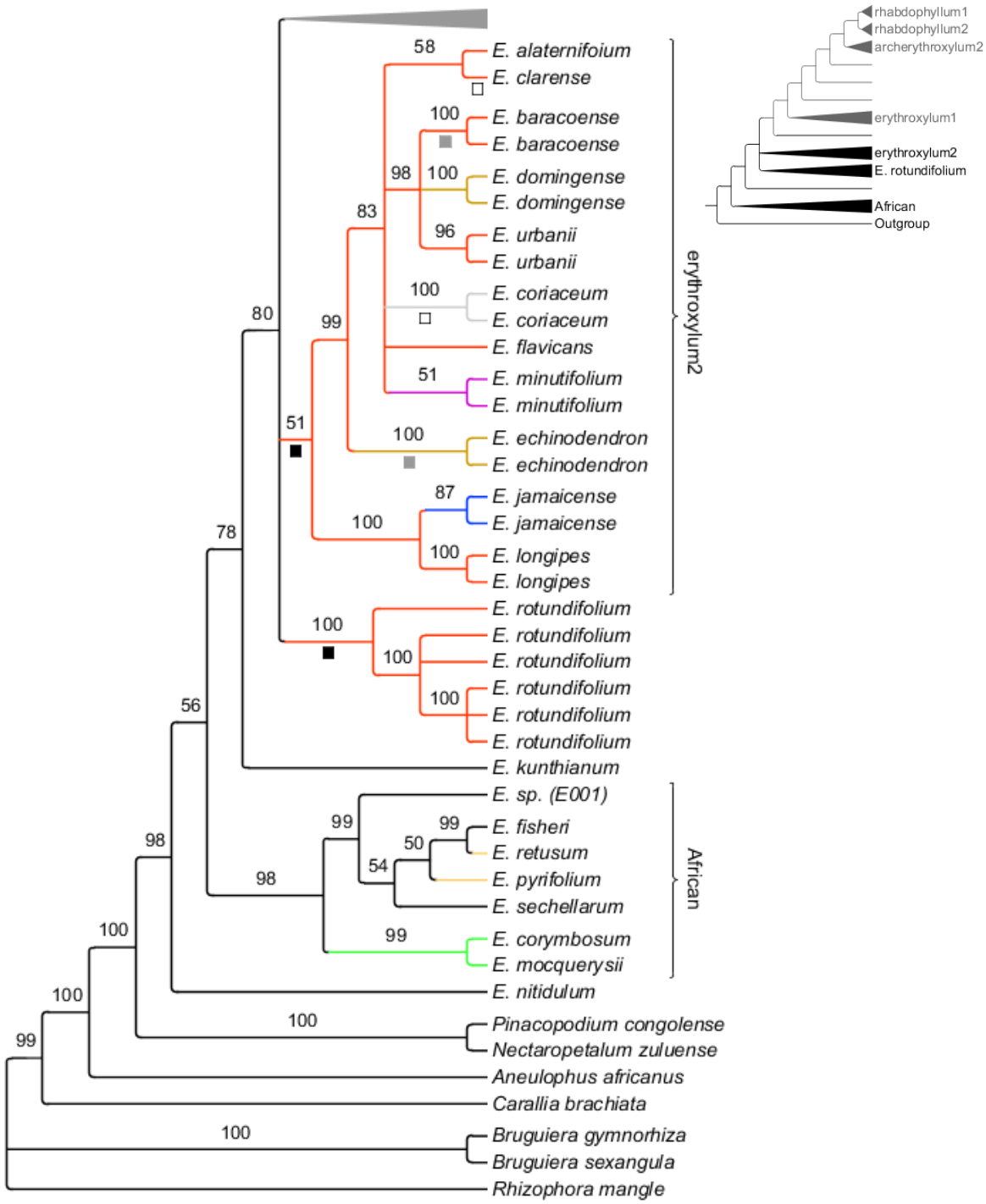


Fig. 1.2b. All character majority-rule parsimony jackknife tree with jackknife values above each branch. Clades annotated with informal clade names. Leaf duration for Neotropical clade marked as follows below the branch, evergreen = dark box, semideciduous = grey box, and deciduous = white box. Branches colored based on Schulz's sections, *Archerythroxylum* (dark blue), *Erythroxylum* (orange), *Eurysepalum* (brown), *Gonocladus* (light green), *Lagynocarpus* (light orange) *Leptogramme* (green), *Macrocalyx* (light blue), *Megalophyllum* (light grey), *Microphyllum* (purple), *Rhabdophyllum* (grey), except for *Pachylobus*, *Schistophyllum*, *Oxystigma* which with just one representative were left black. *E. novo_novo* (*E. novogranatense* var. *novogranatense*), *E. novo_trux* (*E. novogranatense* var. *truxillense*), *E. coca_coca* (*E. coca* var. *coca*), *E. coca_ipadu* (*E. coca* var. *ipadu*).

support and the second for the BS support. The gene trees, ITS and *rpl32-trn*L, are included in Appendix 3 and include support values as above from both methods along with the strict consensus trees and best likelihood trees from each analysis.

The likelihood and parsimony topologies for all the analyses did not strongly (>80 BS/JK) support contradictory hypotheses so no additional analyses were necessary. Summary statistics for the parsimony analyses are included in Table 1.4. Clades are informally named and annotated on all figures.

Table 1.4. Statistics for each of the parsimony analyses. PIC=parsimony informative characters, MPT= most parsimonious tree, CI = ensemble consistency index after removal of the parsimony-uninformative characters, RI=ensemble retention index.

Matrix	**# of terminals**	**# characters analyzed**	**# of PIC**	**% miss.**	**MPT length**	**# of MPTs**	**CI**	**RI**
ITS rDNA	125	737	306	9.5	1,162	13,230	0.52	0.97
*rpl32-trn*L	115	1,744	104	50	317	550	0.80	0.92
all molecular	119	2,481	411	40	1,489	9,750	0.57	0.87
all characters	117	2,511	440	40	1,688	9,440	0.54	0.85

The separate gene trees, ITS and *rpl32-trn*L, differed only in the placement of *E. vacciniifolium, E. jamaicense*, and *E. longipes*. The *rpl32-trn*L trees from both methods lacked structure compared to the ITS trees. In the *rpl32-trn*L trees, *E. vacciniifolium* is sister to all of the New World species except erythroxylum2 and *E. rotundifolium*, but in the ITS tree, *E. vacciniifolium* is sister only to rhabdophyllum1, rhapdophyllum2, *E. ovalifolium*, and *E. pulchrum* (Appendix 3). In the ITS phylogeny, *E. jamaicense* and *E. longipes* form a highly supported clade (100/100) as part of the polytomy within the larger New World *Erythroxylum* clade. In contrast, for the likelihood and parsimony *rpl32-trn*L trees, *E. jamaicense* and *E. longipes* occur within the smaller clade, erythroxylum2.

For the combined analyses of both loci, no strongly supported contradictory clades exist between the two methods. Placement of *E. vacciniifolium* reflects the ITS tree (90/62) as sister to rhabdophyllum1, rhabdophyllum2, *E. ovalifolium*, and *E. pulchrum*, and *E. jamaicense* and *E. longipes* form a strongly

supported clade (100/100) sister to the largely Caribbean clade (74/60). In the combined analysis of all characters, clades are mostly congruent with the molecular only trees and the individual gene trees. In the combined analysis of all characters, *E. vacciniifolium* is supported (93) as sister to rhabdophyllum1, rhabdophyllum2, *E. ovalifolium*, and *E. pulchrum*, as in the combined molecular analysis and ITS gene tree. The *E. jamaicense* and *E. longipes* clade is supported as sister to the largely Caribbean clade (59) in the same position as the combined molecular analysis but with less support. Finally, in the ITS tree *Rhabdophyllum* and *Macrocalyx* species form a polytomy, but in the combined all character tree (Fig. 1.2), *Rhabdophyllum* and *Macrocalyx* species form two strongly supported clades (87 and 99, Fig. 1.2) with *Macrocalyx* species in one clade with some of the *Rhabdophyllum* species.

Erythroxylum squamatum is sister to rhabdophyllum1 clade in the combined analysis of all characters. In the ITS tree, *E. squamatum* forms a polytomy with clades rhabdophyllum1 and rhabdophyllum2. In the combined all molecular analysis *E. squamatum* is sister to rhabdophyllum2 (66) in the parsimony JK tree, but not in the likelihood tree for which *E. squamatum* is sister to rhabdophyllum1 and rhabdophyllum2 clades (BS 72; Fig. 1.1).

The monophyly of the *Rhabdophyllum* section was tested using Wilcoxon rank test and a MPT from the combined analysis of all characters. The MPT from the analysis constraining the monophyly of *Rhabdophyllum* was significantly worse ($p<0.0001$; MPT length 1,715) than the unconstrained MPT (MPT length 1,688). Comparing the best trees from two likelihood analyses for all molecular characters, the unconstrained tree was better (-ln L =10988.29) than the constrained tree (-ln L=11008.76), but no significant difference ($p=0.076$) was found using the Shimodaira-Hasegawa test, with 1,000 bootstrap replicates from the RELL method

To test the monophyly of *E. sauve* with erythroxylum2, I compared the MPTs from a combined analyses of all characters with and without constraining *E. rotundifolium* and erythroxylum2 using the Wilcoxon rank test. The trees from both analyses were the same length and not significantly different ($p=1.00$), and therefore, not conclusive as to the monophyly of *E. rotundifolium* and erythroxylum2.

In all analyses with two exceptions, the monophyly of *Erythroxylum* was supported and the strongest support from the combined molecular analysis parsimony JK tree (98) and the combined analysis of all characters (98). In the combined molecular analysis likelihood BS tree for the molecular only dataset, *E. nitidulum* was supported (31) as sister to *Pinacopodium congolense* and *Nectaropetalum zulense*. Support for the monophyly of the remaining *Erythroxylum* species was low (29). In the ITS likelihood BS tree, *E. nitidulum* was placed sister to *Pinacopodium, Nectaropetalum*. *Erythroxylum nitidulum* in all other analyses (*rpl32-trn*L, ITS parsimony, all molecular parsimony, and all characters) was sister to all other *Erythroxylum* species. In all analyses, the African or Malagasy *Erythroxylum* species (except for *E. nitidulum*) formed a well supported clade (98/93; Fig. 1.1) often sister to the one sampled Asian species and the New World *Erythroxylum* species (79/54; Fig. 1.1).

New World *Erythroxylum* are split into four main clades (Fig. 1.1, Fig. 1.2). The relationships within these clades will be discussed in detail below. Several based on the current sampling and the phylogeny show promise as potentially useful characters for future analyses, in particular leaf duration (*see* Rury 1981 as substitute for wood anatomy characters), leaf areole shape, flower sexuality, phloem fiber pattern, endocarp, and wood crystal type. Two morphological characters are sufficiently sampled and important in supporting particular clades, the presence of striated stipules and the presence of ramified fibro-sclereids. Plowman and Hensold (2004) and Loiola (2001) both noted the potential importance of stipule striation for dividing *Erythroxylum.*

Discussion

Erythroxylaceae systematics. To understand the systematics of *Erythroxylum*, the genus must be placed in the context of the evolutionary history of its family, Erythroxylaceae. This is a small pantropical family of flowering trees and shrubs comprising four genera, *Aneulophus* Benth., *Nectaropetalum* Engl., *Pinacopodium* Exell & Mendonça, and of course, *Erythroxylum* P. Browne (Heywood 1993; Stevens 2001, onwards). *Aneulophus* (2 species), *Nectaropetalum* (6 species), and *Pinacopodium* (2 species) are small genera that are mostly geographically restricted to mesic regions of tropical Africa (*Nectaropetalum* also occurs in Madagascar; Heywood 1993).

Erythroxylaceae is placed in the order Malpighiales sister to Rhizophoraceae (Chase 1993; Setoguchi et al. 1999; Schwarzbach & Ricklefs 2000; Wurdack & Davis 2009). Wurdack and Davis (2009) demonstrated weak to strong support, depending on the method of phylogenetic reconstruction, for the sister relationship of the monogeneric family Ctenolophonaceae to Rhizophoraceae and Erythroxylaceae. Members of all three families have 'opposite leaves with sheathing, interpetiolar stipules' (p.1558, Wurdack and Davis 2009). *Erythroxylum, Pinacopodium,* and *Nectaropetalum* have alternate leaves. Based on the sister relationship of Erythroxylaceae and Rhizophoraceae and the shared morphological characters between *Aneulophus* and 'primitive' Rhizophoraceae, the Angiosperm Phylogeny Group II (p. 408; APGII 2003) proposed the transfer of Erythroxylaceae to Rhizophoraceae, but the two families have not yet been formally combined (APGIII 2009). Four Rhizophoraceae species were included in this analysis and both the ITS and *rpL32-trnL* loci were amplified; however, the chloroplast locus was highly divergent between Rhizophoraceae and Erythroxylaceae species causing the determination of homology between the sequences impossible and exclusion of the sequences from subsequent analyses. The locus was alignable between the Rhizophoraceae species.

Systematists prior to the 1950s debated the generic make-up of Erythroxylaceae and its status as a family. After the 1950s, systematists consistently included *Aneulophus*, *Erythroxylum, Nectaropetalum*, and *Pinacopodium* in the family (for overview see Hegnauer 1981; Rury 1981). The placement of *Aneulophus* in the family was recently confirmed using molecular characters (Wurdack and Davis 2009). The four Erythroxylaceae genera share the following traits: glabrous shrubs and trees, simple, entire leaves, partially (or wholly) intrapetiolar stipules, and heterostylous flowers with a persistent 5-lobed calyx (Rury 1982; Plowman 1991; Stevens 2001, onwards). In the absence of flowers, *Erythroxylaceae*, *Nectaropetalum*, and *Pinacopodium* are morphologically indistinguishable unless combinations of fine-scale anatomical features are considered (Rury 1982). Multiple attempts to sample additional material for *Aneulophus, Nectaropetalum,* and *Pinacopodium* failed or resulted in partial sequences leaving only one representative of each genus in the analyses. Compared to the other three genera *Aneulophus* sequences were highly divergent causing problems with sequence alignment and certain ambiguous regions were

coded as missing. This divergence was expected based on the morphological differences between *Aneulophus* and the other three genera, especially *Erythroxylum* (Rury 1982). Although *Pinacopodium* and *Nectaropetalum* sequences were divergent compared to *Erythroxylum*, the differences among these sequences were not sufficient to impact alignment and were consistent with the degree of divergence between the Old World versus New World *Erythroxylum* species. Based on leaf venation, leaf midvein, and wood anatomy, *Aneulophus* and *Pinacopodium* were hypothesized as sister taxa and likely the most basal genera within Erythroxylaceae (Rury 1982). Some *Erythroxylum* species, especially in section *Rhabdophyllum*, share striated stipules, and similarities in leaf, bud, and stem anatomy with *Nectaropetalum* (Rury 1982). The phylogeny indicates support for *Nectaropetalum* and *Pinacopodium* as sister genera (100; Fig. 1.2) and well supported (100; Fig. 1.2) as sister to *Erythroxylum*; however, because of the depauperate sampling within *Nectaropetalum* and *Pinacopodium*, the phylogeny fails to refute the possibility that *Nectaropetalum* may be sister to *Erythroxylum*.

Monophyly of *Erythroxylum*. The inclusion of these other *Erythroxylaceae* genera along with Rhizophoraceae allowed for testing the monophyly of *Erythroxylum*, which is strongly supported as monophyletic (98, Fig. 1.1; 98, Fig. 1.2) in every analysis with the exception of the placement of the species *E. nitidulum* in two of the likelihood analyses. *Erythroxylum nitidulum*, native to Madagascar, was placed by Schulz (1907a, 1931) as the only species in section *Schistophyllum* based on a few characters unique to this species, such as stipules parted to the base, styles fused along the length (most *Erythroxylum* are free or partially fused), and flowers produced on branched peduncles, which otherwise is found only in some *Pinacopodium* species. In the likelihood analysis of all molecular characters, *E. nitidulum* was sister to *Pinacopodium* and *Nectaropetalum*, and in the ITS likelihood BS tree, the species was sister to *Pinacopodium*, *Nectaropetalum*, and *Erythroxylum*. Only one accession of *E. nitidulum* was sequenced, so the placement of this species with *Pinacopodium* or as a monotypic genus will need to await further anatomical and molecular work. Other African species, from Madagascar, Seychelles, and mainland Africa, form a clade called African, in every analysis with typically moderate to high levels of support (98, Fig. 1.1). In most analyses, the African clade is supported (56) as sister to the remaining

Erythroxylum species including *E. kunthianum*, native to India and Burma. When *E. kunthianum* is excluded from the analyses support for the sister relationship of the African clade to the remaining *Erythroxylum* species is higher. No other species from Asia or Australia were included in the analysis, and if included along with additional accessions of *E. kunthianum* then the placement of *E. kunthianum* may be better supported as either within the non-African *Erythroxylum* species or as a separate clade with species from Asia. The African clade represents species from several of Schulz's sections with one section, *Gonocladus*, as monophyletic. The current pattern of distribution of Erythroxylaceae across the Old and New World tropics is consistent with a possible origin in the Northern Tropical Gondwana biota (Sanmartín & Ronquist 2004). The break-up of this northern tropical region of Gondwana, beginning 130 million years ago (ma; Sanmartín & Ronquist 2004), may account for the current distribution of *Erythroxylum*; however, several recent biogeographical studies of plants with similar distributions to *Erythroxylum* support an alternate hypothesis to the classic vicariant model (review Sanmartín & Ronquist 2004). In these studies, dispersal between continents after the separation of Gondwana explains the current distribution given the estimated ages of the taxa. Bell et al. (2010) estimated the age of divergence of Erythroxylaceae and Rhizophoraceae between 44-79 ma, and the last hypothesized connections between South America and Africa were 105 ma (McLoughlin 2001). In order to robustly investigate the timing of the divergence of *Erythroxylum*, fossil calibration within *Erythroxylum* should be combined with the age estimates for the split between Rhizophoraceae and Erythroxylaceae (Bell et al. 2010). Unfortunately, no *Erythroxylum* fossils are known. The one reported fossil is not reliable (Berry 1938). Based on the phylogeny and the age estimates for the divergence of Erythroxylaceae and the connection between the continents (McLouglin 2001; Bell et al. 2010), the distribution of Erythroxylaceae and *Erythroxylum* was derived from boreotropical dispersal rather than a the break-up of Gondwana (Davis et al. 2002; Feldberg et al. 2010).

Schulz's Sections. As suggested by morphological studies, most of Schulz's sections are not monophyletic (Payens 1958, Rury 1982, Plowman & Rivier 1983; Loiola 2001). Section *Heterogyne* will be referred to as section *Erythroxylum* because it includes the type species of the genus, *E. areolatum*

(Payens 1958; Loiola 2001). For the New World sections sampled, species placed in *Macrocalyx* are supported within the larger section *Rhabdophyllum;* at least two clades of species assigned by Schulz (1907) to *Archerythroxylum* occur in the phylogeny; *Leptogramme* is not monophyletic but forms a grade out of which arises one of the *Archerythroxylum* clades; and section *Erythroxylum* includes species from *Archerythroxylum, Megalophyllum* and *Eurysepalum,* but not the type species of the section (Loiola 2001). Only *Gonocladus* is a monophyletic group based on the current phylogeny.

Several well supported clades correspond to groups proposed by Plowman (Rury 1982). In Rury's dissertation on the systematic anatomy of Erythroxylaceae, he listed a series of taxonomically difficult groups proposed by Plowman for further investigation. Although these groups are encompass possible species complexes, these informal groups correspond well to clades in the phylogeny typically with the addition of other species and were used in part to name the primary clades supported by the phylogenetic analyses (Fig. 1.2).

Clade rhabdophyllum1, includes species from *Rhabdophyllum* and *Macrocalyx* (87; Fig. 1.2). Section *Rhabdophyllum* is comprised of 50 species and *Macrocalyx* includes eight species (Loiola 2001) for which I sampled nine and five, respectively. Both sections share striated stipules or the presence of "fiber ensheathed vascular bundles" (p. 380; Rury 1982), which are easy to see by eye or with a hand lens. The morphological differences used to split these two sections was the shape and margin (frayed or not) of the sepals and the calyx arrangement in the bud (imbricate (*Macrocalyx*) or valvate (*Rhabdophyllum*; Loiola 2001). Compared to most of Schulz's sections these two sections are morphologically homogeneous as originally delimited (Rury 1982; Loiola 2001) especially within the evergreen species of *Rhabdophyllum* (Rury 1982). I sampled four evergreen *Rhabdophyllum* species cited by Rury (1982), *E. citrifolium, E. engleri, E. fimbriatum,* and *E. squamatum,* and all but *E. engleri* occur in clade rhabdophyllum1. *Erythroxylum engleri* occurs in the sister clade with deciduous *Rhabdophyllum* species. One deciduous species, *E. daphnites* occurs in clade rhabdophyllum1 with the evergreen species.

Because of the inclusion of *Macrocalyx* species within *Rhabdophyllum,* I tested the monophyly of *Rhapdophyllum*. In the combined analysis of all characters, *Rhabdophyllum* was not supported as

monophyletic, and the phylogeny resulting from constraint of monophyly was significantly worse than the unconstrained tree (Wilcoxon rank test, $p<0.0001$). Comparing the best trees from two likelihood analyses (one constrained and one unconstrained), the monophyly of *Rhabdophyllum* was not refuted (Shimodaira-Hasegawa, $p<0.076$; Wilcoxon rank test, $p<0.375$). Morphological characters combined with molecular characters fail to support the monophyly of *Rhabdophyllum*, and use of this section without the inclusion of *Macrocalyx* in predicting or understanding the evolution of traits is not warranted. The presence of stipule striations, evergreen leaves (except for *E. daphnites*), and ramified fiber sclereids in their leaves serve as synapomorphies for rhabdophyllum1 making this a well supported group. The well supported clade (98), rhabdophyllum2, includes species from section *Rhabdophyllum* which like rhabdophyllum1 have striated stipules, but differ in the absence of ramified fiber sclereids (except *E. engleri*) and the presence of deciduous leaves (except *E. banaoense* and *E. engleri*).

Rhabdophyllum2 corresponds loosely to Plowman's species complex *Rhabdophyllum* group 2 (*E. deciduum, E. raimondii, E. rufum,* and *E. squamatum*; Rury 1982). Based on anatomical similarities and the phylogeny (Fig. 1.1, Fig. 1.2), *E. squamatum* is a distinct species (Rury 1982; Loiola 2001), but whether the other three species are distinct is difficult to interpret based on the phylogeny and anatomy (Rury 1982).

Loiola (2001) synonymized *E. raimondii* and *E. rufum* and maintained *E. rufum* and *E. deciduum* as separate species based on consistent morphological differences including leaf shape, length of style in long-styled flower form, and presence of bristled setae on the stipules. To complicate the taxonomy of this clade, *E. engleri*, which was not recognized as part of this species complex by Rury (1982), was compared by Loiola (2001) for its close affinity to *E. deciduum*. Loiola (2001) ultimately determined that these two species are distinct. If Loiola's (2001) recommendation for *E. rufum* and *E. raimondii* is followed then in light of the phylogeny, all of the species in clade rhabdophyllum2 would have to be synonymized, which seems unrealistic based on the morphological differences among these species. *Erythroxylum banaoense* differs from *E. rufum* in the number of flowers per inflorescence (2-3 versus 10-20), pedicel size (1-1.5 mm vs 5-13 mm long), and fruit size (13-15 x 4-6 mm versus 8-9 x 4-4.5 mm)

along with a few other traits. The Cuban endemic, *E. banaoense*, is likely an example of recent in-situ speciation. Based on geography, *E. raimondii*, *E. deciduum*, and *E. engleri* do not overlap, which for *E. deciduum* and *E. engleri* supports the morphological argument that they are distinct species (Loiola 2001). *Erythroxylum raimondii* and *E. rufum* do not overlap in distribution, which along with the current phylogeny supports their distinction as two species in contrast to Loiola's (2001) recommendation.

Finally, the two *Rhabdophyllum* lineages could be combined to for a revised section *Rhabdophyllum*. This section would then include all *Erythroxylum* species with striated stipules. The type species of sections *Rhabdophyllum* and *Macrocalyx* were not sampled in this study, and future research sampling these species may shed light on the best way to formally divide and name these clades.

Section *Archerythroxylum* has the highest number of species of any section with approximately 65, for which I sampled 21. Species from this section form two well supported clades, archerythroxylum2 and erythroxylum1, with a few species placed in other clades. Some species from sections *Erythroxylum* and *Leptogramme* occur in these predominately *Archerythroxylum* clades. Clade archerythroxylum2 is named after Plowman's *Archerythroxylum* group 2 species complex (Rury 1982), which will be discussed below. First, archerythroxylum2 includes a species from section *Leptogramme* and a species from *Erythroxylum*. These sections (Schulz 1907a; Loiola 2001) differ in very few characters. Schulz (1907a) placed five species into *Leptogramme* based on the presence of faint stipular striations. Otherwise, species of *Leptogramme* differ little from species of *Archerythroxylum*. A nodal anatomical examination of *E.* (sect. *Leptogramme*) *ulei* did not reveal evidence of striations. In light of this evidence, *E. ulei* is grouped with the other species in archerythroxylum2 (99; Fig. 1.2).

Archerythroxylum2 also includes, *E. bequaertii*, a species placed by Rury (1982) in section *Erythroxylum*. According to Shculz's classification, species in section *Erythroxylum* are subdioecious in contrast to species placed in *Leptogramme* and *Archerythroxylum*, which are hermaphroditic. A subdioecious species has individuals that are either male or female but with a few indivuals that are hermaphroditic. Subdioecy *Erythroxylum bequaertii* is subdioecious (Barrie & Plowman, unpublished) and all the other species in this clade have hermaphroditic flowers suggesting, based on evidence from

throughout the phylogeny that several transitions from hermaphroditic to subdioecious sexuality occurred within the genus *Erythroxylum*. Based on the present morphological sampling, archerythroxylum2 differs from clade erythroxylum1 in the presence of evergreen leaves versus deciduous leaves, although with a few exceptions in both clades (Fig. 1.2).

Within archerythroxylum2 is a species complex proposed by Plowman (Rury 1982) that includes the following species, *E. bequaertii, E. incrassatum, E. spruceanum, E. orinocense, E. oxycarpum,* and *E. roigii*. I sampled *E. bequaertii, E. incrassatum, E. oxycarpum,* and *E. roigii*. All of these species occur in archerythroxylum2 along with five other species. *Erythroxylum roigii*, a Cuban endemic, is strongly supported (100, Fig. 1.2) as sister to *E. bequaertii*, native to Mexico and Belize, in all phylogenies (Fig. 1.1; Fig. 1.2). Unfortunately, as only one accession for each species was sampled it is difficult to interpret the phylogeny in terms of the monophyly of these species, but these two species differ from one another in several morphological traits and a determination of species delimitation should wait for further evidence. In contrast, *E. incrassatum*, endemic to Jamaica, is strongly supported (95) as sister to *E. ligustrinum* and *E. anguifugum*, and *E. oxycarpum* is sister to *E. impressum* (62). The phylogeny provides support for the current species delimitation, which is supported in part by Rury (1982).

Erythroxylum1 is the other well supported (100; Fig. 1.2) clade comprised of *Archerythroxylum* species and one species from section *Erythroxylum*. *Erythroxylum pedicellare*, a Cuban endemic, is sister to erythroxylum1 and is one of two representatives of *Microphyllum*. Except for *E. shatona,* all of the species in this clade have deciduous leaves.

The type species of the genus and section *Erythroxylum*, *E. areolatum,* occurs in the erythroxylum1 clade, hence the informal name. *Erythroxylum areolatum* is subdioecious, and the other species in this clade are hermaphroditic indicating a shift to subdioecy. Other species of *Erythroxylum* have hermaphroditic flowers that are heterostylous (Darwin 1888; Gander 1979). Individual plants have one of two floral morphs with either a long or short style. The main difference between sections *Archerythroxylum* and *Erythroxylum* is flower sexuality. Eight species were placed in section *Erythroxylum* based on subdioecy (Schulz 1907a; Rury 1982), and all but one of these species was

included in the analysis. Subdioecy differs from strict dioecy in that the species may produce some invidual plants with functional hermaphroditic flowers. Species of section *Erythroxylum* produce flowers with either rudimentary pistils or stamens depending on the floral morph of the individual. Hence,

Species that appear morphologically hermaphroditic may be functionally dioecious (Li et al. 2010; Le Péchon et al. 2010). The sexuality of one subdioecious *Erythroxylum* species, *E. rotundifolium*, was studied and found to be dioecious (Abaraca et al. 2008). A typical hermaphroditic species, *E. havanense*, was determined to be gynodioecious (Dominguez et al. 1997) and evolving towards dioecy (Avila-Sakar & Dominguez 2000). Interpreting breeding systems especially between subdioecy versus dioecy based solely on herbarium specimens is not always possible. In the phylogeny, subdioecy or dioecy occurs at least three times, if *E. rotundifolium* is grouped with erythroxylum2 (discussed below).

Most species in section *Erythroxylum* occur in the Caribbean. Evolution of dioecy on oceanic islands is common (Venkatasamy et al. 2007; Li et al. 2010; Le Pénchon et al. 2010) especially for fleshy fruited woody plants. Studies investigating the breeding systems of the subdioecious or dioecious species are underway (Oviedo pers. comm.) and should provide evidence towards the functional breeding system of these plants and evolution of subdioecy (or dioecy) possibly via 'leak dioecy' in the Caribbean (Anderson et al. 2006; Venkatasamy et al. 2007).

Erythroxylum1 includes representatives from two species complexes identified by Plowman (Rury 1982). The first species complex includes *E. brevipes*, *E. cumanense*, and *E. havanense*. Based on Rury (1982) no anatomical differences distinguish these species from each other, and, although, these species differ from one another, the differences can be seen as a morphological gradient across the species' ranges. Based on my phylogeny, *E. brevipes*, which occurs scattered throughout the Caribbean, is a separate species and sister to the clade that includes *E. cumanense* and *E. havanense* among other species. *Erythroxylum cumanense* is widespread throughout northern South America. *Erythroxylum havanense* is a widespread species occurring in Cuba, the Lesser Antilles, Mexico, Central America, and northern South America. Only one accession of *E. cumanense* was sampled, but four accessions of *E. havanense* were sampled from throughout its range. The phylogeny does not refute the findings of Rury (1982) that *E.*

havanense and *E. cumanense* may be the same species, supporting D'Arcy & Schanen (1975), and not Plowman & Hensold (2004). *Erythroxylum havanense* accessions from Central America form a clade with *E. cumanense* (80/92, ITS; 61/87, Fig. 1.1), and in all of the phylogenies the Caribbean accessions form a clade. Further molecular and morphological studies may help elucidate the relationships between the Central American *E. havanense* and *E. cumanense* and the Caribbean *E. havanense*, to determine if they should be treated as one species or if the Caribbean populations should be split into a new species.

To compound the taxonomic complexity within erythroxylum1, another species complex identified by Plowman (Rury 1982) occurs in this clade. This species complex includes species from both sections *Archerythroxylum* and *Erythroxylum*: *E. areolatum, E. confusum, E. guatemalensis, E. impressum*, and *E. reticulatum. Erythroxylum guatemalensis* and *E. reticulatum* were not sampled, but based on anatomy were found to be distinct from the other three species (Rury 1982). Of the remaining species, *E. areolatum* and *E. confusum* both occur in clade erythroxylum1, and *E. impressum* occurs in clade, archerythroxylum2. Based on the phylogeny, all three species are distinct. *Erythroxylum areolatum* forms a strongly supported clade with *E. mexicanum* and *E. glaucum* in all phylogenies (Fig. 1.1, Fig. 1.2) except for the separate analysis of the *rpl32-trn*L data set for which they occur as a polytomy. *Erythroxylum confusum* forms a clade with *E. williamsii* (Caribbean species) in the ITS tree (76/60) or is part of the larger erythroxylum1 polytomy. These three species should be recognized.

The last group of species, erythroxylum2 and *E. rotundifolium*, includes species primarily from section *Erythroxylum.* All species in erythroxylum2 are Caribbean endemics with most species restricted to one island. This clade is moderately to poorly supported (74/60, Fig. 1.1; 51, Fig. 1.2) depending on the analysis. *Erythroxylum rotundifolium* is supported as sister to erythroxylum2 with low support (42/48, Fig. 1.1; 38, Fig. 1.2). Most of the species in erythroxylum2 and *E. rotundifolium* have evergreen leaves and subdioecious flowers.

Origin of cocaine synthesis. Humans have exploited secondary metabolites for thousands of years as poisons, medicines, cosmetics, and insecticides (Croteau et al. 2000). Cocaine is no exception. For the last 8,000 years, the cultivated cocas with high amounts of cocaine have been cultivated for their leaves in

certain regions of South America (Dillehay et al. 2010; Plowman 1984a,b). Today, four taxa are cultivated, *E. coca* var. *coca*, *E. coca* var. *ipadu*, *E. novogranatense* var. *novogranatense*, and *E. novogranatense* var. *truxillense* (United Nations 2007) primarily for the illicit extraction and processing of cocaine to be sold as an illegal, recreational drug (United Nations 2006).

A smaller percentage of coca in Bolivia and Peru is legally cultivated for local markets to be used in beverages or chewed directly. *Erythroxylum novogranatense* var. *truxillense* is also cultivated for the compound methyl salicylate to flavor the soft drink Coca-Cola® (Plowman 1984a) from the decocanized leaves. The cocaine resulting from the process of decocanization is sold for legal uses in modern surgery to manage pain (Cara et al. 2003). Coca species continue to be investigated for possible new drugs to treat diseases such as cancer or herpes (Mi et. al. 2001; Gonzales-Guevara et al. 2004; Rodeiro et al. 2008) or as a tool in neurologic research (Dräger 2002). Mi et al. (2001) discovered a novel chemical similar to cocaine in an *Erythroxylum* species that may prove useful in delivering multi-drug treatments to cancer patients. By understanding the production of cocaine by different species within a phylogenetic framework, researchers can target which species out of 230 are most likely to have medicinal value.

Because production and evolution of secondary metabolites are influenced by natural selection (Wink 2003), increased reproduction by the plant must outweigh the cost to the plant to produce these additional chemicals. Hence, even between closely related species some *Erythroxylum* species may produce cocaine while others will not. By testing for the presence of precursors and not just cocaine, the variable production of precursors should shed light on the possible origin or origins of cocaine biosynthesis in *Erythroxylum*. Secondary chemistry data of four alkaloids were analyzed together with the molecular and other morphological characters (Fig. 1.2; Fig. 1.3) with precursors chosen based on the biosynthetic pathway by Leete (1991). The secondary chemistry data were coded as separate present/absent characters, but for ease of visualization the characters were mapped onto the all character parsimony strict consensus tree as one multistate character (following Barkman 2001; Fig. 1.3) with cocaine and cis-cinnomoylcocaine as the end product of the pathway.

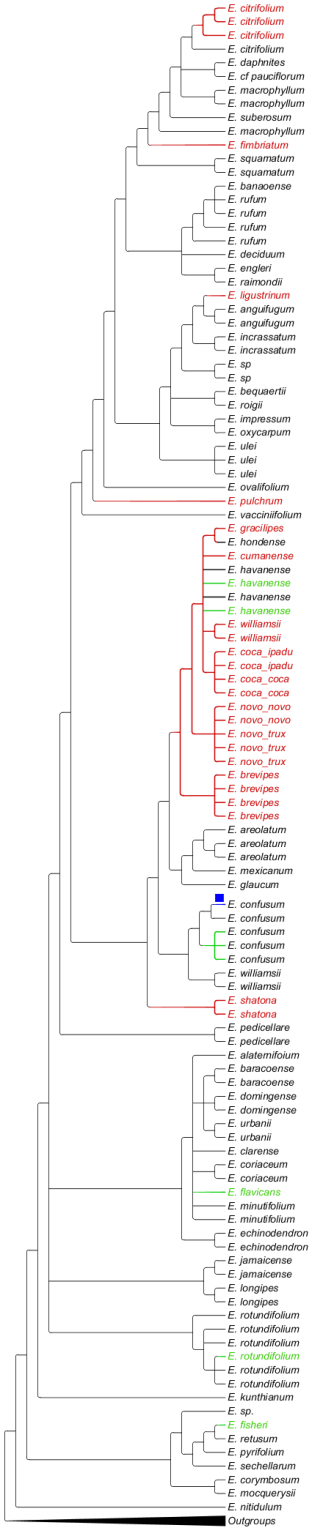


Fig. 1.3. All character majority-rule parsimony strict consensus tree with jackknife values above branch, and presence of hygrine (blue and square), methylecogonine (green), and cocaine and cis-cinnamoylcocaine (red) or absence of any alkaloid (black) mapped onto the tree in Mesquite with parsimony reconstruction. *E. novo_novo* (*E. novogranatense* var. *novogranatense*), *E. novo_trux* (*E. novogranatense* var. *truxillense*), *E. coca_coca* (*E. coca* var. *coca*), *E. coca_ipadu* (*E. coca* var. *ipadu*).

Cocaine, cis-cinnomoylcocaine, and trans-cinnomoylcocaine often occur in different ratios in a plant depending on the presence of certain molecules (Lydon pers. comm.). The actual relationship among these three molecules is unknown except that any one of them can be the end product of the pathway, and a plant may produce all three. For discussion purposes, the presence of cis-cinnomoylcocaine will not be distinguished from the presence of cocaine. Three outgroup species were tested for the presence of the alkaloids, and species from nine of Schulz's sections were examined. Based on Fisher's exact test (two-tailed), herbarium material was not significantly different (p=1.0) than fresh material in being positive for the presence of cocaine. Hence, no obvious bias in our ability to detect the presence of cocaine based on the age of the leaves.

No cocaine or any of its precursors were found in the outgroup taxa supporting the hypothesis that this pathway is restricted to *Erythroxylum*. Production of the cocaine pathway is not restricted to a single clade but occurs throughout the phylogeny (Fig. 1.3). Of the 37 species tested, eight species show presence of at least one alkaloid, and two for the presence of cocaine (Appendix 1). Clade erythroxylum1, which includes the cultivated species, has the most number of species that produce the cocaine pathway with eight species showing production of at least one constituent of the pathway. At least eleven species not included in the phylogeny are known to produce cocaine (Bieri et al 2006), and three species included in the phylogeny, *E. areolatum, E. impressum,* and *E. macrophyllum,* have at least one record indicating the presence of cocaine (see Bieri et al. 2006), but our tests found those species to be negative, and therefore, they were coded as such. The phylogeny predicts multiple origins of the cocaine pathway, however, this pattern may be due to the current sampling of species, characters, and alkaloid precursors tested. Subsequent testing for other basal precursors may indicate if the presence of the pathway occurs throughout *Erythroxylum*. This scenario would appear more parsimonious than that of multiple origins of the cocaine pathway.

Summary. Rhizophoraceae and Erythroxylaceae are sister families and based on the loci (ITS & *rpL32-trnL*) sequenced for this study appear to be highly divergent. Within Erythroxylaceae, the

intrageneric relationships among *Aneulophus, Erythroxylum, Pinacopodium,* and *Nectaropetalum* remain untested because of limited taxon sampling of these genera except for *Erythroxylum*. *Aneulophus* sequences compared to the other three genera were highly divergent as expected based on its divergent morphology. *Nectaropetalum* and *Pinacopodium* were supported as sister genera and well supported as sister to *Erythroxylum*.

Erythroxylum was strongly supported as monophyletic with the exception of the placement of *E. nitidulum* in two of the likelihood analyses. *Erythroxylum nitidulum,* native to Madagascar, differs from other *Erythroxylum* species morphologically and possibly represents a a monotypic genus or a species of *Pinacopodium*. Other African species, from Madagascar, Seychelles, and mainland Africa, are strongly supported in a clade that is sister to the remaining *Erythroxylum* species including *E. kunthianum,* native to India and Burma. The Paleotropical species are not monophyletic. Based on the phylogeny and the age estimates for the divergence of Erythroxylaceae and the connection between the continents, the distribution of Erythroxylaceae and *Erythroxylum* is consistent with boreotropical dispersal rather than the break-up of Gondwana.

Within *Erythroxylum,* only one of Schulz's sections, *Gonocladus*, is monophyletic. For the Neotropical sections sampled, *Macrocalyx* is supported within the larger section *Rhabdophyllum;* at least two clades of *Archerythroxylum* occur in the phylogeny; *Leptogramme* is not monophyletic but forms a grade out of which arises one of the *Archerythroxylum* clades; and section *Erythroxylum* includes species from *Archerythroxylum, Megalophyllum* and *Eurysepalum*, but not the type species of the section. Four strongly supported clades were given informal names with at least one morphological feature identified that is shared by the members of each clade.

Clade rhabdophyllum1, includes species from *Rhabdophyllum* and *Macrocalyx* that share the presence of stipule striations, evergreen leaves (except for *E. daphnites*), and ramified fiber sclereids. Rhabdophyllum2, includes species from section *Rhabdophyllum* which like rhabdophyllum1 has striated stipules, but differs in the absence of ramified fiber sclereids (except *E. engleri*) and the presence of deciduous leaves (except *E. banaoense* and *E. engleri*). Species from section *Archerythroxylum* form two

well supported clades, archerythroxylum2 and erythroxylum1, with a few *Archerythroxylum* species placed in other clades. Species from sections *Erythroxylum* and *Leptogramme* occur in these predominately *Archerythroxylum* clades. Archerythroxylum2 differs from erythroxylum1 in the presence of evergreen leaves versus deciduous leaves with several exceptions. Erythroxylum2 includes only Caribbean endemics. *Erythroxylum rotundifolium,* a widespread Caribbean species, is poorly to moderately supported as sister to erythroxylum2. Most of the species in erythroxylum2 and *E. rotundifolium* have evergreen leaves, and all have subdioecious or dioecious flowers.

Subdioecy or dioecy occurs at least three times, given that section erythroxylum2 includes *E. rotundifolium*. Most subdioecious species occur in the Caribbean. Higher levels of dioecy on islands compared to the mainland especially for fleshy fruited woody plants like *Erythroxylum* is a common pattern.

No cocaine or any of its precursors were found in the outgroup taxa supporting the hypothesis that this pathway is restricted to *Erythroxylum*. Production of cocaine is not restricted to a single clade but occurs scattered throughout the phylogeny. Erythroxylum1, has the highest number of species producing the cocaine or its precursors with eight species showing production of at least one constituent of the cocaine pathway. The phylogeny predicts multiple origins of the cocaine pathway.

Erythroxylum species are morphologically difficult to distinguish even with fine-scale anatomical studies. In contrast to the findings by Rury (1982) or Loiola (2001), the phylogenetic analyses support *E. brevipes, E. raimondii, E. banaoense, E. areolatum, E. confusum,* and *E. impressum* as distinct species. The phylogeny does not refute the findings of Rury (1982) that *E. havanense* and *E. cumanense* may be the same species supporting D'Arcy & Schanen (1975), and not Plowman & Hensold (2004). Further molecular and morphological studies may help elucidate the relationships between the Central American *E. havanense* and *E. cumanense* and the Caribbean *E. havanense*.

CHAPTER TWO: The Evolutionary History of the Cultivated Cocas (*Erythroxylum*)

Abstract

The tropical, flowering shrubs in the genus *Erythroxylum* (230 species) are commonly known as coca and occur throughout the tropics. Two species, each with two varieties, are exploited for their production of the secondary metabolite, cocaine. Most studies of coca are devoted to understanding the economic and medical impacts of the legal and illegal uses of cocaine. Until now, the evolutionary relationships among the cultivated species and their closest wild relatives remained untested in a modern phylogenetic framework. Based on their morphological similarity, the cultivated species were hypothesized to be more closely related to each other than to any other *Erythroxylum* species. Both species have been classified in section *Archerythroxylum*, which should include their closest relatives. This section includes approximately 60 species from the Neotropics. Two nuclear loci (nrITS and *idhB*) and one chloroplast locus (*rpl32-trn*L), along with morphological characters, were sampled from approximately 70 species to infer the evolutionary relationships among the cultivated species and to determine their closest wild relatives. The *idhB* locus was sampled from 35 species. Based on the phylogenetic analyses, the two cultivated species, *E. coca* and *E. novogranatense*, and their varieties form separate well supported clades and are supported as distinct species. Although the two cultivated species are closely related, their monophyly is not supported. Their closest wild relatives are *Archerythroxylum* species with deciduous leaves.

Introduction

Erythroxylum P. Browne (coca) is the most species rich and diverse genus in Erythroxylaceae in ecology, geography, and morphology. This genus of flowering trees and shrubs occurs throughout the world's tropics (Rury 1982; Plowman 1991). For at least 8,000 years, two species, *E. coca* and *E. novogranatense*, have been cultivated in regions of South America for the high amounts of cocaine in their leaves (Dillehay et al. 2010), which are chewed with an alkaline substance and used religiously and medicinally in regions of South America (Plowman 1984a). Each cultivated species is split into two

varieties, *E. coca* var. *coca, E. coca* var. *ipadu, E. novogranatense* var. *novogranatense*, and *E. novogranatense* var. *truxillense*. All four taxa are cultivated either for traditional uses in Bolivia, Peru, and Colombia, or they are cultivated for the extraction of cocaine (United Nations 2007). *Erythroxylum novogranatense* var. *truxillense* or 'Trujillo' coca is also cultivated for use in beverages including Coca-Cola® (Plowman 1984a; Rury 1992). The range of cultivation and the quantity of production has shifted over at least the last five hundred years and even significantly within the last thirty years (Plowman 1984a; United Nations 2007). The great majority of cultivation is for the illicit extraction and production of the illegal recreational drug, cocaine, for which the highest market for the drug is in countries outside of South America (United Nations 2006). *Erythroxylum coca* var. *coca* is still the primary taxon cultivated for extraction of cocaine (Plowman 1979, 1984a,b; United Nations 2006).

Coca plants remain culturally significant (Plowman 1984b) and economically important (Alvarez 1995; DeFranco & Godoy 1992) in both their native and non-native regions. Our understanding of the relationships between the cultivated species and their varieties, and of the events leading to domestication are based on work by Plowman (1979, 1984a, b) and Bohm et al. (1982) and remain untested in a phylogenetic framework. Studies focusing on the economic and medical impact, chemical composition, or cultural usage of the cultivated species and their varieties assume Bohm et al.'s (1982) hypothesis of a linear evolutionary relationship between the cultivated taxa. This hypothesis suggests that initially *E. coca* was domesticated and was the progenitor for subsequent selection of the other cultivated taxa. Recent studies (Johnson et al. 2003, 2005; Emche et al. 2011) fail to support the hypothesis and provide evidence of the possibility of a different series of domestication events. To infer the evolutionary history of the cultivated species, identify their closest wild relatives, and test both Bohm et al.'s (1982) hypothesis of domestication compared to Emche et al. (2011), I conducted a morphological and molecular phylogenetic analysis of *Erythroxylum*.

Materials and Methods

Plant material. Identification of the cultivated cocas is difficult with only leaf material or even herbarium specimens. Therefore, fresh leaves from the living collection at the USDA ARS, Sustainable

Perennial Crops Laboratory, Beltsville, MD, were collected from multiple accessions of mature plants of each of the cultivated cocas. All plants were verified by me and earlier by Rury (1992) as having typical morphology except for the two *E. coca* var. *ipadu* plants, which were not part of the Beltsville collection in 1992. Two accessions of *E. coca* var. *ipadu,* one of *E. novogranatense* var. *novogranatense,* and three of *E. novogranatense* var. *truxillense* accessions were also used by Johnson et al. (2005) for AFLP analyses, and only one *E. novogranatense* var. *truxillense* accession was used in Emche et al. (2011). In both analyses, these accessions clustered with plants from the same taxon as expected based on morphology. The Drug Enforcement Administration (DEA) classifies coca leaves (*E. coca* var. *coca*) as a Schedule II drug, and I obtained a DEA permit necessary for acquiring and keeping coca leaves for DNA extraction (DEA # RU0357930).

Erythroxylum is classified into 19 sections (Schulz 1907a, 1931). Schulz's section *Heterogyne* includes the type of the genus, *E. areolatum*, but the section's name should be *Erythroxylum* to reflect the inclusion of the type species (Payens 1958; Loiola 2001). The cultivated cocas are placed in section *Archerythroxylum.* If Schulz's classification coincided with the evolutionary relationships in the genus, then the closest relatives of the cultivated cocas should also be placed in this section. Although the naturalness of Schulz's sections has been questioned (Payens 1958; Rury 1982; Plowman & Rivier 1983), sampling these sections was useful for placing the cultivated cocas within the context of the genus as a whole. At least two species from each of 13 of Schulz's 19 sections were sampled except for *Megalophyllum, Schistophyllum,* and *Oxystigma,* in which only one species was sampled from each. Five species from the sister family, Rhizophoraceae, were sampled along with three species representing each of the other genera in Erythroxylaceae as outgroups.

My preliminary phylogeny of twenty species based on one locus grouped the cultivated cocas together with other species from section *Archerythroxylum*. Therefore, additional species were sampled within this section to attempt to infer the closest wild relatives of the cultivated cocas and test the relationships among the cultivated cocas. *Erythroxylum coca* var. *coca* is the only cultivated taxon known to occur as feral (and possibly wild) populations in parts of eastern Peru, especially in the departments of

Huánuco and Cuzco (Plowman 1979, 1984a,b). If Bohm et al.'s (1982) hypothesis is correct, then the closest wild relatives of the cultivated coca may occur in the same habitat and geographic area as *E. coca* (Plowman 1984a). To tease out if morphology or geography better estimates the closest wild relatives of the cultivated cocas, I sampled four species native to Peru and in *Archerythroxylum*, six species native to Peru but in Schulz's section *Erythroxylum*, *Macrocalyx*, and *Rhabdophyllum*, and fourteen non-Peruvian *Archerythroxylum* species (Table 2.1). Many non-Peruvian and non-*Archerythroxylum* species were sampled (Appendix 1, for all species sampled). When possible, multiple accessions of a species were sampled for a total of 119 accessions, representing 59 ingroup taxa and seven outgroup species (Appendix 1).

Table 2.1. Species included in the analyses to test if morphology or geography is a better predictor of the closest relatives of the cultivated cocas. All are in *Archerythroxylum* unless otherwise noted. *Species known to occur in either Huánuco or Cuzco, Peru.

Species	**Peru**	**Schulz Section**
E. anguifugum	Yes	
E. brevipes	No	
E. citrifolium	Yes	*Rhabdophyllum*
E. confusum	No	
E. cumanense	No	
E. deciduum	Yes	*Rhabdophyllum*
**E. fimbriatum*	Yes	*Rhabdophyllum*
E. glaucum	Yes	
**E. gracilipes*	Yes	
E. havanense	No	
E. hondense	No	
E. incrassatum	No	
E. impressum	No	
E. jamaicense	No	
E. ligustrinum	No	
**E. macrophyllum*	Yes	*Macrocalyx*
E. mexicana	No	
E. oxycarpum	No	
**E. raimondii*	Yes	*Rhabdophyllum*
E. roigii	No	
E. shatona	Yes	
E. squamatum	No	*Rhabdophyllum*
**E. ulei*	Yes	*Erythroxylum*
E. vacciniifolium	No	
E. williamsii	No	

Specimens for DNA extraction and coding of morphological characters were collected in the field, USDA's living collection, or by acquiring from the following herbaria: Colorado State University (CS), Herbario Nacional Colombiano (COL), The Field Museum (F), New York Botanical Garden Herbarium (NY), Missouri Botanical Garden Herbarium (MO), and Fairchild Tropical Garden Herbarium (FTG). Genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen Inc., Valencia, CA, USA) or the GenCatch Plant Genomic DNA Purification System (EPOCH Biolabs, Sugarland, TX, USA) with modifications following Johnson et al. (2003) for some species. Because of the broad geographic sampling for this study, most species were sampled from herbarium specimens.

Morphological characters. Twenty-eight morphological and alkaloid characters were coded and analyzed (Appendix 2). Two characters, habitat and Schulz's sections (Appendix 2), were not included in analyses but were mapped on to the most parsimonious trees (MPTs). Nine characters were from Rury (1982) and five from Oviedo (2002). I coded four alkaloid characters. The remaining characters were based on features deemed important for delineation of the sections by Schulz (1907a) and/or by other workers (Rury 1982; Plowman & Hensold 2004). Characters were coded from herbarium specimens, freshly collected specimens, and the literature (Baker 1883; Schulz 1907a; Corbishley 1919; Phillips 1935; Machado 1972; Rury 1982; Plowman 1991; Plowman 1984c; Zappi 1995; Plowman & Berry 1999; Felger et al. 2001; Loiola 2001; Plowman 2001; Heald 2002; Plowman & Hensold 2004; Barrie & Plowman, unpublished). Details of the extraction and analysis of alkaloid characters are described in Chapter 1.

Molecular characters. Total genomic DNA was extracted from herbarium and silica-dried material. Two loci were amplified and sequenced for almost all 119 accessions, one from the nuclear genome (nuclear ribosomal internal transcribed spacer (ITS)) and one locus from the chloroplast (plastid) genome (*rpl32-trn*L intergenic spacer (*rpl32-trn*L)). Preliminary phylogenies failed to provide clear support for or against the monophyly of the two cultivated species. Hence, another locus from a region of the *idh* nuclear gene, *idhB* (Weese & Johnson 2005), was amplified for 46 accessions.

Target loci were amplified using standard PCR procedures. For ITS, the PCR profile started with an initial denaturation of 95°C (3 min), followed by 25 cycles of 95°C denaturation (30 sec), 50°C annealing (20 sec), and 72°C extension (1 min). For *idhB*, PCR conditions were 95°C (2 min) initial denaturation, followed by 35 cycles of 95°C (1 min), 52°C (1 min), 72°C (1 min), and a final extension of 72°C (8 min). The PCR profile for *rpl32-trn*L was modified from Shaw et al. 2007 with an initial denaturation of 80°C (5 min) followed by 40 cycles of 95°C denaturation (1 min), 50°C annealing (1 min), and 65°C extension (4 min) and a final extension at 65°C (5 min). *rpl32-trn*L was amplified with a combination of primers, *rpL32F-trn*L$^{(UAG)}$ (whole gene), *1327R-trn*L (whole gene), or *1327R-467F* (5' end ~600 bp) and *467FR-trn*L (3' end ~500 bp) and. Initially, primers *rpL32F* and *trn*L$^{(UAG)}$ (Shaw et al. 2007) were used to amplify the entire region, which varies in *Erythroxylum* from 700-1200 bp. Internal primers and a new 5' primer were developed using *Erythroxylum* sequences and NCBI Blast-Primer3 (S. Rozen & H.J. Skaletsky 2000; Table 2.2). ITS was amplified and sequenced using either ITSA-ITSB (Blattner et al. 1999) or ITS5-ITS4 (White et al. 1990) primers. *Idhb* was amplified and sequenced using idh751F and idh964R primers (Weese & Johnson 2005).

Table 2.2. New *rpl32-trn*L primers.

Primer Name	**Sequence**
1327R	TGG TCA GCG TTG AAA GCT TTT TCG T
467FR	GCG ATC AAT ACC TAA TTG GGT TGC CA
467F	TGG CAA CCC AAT TAG GTA TTG ATC GC

The PCR products were then purified using ExoSapIT® (Amersham, Piscataway, NJ, USA), and DNA sequences were generated by Macrogen (Seoul, South Korea) or the University of Chicago Cancer Research Center, DNA Sequencing Facility (Chicago, IL). The same primers used to amplify the sequences were used for sequencing.

Data analyses. The nucleotide sequences obtained for each locus were aligned separately using the online version of MAFFT (Katoh et al. 2005) after the removal of primer sequences. For *rpl32-trn*L, the E-INS-i -option was selected as recommended by MAFFT for fewer than 200 sequences "with multiple conserved domains and long gaps" (Katoh 2011). The scoring matrix 1PAM/K2 was selected because the

taxa are closely related along with the default gap opening penalty 1.53 and an offset value 0.1 recommended for long gaps. For ITS, the G-INS-i option was selected because of the number of sequences and no long gaps were expected based on preliminary alignments. The scoring matrix 1PAM/K=2 was selected with the default gap opening penalty 1.53 and an offset gap penalty 0.0 because no large gaps were expected.

Manual adjustments to the alignments followed the procedure outlined by Simmons (2004), following Zurawski & Clegg (1987). Outgroup taxa were often highly divergent, especially for the chloroplast region, resulting in large areas that were ambiguously aligned between the ingroup and the outgroup; these unalignable areas were coded as ambiguous for the outgroup only. Ambiguously aligned areas for all sequences were removed prior to the analyses (*rpl32-trn*L 534/1209bp; ITS 14/723bp; *idhB* 63/821bp). Gap characters were coded for the parsimony analyses. If a locus contained parsimony-informative gap characters, these characters were coded in unambiguously aligned regions using modified complex indel coding (Simmons & Ochoterena 2000; Müller 2006). Most gaps coded were simple and did not require a step matrix except for two gaps in ITS. Twenty gap characters were coded for *rpl32-trn*L and ITS, and seven gap characters for *idhB.*

Phylogenies were estimated using morphological and molecular characters that were analyzed separately and in combination (Kluge 1989; Nixon & Carpenter 1996) using maximum parsimony in PAUP* 4.0b10 (Swofford 2002) and maximum likelihood in RAxML 7.2.6 (Stamatakis 2005, 2006). Most of the likelihood bootstrap analyses were conducted either using RAxMLGui 0.9 beta 1 (Silvestro & Michalak 2010) or RAxML in CIPRES (Portal 3, Miller et al. 2009). Four data matrices were analyzed for all 119 accessions: each locus separately, loci together, and a combined analysis of morphological and molecular characters. In the combined analyses of all characters, two accessions of a putative hybrid *E. rotundifolium*X *flavicans* were dropped because no herbarium specimens or literature was available for coding the morphological characters for this taxa. For the smaller taxon data set of 44 (46 for *idhB*) accessions, five data matrices were analyzed: each locus separately, nuclear loci together, and all loci together. Morphological characters and Schulz's classification were mapped onto the tree using Mesquite

2.72 (build 527, Maddison & Maddison 2009) and parsimony ancestral state reconstructions for the analysis of all characters (117 accessions). Schulz's classification was not used in the phylogenetic analyses, but was mapped on for illustrative purposes (see Fig. 1.2). Equally weighted parsimony heuristic searches were conducted using 2,000 tree-bisection-reconnection (TBR) searches with a maximum of 10 trees held per search. Parsimony jackknife analyses (JK; Farris et al. 1996) were conducted with the removal probability set to approximately e^{-1} (36.7879), and "jac" resampling emulated. Two-thousand replicates were performed with 100 TBR searches per replicate and a maximum of 10 trees held per TBR search.

Maximum likelihood analyses (Felsenstein 1973) were conducted for molecular characters only excluding gap characters using the only model available in RAxML, GTR, and four discrete rate categories. One-thousand separate heuristic searches were performed, using randomized stepwise parsimony trees generated by RAxML with the model above to find the best tree. A thorough bootstrap search (Felsenstein 1985) was performed until a sufficient number of bootstrap (BS) replicates was found using the majority rule criterion, autoMR (Pattengale et al. 2010). For the combined molecular analyses, the data were partitioned for the two or three loci to allow for different estimates of model parameters. Results of the bootstrap analyses were drawn on to the best tree for that dataset in RAxML. All tree figures were drawn using TreeGraph2 (Müller & Müller 2004; Stöver & Müller 2008).

Monophyly of the cultivated species was tested for the combined molecular analysis of the 119 accession (2 loci) and the 44 accession data sets (3 loci). Likelihood heuristic searches for the best tree were conducted as above but with the monophyly of the cultivated species constrained. The constrained and unconstrained best trees from the likelihood analyses were compared with the Shimodaira-Hasegawa test in PAUP.

Results

The majority-rule parsimony JK tree for the combined analysis of all characters (117 accessions, Fig. 2.1) and all molecular characters for the smaller data set (44 accessions, Fig. 2.2) are presented. The parsimony JK support values are above the branches, and the likelihood BS support values are below each

branch. If a contradictory clade was present in the likelihood analysis compared to the parsimony analysis and was supported more than 50%, then the support value is in square brackets below the node containing all the taxa in the contradictory clade. When two support values are listed in the text, the first value is for JK support and the second for the BS support. The trees from the larger data set analyses are included in Appendix 3 and includes majority-rule parsimony jackknife trees with support values as above for both methods along with the strict consensus trees and best likelihood trees from each analysis.

The likelihood and parsimony topologies for all the analyses did not strongly (>80 BS/JK) support contradictory hypotheses so no additional analyses were necessary. Summary statistics for the parsimony analyses are included in Table 2.3. Clades are informally named and annotated (Fig. 2.1, Fig. 2.2).

Table 2.3. Statistics for each of the parsimony analyses. PIC=parsimony informative characters, MPT= most parsimonious tree, CI = ensemble consistency index after removal of the parsimony-uninformative characters, RI=ensemble retention index.

Matrix	**# of terminals**	**# characters analyzed**	**# of PIC**	**% miss.**	**MPT length**	**# of MPTs**	**CI**	**RI**
ITS rDNA	125	737	306	9.5	1,162	13,230	0.52	0.97
*rpl32-trn*L	115	1,744	104	50	317	550	0.80	0.92
all molecular	119	2,481	411	40	1,489	9,750	0.57	0.87
all characters	117	2,511	440	40	1,688	9,440	0.54	0.85
idhB	46	821	136	27	278	9,430	0.84	0.90
*rpl32-trn*L	44	1,098	55	28	115	14,970	0.86	0.95
ITS	44	715	179	4.5	429	101	0.67	0.86
ITS + *idhB* (nuclear)	44	1,536	313	16	715	42	0.72	0.86
all molecular	44	2,634	368	21	847	96	0.72	0.87

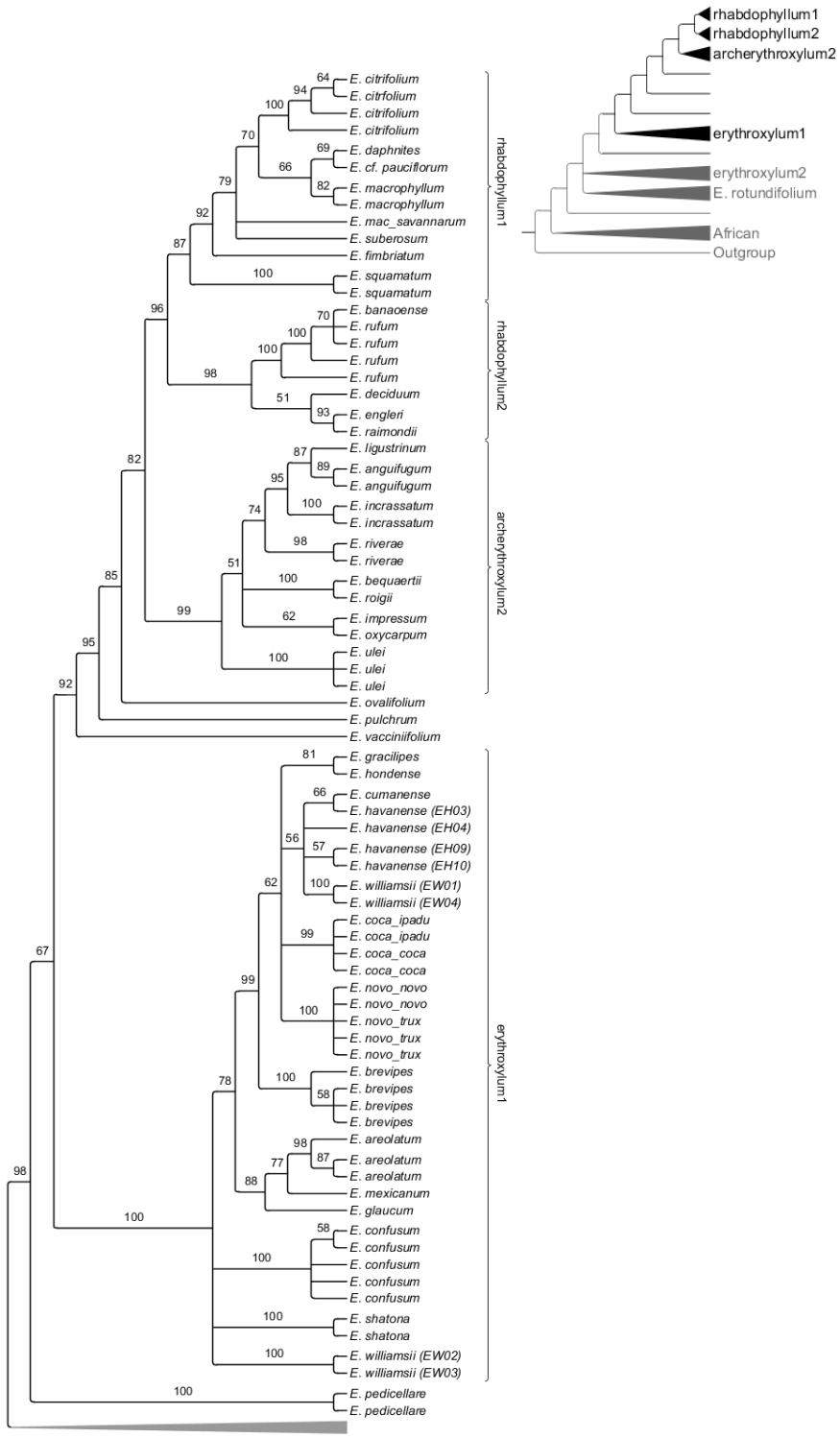


Fig. 2.1a. All character majority-rule parsimony jackknife tree with jackknife values above each branch. Clades annotated with informal clade names. *E. mac_savannarum*, *E. novo_novo* (*E. novogranatense* var. *novogranatense*), *E. novo_trux* (*E. novogranatense* var. *truxillense*), *E. coca_coca* (*E. coca* var. *coca*), *E. coca_ipadu* (*E. coca* var. *ipadu*).

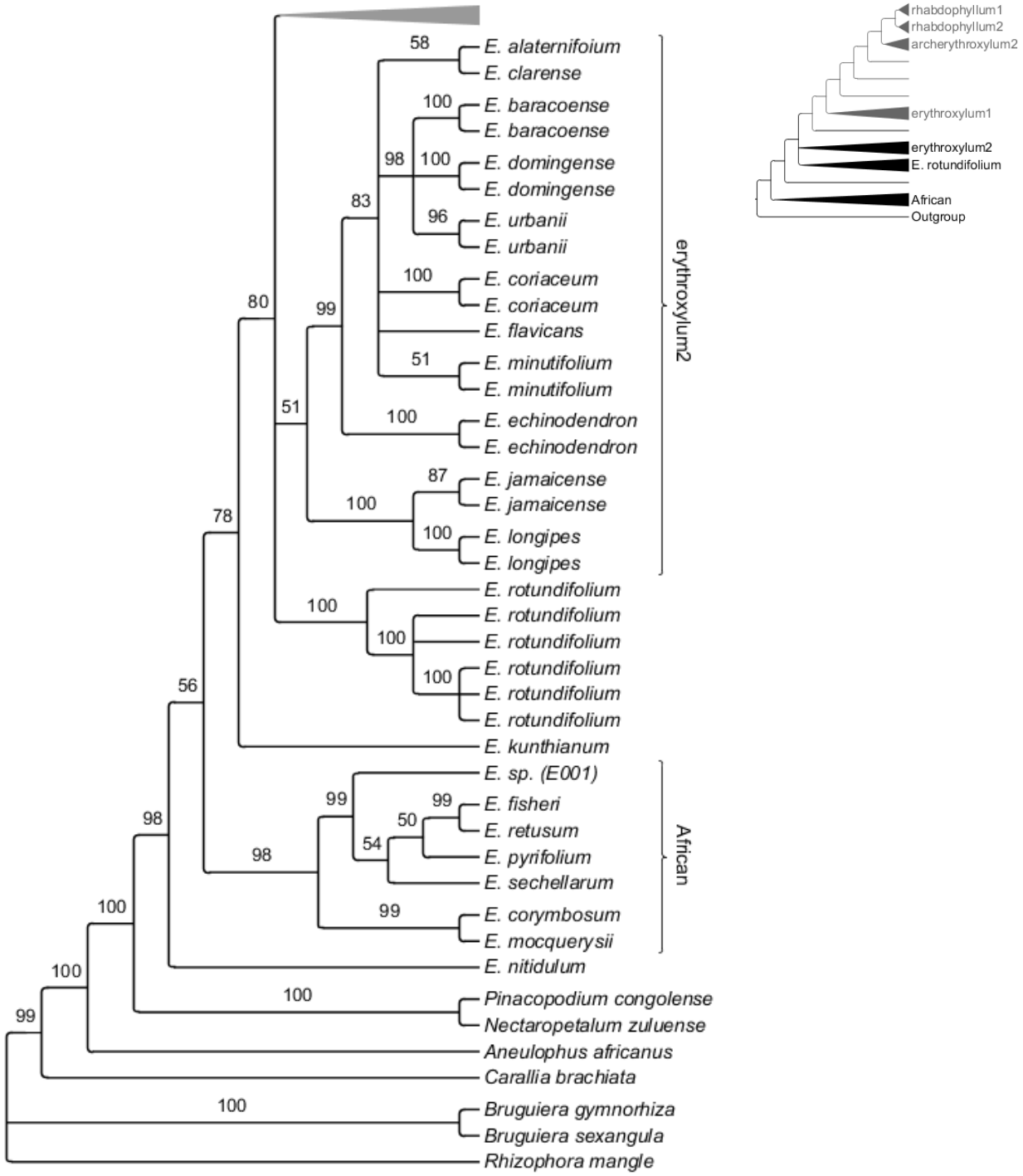


Fig. 2.1b. All character majority-rule parsimony jackknife tree with jackknife values above each branch. Clades annotated with informal clade names. *E. mac_savannarum*, *E. novo_novo* (*E. novogranatense* var. *novogranatense*), *E. novo_trux* (*E. novogranatense* var. *truxillense*), *E. coca_coca* (*E. coca* var. *coca*), *E. coca_ipadu* (*E. coca* var. *ipadu*).

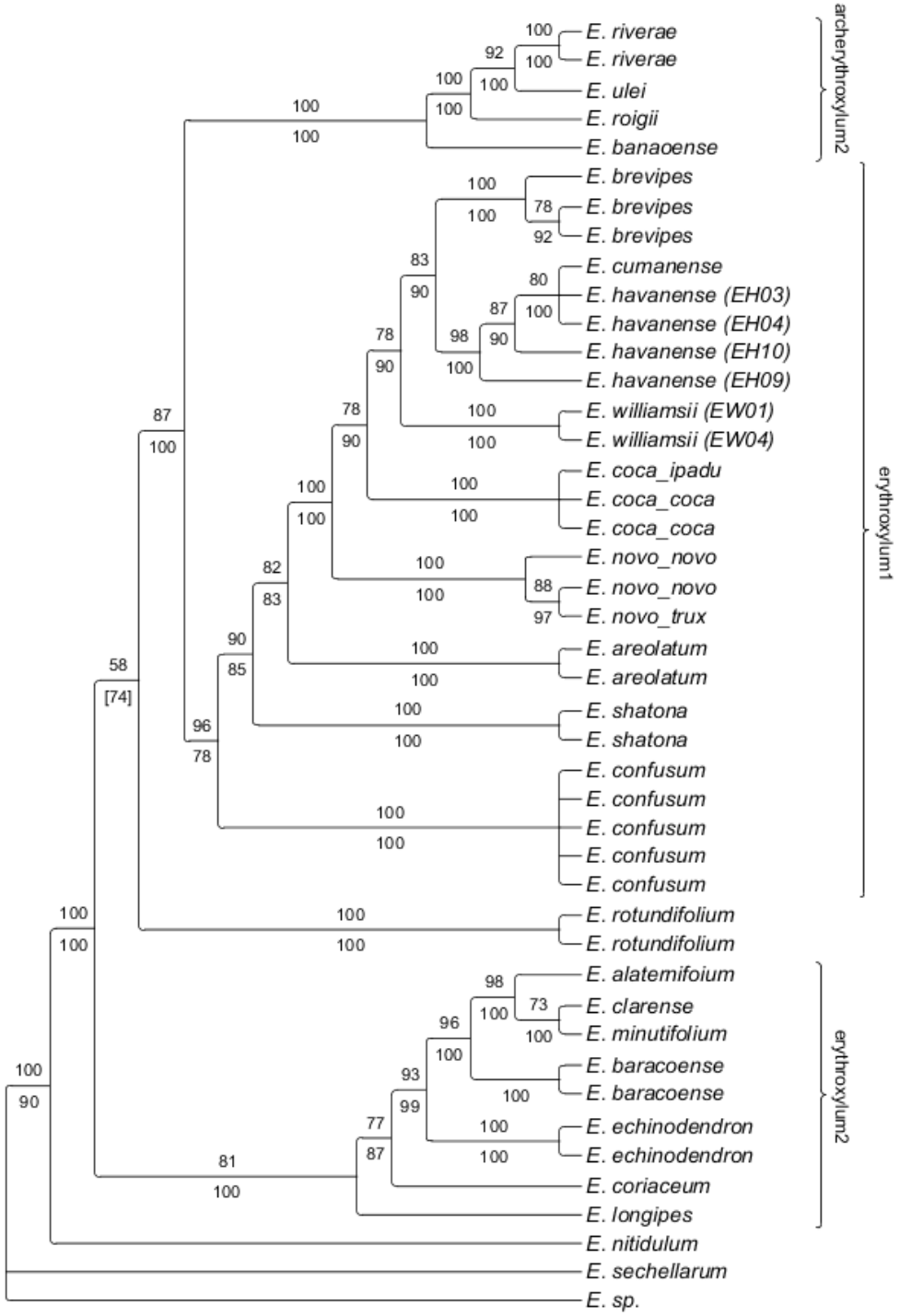


Fig. 2.2. All molecular (44 accessions) majority-rule parsimony jackknife tree with jackknife values above each branch and likelihood values below each branch. Support for incongruent clades in the best likelihood tree were placed in []. Clades annotated with informal clade names.

Erythroxylum coca and *Erythroxylum novogranatense* accessions form two separate well supported clades (Fig. 2.1, Fig. 2.2) supporting the circumscription of the cultivated cocas as two species (Plowman 1979; Johnson et al. 2005; Emche et a. 2011). The cultivated cocas are placed in the informally named clade erythroxylum1 with other species from section *Archerythroxylum* and one species, *E. areolatum,* from section *Erythroxylum* (Fig. 2.1, Fig. 2.2). Other species from section *Archerythroxylum* are placed in the informally named clade, archerythroxylum2. In clade erythroxylum1, most species have deciduous leaves except the cultivated cocas, *E. gracilipes,* and *E. shatona*. In contrast, clade archerythroxylum2, has evergreen leaves except for *E. impressum* and *E. oxycarpum*.

In the combined analysis of all characters (117 accessions), the cultivated cocas are supported (63, Fig. 2.1) in a polytomy with *E. gracilipes, E. hondense, E. havanense,* and *E. williamsii*. In the combined molecular analyses for the smaller data set (44 accessions), *Erythroxylum novogranatense* is supported (100/100, Fig. 2.2) as sister to a clade comprising *E. coca, E. williamsii, E. havanense*, and *E. brevipes* (78/90, Fig. 2.2). *Erythroxylum hondense* and *E. gracilipes* were not sampled for *idhB*. Of the four species that occur in Peru and are placed in *Archerythroxylum*, three species are placed in erythroxylum1 with the cultivated species. The other species of *Archerythroxylum* in clade erythroxylum1 are distributed in the Caribbean, northern South America, or Central America. Using the Bayesian reconstruction of ancestral geography (RASP, Yu 2011; Yu et al. 2010; see Chapter 3 for details), I infer from the all molecular data set (119 accessions, 2 loci) that *E. coca* originated from northern South American (64%) or northern South American/South American (34%) ancestor, with nearly the same result for *E. novogranatense* (51% northern South America, 28% northern South America/South America, and 18% South America, only).

Monophyly of the cultivated cocas was tested with both the 119 accessions (2 loci) phylogeny and the 44 accessions (3 loci) phylogeny. In RAxML, a heuristic ML search was preformed to find the best tree out of 50 searches (119 accessions) or 1000 searches (44 accessions) constraining the monophyly of the cultivated taxa. The best trees from the unconstrained ML searches (RAxML) for both data sets were then compared to the best trees from the constrained ML searches using the Shimodaira-Hasegawa test and

1,000 RELL bootstrap replicates in PAUP. For both data sets, the constrained tree was not significantly different than the unconstrained tree (Table 2.4; p value = 0.052 (119 accessions), and 0.063 (44 accessions)).

Table 2.4. Results of the Shimodaira-Hasegawa Test, using 1,000 RELL bootstrap replicates. *Cultivated coca species constrained as monophyletic.

	-log likelihood Constrained Tree*	-log likelihood Unconstrained Tree	P-value (one-tailed test)
all molecular (119 accessions)	11178.65429	11161.71599 (best)	0.052
all molecular (44 accessions)	8867.11016	8852.08054 (best)	0.063

Discussion

Erythroxylum is a genus of tropical shrubs and trees. For at least 8,000 years, two *Erythroxylum* species, *E. coca* and *E. novogranatense*, have been cultivated for their cocaine bearing leaves (Dillehay et al. 2010). These two species have had enormous impacts on human history and the environment (DeFranco & Godoy 1992; Alvarez 1995; Streatfeild 2001; Fjeldså et al. 2005). The legendary ethnobotanist, Richard Schultes, stated that "*Coca* must be rated as the most important and culturally significant narcotic of South America…[and]…of the New World flora, has played the greatest role in Western Medicine" (p. 119; Schultes 1979). Since the Spanish arrived in South America, the tradition of using coca leaves has been controversial, which when coupled with increased cultivation for the legal and then illegal extraction of cocaine has sparked violence, heated debate, and confusion between coca, the plants, and cocaine (Henman et al. 2009). Despite a wealth of literature, our understanding of the relationships among the cultivated species, their origin, the possible number of domestication events, and their wild relatives relies heavily on a few studies from the 1980s (Gander 1979; Bohm et al. 1982). Several recent genetic studies have shed light on the relationships among the cultivated cocas (Johnson et al. 2005; Emche et al. 2011), but not within the wider context of the evolutionary history of the genus.

Wider context. *Erythroxylum* (Schulz 1907a, 1931) was divided into 19 sections based on stipular and floral characteristics. The cultivated species were placed within the section *Archerythroxylum* with about 60 other species native to the Neotropics. My phylogenetic analyses show that the cultivated species are placed in the same moderate to well supported clade with some species from section *Archerythroxylum* in the informally named clade, erythroxylum1 (Fig. 2.1, Appendix 3). Section *Archerythroxylum* is not monophyletic, but most *Archerythroxylum* species with deciduous leaves are placed in erythroxylum1 with the cultivated species. The evergreen-leaved *Archerythroxylum* species are placed in distantly related clade, archerythroxylum2 (Fig. 2.1).

Relationships among the cultivated cocas. Based on hybrid vigor, flavonoid profiles, and ecogeography, Bohm et al. (1982) proposed a linear evolutionary hypothesis of the domestication of cocas. In the 1970s and 80s, the four taxa were cultivated allopatrically (Plowman 1984a,b); however, sparse archeological evidence points to a wider pre-Colombian cultivation (Plowman 1984b) making it difficult to interpret domestication of these taxa. The estimation of the origin of cultivated cocas by Bohm et al. (1982) depended in part on the known ranges. *Erythroxylum coca* var. *coca* occurs in cultivation from Ecuador (sparingly) to Bolivia in the cool, moist montaña zone of the Andes; only *E. coca* var. *coca* exists in the wild (Plowman 1984a). The other cultivated taxa occur only in association with humans either because they rarely produce seed, *E. coca* var. *ipadu* (Plowman 1981), or their seeds, *E. novogranatense*, quickly die in the hot, arid conditions surrounding the irrigated fields (Bohm et al. 1982). Once established, feral plants of all varieties (less so for *E. coca* var. *ipadu*) may survive, but seedling establishement has only been seen for *E. coca* var. *coca* (Plowman 1979, 1984a).

Erythroxylum coca var. *ipadu* is cultivated in the lowlands of the upper Amazon basin. The plant is ill adapted to the hot, wet climate and does not persist long after a field is abandoned (Plowman 1981). However, this may no longer be the case given the new cultivar growing in Colombia (see below). *Erythroxylum coca* var. *ipadu* was once a little known plant cultivated by a few remote Amazonian tribes in Brazil, Colombia, and Peru. Today, this variety produces almost 20% of the coca crop in Colombia (United Nations 2006; Johnson et al. 2005). Johnson et al. (2005) discovered a putative hybrid, which

looks similar morphologically to *E. coca* var. *ipadu* but based on flavonoid chemistry resembles a hybrid between *E. coca* var. *coca* and *E. novogranatense* var. *truxillense*. The phenetic analysis of AFLP data by Johnson et al. (2005) did not support the putative hybrid origin but did provide evidence for a unique, newly developed cultivar of *E. coca* var. *ipadu*.

Both varieties of *Erythroxylum novogranatense* survive in dry, hot lower elevation climates than *E. coca* (Plowman 1979, 1981). In the 1980s, *Erythroxylum novogranatense* var. *novogranatense* was cultivated on a small scale for medicinal use in remote areas of the Colombian Andes, but now approximately 20% of the coca crop in Colombia is from this variety (United Nations 2006). This greater range of cultivation may have been common during pre-Colombian times reaching into Venezuela and Panama (Plowman 1981). *Erythroxylum novogranatense* var. *truxillense* is cultivated in coastal deserts of Peru and has a better documented history of cultivation that during pre-Colombian times reached northern Chile (Plowman 1984b; Dillehay et al. 2010). Although both varieties of *E. novogranatense* are adapted to dry, hot conditions their seeds, like those of *E. coca*, only last one or two weeks before drying out (Plowman 1984a,b) and, therefore, without human intervention *E. novogranatense* would not persist in the wild.

Bohm et al. (1982) took into account the ecology and geographic preferences of the cultivated cocas and the results of breeding experiments, which showed successful crosses only between *E. coca* var. *coca* and *E. novogranatense* var. *truxillense* and breakdown of self-incompatibility in *E. novogranatense* var. *novogranatense*, to propose the following scenario of domestication. Humans first domesticated *E. coca* var. *coca* in the eastern Andes. *Erythroxylum coca* var. *coca* was introduced to the Amazonian lowlands and through human selection gave rise to *E. coca* var. *ipadu*. In a different bout of human migration or trade, *E. coca* var. *coca* was taken to drier areas on the western side of the Andes giving rise to *E. novogranatense* var. *truxillense* and then with further selection to *E. novogranatense* var. *novogranatense*. *Erythroxylum novogranatense* var. *novogranatense* was thought to be a later cultivar based on its more derived character of self-compatibility (Bohm et al. 1982). Wild populations of *E. coca* may have become extinct because of overharvesting (Plowman 1979). Plowman (1979) cited an unpublished study that

transplanted material of *E. coca* var. *coca* and *E. novogranatense* var. *truxillense* into the other taxon's habitat, and neither taxon survived in the new habitat. For Plowman (1979), this supported the linear evolutionary hypothesis, but this could alternatively support the independent domestication of two species.

If Bohm et al. (1982) witnessed the recent shifts in cultivation, then they may have favored one of their alternate hypotheses. For example, *E. coca* var. *coca* was absent from Colombia and Venezuela (Plowman 1979) leading to the hypothesis that the origin of the species was in southern South America. Today, *E. coca* var. *coca* produces almost 60% of the coca leaves in Colombia (United Nations 2006). The absence of *E. coca* in northern South America during Plowman's investigations was likely the result of colonial pressures and not a reflection of pre-Colombian cultivation or origin as alluded to by Plowman (1979, 1981). No clear data on historical ranges of cultivation for the four taxa exist except for *E. novogranatense* var. *truxillense*. Archeological plant remains showing usage of coca leaves are all from one taxon, *E. novogranatense* var. *truxillense* (Rury & Plowman 1983; Dillehay et al. 2010). This taxon was and is cultivated in the drier areas of eastern Peru, which is conducive to preservation of plant remains. For this area only one taxon appears to have been used. Other types of archeological evidence of coca use do not provide evidence for which taxa were used (Plowman 1984b).

Monophyly of cultivated cocas. If *E. novogranatense* was recently derived from *E. coca* then one would expect a lower level of genetic diversity in *E. novogranatense* than in *E. coca* based on what is usually found in putative progenitor-derivative species pairs (e.g., Ranker and Schnabel 1986). However, Johnson et al. (2005) found equally high levels of genetic diversity in *E. coca* and *E. novogranatense*, thus challenging Bohm et al.'s hypothesis of a linear evolutionary series. Johnson et al.'s (2005) study was limited to the cultivated taxa and included only one other *Erythroxylum* species. Emche et al. (2011) expanded the sampling of AFLPs to 34 additional wild species and showed the cultivated species as sister taxa. In my analyses, the two cultivated species formed two distinct clades in nearly all analyses (except *idhB,* Appendix 3) supporting the assertion that the cultivated species are two separate species (Plowman

1979; Johnson et al. 2005; Emche et al. 2011). The monophyly of the cultivated species and their origins is not as clear, however, from my results.

The cultivated species either form a polytomy with several wild species, or *E. novogranatense* is sister to *E. coca* and those wild species (Fig. 2.1, Fig. 2.2, respectively). I tested the monophyly of the cultivated species using the Shimodaira-Hasegawa test for the combined molecular large (119 accessions) and small (44) likelihood analyses. For both analyses, the constrained tree was not significantly different than the unconstrained tree (Table. 2.4; p value = 0.052 (119 accessions), and 0.063 (44 accessions)).

Bohm et al.'s (1982) hypothesis of a linear evolutionary series is not supported. Instead, the alternative hypothesis that the cultivated species are derived from two separate domestication events is consistent with the phylogenies and high levels of genetic diversity (Johnson et al. 2005). In light of the high levels of genetic diversity, the long period of cultivation (Plowman 1984b; Dillehay et al. 2010), and phylogenetic relationships among the cultivated taxa (Fig. 2.1, Fig. 2.2), the morphological similarity between the two species especially in their high levels of cocaine production may in part be attributed to artificial selection rather than shared ancestry with *E. coca* as the progenitor. Future analyses are needed that sample additional *Archerythroxylum* species with deciduous leaves, to support or reject monophyly of the cultivated species.

Origin and wild relatives. As *Erythroxylum coca* var. *coca* is the only cultivated taxon known to occur as feral populations, I used the hypothesized origin in parts of eastern Peru especially the departments of Huánuco and Cuzco (Plowman 1979, 1984a,b) to tease out if morphology or geography better estimates the closest wild relatives of *E. coca*. I sampled four species native to Peru and in *Archerythroxylum*, six species native to Peru but in different sections, and fourteen non-Peruvian *Archerythroxylum* species (Table 2.1). Of the four species that occur in Peru and are placed in *Archerythroxylum*, three species are placed in erythroxylum1 with the cultivated species. The closest relatives of the cultivated species are *E. gracilipes*, *E. hondense*, *E. williamsii*, and *E. havanense*. Emche et al. (2011) found the same species as closely related in addition to *E. impressum* (Emche et al. 2011). They did not sample *E. gracilipes*. In my phylogenies, *E. impressum* is placed within the other

Archerythroxylum clade, archerythroxylum2. Of these species, only *E. gracilipes* occurs in Peru. The other species have a similar distribution as the species in clade erythroxylum1, and are distributed in the Caribbean, northern South America, or Central America. Species native to Peru but in different sections are not closest relatives discounting geography only as an indicator of lineage, and one group of species in section *Archerythroxylum* are closely related, providing support for morphology as a good indicator of lineage.

Alternatively, the hypothesis that *E. coca* var. *coca* originated in eastern Peru is not supported by the phylogenetic analyses. Bayesian reconstruction of ancestral states for geography (RASP, Yu 2011; Yu et al. 2010; see Chapter 3) on the MPTs from the all molecular data set (119 accessions, 2 loci), supports an origin of *E. coca* from a northern South American (64%) or northern South American/South American (34%) ancestor, with nearly the same result for *E. novogranatense* (51% northern South America, 28% northern South America/South America, and 18% South America, only). Based on floristic communities, fossils, and geology, regions of South America experienced different geologic and climatic histories in addition to differing histories of mountain building throughout the Andes (Taylor 1991). For this analysis, northern South America is the region from 3°S to 10°N (Taylor 1991; Antonelli et al. 2009) and includes the North Andes, Guiana subregions, and Northern Venezuela-Colombia phytogeographic regions (Gentry 1982), and South America refers to the area south of 3°S.

Erythroxylum coca has been assumed to have originated from South America, but this assumption may have been biased because the majority of cultivation was in eastern Peru and Bolivia. Investigating the wild areas near the new sites of cultivation in Colombia may also lead to discovery of feral or wild populations distributed by birds (Plowman 1979). Hence, wild populations in eastern Peru may only indicate high areas of cultivation and not origin. The origin of the domestication events for both species is likely to be northern South America.

Based on the taxon sampling, the closest wild relatives of the cultivated species are deciduous species from section *Archerythroxylum*, the majority of which occur in northern South America. Species similar to these species should be included in future analyses to discover other wild relatives of coca. In addition,

by including a more intense taxon sampling of deciduous *Archerythroxylum* species, the inferred origin of the cultivated species from northern South America may be supported or refuted.

Summary. The cultivated species are placed in the moderate to well supported clade, erythroxylum1, with other species from section *Archerythroxylum*. Section *Archerythroxylum* is not monophyletic, but most *Archerythroxylum* species with deciduous leaves are placed in erythroxylum1 with the cultivated species. Most of the evergreen-leaved *Archerythroxylum* species are placed in distantly related clade, archerythroxylum2. The cultivated species either form a polytomy with several wild species, or *E. novogranatense* is sister to *E. coca* and those wild species. The monophyly of the cultivated species was tested and was not supported or refuted. However, Bohm et al.'s (1982) hypothesis of a linear evolutionary series is not supported. Instead, the alternative hypothesis that the cultivated species are derived from two separate domestication events is consistent with the phylogenies and high levels of genetic diversity. In light of the high levels of genetic diversity, the long period of cultivation, and phylogenetic relationships among the cultivated taxa, the morphological similarity between the two species especially in their high levels of cocaine production may in part be attributed to artificial selection rather than shared ancestry with *E. coca* as the progenitor.

Erythroxylum coca was hypothesized to have been domesticated in eastern Peru. In the phylogeny, species native to Peru but in different sections are not closest relatives. Instead, a group of *Archerythroxylum* species are closely related to the cultivated species including *E. coca* indicating that morphology is a better predictor of lineage than geography. The closest relatives of the cultivated species are *E. gracilipes*, *E. hondense*, *E. williamsii*, and *E. havanense*. Of these species, only *E. gracilipes* occurs in Peru, the putative origin of *E. coca*. The other species have a similar distribution as the species in clade erythroxylum1, and are distributed in the Caribbean, northern South America, or Central America. The Bayesian geographic reconstruction supports an origin of *E. coca* from a northern South American or northern South American/South American ancestor, with nearly the same result for *E. novogranatense*. *Erythroxylum coca* has been assumed to have originated from South America, but this assumption may have been biased because the majority of cultivation was in eastern Peru and Bolivia.

The origin of the domestication events for both species is likely to be northern South America. Based on the taxon sampling, the closest wild relatives of the cultivated species are deciduous species from section *Archerythroxylum*, the majority of which occur in northern South America.

CHAPTER THREE: Phylogeography of Caribbean *Erythroxylum*

Abstract

The Caribbean islands have a geologically complex history. The islands are diverse in their habitats, soils, and elevations both within an island and among islands leading to a high number of endemic species. Within the context of a broader investigation of *Erythroxylum* systematics, I investigated the evolutionary relationships among the Caribbean-island and continental *Erythroxylum* species within a phylogenetic framework using both molecular (ITS and *rpl32-trn*L) and morphological characters. *Erythroxylum*, in the family Erythroxylaceae, is distributed across the tropics with a typical Northern Tropical Gondwanan distribution. The age of divergence of Erythroxylaceae along with my hypothesized origin of the South America species supports a boreotropical dispersal rather than a Gondwanan origin. The Caribbean *Erythroxylum* species arose from multiple colonization events, including three long-distance dispersal events from northern South America and one from Mexico or Central America. Species from the same island were not monophyletic indicating species on the same island originated independently from either other islands or continental species, except for a large Cuban and Hispañiolan clade. This clade indicates a radiation within Cuba with one or two dispersal events to Hispañiola.

Introduction

The Caribbean islands are comprised of 1000 islands that are grouped into three archipelagos, the Greater Antilles (Cuba, Jamaica, Hispañiola, Puerto Rico, Virgin Islands, and Cayman Islands), the Lesser Antilles (two volcanic arcs of islands from Sombrero to Grenada), and the Bahamas including Turk and Caicos (Fig. 3.1; Acevedo-Rodriguez & Strong 2008). The southern Caribbean islands, Aruba, Bonaire, Curaçao, Trinidad, and Tobago, are not considered part of the Caribbean islands based on geology, flora, and fauna, but rather extensions of South America (Iturralde-Vinent & MacPhee 1999; Hedges 2006; Acevedo-Rodriguez & Strong 2008; Crew & Gillespie 2010). Each of these archipelagos has a dynamic geologic history. Species diversity in the Caribbean is typically attributed to the complex geological history, close proximity to the Americas, the diverse habitats among and within islands, and the diverse soils including serpentine, volcanic, and limestone (Acevedo-Rodriguez & Strong 2008;

Santiago-Valentín & Olmstead 2004). A robust model of the physical history of the Caribbean Basin (Iturralde-Vinent & MacPhee 1999; Graham 2003; Iturralde-Vinent 2006), including sea level fluctuations (Iturralde-Vinent & MacPhee 1999) and climate change (Curtis et al. 2001), has aided the interpretation of the processes involved in the current patterns of colonization and diversity in the Caribbean.

The elevation of the Caribbean islands and their geographic positions has shifted over time. Hypotheses of the physical history of the region continue to be debated (Iturralde-Vinent & MacPhee 1999; Hedges 2001; Francisco-Ortega et al. 2007; Pindell & Kennan 2009). For the purposes of understanding the modern flora, no permanently exposed land areas occurred until the Late Eocene uplift approximately 40ma. Land or shallow marine waterways may have connected the Greater Antilles to northern South America along the Aves Ridge (Greater Antilles + Aves Ridge or Gaarlandia; Iturralde-Vinent & MacPhee 1999) during the Early Oligocene (35-33ma). Rising sea levels in the Late Oligocene (~29-27ma) along with tectonic movements of the Caribbean plate (Miocene-Pliocene) caused isolation of the island flora by submerging land connections between the islands and the creation of deep trenches (Iturralde-Vinent & MacPhee 1999; Iturralde-Vinent 2006).

Fig. 3.1. Map of Caribbean Archipelagos

The Greater Antilles may have also been colonized by plants from the Yucatán Peninsula (Judd 2001; Chiappy-Jhones et al. 2001) based on the proximity of the Maya Block and islands off the Chortis Block (Iturralde-Vinent & MacPhee 1999; Iturralde-Vinent 2006). Cuba, Hispañiola, and Puerto Rico (with Hispañiola and Puerto Rico connected until the Middle Miocene) are hypothesized to have permanently exposed land since the Late Eocene (Iturralde-Vinent & MacPhee 1999; Iturralde-Vinent 2006). In contrast, Jamaica was not permanently emergent until 16-14ma (Iturralde-Vinent 2006).

The Lesser Antilles is comprised of two volcanic arcs that were formed in the Eocene (~45ma) for one arc and the Oligocene (~30ma) for the other arc. During the hypothesized formation of Gaarlandia,

the Lesser Antilles remained submerged, and land was not permanently exposed until after the Middle Miocene (16-14ma; Iturralde-Vinent 2006).

Studies on the origins of the Caribbean flora support dispersal to the islands but from different continental areas (reviewed by Santiago-Valentín & Olmstead 2004). Of the plant phylogeographic studies on endemic plant genera, biogeographic origins of only 20 genera of the 63 studied could be inferred (Francisco-Ortega et al. 2007). More studies evaluating the relationships between island and continental taxa within a molecular phylogenetic context are needed for the Caribbean (Francisco-Ortega et al. 2007; Acevedo-Rodriguez & Strong 2008). The flowering plant genus *Erythroxylum* P. Browne (230 species) is ecologically and geographically diverse throughout the world's tropics with 30 species dispersed throughout the Caribbean archipelagos (Table 3.1; Oviedo 2002). Six species are widespread but with differing distributions. The remaining 24 species are either endemic to one island, endemic to two islands (*E. urbanii*), or endemic to three islands (*E. brevipes*). No species are endemic to the Cayman Islands, Puerto Rico, or the Lesser Antilles.

Table 3.1**.** List of Caribbean *Erythroxylum* species by island, adopted from Oviedo (2002) with refinements. *Species not sampled. PR-Puerto Rico

Species	Notes on distribution	Species	Notes on distribution
Bahamas		**Puerto Rico**	
E. areolatum	Widespread	*E. areolatum*	Widespread
E. confusum	Widespread	*E. brevipes*	3 islands, Cuba, Hispañiola, PR
**E. reticulatum*	Endemic	*E. rotundifolium*	Widespread
E. rotundifolium	Widespread	*E. rufum*	Widespread
Cuba		*E. urbanii*	2 islands, Hispañiola, PR
E. alaternifolium	Endemic	**Jamaica**	
E. areolatum	Widespread	*E. areolatum*	Widespread
**E. armatum*	Endemic	*E. confusum*	
E. banaoense	Endemic	*E. incrassatum*	Endemic
E. baracoense	Endemic	*E. jamaicense*	Endemic
E. brevipes	3 islands, Cuba, Hispañiola, PR	*E. rotundifolium*	Widespread
E. clarense	Endemic	**Lesser Antilles**	
E. confusum	Widespread	*E. havanense*	Widespread
E. coriaceum	Endemic	*E. squamatum*	Widespread
**E. dumosum*	Endemic	**Cayman Islands**	
E. echinodendron	Endemic	*E. areolatum*	Widespread
E. flavicans	Endemic	*E. confusum*	Widespread
E. havanense	Widespread	*E. rotundifolium*	Widespread
**E. lineolatum*	Endemic		
E. longipes	Endemic		
E. minutifolium	Endemic		
**E. mogotense*	Endemic		
E. pedicellare	Endemic		
E. roigii	Endemic		
E. rotundifolium	Widespread		
E. rufum	Widespread		
**E. spinescens*	Endemic		
Hispañiola			
E. areolatum	Widespread		
**E. barahonense*	Endemic		
E. brevipes	3 islands, Cuba, Hispañiola, PR		
E. domingense	Endemic		
E. rotundifolium	Widespread		
E. rufum	Widespread		
E. urbanii	2 islands, Hispañiola, PR		
?E. williamsii	Endemic		

If the hypotheses of Iturralde-Vinent (2006) on the geologic history of the Caribbean are correct, then a number of possibilities exist for colonization of the Caribbean islands by continental *Erythroxylum.* *Erythroxylum* may have dispersed (over land) from northern South America to the Greater Antilles across the exposed land connecting northern South America to the Greater Antilles (Iturralde-Vinent & MacPhee 1999). The ancestral and sister species to the Greater Antilles would be native to northern South America

or South America (Fritsch 2003; McDowell et al. 2003; Santiago-Valentín & Olmstead 2003; van Ee et al. 2008; Crews & Gillespie 2010; Oneal et al. 2010). Subsequent isolation by rising sea levels may have lead to isolation of populations between islands and vicariant speciation (Iturralde-Vinent 2006). Species divergence time estimates along with their phylogenetic relationships would be expected to mimic the hypothesized isolation of the Greater Antilles (Iturralde-Vinent 2006; Oneal et al. 2010).

Alternatively, dispersal (over water) from the Yucatán Peninsula to the exposed land of Cuba or Hispañiola may have occurred. If this dispersal event occurred, then Cuban, Jamaican, and/or Hispañiolan species would be closely related to species from the Yucatán Peninsula (Chiappy-Jhones et al. 2001; Fritsch 2003; McDowell et al. 2003; van Ee et al. 2008). Other possible over-water dispersal events may have occurred. Once the Lesser Antilles were exposed permanently during the Miocene, dispersal from the Greater Antilles to the Lesser Antilles could occur (Glor et al. 2005; Michelangeli et al. 2008; Crews & Gillespie 2010), or from northern South America to the Lesser Antilles (Fritsch 2003; Michelangeli et al. 2008). In addition, with the establishment of a Caribbean flora long-distance dispersal events may have occurred from the Caribbean to the mainland (Santiago-Valentín & Olmstead 2003; Nicholson et al. 2005; Michelangeli et al. 2008; Roalson et al. 2008; Antonelli et al. 2009).

Erythroxylum fruits are red drupes probably dispersed by birds (Oviedo 2002) as demonstrated for *E. havanense* (Gryj & Domínguez 1996). *Erythroxylum havanense* occurs throughout the Caribbean Basin including Mexico, Central America, Cuba, and Grenada. (Plowman & Hensold 2004). Of the six bird species identified as dispersing *E. havanense*, only one species, *Vireo olivaceus* L., also occurs in the Caribbean as a migratory bird (Lepage 2011) providing a possible opportunity for long distance dispersal.

Based on the hypotheses above, I investigated the evolutionary relationships among the Caribbean and continental *Erythroxylum* species to test alternative hypotheses of the origin(s) of island taxa. The phylogeographic analyses used both morphological and molecular characters obtained from Caribbean and continental *Erythroxylum* species to reconstruct the geographic origin of the Caribbean species to infer 1) the number of colonization events, 2) the geographic origin of colonizing ancestors, and 3) relationships among island taxa.

Materials and Methods

Plant material. This study was conducted within the broader context of the evolutionary relationships of the continental *Erythroxylum* species, which included phylogenetic hypotheses for the genus throughout its range. To represent the outgroup, five species from the sister family, Rhizophoraceae Persoon, were sampled along with three species representing each of the other three genera in Erythroxylaceae Kunth (Appendix 1). Of the 30 species that occur in the Caribbean archipelagos, I sampled 23 species including all six widespread species (Table 3.2; Appendix 1). The seven species not sampled are endemic to Cuba (5), Hispañiola (1), or the Bahamas (1) (Tables 3.1, 3.2). When possible multiple accessions of each species were sampled for a total of 54 accessions. To determine the origin of the Caribbean species, 29 species from Mexico (1), northern South America (8), and South America (11) were sampled. The other nine species are widespread across several of the geographic regions above and Central America (Appendix 4).

Table 3.2. Summary of the total # of species on each island or archipelago, # of species sampled from the island, # of endemic species that occur on the island, and # of endemic species sampled. The number of species represented may not be the same as sampled, for example, a widespread species sampled from other islands is represented in the analyses but no accessions were from that island.

Island/Archipelago	***# of species***		***# of endemic species***	
	total	**sampled**	**total**	**sampled**
Bahamas	4	2	1	0
Cuba	22	14	16	11
Hispañiola	8	7	3	2
Puerto Rico	5	2	0	0
Jamaica	5	5	2	2
Cayman Islands	3	0	0	0
Lesser Antilles	2	2	0	0
	----	----	22	15

Specimens for DNA extraction and coding of morphological characters were collected in the field or by acquiring specimens from the following herbaria: Colorado State Herbarium (CS); Herbario Nacional Colombiano (COL), The Field Museum (F), New York Botanical Garden Herbarium (NY), Missouri Botanical Garden Herbarium (MO), and Fairchild Tropical Garden Herbarium (FTG). The cultivated

species were sampled with permission (DEA # RU0357930) from the USDA ARS, Sustainable Perennial Crops Laboratory, Beltsville, MD. Genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen Inc., Valencia, CA, USA) or the GenCatch Plant Genomic DNA Purification System (EPOCH Biolabs, Sugarland, TX, USA) with modifications following Johnson et al. (2003) for some species. Because of the broad geographic sampling for this study, most species were sampled from herbarium specimens.

Morphological characters. Twenty-eight morphological and alkaloid characters were coded and analyzed (Appendix 2). Characters were coded from herbarium specimens, freshly collected specimens, and the literature (Baker 1883; Schulz 1907a; Corbishley 1919; Phillips 1935; Machado 1972; Rury 1982; Plowman 1991; Plowman 1984c; Zappi 1995; Plowman & Berry 1999; Felger et al. 2001; Loiola 2001; Plowman 2001; Heald 2002; Plowman & Hensold 2004; Barrie & Plowman, unpublished).

Molecular characters. Total genomic DNA was extracted from herbarium and silica-dried material. Two loci were amplified and sequenced, one from the nuclear genome (nuclear ribosomal internal transcribed spacer (ITS)) and one locus from the chloroplast (plastid) genome (*rpl32-trn*L intergenic spacer (*rpl32-trn*L)).

Target loci were amplified using standard PCR procedures. The PCR profile for each locus differed. For the ITS locus, the profile started with an initial denaturation of 95°C (3 min), followed by 25 cycles of 95°C denaturation (30 sec), 50°C annealing (20 sec), and 72°C extension (1 min). The PCR profile for *rpl32-trn*L was modified from Shaw et al. 2007 with an initial denaturation of 80°C (5 min) followed by 40 cycles of 95°C denaturation (1 min), 50°C annealing (1 min), and 65°C extension (4 min) and a final extension at 65°C (5 min). *rpl32-trn*L was amplified with a combination of primers, *rpL32F-trn*L$^{(UAG)}$ (whole gene), *1327R-trn*L (whole gene), or *1327R-467F* (5' end ~600 bp) and *467FR-trn*L (3' end ~500 bp) and. Initially, primers *rpL32F* and *trn*L$^{(UAG)}$ (Shaw et al. 2007) were used to amplify the entire region, which varies in *Erythroxylum* from 700-1200 bp. Internal primers and a new 5' primer were developed using *Erythroxylum* sequences and NCBI Blast-Primer3 (S. Rozen & H.J. Skaletsky 2000; Table 3.3). ITS was amplified using either ITSA-ITSB (Blattner et al. 1999) or ITS5-ITS4 (White et al. 1990) primers.

Table 3.3. New *rpl32-trn*L primers.

Primer Name	Sequence
1327R	TGG TCA GCG TTG AAA GCT TTT TCG T
467FR	GCG ATC AAT ACC TAA TTG GGT TGC CA
467F	TGG CAA CCC AAT TAG GTA TTG ATC GC

The PCR products were then purified using ExoSapIT® (Amersham, Piscataway, NJ, USA), and DNA sequences were generated by Macrogen (Seoul, South Korea) or the University of Chicago Cancer Research Center, DNA Sequencing Facility (Chicago, IL). The same primers used to amplify the sequences were used for sequencing.

Data analyses. The nucleotide sequences obtained for each locus were aligned separately using the online version of MAFFT (Katoh et al. 2005) after the removal of primer sequences. For *rpl32-trn*L, the E-INS-i option was selected as recommended by MAFFT for fewer than 200 sequences "with multiple conserved domains and long gaps" (Katoh 2011). The scoring matrix 1PAM/K2 was selected because the taxa are closely related along with the default gap opening penalty 1.53 and an offset value 0.1 recommended for long gaps. For ITS, the G-INS-i option was selected because of the number of sequences and no long gaps were expected based on preliminary alignments. The scoring matrix 1PAM/K=2 was selected with the default gap opening penalty 1.53 and an offset gap penalty 0.0 because no large gaps were expected.

Manual adjustments to the alignments followed the procedure outlined by Simmons (2004), following Zurawski & Clegg (1987). Non-*Erythroxylum* outgroup taxa were often highly divergent, especially for the chloroplast region, resulting in large areas that were ambiguously aligned between the ingroup and the outgroup; these unalignable areas were coded as ambiguous for the outgroup only. Ambiguously aligned areas for all sequences were removed prior to the analyses (*rpl32-trn*L 534/1209bp; ITS 14/723bp). Gap characters were coded for the parsimony analyses. If a locus contained parsimony-informative gap characters, these characters were coded in unambiguously aligned regions using modified complex indel

coding (Simmons & Ochoterena 2000; Müller 2006). Most gaps coded were simple and did not require a step matrix except for two gaps in ITS. Twenty gap characters were coded for each locus.

Phylogenies were estimated using morphological and molecular characters that were analyzed separately and in combination (Kluge 1989; Nixon & Carpenter 1996) using maximum parsimony in PAUP* 4.0b10 (Swofford 2002) and maximum likelihood in RAxML 7.2.6 (Stamatakis 2005, 2006). Most of the likelihood bootstrap analyses were run either using RAxMLGui 0.9 beta 1 (Silvestro & Michalak 2010) or RAxML in CIPRES (Portal 3, Miller et al. 2009). Four data matrices were analyzed: each locus separately, loci together, and a combined analysis of morphological and molecular characters. In the combined analyses of all characters, two accessions of a putative hybrid *E. rotundifolium*X*flavicans* were dropped because no herbarium specimens or literature was available for coding the morphological characters for this taxa. Morphological characters and Schulz's intrageneric classification (1907a) were mapped onto the tree using Mesquite 2.72 (build 527, Maddison & Maddison 2009) and parsimony ancestral state reconstructions. Schulz's classification was not used in the phylogenetic analyses. Equally weighted parsimony heuristic searches were conducted using 2,000 tree-bisection-reconnection (TBR) searches with a maximum of 10 trees held per search. Parsimony jackknife analyses (JK; Farris et al. 1996) were conducted with the removal probability set to approximately e^{-1} (36.7879), and "jac" resampling emulated. Two-thousand replicates were performed with 100 TBR searches per replicate and a maximum of 10 trees held per TBR search.

Maximum likelihood analyses (Felsenstein 1973) were conducted for molecular characters only excluding gap characters using the only model available in RAxML, GTR, and four discrete rate categories. One-thousand separate heuristic searches were performed, using randomized stepwise parsimony trees generated by RAxML with the GTR model to find the best tree. A thorough bootstrap search (Felsenstein 1985) was performed until a sufficient number of bootstrap (BS) replicates was found using the majority rule criterion, autoMR (Pattengale et al. 2010). For the combined molecular analyses, the data were partitioned for the two loci to allow for different estimates of model parameters. Results of

the bootstrap analyses were drawn on to the best tree for that dataset in RAxML. All tree figures were drawn using TreeGraph2 (Müller & Müller 2004; Stöver & Müller 2008).

One clade was of particular interest based on the results from the analyses and its possible monophyly was tested using the Wilcoxon rank test (Templeton 1983) in PAUP. The placement of *E. rotundifolium* as sister to the majority of Caribbean endemic species was tested. A combined heuristic search of all characters constraining the monophyly of *E. rotundifolium* with the informally named clade, erythroxylum2, was conducted as above, with the results compared to the unconstrained heuristic search for the same dataset. The clade, erythroxylum2, includes most of the Caribbean endemics.

The program Reconstruct Ancestral State in Phylogenies or RASP (Yu 2011) was used to reconstruct the origins of the Caribbean *Erythroxylum*. RASP is the latest version of S-DIVA (Yu et al. 2010) a tool developed to estimate biogeography using the non-parametric method, Dispersal-Vicariance-Analysis (DIVA; Ronquist 1997), based on a cost-matrix analysis. RASP does not require a robust understanding of the geography of the area or dated clades. The backbone of the reconstruction is DIVA, but RASP incorporates phylogenetic uncertainty (Nylander et al. 2008) and uncertainty in DIVA optimization (Harris and Xiang 2009).

Coding of the phytogeographic areas followed Borhindi (1996) with slight modification based on Gentry's (1982) phytogeographic regions (Table 3.1). Based on floristic communities, fossils, and geology, regions of South America experienced different geologic and climatic histories in addition to differing histories of mountain building throughout the Andes (Taylor 1991). For this analysis, northern South America is the region from 3°S to 10°N (Taylor 1991; Antonelli et al. 2009) and includes the North Andes, Guiana subregions, and Northern Venezuela-Colombia phytogeographic regions (Gentry 1982), and South America refers to the area south of 3°S.

In RASP, the Bayesian MCMC analysis was used with the 8,800 most parsimonious trees from a heuristic maximum parsimony analysis conducted as above except that the trees were constrained as binary. The following commands in RASP were used for the analyses, number of cycles (=Ngen for Mr. Bayes) =1,000,000, number of chains = 10 (default), frequency of samples = 1,000 (=SwapFreq for Mr.

Bayes), 880 trees were discarded, temperature = 0.1 (default), maximum number of areas was four, root distribution–outgroup, and state frequency model was F81 with the default gamma distribution.

Results

The majority-rule parsimony JK tree for the combined analysis of all characters is presented (Fig. 3.2). The only node with less than 50% support shown on the tree is for the clade *E. rotundifolium* and erythroxylum2. The majority-rule parsimony JK tree for the combined molecular only analysis is in Appendix 3. The gene trees, ITS and *rpl32-trn*L, are included in Appendix 3. The likelihood and parsimony topologies for all the analyses did not strongly (>80 BS/JK) support contradictory hypotheses so no additional analyses were necessary. Summary statistics for the parsimony analyses are included in Table 3.2. Clades are informally named and annotated on all figures under the branch for the subtree.

Table 3.2. Statistics for each of the parsimony analyses. PIC=parsimony informative characters, MPT= most parsimonious tree, CI = ensemble consistency index after removal of the parsimony-uninformative characters, RI=ensemble retention index.

Matrix	# of terminals	# characters analyzed	# of PIC	% miss.	MPT length	# of MPTs	CI	RI
ITS nrDNA	125	737	306	9.5	1,162	13,230	0.52	0.97
*rpl32-trn*L	115	1,744	104	50	317	550	0.80	0.92
all molecular	119	2,481	411	40	1,489	9,750	0.57	0.87
all characters	117	2,511	440	40	1,688	9,440	0.54	0.85

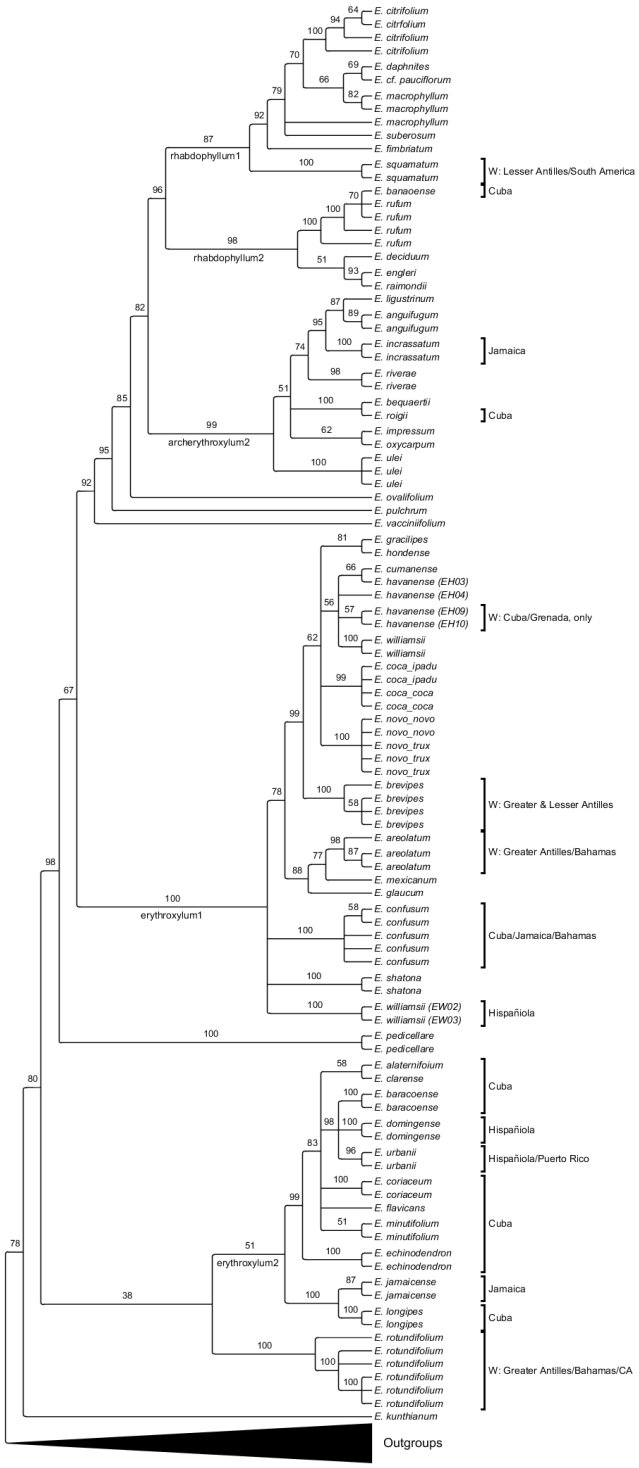


Fig. 3.2. The majority-rule parsimony jackknife tree for all characters with jackknife support values above the branches, and the informal names of well-supported clades below the branches. Caribbean species distributions are in brackets. W=Widespread; CA=Central America.

Four clades were well supported in the combined phylogeny of all characters. Caribbean species did not form a monophyletic group. The Caribbean species occurred throughout the phylogeny with one large Caribbean clade informally named, erythroxylum2 (51; Fig. 3.2.). *Erythroxylum rotundifolium*, a widespread Caribbean species, was sister to this clade with low support (38; Fig. 3.2). This relationship is shown (Fig. 3.2) because the RASP analyses reconstructing the biogeography recovered this relationship and estimated the geography of the node. To test the monophyly of *E. rotundifolium* plus erythroxylum2, a Wilcoxon rank test was conducted comparing the MPTs for the combined all character data set with and without the constraint enforced. The trees from both analyses were the same length and not significantly different from each other (p=1.00), and therefore, not conclusive as to the monophyly of *E. rotundifolium* and erythroxylum2.

All the species in clade erythroxylum2, occur in the Greater Antilles. Other species from the Greater Antilles were scattered throughout the phylogeny, but none were closest relatives to the two species that occur in the Lesser Antilles. The species from the Lesser Antilles were not monophyletic. Endemic species to Cuba, Hispañiola, and Jamaica were not monophyletic.

Based on the phylogeny, at least 10 colonizations from the mainland could be inferred. Because some Caribbean species were placed in polytomies, the origin of these species or clades cannot be determined (Fig. 3.3). Origin of four dispersal events in association with mainland species was inferred with some certainty to a geographic region. The colonization of species from other islands was inferred for three species, *E. jamaicense, E. urbanii,* and *E. domingense*. *Erythroxylum jamaicense*, endemic to Jamaica, was sister to *E. longipes*, endemic to Cuba, and based on the RASP analysis *E. jamaicense* was derived from a dispersal event from Cuba to Jamaica (92%; Fig. 3.3). *Erythroxylum domingense* and *E. urbanii* are endemic to Hispañiola and Puerto Rico for *E. urbanii*. Both species were placed within the clade of Cuban endemics and were likely derived from one or two dispersal events from Cuba (96%; Fig. 3.3). Because of the polytomy of *E. domingense, E. urbanii,* and *E. baracoense* (Cuba) the number of dispersal events cannot be inferred (Fig. 3.3). One in-situ speciation event can be inferred, *E. banaoense* in Cuba from *E. rufum*.

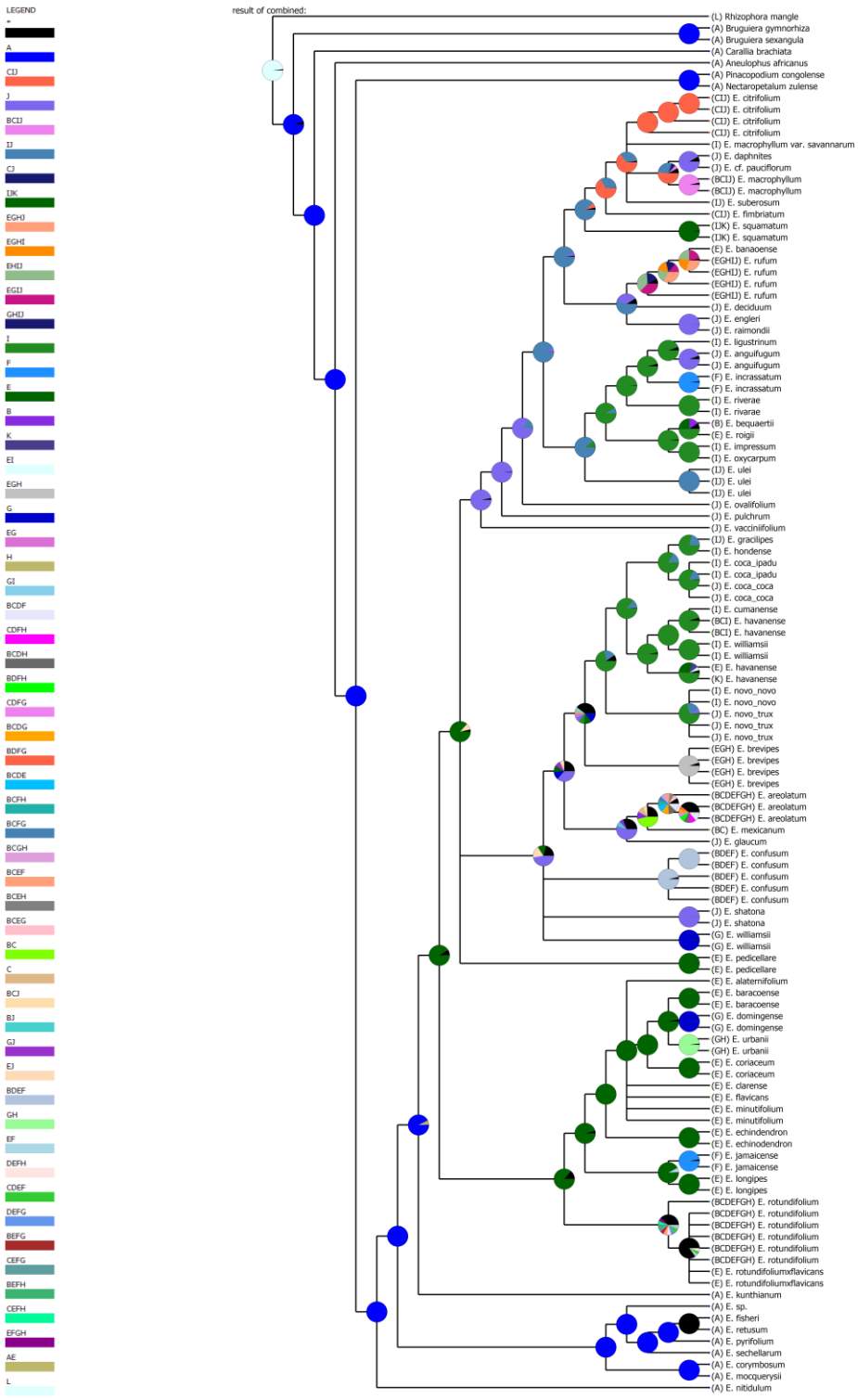


Fig. 3.3. The consensus tree of the combined analysis of all molecular characters with the Bayesian MCMC geographic reconstruction at each node estimated for 7,920 MPTs. The distribution of each species is represented by one or more of the following letters in parentheses. A=Africa; B=Mexico (Guatemala & Costa Rica); C=Central America; D=Bahamas; E=Cuba; F=Jamaica; G=Hispañiola; H=Puerto Rico; I=northern South America; J=South America; K=Lesser Antilles.

Discussion

The Caribbean *Erythroxylum* have a long history of misidentification and nomenclatural confusion. Lack of sufficient material for workers in the Caribbean has led to the merging and splitting of species, especially for several of the widespread species (Oviedo 2002, 2003). The number of species documented to occur in the Caribbean has fluctuated over the last hundred years, and no comprehensive flora yet exists for the Caribbean islands (Acevedo-Rodriguez & Strong 2008). Based on the recent review by Oviedo (2002), fieldwork by me and Ramona Oviedo Prieto, and my phylogenetic analyses, 30 *Erythroxylum* species occur in the Caribbean (Table 3.1). Twenty-three of these species were sampled for the phylogenetic analyses. Of the 22 endemic species in the Caribbean, I sampled 15 species (68%). *Erythroxylum* shows a similar distribution of species across the Caribbean as other groups of taxa with the highest number of species in the Greater Antilles compared to the Bahamas and the Lesser Antilles (Crews & Gillespie 2010), and the typical pattern of species number decreasing as island size decreases, Cuba>Hispañiola>Puerto Rico-Jamaica>Cayman Islands>Lesser Antilles (Acevedo-Rodriguez & Strong 2008; Crews & Gillespie 2010). Endemic species occur only on the larger islands, with one species endemic to the Bahaman archipelago. With one exception, *E. havanense*, widespread species are restricted to the Greater Antilles and Bahamas. *Erythroxylum havanense* has been documented in Cuba and in the Lesser Antilles.

Origin of *Erythroxylum*. To understand the Caribbean *Erythroxylum* and infer the direction of colonization, the Caribbean *Erythroxylum* must be placed within the larger context of the Old World *Erythroxylum* and Erythroxylaceae. Erythroxylaceae has four genera including *Erythroxylum* with the majority of the 240 species in *Erythroxylum* (230 species). The other three genera occur only in Africa and/or Madagascar. Erythroxylaceae's geographic distribution across the tropics reflects a possible Northern Tropical Gondwanan origin (Sanmartín & Ronquist 2004) or possibly long-distance dispersal events between Africa and the Neotropics.

If Erythroxylaceae's distribution was derived from vicariant events in association with the break-up of Gondwana, one would predict that the split between Rhizophoraceae and Erythroxylaceae would have

occurred at a minimum of 100ma (McLoughlin 2001; Antonelli et al. 2009), and that the Paleo- and Neotropical clades were reciprocally monophyletic (Antonelli et al. 2009). An alternative explanation for the typical Gondwanan distribution is over-land dispersal during the peak temperature in Eocene by establishment in the tropical habitats of the boreotropics (48-52 ma; Tiffney 1985; Zachos et al. 2001). Several plant groups support this alternative explanation (Melastomeae, Renner et al. 2001; Malpighiaceae, Davis et al. 2002; Annonaceae, Richardson et al. 2004 and Pirie et al. 2006; *Croton* L. (Euphorbiaceae *s.s.*), van Ee et al. 2008; Rubiaceae, Antonelli et al. 2009). The origin of these groups and direction of dispersal varies by plant family with Rubiaceae and Annonaceae hypothesized as originating in the Paleotropics and dispersing to the Neotropics via the boreotropics (Richardson et al. 2004; Pirie et al. 2006; Antonelli et al. 2009). Malpighiaceae originated in South America and dispersed across North America and Eurasia and into Africa (Davis et al. 2002). As outlined by Antonelli et al. (2009), the boreotropics dispersal hypothesis would predict that the age of divergence would have been in the Eocene, Paleotropical and Neotropical species would not be reciprocally monophyletic, and "Early Tertiary fossils have been found in North America, Europe, or Asia" (p. 9750).

The recent age estimate of the divergence between Rhizophoraceae and Erythroxylaceae of between 44-79 ma (Bell et al. 2010) refutes the western Gondwanan origin of Erythroxylaceae based on the last estimated age for the split between Africa and South America (McLoughlin 2001). For the current taxon sampling, the Neotropical species are monophyletic but not the Paleotropical species (Fig. 3.3). The only Asian species sampled is sister to the Neotropical *Erythroxylum* (Fig. 3.3). The South American species originated from Cuba. No fossils of Erythroxylaceae exist to support the boreotropical hypothesis, however, age of divergence along with the hypothesized origin of the South American species support the boreotropical dispersal hypothesis for Erythroxylaceae and in particular *Erythroxylum*. Similar to Malpighiaceae with the majority of species in Erythroxylaceae and *Erythroxylum* in the Neotropics, Erythroxylaceae may have originated in South America and dispersed with a possible second radiation in Madagascar (Davis et al. 2002).

Based on the biogeographical reconstruction (Fig. 3.3), the Neotropical *Erythroxylum* species were derived from a Cuban or Antillean ancestor. A similar colonization event from the Antilles to the mainland was inferred within the anoles (Nicholson et al. 2005). Based on the hypothesis of the boreotropical distribution of *Erythroxylum*, *Erythroxylum* moved between South America and 'Laurasia' (or in the other direction) possibly via the island arc between South America and Central America during the Cretaceous (~70 ma; Iturralde-Vinent 2006), and/or through Gaarlandia or the Caribbean islands (van Ee et al. 2008; Antonelli et al. 2010).

Number of colonization events. The Caribbean islands have been and are in close proximity to the mainland regions of southern North America, Mexico, Central America, and northern South America (Fig. 3.1). Grenada is only 150 km from northern South America, and Cuba is only 210 km from the Yucatán Peninsula. As expected based on the proximity of the Caribbean islands to the mainland, most studies of Caribbean species support multiple colonizations events (Lavin et al. 2003; McDowell et al. 2003; Fritsch 2003; Michelangeli et al. 2008; van Ee et al. 2008; Crews & Gillespie 2010), and single colonization events are rare (Santiago-Valentín & Olmstead 2003; Motley et al. 2005 (1-2); Judd 2001). The Caribbean *Erythroxylum* species were not monophyletic, and at least ten colonization events between the mainland and the Caribbean islands may be inferred.

Origin of Caribbean *Erythroxylum*. Several hypotheses of the likely origin of the Caribbean species of *Erythroxylum* are derived from the hypothesized geologic history of the Caribbean (Iturralde-Vinent 2006). First, species may have dispersed across exposed land connecting northern South America and the Greater Antilles during the middle Eocene (~40 ma; Fritsch 2003; McDowell et al. 2003; Heinicke et al. 2007; van Ee et al. 2008; Crews & Gillespie 2010). This hypothesis would predict that the closest relatives of the contemporary Greater Antillean species would be from northern South America or South America, and that the evolutionary history (phylogeny with age estimates) of the species would coincide with the estimated isolation of the islands (Iturralde-Vinent 2006). Alternatively, Greater Antillean species may be closest relatives to species from northern South America because of long-distance dispersal events (McDowell et al. 2003; Francisco-Ortega et al. 2008; Oneal et al. 2010). For this

analysis, northern South America is the region from 3°S to 10°N, and South America refers to the area south of 3°S (Taylor 1991; Antonelli et al. 2009). Without age estimates of cladogenesis, support for the vicariant hypothesis (Iturralde-Vinent 2006; *not* Rosen 1975, 1985) versus long-distance dispersal is often difficult to infer. Of the 10 possible colonization events inferred for *Erythroxylum* (Fig. 3.2), the origin for only four events can be inferred. For the other colonization events, either for single species, *E. havanense, E. brevipes, E. confusum, E. williamsii,* and *E. pedicellare,* or for the large clade, erythroxylum2, these species and the clade, erythroxylum2, lack immediate sister species making inference of the origin ambiguous. The widespread Caribbean species *E. rotundifolium* is shown as sister to erythroxylum2 (Fig. 3.2) but this relationships is not well supported (38).

The four colonization events that can be inferred are consistent with long-distance dispersal events rather than the vicariant hypothesis. *Erythroxylum squamatum* is native to northern South America, South America and the Lesser Antilles. This species occurs within the largely South American clade (87; Fig. 3.2) and originated from northern South America and South America (89%; Fig. 3.3) consistent with the proximity of the Lesser Antilles to northern South America and the recent origins of the islands (Iturralde-Vinent 2006). A South American floristic association with the Lesser Antilles is well documented (Acevedo-Rodriguez & Strong 2008). *Erythroxylum incrassatum*, endemic to Jamaica, shows a similar pattern to *E. squamatum* and originated from a northern South America ancestor (96%; Fig. 3.3; Acevedo-Rodriguez & Strong 2008). The other two species, *E. roigii* and *E. areolatum*, originated from two separate long-distance dispersal events but with less certainty as to which region they originated from compared to *E. squamatum* and *E. incrassatum. Erythroxylum roigii*, endemic to Cuba, is sister to *E. bequaertii,* which occurs in Mexico, and their shared lineage may have originated from northern South America (50%; Fig. 3.3). Both species were placed in a South American clade. The ancestral species that dispersed to Cuba may have been a widespread species occurring in the Yucatan Peninsula but originating in northern South America. Alternatively, the geographic reconstruction of a northern South American origin may indicate dispersal from northern South America to Cuba and then another dispersal event to Mexico.

Erythroxylum areolatum is a widespread species distributed in Mexico, Central America, Bahamas, and the Greater Antilles. As expected based on this distribution, *E. areolatum* arrived in the Caribbean from Mexico or Central America (68%; Fig. 3.3).

Relationships among Caribbean *Erythroxylum*. *Erythroxylum* species from individual islands were not monophyletic, and most Caribbean species' sister species, if known, were not Caribbean, except for species within clade erythroxylum2. This pattern indicates multiple dispersal events to individual islands. For the Greater Antilles, Jamaica is a younger island in terms of its contemporary flora (Iturralde-Vinent 2006), and the island was not hypothesized to be part of Gaarlandia. The two Jamaican endemic species, *E. jamaicense* and *E. incrassatum*, were not sister species as expected based on other studies, which cite the isolation of this island giving rise to monophyletic taxa (Hedges 1996; Nicholson et al. 2005; Crews & Gillespie 2010). Both endemic species originated from relatively recent over-water dispersal events, *E. jamaicense* from Cuba, and *E. incrassatum* from northern South America. *Erythroxylum jamaicense* was sister to *E. longipes*, endemic to eastern Cuba, and *E. jamaicense* was derived from a dispersal event from Cuba to Jamaica (92%; Fig. 3.3).

In light of the fact that Cuba is the largest island in the Caribbean, it is not surprising that Cuba has the highest number of plant families, species, and endemic species of angiosperms overall (Acevedo-Rodriquez & Strong 2008). *Erythroxylum* shows the same pattern with 22 of 30 Caribbean species occurring in Cuba, and of 16 of 22 Caribbean endemic species occurring in Cuba (Table 3.2). The Cuban species were not monophyletic although the majority of endemic species were placed in the erythroxylum2 clade. Widespread species that occur in Cuba were scattered throughout the phylogeny. Only two of the six widespread species have sister species in the phylogeny making it difficult to infer the origin of the remaining four. *Erythroxylum rufum*, a widespread species occurring in the Greater Antilles, northern South America, and South America, is closely related to *E. banaoense* a central Cuban endemic. Based on the phylogeny, *E. banaoense* may have recently diverged in-situ from populations of *E. rufum* in Cuba without any subsequent radiation and little genetic divergence (Michelangeli et al. 2008). *Erythroxylum banaoense* differs from *E. rufum* in the number of flowers per inflorescence (2-3 versus 10-

20), pedicel size (1-1.5 mm vs 5-13 mm long), and fruit size (13-15 x 4-6 mm versus 8-9 x 4-4.5 mm) along with a few other traits. *Erythroxylum rufum* occurs in a large South America clade, but its inferred geographic origin is uncertain (Fig. 3.3).

The large Caribbean clade, erythroxylum2, includes eight of the 11 Cuban endemics. No continental species is sister to this group leaving the possible continental origin for the clade unknown (Fig. 3.3), and the inferred geographic origin is Cuba (89%; Fig. 3.3). This clade demonstrates a high level of diversification within Cuba with two or three dispersal events to other islands. Dispersal to Jamaica, as discussed above, gave rise to the Jamaican endemic, *E. jamaicense*, and the dispersal event(s) to Hispañiola is discussed below. Lack of structure within the clade precludes any prediction of relationships among Cuban endemics including how these species may have been influenced by the changing geology of Cuba over the last ~30 ma (Iturralde-Vinent & MacPhee 1999).

Three endemic species occur on the island of Hispañiola. One species was previously identified as *E. williamsii* (Oviedo 2003), a species native to a small region of Venezuela (Plowman 1982). The two accessions of *E. williamsii* from Venezuela occur as a well supported clade (100; Fig. 3.2) as part of the *E. havanense* polytomy with low support (56; Fig. 3.2), and removed from the Caribbean *E. williamsii* within the large clade that includes the cultivated species, *E. coca* and *E. novogranatense* (78; Fig. 3.2). The Caribbean collections of *E. williamsii* form a well supported clade (100; Fig. 3.2) within the larger erythroxylum1 polytomy. Based on the phylogeny, I consider the accessions from Hispañiola an endemic species that will be named in a future publication. Of the other two species endemic to Hispañiola, *E. domingense* was collected in the field and *E. barahonense*, is known only from one herbarium specimen (Oviedo 2003). Recent fieldwork in the Dominican Republic failed to locate this species. *Erythroxylum domingense* is endemic to Hispañiola and *E. urbanii* is native to Hispañiola and Puerto Rico. Both species were placed within the clade of Cuban endemics, erythroxylum2, and were likely derived from one or two dispersal events from Cuba (96%; Fig. 3.3). Because of the polytomy of *E. domingense, E. urbanii,* and *E. baracoense* (Cuba) the number of dispersal events cannot be inferred (Fig. 3.3). Based on either the physical proximity of Cuba, Hispañiola, and Puerto Rico and/or the islands once part of Gaarlandia, the

relative occurrences of over-water versus over-land dispersals cannot be inferred without a dated phylogeny.

Dispersal events between the Greater Antilles and the Lesser Antilles have been documented (Glor et al. 2005; Rickleffs & Bermingham 2008; Crews & Gillespie 2010). Only two species of *Erythroxylum* occur on the Lesser Antilles. *Erythroxylum havanense*, a widespread species throughout Mexico, Central America, and northern South America, also occurs in the dry, deciduous forests of Grenada and in Cuba. Two *E. havanense* accessions from Panama and Mexico did not form a monophyletic clade with the two Caribbean accessions (Cuba and Grenada). The Caribbean species typically formed a clade with weak support in the combined all character analysis (57, Fig. 3.2), and stronger support in the combined all molecular analysis (70 parsimony jackknife, 90 likelihood bootstrap; Appendix 3). In order to better understand the relationship between island and continental populations, more populations from across *E. havanense*'s range need to be sampled. Because the Caribbean species typically form a clade, two dispersal events from the mainland, one from Mexico or Central America to Cuba and one from northern South America to Grenada, seems unlikely. Another hypothesis is human introduction of *E. havanense* to Cuba (Rickleffs & Bermingham 2008). The other species of *Erythroxylum* in the Lesser Antilles, *E. squamatum*, was discussed above and originated from northern South America. Hence, for *Erythroxylum* no dispersal events between the Greater Antilles and Lesser Antilles are unambiguously inferred.

Summary. Based on the recent review by Oviedo (2002), fieldwork with Ramona Oviedo Prieto, and my phylogenetic analyses, 30 *Erythroxylum* species occur in the Caribbean. *Erythroxylum* is distributed across the Caribbean with the highest number of species in the Greater Antilles compared to the Bahamas and the Lesser Antilles, and number of species decreasing as island size decreases, Cuba>Hispañiola>Puerto Rico-Jamaica>Cayman Islands>Lesser Antilles. Endemic species occur only on the larger islands, with one species endemic to the Bahaman archipelago. With one exception, *E. havanense*, widespread species are restricted to the Greater Antilles and Bahamas. *Erythroxylum havanense* has been documented in Cuba and in the Lesser Antilles.

The age of divergence of Erythroxylaceae and Rhizophoraceae along with the hypothesized origin of the South American *Erythroxylum* species is consistent with the boreotropical dispersal hypothesis for the origin of Erythroxylaceae and in particular *Erythroxylum* compared to the Gondwanan break-up hypothesis. The Neotropical *Erythroxylum* species are monophyletic but not the Paleotropical species. The only Asian species sampled is sister to the Neotropical *Erythroxylum*. Based on the biogeographical reconstruction, the Neotropical *Erythroxylum* species were derived from a Cuban or Antillean ancestor. Based on the hypothesis of the boreotropical distribution of *Erythroxylum*, *Erythroxylum* moved between South America and 'Laurasia' (or in the other direction) possibly via the island arc between South America and Central America during the late Cretaceous, and/or through Gaarlandia or the Caribbean islands.

The Caribbean *Erythroxylum* species were not monophyletic, and at least ten colonization events between the mainland and the Caribbean islands may be inferred. Of the 10 possible colonization events inferred for *Erythroxylum*, the origin for only four events can be inferred. For the other colonization events, either for single species, *E. havanense, E. brevipes, E. confusum, E. williamsii,* and *E. pedicellare,* or for the large clade, erythroxylum2, these species and the clade, erythroxylum2, lack immediate sister species making inference of the origin ambiguous. The widespread Caribbean species *E. rotundifolium* is shown as sister to erythroxylum2 but this relationships is not well supported.

The four colonization events that can be inferred are consistent with long-distance dispersal events rather than the vicariant hypothesis. In two separate colonization events, *Erythroxylum squamatum* and *E. incrassatum* originated from northern South America. The other two species, *E. roigii* and *E. areolatum*, originated from two separate long-distance dispersal events but with less certainty as to which region they originated from. *Erythroxylum roigii*, endemic to Cuba, is sister to *E. bequaertii*, which occurs in Mexico, and their shared lineage may have originated from northern South America. The ancestral species that dispersed to Cuba may have been a widespread species occurring in the Yucatan Peninsula but originating in northern South America. Alternatively, the geographic reconstruction of a northern South American origin may indicate dispersal from northern South America to Cuba and then another dispersal event to

Mexico. *Erythroxylum areolatum* is a widespread species is inferred to have originated from Mexico or Central America.

Erythroxylum species from individual islands were not monophyletic, and most Caribbean species' sister species, if known, were not Caribbean, except for species within clade erythroxylum2. This pattern indicates multiple dispersal events to individual islands. The two Jamaican endemic species, *E. jamaicense* and *E. incrassatum*, were not sister species as expected based on other studies. Both endemic species originated from relatively recent over-water dispersal events, *E. jamaicense* from Cuba, and *E. incrassatum* from northern South America.

In light of the fact that Cuba is the largest island in the Caribbean, it is not surprising that 22 of 30 Caribbean *Erythroxylum* species occur in Cuba, and 16 of 22 Caribbean endemic species occur in Cuba. The Cuban species were not monophyletic although the majority of endemic species were placed in the erythroxylum2 clade. Widespread species that occur in Cuba were scattered throughout the phylogeny. Only two of the six widespread species have sister species in the phylogeny making it difficult to infer the origin of the remaining four. *Erythroxylum rufum*, a widespread species is closely related to *E. banaoense* a central Cuban endemic. Based on the phylogeny, *E. banaoense* may have recently diverged in-situ from populations of *E. rufum* in Cuba without any subsequent radiation and little genetic divergence. *Erythroxylum rufum* occurs in a large South America clade, but its inferred geographic origin is uncertain. The large Caribbean clade, erythroxylum2, includes eight of the 11 Cuban endemics. No continental species is sister to this group leaving the possible continental origin for the clade unknown, and the inferred geographic origin is Cuba. This clade demonstrates a high level of diversification within Cuba with two or three dispersal events to other islands. Dispersal to Jamaica gave rise to the Jamaican endemic, *E. jamaicense*, and one or two dispersal event(s) to Hispañiola gave rise to *E. urbanii* and *E. domingense*. Lack of structure within the clade precludes any prediction of relationships among Cuban endemics including how these species may have been influenced by the changing geology of Cuba over the last ~30 ma.

Three endemic species occur on the island of Hispañiola. *Erythroxylum williamsii* was believed to have a disjunct distribution in Hispañiola and Venezuela. The Caribbean *E. williamsii* were not most closely related to the Venezuelan *E. williamsii*. Because of the phylogenetic analyses, the Caribbean *E. williamsii* is considered an endemic species and will be named in a future publication. Of the other two species endemic to Hispañiola, *E. domingense* was collected in the field and *E. barahonense*, known only from one herbarium specimen, was not located in the field. *Erythroxylum domingense* and *E. urbanii* (native to Hispañiola and Puerto Rico) were likely derived from one or two dispersal events from Cuba. Because of the polytomy of *E. domingense, E. urbanii,* and *E. baracoense* (Cuba) the number of dispersal events cannot be inferred.

Only two species of *Erythroxylum* occur on the Lesser Antilles. *Erythroxylum havanense* and *E. squamatum*. Continental and Caribbean accessions of *E. havanense* do not form a well supported clade. In order to better understand the relationship between island and continental populations, more populations from across *E. havanense*'s range need to be sampled. *Erythroxylum squamatum* originated from northern South America. No dispersal events between the Greater Antilles and Lesser Antilles are unambiguously inferred for *Erythroxylum*.

CHAPTER FOUR: Taxonomy and occurrence Caribbean *Erythroxylum*

Abstract

Within the context of a broader investigation of *Erythroxylum* systematics, I investigated the evolutionary relationships among the Caribbean-island and continental *Erythroxylum* species with phylogenetic analyses of both molecular (ITS and *rpl32-trn*L) and morphological characters. A number of the newly defined or recircumscribed Caribbean *Erythroxylum* species were included in the analyses to test species boundaries for nine species. *Erythroxylum aristigerum* should remain a synonym of *E. squamatum*. Segregation of *E. banaoense* from *E. rufum* is consistent with the morphology but was neither supported nor refuted by the phylogenetic analyses. *Erythroxylum brevipes, E. minutifolium,* and *E. domingense* are supported as distinct species. The designation of all species of *E. rotundifolium* as *E. sauve* except for the Jamaican populations was not supported in the phylogenetic analyses or by field observations of the morphological characters. The recent citation of *E. williamsii* in Hispañiola was not supported, as these populations are a previously undescribed endemic species. Species occurrence records from the Lesser Antilles were reviewed, and the number of species naturally occurring on these islands was revised from six to two, *E. squamatum* and *E. havanense*.

Introduction

One of the major problems facing the planet is the rapid loss of biodiversity caused by the direct and indirect activities of humans. This loss directly impacts ecosystem services such as pollution absorption and water quality, and biological resources such as new foods and medicines. Sound conservation and sustainable development strategies lessen the loss of biodiversity. The development of these strategies requires input from land managers, political leaders, economists, ecologists, and biologists, among others. Biologists, specifically systematists, advance basic knowledge of the Earth's biota by identifying species, determining the origin of diversity, and revealing the evolutionary patterns and processes that create diversity. This basic information is vital to workers developing conservation strategies (Dubois 2003).

Based on the number of endemic taxa and the threats to their survival, the Caribbean islands were designated a biodiversity hotspot (Myers et al. 2000). The Caribbean islands are comprised of 1000

islands that are grouped into three archipelagos, the Greater Antilles (Cuba, Jamaica, Hispañiola, Puerto Rico, Virgin Islands, and Cayman Islands), the Lesser Antilles (two volcanic arcs of islands from Sombrero to Grenada), and the Bahamas including Turk and Caicos (Acevedo-Rodriguez & Strong 2008). Species diversity in the Caribbean is typically attributed to the complex geological history, close proximity to the Americas, and the diverse habitats (Santiago-Valentín & Olmstead 2004; Acevedo-Rodriguez & Strong 2008). After Madagascar and New Caledonia, the Caribbean islands rank highest in the number of endemic species and genera (Maunder et al. 2008). Shi et al. (2005) prioritized conservation of these hotspots by adding the impacts of human population pressure and protection of the areas and ranked the Caribbean islands as one of six "hottest" hotspots.

A greater knowledge of the origins of island taxa and their relation to continental relatives aids in conservation programs within the islands (Francisco-Ortega et al. 2007; Maunder et al. 2008). The first step is to simply understand the distribution and occurrence of plant species. No comprehensive flora of the Caribbean has been completed. A checklist has been compiled but largely relies on floras of each island without additional ground truthing or checking of the taxonomy (Acevedo-Rodriguez & Strong 2008).

Erythroxylum P. Browne is a genus of tropical flowering shrubs and trees that occurs throughout the tropics including the Caribbean. Many of the Caribbean *Erythroxylum* species are restricted to particular areas on an island either because of their soil specificity, i.e., serpentine or limestone outcrops, or because of the impacts of human development on islands in their preferred habitat. Many of the populations of *Erythroxylum* I encountered or collected in the Caribbean survive only in protected forests, in the undeveloped public roadsides, or limestone outcrops within fields. Those plants occurring near pasturelands, fields, or roads often show evidence of multiple years of stem removal by humans and resprouting. Stems are used for firewood or as plant stakes for vine crops such as yams. *Erythroxylum* species grow in habitats heavily impacted by humans, and thus in a region already under immense human pressure this genus often occurs at the forefront of land conversion from dry, deciduous forests to pastures and fields.

The original description and first publication of the name *Erythroxylum* describes two species from Jamaica (Browne 1756; Plowman 1976), which were later named *E. areolatum* by Linnaeus (1759) and *E. rotundifolium* by Lunan (1814). This genus is often under-collected and notoriously difficult to identify based on herbarium specimens alone (Plowman 1984c; Oviedo 2003), which has delayed taxonomic revisions of the group especially in the Caribbean (Oviedo 2002). Floras have perpetuated synonyms or repeated species occurrence based on misidentifications adding to the confusion (Britton 1907; Schulz 1907b; Barker & Dardeau 1930; Plowman 1988). As an example, *E. rotundifolium*, a widespread species, was first described in 1756 (Browne 1756), but no name was published until 1814 (Lunan 1814), and no type specimen was designated for the name until 2004 (Plowman & Hensold 2004). *Erythroxylum rotundifolium* has eight synonyms not including varieties.

Oviedo (2002) reviewed thousands of herbarium specimens of Caribbean *Erythroxylum* and conducted nearly thirty years of fieldwork resulting in several nomenclatural changes (Oviedo 2002; 2003). I tested her nomenclatural recommendations by sampling Caribbean species and analyzing them in a broader study of the evolutionary relationships of *Erythroxylum*. I will discuss the findings of the phylogeny as it impacts the taxonomy of nine Caribbean *Erythroxylum* species, and then review occurrence of species in the Lesser Antilles.

Taxonomy

Within the context of a broader investigation of *Erythroxylum* systematics, I investigated the evolutionary relationships among the Caribbean-island and continental *Erythroxylum* species with phylogenetic analyses of both molecular (ITS and *rpl32-trn*L) and morphological characters. The Caribbean *Erythroxylum* species arose from multiple colonization events, including three long-distance dispersal events from northern South America and one from Mexico or Central America. Species from the same island were not monophyletic indicating species on the same island originated independently from either other islands or continental species, except for a large Cuban and Hispañiolan clade. This clade indicates a radiation within Cuba with one or two dispersal events to Hispañiola. A number of the newly defined or recircumscribed species of Oviedo (2002) were included in this analysis.

Erythroxylum squamatum. *Erythroxylum squamatum* Sw. is unusual in being variable for the presence or absence of stipule striations, depending on geography. In his monograph of Erythroxylaceae, Schulz (1907a) split this species into two species based on the presence of striations in the Caribbean and Guianan specimens, which he referred to as *E. squamatum,* and the absence of striations in the South American specimens, *E. aristigerum* Peyr. These species were placed in different sections. *Erythroxylum squamatum* was placed in *Rhabdophyllum* because of the presence of striations, and *E. aristigerum* was placed in section *Archerythroxylum* because of the lack of striated stipules (Plowman 1988). I included two specimens of *E. squamatum* in my analyses, one from Grenada and the other from Brazil. These two specimens were strongly supported as a clade in the phylogenetic analyses. *Erythroxylum aristigerum* should continue to be treated as a synonym of *E. squamatum*.

***Erythroxylum rufum* and *E. banaoense*.** *Erythroxylum rufum* Cav. is a widespread Caribbean species occurring in the Greater Antilles (except in Jamaica) and South America. Oviedo (2002; 2003) segregated populations of *E. rufum* from 'the mountains of Banao, Sancti Spíritus' into a new species *E. banaoense* R. Oviedo based on several morphological characters (Table 4.1). Based on the phylogeny, *E. banaoense* may have recently diverged in-situ from populations of *E. rufum* in Cuba without any subsequent radiation and little genetic divergence. However, the phylogenetic analyses failed to support or reject the species delineation of *E. banaoense*. *Erythroxylum banaoense* differs from *E. rufum* in leaf duration, number of flowers per inflorescence, pedicel size, and fruit size (Table 4.1) and should be treated as a separate species based on the morphology until additional evidence sheds light on the relationship between *E. rufum* and *E. banaoense*.

Table 4.1. Comparison of morphological traits between *E. rufum* and *E. banaoense*. Using descriptions by Oviedo (2002; 2003) and Schulz (1907b) along with herbarium specimens.

Character	***E. rufum***	***E. banaoense***
leaf duration	deciduous	Evergreen
leaf shape	elliptic	ovate-elliptic
leaf size	4.2-9.8 x 3.2-5.3 cm	5-7 x 2.5-3.5 cm
petiole	4-5 mm	5-7 mm
stipule shape	lanceolate	Truncate
pedicels	5-13, 5 mm	1-1.5 x 1 mm
# of flowers/ inflorescence	10-20	2-3 (4)
drupe size	8-9 x 4-4.5 mm	13-15 x 4-6 mm
drupe shape	oblong-ovate	Oblong

Erythroxylum brevipes. *Erythroxylum brevipes* DC. is one of the few endemic Caribbean species that occurs on more than one island (Cuba, Hispañiola, and Puerto Rico). Liogier (1985) considered this species a synonym of *E. rotundifolium*, but Oviedo (2002) recommended the recognition of the two species. *Erythroxylum brevipes* is generally multi-stemmed shrub, 2 m (7 m, Bisse 1981) in height, whereas *E. rotundifolium* is a small tree 7 m tall. Otherwise, these species only differ in petiole length (1-5 mm versus 0.8-2 mm) and stipule length (0.8-1.5 mm versus 1.5-3 mm) for *E. rotundifolium* and *E. brevipes*, respectively. Based on the phylogenetic analyses, *E. brevipes* and *E. rotundifolium* are distantly related and although morphologically similar should continue to be considered distinct species.

***Erythroxylum minutifolium* and *E. domingense*.** *Erythroxylum minutifolium* Griseb. was first described from western Cuba (Schulz 1907b) and was treated in various floras as occurring in both Cuba and Hispañiola (Briton 1907; Barker & Dardeau 1930; Liogier 1985). Oviedo (2002; 2003) split the species into two species based on several morphological characters that differ between Cuba and Hispañiola (Table 4.2). The Cuban populations remain as *E. minutifolium*, and the Hispañiolan populations are *E. domingense* R. Oviedo. In the phylogenetic analyses, *E. domingense* forms a well supported clade with *E. urbanii* O.E. Schulz, endemic to Hispañiola and Puerto Rico, and *E. baracoense* Borhindi, endemic to Cuba. The phylogenetic analysis supports Oviedo's (2002; 2003) separation of *Erythroxylum minutifolium* and *E. domingense*.

Table 4.2. Comparison of morphological traits between *E. minutifolium* and *E. domingense*. Using descriptions by Schulz (1907b), Britton (1907), and Oviedo (2002; 2003) along with herbarium specimens. *Angle of secondary veins from midrib.

Character	***E. minutifolium***	***E. domingense***
leaf shape	orbicular	ovate
leaf size	3.5-6 mm x 3.5-8 mm	5-7 mm x 2-3 mm wide
leaf apex	retuse	deeply emarginate
leaf venation	craspedromous-semicrapedromous	actinomorphic-perfect
*secondary leaf venation	wide	acute
petiole	0.75-1 mm	1-1.2 mm
stipule shape	triangular truncate	triangular hastate
stipule apex	bisected-setulose	acuminate, not bisected
pedicels	1-2 mm	5 mm
drupe shape	ovate	oblong

***Erythroxylum rotundifolium* and *E. suave*.** *Erythroxylum rotundifolium* is a widespread, polymorphic species occurring in the Bahamas, Greater Antilles, Mexico, and Central America. Oviedo (2002) segregated the Jamaican populations of *E. rotundifolium* from the other populations of *E. rotundifolium* based on a few morphological characters (Table 4.3). The first description of the species was by Browne (1756) but he didn't provide a name. Lunan named the Jamaican specimens of Browne's description in 1814. So the species on Jamaica remained *E. rotundifolium*, and for the other populations outside of Jamaica Oviedo (2002) applied the name *E. suave* O.E. Schulz (2002). Six accessions of *E. rotundifolium* s.l. were included in the final phylogenetic analyses and were sampled from Costa Rica, Jamaica, Cuba, and Hispañiola. The two accessions from Jamaica are supported in the *E. rotundifolium* clade. *Erythroxylum rotundifolium* and *E. suave* are not supported as distinct species. *Erythroxylum rotundifolium* is the valid name, and *E. suave* remains a synonym. My recent observations in the field (Hispañiola and Jamaica) support the phylogenetic conclusions, and the morphological differences cited in Oviedo (2002) are not consistently different between the populations in Jamaica from populations outside of Jamaica.

Most species within *Erythroxylum* differ by only a few characters whether the species are closely or distantly related, and the characters are usually consistent throughout the entire range of the species.

Therefore, the few widespread and polymorphic *Erythroxylum* species, such as *E. macrophyllum* and *E. rotundifolium*, have led to questions of the distinctiveness of disjunct populations and to many synonyms in attempts to segregate the variability. The molecular data used in my phylogenetic analyses are consistent with a recent (Bell et al. 2010) rapid radiation, so the molecular markers used in the study may not be sufficient to separate two closely related sister species. The current evidence for *E. rotundifolium* is that it is one species, albeit polymorphic, although a fine-scale genetic study may provide evidence for or against splitting this species.

Table 4.3. Comparison of morphological traits between *E. rotundifolium* and *E. suave* (Oviedo 2002).

Character	***E. rotundifolium***	***E. suave***
leaf shape	obovate-spatulate	obovate-rounded
leaf size	2.5-4 cm x 1.5-2.5 cm	0.5-2 cm x 0.5-1.5 cm
leaf apex	rounded	obtuse
leaf venation	craspedromous-semicraspedromus	camptodromus-brochiodromous
stipule apex	no setae	bisetate
# of flowers in inflorescence	3 to 5	1 to 2
drupe size	1-1.2 cm long	5-8 mm long

***Erythroxylum williamsii*.** The endemic Venezuelan species, *E. williamsii* Standl. ex Plowman, was recently reported as being collected in Hispañiola (Oviedo 2003) based on earlier collected accessions. These Hispañiolan specimens were previously misidentified as *E. brevipes, E. rotundifolium,* or *E. minutifolium*. The only morphological differences between the Caribbean and Venezuelan specimens are stipule length (1.5-3 mm versus 1-1.5 mm, respectively) and shape of the stipule base (sagittate versus truncate, respectively). For the phylogenetic analysis, two specimens from each locality were sampled. The accessions from each locality formed well supported clades that were not closely related to each other. *Erythroxylum williamsii* specimens from Venezuela were most closely related to *E. havanense* Jacq. and the cultivated cocas, *E. coca* Lam. and *E. novogranatense* (D. Morris) Hieron. The Hispañiola specimens were placed either in a large polytomy that included the cultivated coca clade with *E. williamsii* (Venezuela) or as sister to *E. confusum* Britton. *Erythroxylum confusum* is distributed in the

Bahamas, Cuba, Jamaica and Mexico. The phylogeny does not support a close relationship between the Hispañiolan specimens and the species for which it has been misidentified, *E. brevipes*, *E. minutifolium*, and *E. rotundifolium.* The Hispañiolan specimens are not *E. williamsii* or any other previously identified species, and in collaboration with Ramona Oviedo Prieto will be given a new name and a detailed description based on a review of the few Caribbean specimens.

Erythroxylum in the Lesser Antilles

The actual occurrence of *Erythroxylum* species in the Lesser Antilles is less well understood compared to the Greater Antilles and Bahamas, and reports of several species on these islands are based on a few collections that are 100 years old. To sample all the species that occur on a particular island or an archipelago for the phylogenetic analyses, I reviewed herbarium specimens, current floras, and the taxonomic review by Oviedo (2002) to determine the geographic distribution of the Caribbean *Erythroxylum*. For the Lesser Antilles, which are comparatively depauperate of *Erythroxylum* species compared to the Greater Antilles, six species, *E. brevipes*, *E. havanense*, *E. lineolatum* DC., *E. novogranatense*, *E. oxycarpum* O.E. Schulz, and *E. squamatum* Sw., were reported to occur (Plowman 1988). *Erythroxylum novogranatense* and *E. lineolatum* were likely introduced by humans (Plowman 1988).

Erythroxylum lineolatum was reported to be cultivated in Martinique (Schulz 1907b) although the only cited specimens were housed in Berlin and destroyed. This species is reported to also occur in Cuba based on a few older specimens (Oviedo 2002). The report for Cuba is either based on a misidentification or the species was introduced as an ornamental plant in Cuba like in Martinique. This species is native to northern South America and Trinidad, which is not considered part of the Lesser Antilles, and was likely never naturally occurring in the Caribbean.

Erythroxylum novogranatense was introduced to several islands as a possible source of cocaine and was reported as naturalized in Martinique and Guadeloupe (citation in Plowman 1988). Naturalization of this species would be highly unlikely (Bohm et al. 1982; Plowman 1984a,b).Cultivated species are not

useful for understanding the phylogeography of the Caribbean *Erythroxylum*, and therefore, no attempt to include specimens of these species from the Lesser Antilles was made.

Erythroxylum brevipes was described as occurring in the Lesser Antilles (Britton 1907; Schulz 1907b; Plowman 1988; Plowman & Hensold 2004), but this is based on only two Guadeloupe specimens. Only one specimen is recorded in Plowman's database (2011), and this specimen was annotated as *E. havanense* (Plowman 1988). No new specimens of *E. brevipes* from the Lesser Antilles have been found. Of the three remaining species, *E. squamatum* and *E. havanense* occur on a number of islands, and I collected vouchers of both species from Grenada. The third species, *E. oxycarpum*, has only been collected in the hillsides around or in the city of St. George, Grenada, at various times between 1891-1906. With colleagues, I searched for *E. oxycarpum*. Most of these sites are urban areas and were likely urban at the time of the original collections. *Erythroxylum oxycarpum* was not found in the undeveloped areas within St. George. *Erythroxylum havanense* occurs in dry areas similar to *E. oxycarpum* and was found in these undeveloped or neglected areas in the city. If *E. oxycarpum* did occur naturally in Grenada and was not an artifact of human introduction, then it is likely extinct from the island as its preferred habitat of 'open places by the sea' have been heavily impacted by human development (Hawthorne et al. 2004).

Literature cited

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

Acevedo-Rodriguez, P., M.T. Strong. 2008. Floristic richness and affinities in the West Indies. *Botanical Review* 74: 5-36.

Alvarez, E.H. 1995. Economic development, restructuring and the illicit drug sector in Bolivia and Peru: current policies. *Journal of Interamerican Studies and World Affairs* 37: 125-149.

Anderson, G.J., G. Bernardello, M.R. Opel, A. Santos-Guerra, M. Anderson. Reproductive biology of the dioecious Canary Islands endemic *Withania aristata* (Solanaceae). *American Journal of Botany* 93: 1295-1305.

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

APGII. Angiosperm Phylogeny Group II. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of Linnean Society* 141: 399-436.

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

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

Aynilian, G.H., J.A. Duke, W.A. Gentner, N.R. Farnsworth. 1974. Cocaine content of *Erythroxylum* species. *Journal of Pharmaceutical Sciences* 63: 1938-1939.

Baker, J.G. 1883. Contributions and further contributions to the Flora of Madagascar. *Journal of Linnean Society* 20: 1883.

Barker, H.D., W.S. Dardeau. 1930. *Flore D'Haïti*. Port-Au-Prince: Départment de l'Agriculture.

Barkman, T.J. 2001. Character coding of secondary chemical variation for use in phylogenetic analyses. *Biochemical Systematics and Ecology* 29: 1-20.

Barrie, F.R., T. Plowman. unpublished. 136. Erythroxylaceae. *Flora of Mesoamerica*.

Bell, C.D., D.E. Soltis, P.S. Soltis. 2010. The age and diversification of the angiosperms re-revisited. *American Journal of Botany* 97: 1296-1303.

Blattner, F.R. 1999. Direct amplification of the entire ITS region from poorly preserved plant material using recombinant PCR. *Biotechniques* 27: 1180-1186.

Berry, E.W. 1938. Tertiary flora from the Rio Pichileufu, Argentina. *Geological Society of America Special Paper* 12:1-149.

Bieri, S., A. Brachet, J.L. Veuthey, P. Christen. 2006. Cocaine distribution in wild *Erythroxylum* species. *Journal of Ethnopharmacology* 103: 439-447.

Bisse, J. 1981. *Arboles de Cuba*. La Habana: Ministerio de Cultura Editorial Cientifico-Tecnica.

Bohm, B.A., F.R. Ganders, T. Plowman. 1982. Biosystematics and evolution of cultivated coca (Erythroxylaceae). *Systematic Botany* 7: 121-133.

Borhidi, A. 1996. *Phytogeography and vegetation ecology of Cuba*, 2 ed. Budapest.

Britton, N.L. 1907. Family 3. Erythroxylaceae. *North American Flora* (Part 1) 25: 59-66.

Browne, P. 1765. *Civil and Natural History of Jamaica*. London: Printed for the Author; and fold by T. Osborne, and J. Shipton.

Cara, D.M., A.M. Norris, and L.J. Neale. Pain during awake nasal intubation after topical cocaine or phenylephrine/lidocaine spray. *Anaesthesia* 58: 777-780.

Chase, M.W., D.E. Soltis, R.G. Olmstead, D. Morgan, D.H. Les, B.D. Mishler, M.R. Duvall, R.A. Price, H.G. Hills, Y.-L. Qui, K.A. Kron, J.H.. Rettig, E. Conti, J.D. Palmer, J.R. Manhart, K.J. Sytsma, H.J. Michaels, WJ. Kress, K.G. Karol, W.D. Clark, M. Hedren, B.S. Gaut, R.K. Jansen, K.-J. Kim, C.F. Wimpee, et al. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. *Annals of the Missouri Botanical Garden* 80: 528-580.

Chiappy-Jhones, C., V. Rico-Gray, L. Gama, L. Giddings. 2001. Floristic affinities between the Yucatán Peninsula and some karstic areas of Cuba. *Journal of Biogeography* 28: 535-542.

Corbishley, A.G. 1919. Nectaropetalum zulense. *Bulletin of Miscellaneous Information/Royal Botanical Gardens, Kew* 1919: 449-450.

Crews, S.C., R.G. Gillespie. 2010. Molecular systematics of *Selonops* spiders (Araneae: Selenopidae) from North and Central America: implications for Caribbean biogeography. *Biological Journal of the Linnean Society* 101: 288-322.

Croteau, R., T.M. Kutchan, N.G. Lewis. 2000. Chapter 24: Natural Products (Secondary Metabolites). *In*: B. Buchanan, W. Gruissen, R. Jones, editors. *Biochemistry & Molecular Biology of Plants*, Rockville, MD: American Society of Plant Physiologists.

Curtis, J.H., M. Brenner, D.A. Hodell. 2001. Climate change in the circum-Caribbean (Late Pleistocene to present) and implications for regional biogeography. *In*: C.A. Woods & F.E. Sergile, editors. *Biogeography of the West Indies. Patterns and Perspectives*, pp. 35-54. ed. 2. New York: CRC Press.

D'Arcy, W.G., N. Schanen. 1975. Erythroxylaceae. *In*: R.E. Woodson, R.W. Schery and Collaborators., editors. Flora of Panama. Part VI. Family 87. *Annals of Missouri Botanical Garden* 62: 21-33.

Darwin, C. 1888. *The different forms of flowers on plants of the same species.* London: John Murray.

Davis, C.C., W.R. Anderson, M.J. Donoghue. 2001. Phylogeny of Malpighiaceae: evidence from chloroplast *ndhF* and *trnL-F* nucleotide sequences. *American Journal of Botany* 88: 1830-1846.

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

Davis, C.C., K.J. Wurdack. 2004. Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. *Science* 305: 676-769.

Davis, C.C., M. Latvis, D.L. Nickrent, K.J. Wurdack, D.A. Baum. 2007. Floral gigantism in Rafflesiaceae. *Science* 315: 1812.

DeFranco, M. and R. Godoy. 1992. The economic consequences of cocaine production in Bolivia: historical, local and macroeconomic perspectives. *Journal of Latin American Studies* 24: 375-406.

Dillehay, T.D., J. Rossen, D. Ugent, A. Karathansis, V. Vasquez, P.J. Netherly. 2010. Early Holocene coca chewing in northern Peru. *Antiquity* 84: 939-953.

Domínguez, C.A., G. Ávila-Sakar, S. Vásquez-Santana, J. Márquez-Guzmán. 1997. Morph-biased sterility in the tropical distylous shrub *Erythroxylum havanense* (Erythroxylaceae). *American Journal of Botany* 84: 626-632.

Dräger, B. 2002. Analysis of tropane and related alkaloids. *Journal of Chromatography, A* 978: 1-35.

Dubois, A. 2003. The relationships between taxonomy and conservation biology in the century of extinctions. *Comptes Rendus Biologies* 326: S9-S21.

Emche, S.D., D. Zhang, M.B. Islam, B.A. Bailey, L.W. Meinhardt. 2011. AFLP phylogeny of 36 *Erythroyxlum* species. *Tropical Plant Biology* online February 7, 2011. Url. http://dx.doi.org/10.1007/s12042-011-9070-9

Fan, C., M.D. Purugganan, D.T. Thomas, B.M. Wiegmann, Q.-Y. Xiang. 2004. Heterogeneous evolution of the *Myc-like* anthocyanin regulatory gene and its phylogenetic utility in *Cornus* L. (Cornaceae). *Molecular Phylogenetics and Evolution* 33: 580-594.

Farris, J.S., V.A. Albert, M. Källersjö, D. Libscomb, A.G. Kluge. 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* 12: 99-124.

Feldberg, K., J. Váňa, D.G. Long, A.J. Shaw, J. Hentschel, J. Heinrichs. 2010. A phylogeny of Adelanthaceae (Jungermanniales, Marchantiophyta) based on nuclear and chloroplast DNA markers, with comments on classification, cryptic speciation and biogeography. *Molecular Phylogenetics and Evolution* 55: 293-304.

Felger, R. S., M.B. Johnson, M.F. Wilson. 2001. *The Trees of Sonora, Mexico*. New York: Oxford University Press.

Felsenstein, J. 1973. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic Zoology* 22: 240-249.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783-791.

Fjeldså, J., M. D. Álvarez, J.M. Lazcano, B. León. 2005. Illicit crops and armed conflict as constraints on biodiversity conservation in the Andes region. *Ambio* 34: 205-211.

Francisco-Ortega, J., I. Ventosa, R. Oviedo, F. Jiménez, P. Herrera, M. Maunder, J.L. Panero. 2008. Caribbean island Asteraceae: systematics, molecules, conservation on a biodiversity hotspot. *Botanical Review* 74: 112-131.

Francisco-Ortega, J., E. Santiago-Valentín, P. Acevedo-Rodríguez, C. Lewis, J. Pipoly III, A.W. Meerow, M. Maunder. 2007. Seed plant genera endemic to the Caribbean Island biodiversity hotspot: A review and a molecular phylogenetic perspective. *Botanical Review* 73: 183-234.

Fritsch, P.W. 2003. Multiple geographic origins of Antillean *Styrax*. *Systematic Botany* 28: 421-430.

Gander, F.R. 1979. Heterostyly in *Erythroxylum coca* (Erythroxylaceae). *Botanical Journal of the Linnean Society* 78: 11-20.

Gentry, A.H. Neotropical floristic diversity: phytogeographic 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.

Glor, R.E., J.B. Losos, A. Larson. 2005. Out of Cuba: overwater dispersal and speciation among lizards in the *Anolis carolinensis* subgroup. *Molecular Ecology* 14: 2419-2432.

Gonzáles-Guevara, J.L., J.A. Gonzáles-Lavaut, S. Pino-Rodríguez, M. García-Torres, M.T. Carballo-Gonzáles, O.A. Echemendia-Arana, J. Molina-Torres, S. Prieto-Gonzáles. 2004. Phytochemical screening and in vitro antiherpetic activity of four *Erythroxylum* species. *Acta Farmacia Bonaerense* 23: 506-509.

Graham, A. 2003. Geohistory models and Cenozoic paleoenvironments of the Caribbean region. *Systematic Botany* 28: 378-386.

Harris, A.J., Q-Y. Xiang. 2009. Estimating ancestral distributions of lineages with uncertain sister groups: a statistical approach to dispersal-vicariance analysis and a case using *Aesculus* L. (Sapindaceae) including fossils. *Journal of Systematics and Evolution* 47: 349-368.

Hawtorne, W.D., D. Jules, G. Marcelle. 2004. *Caribbean Spice Island Plants: Trees, Shrubs and Climbers of Grenada, Carriacou and Petit Martinique: a picture gallery with notes on identification, historical and other trivia*. Oxford: Oxford Forestry Institute.

Heald, S.V. 2002. Erythroxylaceae (Coca Family). *In*: S.A. Mori, G. Cremers, C.A. Gracie, J.J. de Granville, S.V. Heald, M. Hoff, J.D. Mitchell, editors. Guide to the Vascular Plants of Central French Guiana, Part 2, Dicotyledons. *Memoirs of The New York Botanical Garden* 76: 263-265.

Hedges, S.B. 1996. The origin of the West Indian vertebrates. *Annual Review of Ecology and Systematics* 27: 163-196.

Hedges, S.B. 2001. Biogeography of the West Indies: An Overview. *In:* C.A. Woods & F.E. Sergile, editors. *Biogeography of the West Indies: Patterns and Perspectives*, pp. 15-33. Boca Raton: CRC Press.

Hedges, S.B. 2006. Paleogeography of the Antilles and the origin of the West Indian amphibians and reptiles. *Applied Herpetology* 3: 281-292.

Hegnauer, R. 1981. Chemotaxonomy of Erythroxylaceae (including some ethnobotanical notes on the Old World species). *Journal of Ethnopharmacology* 3: 279-292.

Heinicke, M.P., W.E. Duellman, S.B. Hedges. 2007. Major Caribbean and Central American frog faunas originated by ancient oceanic dispersal. *Proceedings of the National Academy of Sciences of the United States of America* 104: 10092-10097.
Henman, P. Metaal. 2009. Coca myths. *Drugs & Conflict Debate Papers* 17: 1-24.
Heywood, V.H. (ed.). 1993. *Flowering Plants of the World*. Oxford: Oxford University Press.
Holmstedt, B., E. Jäätmaa, K. Leander, T. Plowman. 1977. Determination of cocaine in some South America species of *Erythroxylum* using mass fragmentography. *Phytochemistry* 16: 1753-1755.
Howarth, D.G., D.A. Baum. 2002. Phylogenetic utility of a nuclear intron from nitrate reductase for the study of closely related plant species. *Molecular Phylogenetics and Evolution* 23: 525-528.
Iturralde-Vinent, M.A. 2006. Meso-Cenozoic Caribbean Paleogeography: implications for the historical biogeography of the region. *International Geology Review* 48: 791-827.
Iturralde-Vinent, M.A., R.D.E. MacPhee. 1999. Paleogeography of the Caribbean region: implications for Cenozoic biogeography. *Bulletin of the American Museum of Natural History* 238: 1-95.
Johnson, E.L., D. Zhang, S.D. Emche. 2005. Inter- and Intra-specific variation among five *Erythroxylum* taxa assessed by AFLP. *Annals of Botany* 95: 601-608.
Johnson, E.L., J.A. Saunders, S. Mischke, C.S. Helling, S.D. Emche. 2003. Identification of *Erythroxylum* taxa by AFLP DNA analysis. *Phytochemistry* 64: 187-197.
Johnson EL. 1996. Alkaloid content in E*rythroxylum coca* tissue during reproductive development. *Phytochemistry* 42: 35-38.
Judd, W.S. 2001. Phylogeny and biogeography of *Lyonia* sect. *Lyonia* (Ericaceae). *In*: C.A. Woods & F.E. Sergile, editors. *Biogeography of the West Indies. Patterns and Perspectives*, pp. 63-75. ed. 2. New York: CRC Press.
Katoh, K., K. Kuma, H. Toh, T. Miyata. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* 33: 511-518.
Katoh, K. 2011. MAFFT version 6, online. http://mafft.cbrc.jp/alignment/server/.
Kluge, A. G. 1989. A concern for evidence and a phylogenetic hypothesis for relationships among Epicrates (Boidae, Serpentes). *Systematic Zoology* 38: 7-25.
Kodandaramaiah, U. 2010. Use of dispersal-vicariance analysis in biogeography-a critique. *Journal of Biogeography* 37: 3-11.
Kulji, K.K.M., S.E.C. Sierra, S.G.A. Draisma, R. Samuel, P.C. van Welzen. 2007. Molecular phylogeny of *Macaranga, Mallotus*, and related genera (Euphorbiaceae s.s.): insights from plastid and nuclear DNA sequence data. *American Journal of Botany* 94: 1726-1743.
Kuzoff, R.K., J.A. Sweere, D.E. Soltis, P.S. Soltis, E.A. Zimmer. 1998. The phylogenetic potential of entire 26S rDNA sequences in plants. *Molecular Biology & Evolution* 15: 251-263.
Lavin, M., M.F. Wojciechowski, P. Gasson, C. Hughes, E. Wheeler. 2003. Phylogeny of robinioid legumes (Fabaceae) revisited: *Coursetia* and *Gliricidia* recircumscribed, and a biogeographical appraisal of the Caribbean endemics. *Systematic Botany* 28: 387-409.
Leete, E., J.A. Bjorklund, M.M. Couladis, S.H. Kim. 1991. Late intermediates in the biosynthesis of cocaine: 4-(1-methyl-2-pyrrolidinyl-)3-oxobutanoate and methyl ecogonine. *Journal of American Chemistry Society* 113: 9286-9292.
Lepage, D. 2011. Avibase-bird checklist of the world, Caribbean (West Indies). Url: http://avibase.bsc-eoc.org/checklist.jsp?region=car&list=howardmoore
Le Péchon, T., J.-Y. Dubuisson, T. Haevermans, C. Cruaud, A. Couloux, L.D.B. Gigord. 2010. Multiple colonizations from Madagascar and converged acquisition of dioecy in the Mascarene Dombeyoideae (Malvaceae) as inferred from chloroplast and nuclear DNA sequence analyses. *Annals of Botany* 106: 343-357.
Li, A.-M., X.-Q. Wu, D.-X. Zhang, S.C.H. Barrett. Cryptic dioecy in *Mussaenda pubescens* (Rubiaceae): a species with stigma-height dimorphism. *Annals of Botany* 106: 521-531.
Linnaeus, C. 1759. *Systema Naturae*. Stockholm: Kiesewetter.
Liogier, A.H. 1985. Familia 108. – *Erythroxylaceae*. La Flora de la Española. III. *San Pedro de Marcorís*: 319-322.

Loiola, M. I. B. 2001. *Revisão taxonômica de* Erythroxylum *P. Browne sect.* Rhabdophyllum *O. E. Schulz (Erythroxylaceae Kunth)*. Ph.D. Thesis. Recife, Brasil: Universidade Federal Rural de Pernambuco.

Lunan, J. 1814. *Hortus jamaicensis.* Jamaica: St. Jago de la Vega Gazette.

Machado, E. 1972. El genero *Erythroxylon* en el Peru. *Raymondiana* 5: 5-101.

Maddison, W. P. and D.R. Maddison. 2009. Mesquite: a modular system for evolutionary analysis. version 2.73 http://mesquiteproject.org

Mathews, S., M.J. Donoghue. 1999. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* 286: 947-950.

Maunder, M., A. Leiva, E. Santiago-Valentín, D.W. Stevenson, P. Acevedo-Rodríguez, A.W. Meerow, M. Mejía, C. Clubbe, J. Francisco-Ortega. 2008. Plant Conservation in the Caribbean Island Biodiversity Hotspot. *Botanical Review* 74: 197-207.

McDowell, T., M. Volovsek, P. Manos. 2003. Biogeography of *Exostema* (Rubiaceae) in the Caribbean region in light of molecular phylogenetic analyses. *Systematic Botany* 28: 431-441.

McLoughlin, S. 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian Journal of Botany* 49: 271-300.

Meyers, N., R.A. Mittermeier, C.G. Mittermeier, G.A.B. da Fonseca, J. Kent. Biodiversity hotspots for conservation priorities. *Nature* 403: 853-858.

Mi, Q., B. Cui, G.L. Silva, D. Lantvit, E. Lim, H. Chai, M. You, M.G. Hollingshead, J.G. Mayo, A.D. Kinghorn, and J.M. Pezzuto. 2001. Pervilleine A, a Novel Tropane Alkaloid that Reverses the Multidrug-resistance Phenotype. *Cancer Research* 61: 4030-4037.

Michelangeli, F.A., W.S. Judd, D.S. Penneys, J.D. Skean Jr., E.R. Bécquer-Granados, R. Goldenberg, C.V. Martin. 2008. Muliple events of dispersal and radiation of the Tribe Miconieae (Melastomataceae) in the Caribbean. *Botanical Review* 74: 53-77.

Miller, M.A., M.T. Holder, R. Vos, P.E. Midford, T. Liebowitz, L. Chan, P. Hoover, T. Warnow. The CIPRES Portals. CIPRES. 2009-08-04. Url: http://www.phylo.org/sub_sections/portal.

Motley, T.J., K.J. Wurdack, P.G. Delprete. 2005. Molecular systematics of the Catesbaeeae-Chiococceae complex (Rubiaceae): flower and fruit evolution and biogeographic implications. *American Journal of Botany* 92: 316-329.

Müller, J., K. Müller. 2004. Treegraph: automated drawing of complex tree figures using an extensible tree description format. *Molecular Ecology Notes* 4: 786–788.

Müller, K. 2006. Incorporating information from length-mutational events into phylogenetic analysis. *Molecular Phylogenetics and Evolution* 38: 667-676.

Nicholson, K.E., R.E. Glor, J.J. Kolbe, A. Larson, S.B. Hedges, J.B. Losos. 2005. Mainland colonization by island lizards. *Journal of Biogeography* 32: 929-938.

Nixon, K. C. and J. M. Carpenter. 1996. On simultaneous analysis. *Cladistics* 12: 221-242.

Novák, M., C.A. Salemink, I. Khan. 1984. Biological activity of the alkaloids of *Erythroxylum coca* and *Erythroxylum novogranatense*. *Journal of Ethnopharmacology* 10: 261-274.

Nylander, J.A.A., U. Olsson, P. Alström, I. Sanmartín. 2008. Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to dispersal-vicariance analysis of the Thrushes (Aves: *Turdus*). *Systematic Biology* 57: 257-268.

Oliveira, S.L., M.S. da Silva, J.F. Tavares, J.G. Sena-Filho, H.F.S. Lucena, M.A.V. Romero, J.M. Barbosa-Filho. 2010. Tropane alkaloids from *Erythroxylum* genus: distribution and compliation of ^{13}C-NMr spectral data. *Chemistry & Biodiversity* 7: 302-326

Oneal, E., D. Otte, L.L. Knowles. 2010. Testing for biogeographic mechanisms promoting divergence in Caribbean crickets (genus *Amphiacusta*). *Journal of Biogeography* 37: 530-540.

Oviedo, R.P. 2002. *Contribución al conocimiento de Erythroxylaceae Kunth en Cuba y las Antillas.* Master's Thesis. Habana, Cuba: Universidad de la Habana.

Oviedo, R. 2003. Novelties in *Erythroxylum* (Erythroxylaceae) of the Greater Antilles. *Willdenowia* 33: 187-195.

Pattengale, N.D., M. Alipour, O.R.P. Binida-Emonds, B.M.E. Moret, A. Stamatakis. 2010. How many bootstrap replicates are necessary? *Journal of Computational Biology* 17: 337-354.
Payens, J.P.D.W. 1958 (1955-1958). Erythroxylaceae. *In*: CGGJ van Steenis, editor. Flora of Malesiana, Series I Spermatophyta, Vol. 5. pp. 543-552. Djakarta: P. Noordhoff Ltd.
Phillips, E.P. 1935. The genera *Erythroxylon* L. and *Nectaropetalum* Engl. *South African Journal of Science* 32: 305-312.
Pirie, M.D., L.W. Chatrou, J.B. Mols., R.H.J. Erkens, J. Oosterhof. 2006. Andean-centered genera in the short-branch clade of Annonaceae: Testing biogeographical hypothesis using phylogeny reconstruction and molecular dating. *Journal of Biogeography* 33: 31-46.
Pindell, J., L. Kennan. 2009. Tectonic evolution of the Gulf of Mexico, Caribbean and northern South America in the mantle reference frame: an update. *In*: K. James, M.A. Lorente, J. Pindell, editors. *The geology and evolution of the region between North and South America*. Geological Society of London, Special Publication.
Plowman, T. 1976. Orthography of *Erythroxylum* (Erythroxylaceae). *Taxon* 25: 141-144.
Plowman, T. 1979. The identity of Amazonian and Trujillo coca. *Botanical Museum Leaflets* 27: 45-68.
Plowman, T. 1981. Amazonian coca. *Journal of Ethnopharmacology* 3: 195-225.
Plowman, T. 1982. Three new species of *Erythroxylum* (Erythroxylaceae) from Venezuela. *Brittonia* 34: 442-457.
Plowman, T. 1984a. The ethnobotany of coca (*Erythroxylum* spp., Erythroxylaceae). *Advances in Economic Botany* 1: 62-111.
Plowman, T. 1984b. The origin, evolution, and diffusion of coca (*Erythroxylum* spp.) in South and Central America. *In*: D. Stone, editor. *Pre-Columbian plant migration*. Papers of the Peabody Museum of Archaeology and Ethnology, 76: 126-157. Cambridge, MA: Harvard University.
Plowman, T. 1984c. New taxa of *Erythroxylum* (Erythroxylaceae) from the Amazon basin. *Acta Amazonica* (supl.) 14: 117-143.
Plowman, T. 1988. Erythroxylaceae. *In*: R.A. Howard, editor. *Flora of the Lesser Antilles* Vol. 4. Dicotyledoneae – Part 1. pp. 543-551. Jamaica Plain, Massachusetts: Arnold Arboretum, Harvard University.
Plowman, T. 1989. Erythroxylaceae. *In:* G. Harling & L. Andersson, editors. *Flora of Ecuador* 36: 1–32. Göteborg, Sweden: University of Göteborg.
Plowman, T. 1991. Erythroxylaceae Kunth. *In*: W. Burger, editor. Flora Costaricensis. *Fieldiana Botany New Series* 28: 30-36.
Plowman, T., P.E. Berry. 1999. Erythroxylaceae. *In*: P.E. Berry, K. Yatskievych, B.K. Holst, editors. *Flora of the Venezuelan Guayana*. Vol. 5 Eriocaulaceae-Lentibulariaceae. pp. 59-71. St. Louis: Missouri Botanical Garden Press.
Plowman, T. 2001. Erythroxylaceae Kunth. *In*: W.D. Stevens, C. Ulloa Ulloa, A. Pool & O.M. Montiel, editors. Flora de Nicaragua. *Monographs Systematic Botany Missouri Botanical Garden* 85: 834-838.
Plowman, T. and Rivier, L. 1983. Cocaine and cinnamoylcocaine content of *Erythroxylum* Species. *Annuals of Botany* 51: 641-659.
Plowman, T., P.E. Berry. 1999. Erythroxylaceae. *In*: P.E. Berry, K. Yatskievych, B.K. Holst, editors. *Flora of the Venezuelan Guayana*. Vol. 5 Eriocaulaceae-Lentibulariaceae. pp. 59-71. St. Louis: Missouri Botanical Garden Press.
Plowman, T., N. Hensold. 2004. Names, types, and distribution of neotropical species of *Erythroxylum* (Erythroxylaceae). *Brittonia* 56: 1-53.
Plowman, T. 2011. Erythroxylum Collection Database, Dept. of Botany, Field Museum. url: http://fm1.fieldmuseum.org/collections/search.cgi?dest=erthroy&action=form
Ranker, T.A. and A.F. Schnabel. 1986. Allozymic and morphological evidence for a progenitor-derivative species pair in *Camassia* (Liliaceae). *Systematic Botany* 11: 433-445.
Ree, R.H., S.A. Smith. 2008. Maximum-likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* 57: 4-14.

Renner, S.S., K. Meyer. 2001. Melastomeae come full circle: biogeographic reconstruction and molecular clock dating. *Evolution* 1315-1324.
Richardson, J.E, L.W. Chatrou, J.B. Mols, R.H.J. Erkens, M.D. Pirie. 2004. Historical biogeography of two cosmopolitan families of flower plants: Annonaceae and Rhamnaceae. *Philosophical Transactions of the Royal Society of London B* 359: 1495-1508.
Ricklefs, R., E. Bermingham. 2008. The West Indies as a laboratory of biogeography and evolution. *Philosophical Transactions of The Royal Society B.* 363: 2393-2413.
Rivier, L. 1981. Analysis of alkaloids in leaves of cultivated *Erythroxylum* and characterization of alkaline substances used during coca chewing. *Journal of Ethnopharmacology* 3: 313-335.
Roalson, E.H., L.E. Skog, E.A. Zimmer. 2008. Untangling Gloxinieae (Gesneriaceae). II. Reconstructing biogeographic patterns and estimating divergence times among New World continental and island lineages. *Systematic Botany* 33: 159-175.
Rodeiro, I., M.T. Donato, I. Martínez, Hernández, G. Garrido, J.A. González-Lavaut, R. Menéndez, A. Laguna, J.V. Castell, and M.J. Gómez-Lechón. 2008. Potential hepatoprotective effects of new Cuban natural products in rat hepatocytes culture. *Toxicology in Vitro* 22: 1242-1249.
Ronquist, F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology* 46: 193-201.
Rosen, D.E. 1975. A vicariance model of Caribbean biogeography. *Systematic Zoology* 24: 431-464.
Rosen, D.E. 1985. Geological hierarchies and biogeographic congruence in the Caribbean. *Annals of the Missouri Botanical Garden* 72: 636-659.
Rozen, S., H.J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. *In*: Krawetz S., Misener S., editors. *Bioinformatics Methods and Protocols: Methods in*
Rury, P.M. 1982. *Systematic anatomy of Erythroxylaceae.* Ph.D. Thesis. Chapel Hill, NC: University of North Carolina.
Rury, P.M., T. Plowman. 1983. Morphological studies of archeological and recent coca leaves (*Erythroxylum* spp.). *Botanical Museum Leaflets* 29: 297-341.
Rury, P.M. 1992. Report on *Erythroxylum* Collections at USDA-ARS in Beltsville, Maryland. Report to U.S. Department of Agriculture, Agricultural Research Service.
Samuel, R., H. Kathriarachchi, P. Hoffmann, M.H.J. Barfuss, K.J. Wurdack, C.C. Davis, M.W. Chase. 2005. Molecular phylogenetics of Phyllanthaceae: evidence from plastid *matK* and nuclear *phyC* sequences. *American Journal of Botany* 92: 132-141.
Sang, T., D.J. Crawford, T.F. Stuessy. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* 84: 1120-1136.
Sanmartín, I., F. Ronquist. 2004. Southern Hemisphere biogeography analyzed with event-based models: plant versus animal patterns. *Systematic Biology* 53: 216-243.
Sanmartín, I., Van Der Mark, P., Ronquist, F., 2008. Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography with special reference to the Canary Islands. *Journal of Biogeography* 35: 428–449.
Santiago-Valentín, E., R.G. Olmstead. 2003. Phylogenetics of the Antillean Goetzeoideae (Solanaceae) and their relationships within Solanaceae based on chloroplast and ITS DNA sequence data. *Systematic Botany* 28: 452-460.
Santiago-Valentín, E., R.G. Olmstead. 2004. Historical biogeography of Caribbean plants: introduction to current knowledge and possibilities from a phylogenetic perspective. *Taxon* 53: 299-319.
Schultes, R.E. 1979. Evolution of the Identification of the Major South American Narcotic Plants. *Journal of Psychedelic Drugs* 11: 119-134.
Schulz, O.E. 1907a. Erythroxylaceae. *In:* A. Engler, editor. *Das Pflanzenreich IV*. 133: 1-164. Leipzig: Verlag von Wilhelm Engelman.
Schulz, O.E. 1907b. VI. Erythroxylaceae. *In:* Ignatius Urban, editor. *Symbolae Antillanae seu Fundamenta Florae Indiae Occidentalis*. Vol. 5. pp. 188-211. Lipsiae: Fratres Borntraeger.
Schulz, O.E. 1931. Erythroxylaceae. *In* A. Engler and Prantl, editor. *Die Naturlichen Pflanzenfamilien,* 2ed. 19a: 130-143. Leipzig: Verlag von Wilhelm Engelman.

Schwarzbach, A.E. and R.E. Ricklefs. 2000. Systematic affinities of Rhizophoraceae and Anisophylleaceae, and intergeneric relationships within Rhizophoraceae, based on chloroplast DNA, nuclear ribosomal DNA, and morphology. *American Journal of Botany* 87: 547-564.
Setoguchi, H., K. Kosuge, H. Tobe. 1999. Molecular phylogeny of Rhizophoraceae based on *rbcL* gene sequences. *Journal of Plant Research* 112: 443-455.
Shaw, J., E.B. Lickey, J.T. Beck, S.B. Farmer, W. Liu, J. Miller, K.C. Siripun, C.T. Winder, E.E. Schilling, R.L. Small. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92: 142-166.
Shaw, J., E.B. Lickey, E.E. Schilling, R.L. Small. 2007. Comparison of whole chloroplast genome sequences to choose non-coding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* 94: 275-288.
Shi, H., A. Singh, S. Kant, Z. Zhu, E. Waller. Integrating habitat status, human population pressure, and protection status into biodiversity conservation priority setting. *Conservation Biology* 19: 1273-1285.
Silvestro, D., I. Michalak. 2010. RAxMLGUI: a graphical front-end for RAxML. url: http://sourceforge.net/projects/raxmlgui/.
Simmons, M. P. 2004. Independence of alignment and tree search. *Molecular Phylogenetics and Evolution* 31: 874-879.
Simmons, M.P., H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369-381.
Stamatakis, A. 2005. Phylogenetics: applications, software and challenges. *Cancer Genomics & Proteomics* 2: 301-305.
Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688-2690.
Stevens, P. F. 2001, onwards. Angiosperm Phylogeny Website. http://www.mobot.org/MOBOT/research/APweb/.
Stöver, B.C., Müller, K.F. 2010. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics* 11: 7.
Strand, A.E., J. Leebens-Mack, B.G. Milligan. 1997. Nuclear DNA-based markers for plant evolutionary biology. *Molecular Ecology* 6: 113-118.
Streatfeild, D. 2001. *Cocaine: an unauthorized biogeography.* New York: Picador.
Swofford, D. L. 2002. PAUP*: *Phylogenetic analysis using parsimony (*and other methods).* Sunderland: Sinauer Associates.
Taberlet P., L. Gielly, G. Pautou, J. Bouvet. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105-1109.
Taylor, D.W. 1991. Paleobiogeographic relationships of Andean angiosperms of Cretaceous to Pliocene age. *Palaeogeography, Palaeoclimatology, Palaeoecology* 88: 69-84.
Templeton, A. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* 37: 221-224.
Tiffney, B.H. 1985. The Eocene North Atlantic land bridge: Its importance in Tertiary and modern phytogeography of the Northern Hemisphere. *Journal of Arnold Arboretum* 66: 243-273.
van Ee, B.W., P.E. Berry, R. Riina, J.E. Gutiérrez Amaro. 2008. Molecular phylogenetics and biogeography of the Caribbean-Centered *Croton* subgenus *Moacroton* (Euphorbiaceae s.s.). *Botanical Review* 2008. 74:132-165.
United Nations. Office of Drug and Crimes. 2006. *Colombia: Coca Cultivation Survey.* http://www.unodc.org/pdf/research/icmp/colombia_2006_en_web.pdf. [accessed: 10/01/2007]
United Nations. Office of Drug and Crimes. 2007. *Coca Cultivation in the Andean region.* http://www.unodc.org/pdf/andean/Part1_excutive_summary.pdf. [accessed: 10/01/2007]
Venkatasamy, S., G. Khitoo, S. Keeley. 2007. Leaky dioecy in *Diospryros* (Ebenaceae) endemic to the Island of Mauritius. *Plant Ecology* 189: 139-146.

Weese, T.L., L.A. Johnson. 2005. Utility of NADP-dependent isocitrate dehydrogenase for species-level evolutionary inference in angiosperm phylogeny: a case study in *Saltugilia*. *Molecular Phylogenetics and Evolution* 36: 24-41.

Wink, M. 2003. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* 64: 3-19.

White, T.J., T. Bruns, S. Lee, J. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *In*: M.A. Innis, D.H. Gelfand, J.J. Sninksy, T.J. White, editors. *PCR Protocols: A guide to methods and applications*. pp. 315-322 New York: Academic Press, Inc.

Wurdack, K.J., C.C. Davis. 2009. Malpighiales phylogenetics: gaining ground on one of the most recalcitrant clades in the angiosperm tree of life. *American Journal of Botany* 96: 1551-1570.

Youssefi, M., R.G. Cooks, J.L. McLaughlin 1979. Mapping of cocaine and cinnamoylcocaine in whole coca plant tissues by MIKES. *Journal of the American Chemical Society* 101: 3400-3402.

Yu Y., A.J. Harris, X.J. He. 2011. RASP (Reconstruct Ancestral State in Phylogenies) [beta1 build 110126]. url: http://mnh.scu.edu.cn/soft/blog/RASP

Yu, Y., A.J. Harris, X.J. He. 2010. S-DIVA (Statistical Dispersal-Vicariance Analysis): a tool for inferring biogeographic histories. *Molecular Phylogenetics and Evolution* 56: 848-850.

Zachos, J., M. Pagani, L. Sloan, E. Thomas, K. Billups. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292: 686-692.

Zappi, D.C. Erythroxylaceae. *In*: Stannard, B.L., Y.B. Harvey, and R.M. Harley, editors. *Flora of the Pico das Almas*. pp. 297-299. Kew, England: Royal Botanical Gardens.

Zurawski, G., M. T. Clegg. 1987. Evolution of higher-plant chloroplast DNA-encoded genes: implications for structure-function and phylogenetic studies. *Annual Review of Plant Physiology* 38: 391-418.

Appendix 1. Species sampled for both loci. Schulz section, species name, collector, collector number, year collected, herbarium (code Index Herbariorum), collector, collector number, (herbarium code), country collected, and if tested for presence of alkaloids (X), which alkaloids were present. ¤Accession only in ITS analysis. *Rury (1981); [§]Oviedo (2002)

Schulz Section	Species	Collector	Collector #	Year	Herb.	Country	Alkaloid	Alkaloids Present
outgroup	*Aneulophus africanus* Benth.	G. McPherson	16920	1997	MO	Gabon	X	
outgroup	*Bruguiera gymnorhiza* (L.) Sav.	M. Islam	6	2007	COLO	USA	X	
outgroup	*Bruguiera sexangula* (Lour.) Poir.	Genbank: AF130333						
outgroup	*Carallia brachiata* (Lour.) Merr.	J.S. Miller et al.	10644	2000	MO	Madagascar		
Erythroxylum	*E. alaternifolium* Rich. in Sagra	R. Oviedo	s.n.	2007	HAC	Cuba	X	
Archerythroxylum	*E. anguifugum* Mart.	A. Krapovickas et al.	46093	1994	F	Paraguay		
Archerythroxylum	*E. anguifugum* Mart.	St.G. Beck	16666	1988	F	Bolivia		
Erythroxylum	*E. areolatum* L.	T. Plowman	14431	1986	F	Bahamas	X	
Erythroxylum	*E. areolatum* L.	M. Islam	09-11	2009	COLO	Jamaica	X	
Erythroxylum	*E. areolatum* L.	M. Islam	09-24	2009	COLO	Dominican Republic	X	hygrine
Rhabdophylllum	*E. banaoense* Oviedo	R. Oviedo	s.n.	2007	HAC	Cuba	X	
Erythroxylum	*E. baracoense* Borhidi	R. Oviedo & L.N. Sánchez	s.n. (M1)	2009	HAC	Cuba	X	
Erythroxylum	*E. baracoense* Borhidi	R. Oviedo & L.N. Sánchez	s.n. (M2)	2009	HAC	Cuba	X	
Erythroxylum	*E. bequaertii* Standl.	M.A. Vincent, R.J. Hickey, & D. Osborne	6118	1993	F	Belize	X	
Archerythroxylum	*E. brevipes* DC.	S. Mori & R. Woodbury	17010	1984	F	U.S. Virgin Islands	X	hygrine, methyl ecogonine, cocaine
Archerythroxylum	*E. brevipes* DC.	R. Gracía & F. Jiménez	6024	1995	F	Dominican Republic	X	hygrine, methyl ecogonine, cocaine
Archerythroxylum	*E. brevipes* DC.	M. Islam	09-36	2009	COLO	Dominican Republic	X	hygrine, methyl ecogonine, cocaine, cinnamoylcocaine
Archerythroxylum	*E. brevipes* DC.	J.C. Trejo et al.	1052	1997	UPRRP	Puerto Rico	X	hygrine, methyl ecogonine, cocaine
Macrocalyx	*E.* cf. *pauciflorum* Rusby	M. Nee & G. Coimbra S.	35880	1987	MO	Bolivia	X	
¤*Microphyllum*	*E. cincinnatum* Mart.	T. Plowman	10174	1980	F	Brazil		
Rhabdophylllum	*E. citrifolium* A.St.-Hil.	A. Jara Muñoz	M633	?	COL	Colombia		
Rhabdophylllum	*E. citrifolium* A.St.-Hil.	A. Jara Muñoz	M633		COL	Colombia		
Rhabdophylllum	*E. citrifolium* A.St.-Hil.	R. Cortés	RC2782	2010	UDBC	Colombia		
Rhabdophylllum	*E. citrifolium* A.St.-Hil.	I.A. Valdespino & P. Soltis	157	1986	F	Panama	X	

Schulz Section	Species	Collector	Collector #	Year	Herb.	Country	Alkaloid	Alkaloids Present
Erythroxylum	*E. clarense* Borhidi	R. Oviedo	s.n.	2007	HAC	Cuba	X	
Archerythroxylum	*E. coca* Lam. var. *coca*	M. Islam	DNA665	2008	COLO	Unknown	X	hygrine, methyl ecogonine, cocaine, cinnamoylcocaine
Archerythroxylum	*E. coca* Lam. var. *coca*	M. Islam	DNA583	2008	COLO	Unknown	X	hygrine, methyl ecogonine, cocaine, cinnamoylcocaine
Archerythroxylum	*E. coca* var. *ipadu* Plowman	M. Islam	DNA910	2008	COLO	Unknown	X	hygrine, methyl ecogonine, cocaine, cinnamoylcocaine
Archerythroxylum	*E. coca* var. *ipadu* Plowman	M. Islam	DNA628	2008	COLO	Unknown	X	hygrine, methyl ecogonine, cocaine, cinnamoylcocaine
Archerythroxylum	*E. confusum* Britton in Britton & Millsp.	I. Olmstead & M.A. Marmolejo	27	1985	F	Mexico	X	hygrine
Archerythroxylum	*E. confusum* Britton in Britton & Millsp.	M. Islam	10	2007	MBC	Florida	X	hygrine, methyl ecogonine
Archerythroxylum	*E. confusum* Britton in Britton & Millsp.	R. Oviedo	s.n.	2007	HAC	Cuba	X	hygrine, methyl ecogonine
Archerythroxylum	*E. confusum* Britton in Britton & Millsp.	M. Islam et al	09-05	2009	COLO	Jamaica	X	
Archerythroxylum	*E. confusum* Britton in Britton & Millsp.	T. Wayt 14773 & R. Oviedo	s.n. (M1)	2009	HAC	Cuba	X	hygrine, methyl ecogonine
Megalophyllum	*E. coriaceum* Britton & P.Wils.	R. Oviedo & L.N. Sánchez	s.n. (M1)	2009	HAC	Cuba	X	
Megalophyllum	*E. coriaceum* Britton & P.Wils.	R. Oviedo & L.N. Sánchez	s.n. (M2)	2009	HAC	Cuba	X	
Gonocladus	*E. corymbosum* Boivin ex Baill.	J. Rabenantoandro et al.	250	2000	F	Madagascar	X	
Archerythroxylum	*E. cumanense* Kunth in Humb.	T. Plowman	10948	1981	F	Venezuela		
Rhabdophylllum	*E. daphnites* Mart.	R. F. Vieira et al.	1662	1993	F	Brazil		
Rhabdophylllum	*E. deciduum* A.St.-Hil.	A. Krapovickas et al.	45842	1994	F	Argentina		
Eurysepalum	*E. domingense* Oviedo	M. Islam	09-14	2009	COLO	Dominican Republic		
Eurysepalum	*E. domingense* Oviedo	R. Gracía et al.	4397	1993	FTG	Dominican Republic		
Eurysepalum	*E. echinodendron* Ekman ex. Urb.	R. Oviedo	s.n.	2007	HAC	Cuba	X	
Eurysepalum	*E. echinodendron* Ekman ex. Urb.	R. Oviedo	s.n.	2007	HAC	Cuba	X	
Rhabdophylllum	*E. engleri* O.E. Schulz in Engl.	B. M.T. Walter et al.	3223	1996	F	Brazil		
Rhabdophylllum	*E. fimbriatum* Peyr.	G.S. Hartshorn	1243	1973	F	Costa Rica		
Pachylobus	*E. fischeri* Engl.	L. Festo et al.	1000	2001	MO	Tanzania	X	hygrine, methyl ecogonine
Erythroxylum	*E. flavicans* Borhidi	R. Oviedo & L.N. Sánchez	s.n. (M2)	2009	HAC	Cuba	X	hygrine, methyl ecogonine

Schulz Section	Species	Collector	Collector #	Year	Herb.	Country	Alkaloid	Alkaloids Present
Archerythroxylum	*E. glaucum* O.E. Schulz	T. Plowman	5435	1976	F	Peru		
Archerythroxylum	*E. gracilipes* Peyr. In Mart.	G. Silva	1122	?	COL	Colombia		
Archerythroxylum	*E. havanense* Jacq. (EH03)	A. Ibáñez Bastidas et al.	3334	2004	F	Panama	X	
Archerythroxylum	*E. havanense* Jacq. (EH04)	H.M. Hernández & R. Torres	297	1984	F	Mexico	X	hygrine, methyl ecogonine
Archerythroxylum	*E. havanense* Jacq. (EH10)	M. Islam	09-49	2009	COLO	Grenada	X	hygrine, methyl ecogonine
Archerythroxylum	*E. havanense* Jacq. (EH09)	R. Oviedo	s.n.	2007	HAC	Cuba	X	
Archerythroxylum	*E. hondense* Kunth in Humb.	A. Jara Muñoz	?	2010	?	Colombia		
Archerythroxylum	*E. impressum* O.E. Schulz in Urb.	T. Plowman & F. Guánchez	13504	1984	F	Venezuela	X	
Archerythroxylum	*E. incrassatum* O.E. Schulz in Urb.	M. Islam	09-02	2009	COLO	Jamaica	X	
Archerythroxylum	*E. incrassatum* O.E. Schulz in Urb.	M. Islam	09-01	2009	COLO	Jamaica	X	
*Archerythroxylum (*Erythroxylum)*	*E. jamaicense* Fawc. & Rendle	T. Parker	3509	2000	IJ	Jamaica	X	
*Archerythroxylum (*Erythroxylum)*	*E. jamaicense* Fawc. & Rendle	A. Gentry et al.	28451	1980	F	Jamaica	X	
Oxystigma	*E. kunthianum* Kurz	H. Guosheng	15093	2003	MO	China		
Archerythroxylum	*E. ligustrinum* DC.	G. Silva	1114	?	COL	Colombia		
Erythroxylum	*E. longipes* O.E. Schulz in Urb.	R. Oviedo & L.N. Sánchez	s.n. (M1)	2009	HAC	Cuba	X	
Erythroxylum	*E. longipes* O.E. Schulz in Urb.	R. Oviedo & L.N. Sánchez	s.n. (M2)	2009	HAC	Cuba	X	
¤*Macrocalyx*	¤*E. macrophyllum* Cav.	M. Nee et al.	25112	1982	F	Mexico		
Macrocalyx	*E. macrophyllum* Cav.	D.A. Hayworth	343	1988	F	Costa Rica	X	
Macrocalyx	*E. macrophyllum* Cav.	C.L. Lundell & E. Contreras	20694	1977	F	Guatemala		
Macrocalyx	*E. macrophyllum* var. *savannarum* Plowman	G. Silva	549	?	COL	Colombia		
Archerythroxylum	*E. mexicanum* Kunth in Humb.	M. Nee et al.	24756	1982	F	Mexico		
Erythroxylum (§Microphyllum)	*E. minutifolium* Griseb.	R. Oviedo	s.n.	2007	HAC	Cuba		
Erythroxylum (§Microphyllum)	*E. minutifolium* Griseb.	R. Oviedo	s.n.	2007	HAC	Cuba		
Gonocladus	*E. mocquerysii* Aug. DC.	J. Rabenantoandro et al.	1440	2003	MO	Madagascar		
Schistophyllum	*E. nitidulum* Baker	J.L. Zarucchi et al.	7425	1991	F	Madagascar		
Archerythroxylum	*E. novogranatense* (D. Morris) Hieron. var. *novogranatense*	M. Islam	DNA608	2008	COLO	Unknown	X	hygrine, methyl ecogonine, cocaine, cinnamoylcocaine
Archerythroxylum	*E. novogranatense* (D. Morris)	M. Islam	DNA839	2008	COLO	Unknown	X	hygrine, methyl ecogonine,

Schulz Section	Species	Collector	Collector #	Year	Herb.	Country	Alkaloid	Alkaloids Present
	Hieron. var. *Novogranatense*							cocaine, cinnamoylcocaine
Archerythroxylum	*E. novogranatense* var. *truxillense* (Rusby) Plowman	M. Islam	DNA658	2008	COLO	Unknown	X	hygrine, methyl ecogonine, cocaine, cinnamoylcocaine
Archerythroxylum	*E. novogranatense* var. *truxillense* (Rusby) Plowman	M. Islam	DNA657	2008	COLO	Unknown	X	hygrine, methyl ecogonine, cocaine, cinnamoylcocaine
Archerythroxylum	*E. novogranatense* var. *truxillense* (Rusby) Plowman	M. Islam	DNA645	2008	COLO	Unknown	X	hygrine, methyl ecogonine, cocaine, cinnamoylcocaine
Leptogramme	*E. ovalifolium* Peyr. in Mart.	T. Plowman et al.	12841	1983	F	Brazil	X	
Archerythroxylum	*E. oxycarpum* O.E. Schulz in Urb.	P.E. Berry	3478	1979	F	Venezuela		
Microphyllum	*E. pedicellare* O.E. Schulz in Urb.	R. Oviedo	s.n. (F1)	2009	HAC	Cuba	X	
Microphyllum	*E. pedicellare* O.E. Schulz in Urb.	R. Oviedo	s.n. (F2)	2009	HAC	Cuba	X	
Leptogramme	*E. pulchrum* A.St.-Hil.	R. Marquete	1329	1993	F	Brazil	X	cocaine
Lagynocarpus	*E. pyrifolium* Baker	L.J. Dorr et al.	4620	1986	F	Madagascar		
Rhabdophylllum	*E. raimondii* O.E. Schulz	St.G. Beck	22277	1994	F	Bolivia		
Lagynocarpus	*E. retusum* Baill. ex O.E. Schulz	P.B. Phillipson	2878	1988	F	Madagascar		
Archerythroxylum	*E. roigii* Britton & P. Wilson	R. Oviedo	s.n.	2007	HAC	Cuba	X	
Erythroxylum	*E. rotundifolium* Lunan	E.A. Montalvo & R. Villacorta	6495	1996	F	El Salvador		
Erythroxylum	*E. rotundifolium* Lunan	F. Gonzáles et al.	12721	1982	F	Mexico		
Erythroxylum	*E. rotundifolium* Lunan	J. Gomez-Laurito	8953	1982	F	Costa Rica		
Erythroxylum	*E. rotundifolium* Lunan	J. Gomez-Laurito	8953	1982	F	Costa Rica		
Erythroxylum	*E. rotundifolium* Lunan	M. Islam	09-08	2009	COLO	Jamaica		
Erythroxylum	*E. rotundifolium* Lunan	M. Islam	09-09	2009	COLO	Jamaica	X	
Erythroxylum	*E. rotundifolium* Lunan	R. Oviedo	s.n.	2007	HAC	Cuba	X	
Erythroxylum	*E. rotundifolium* Lunan	M. Islam	09-29a	2009	COLO	Dominican Republic	X	
Erythroxylum	*E. rotundifolium* Lunan	R. Oviedo	s.n.	2007	HAC	Cuba	X	hygrine, methyl ecogonine
unplaced	*E. rotundifolium* X *flavicans*	R. Oviedo & L.N. Sánchez	s.n. (M1)	2009	HAC	Cuba		
unplaced	*E. rotundifolium* X *flavicans*	R. Oviedo & L.N. Sánchez	s.n. (M2)	2009	HAC	Cuba		
Rhabdophylllum	*E. rufum* Cav.	A. Liogier et al.	36790	1989	F	Puerto Rico	X	
Rhabdophylllum	*E. rufum* Cav.	F. Jiménez & A. Veloz	1546	1994	F	Dominican Republic	X	
Rhabdophylllum	*E. rufum* Cav.	M. Islam	09-18	2009	COLO	Dominican Republic	X	
Rhabdophylllum	*E. rufum* Cav.	T. Plowman	7750	1979	F	Venezuela		

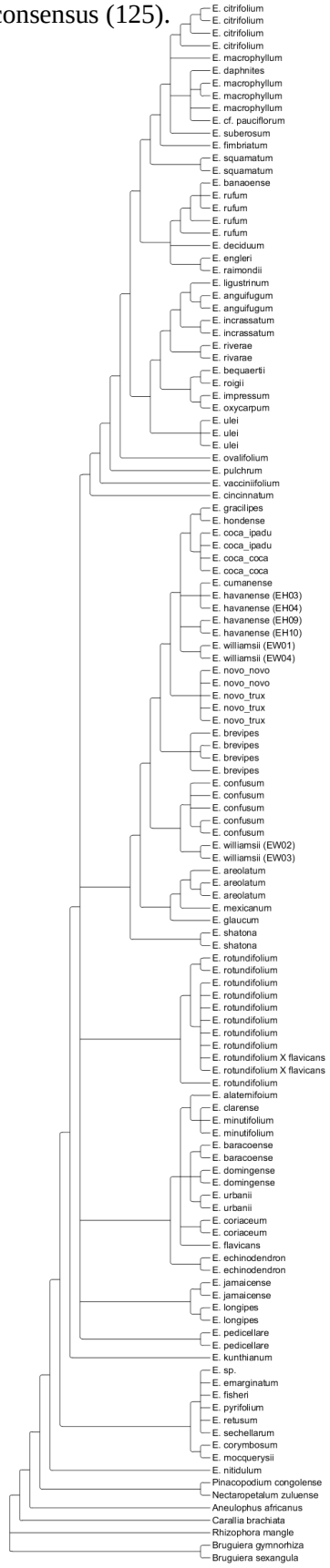
Schulz Section	Species	Collector	Collector #	Year	Herb.	Country	Alkaloid	Alkaloids Present
Pachylobus	*E. sechellarum* O.E. Schulz	S.A. Robertson	2704	1978	F	Seychelles		
Archerythroxylum	*E. shatona* J.F. Macbr.	T. Plowman	10940	1981	F	Peru		
Archerythroxylum	*E. shatona* J.F. Macbr.	A. Gentry et al.	37827	?	MO	Peru		
unplaced	E. sp.	R.H. Archer et al.	2920	2007	CS	Madagascar	X	
Archerythroxylum	*E. riverae*	A. Jara Muñoz	AJM-410	?	COL	Colombia		
Archerythroxylum	*E. riverae*	A. Jara Muñoz	AJM-410	?	COL	Colombia		
Rhabdophyllum	*E. squamatum* Sw.	G.T. Prance	28526	1983	F	Brazil	X	
Rhabdophyllum	*E. squamatum* Sw.	M. Islam	09-52	2009	COLO	Grenada	X	
Macrocalyx	*E. suberosum* c.f. var. *denudatum* O.E. Schulz in Engl.	G. Gottsberger	14-29782	1982	F	Brazil	X	
Leptogramme	*E. ulei* O.E. Schulz in Engl.	JAI	JAI-872	?	COL	Colombia		
Leptogramme	*E. ulei* O.E. Schulz in Engl.	A. Araujo-M et al.	1353	2004	F	Bolivia	X	
Leptogramme	*E. ulei* O.E. Schulz in Engl.	T. Plowman	8023	1979	F	Colombia		
Erythroxylum (*[§]Eurysepalum*)	*E. urbanii* O.E. Schulz in Urb.	P. Acevedo-Rdgz	8557	1996	FTG	Dominican Republic		
Erythroxylum (*[§]Eurysepalum*)	*E. urbanii* O.E. Schulz in Urb.	T. Clase	5386	2008	COLO	Dominican Republic		
Archerythroxylum	*E. vacciniifolium* Mart.	T. Plowman	10022	1980	F	Brazil		
Archerythroxylum	*E. williamsii* Standl. ex Plowman (EW01)	T. Plowman	8024	1979	F	Venezuela		
Archerythroxylum	?E. williamsii (EW02)	M. Islam	09-26	2009	COLO	Dominican Republic	X	
Archerythroxylum	?E. williamsii (EW03)	M. Islam	09-27	2009	COLO	Dominican Republic		
Archerythroxylum	*E. williamsii* Standl. ex Plowman (EW04)	T. Plowman	7760	1979	F	Venezuela		
outgroup	*Nectaropetalum zuluense* (Schönland) Corbishley	K. Balkwill	343	1982	MO	South Africa		
outgroup	*Pinacopodium congolense* (S.Moore) Exell & Mendonça	G. McPherson	15533	1991	MO	Gabon	X	
outgroup	*Rhizophora mangle* L.	Genbank : AF130332.1						

Appendix 2. Morphological Characters. 1-4 and 20-25 (Rury 1982), 5-10 (Oviedo 2002). Characters 22 and 26 were not included in analyses but mapped on to the MPTs using parsimony ancestral state reconstruction in Mesquite.

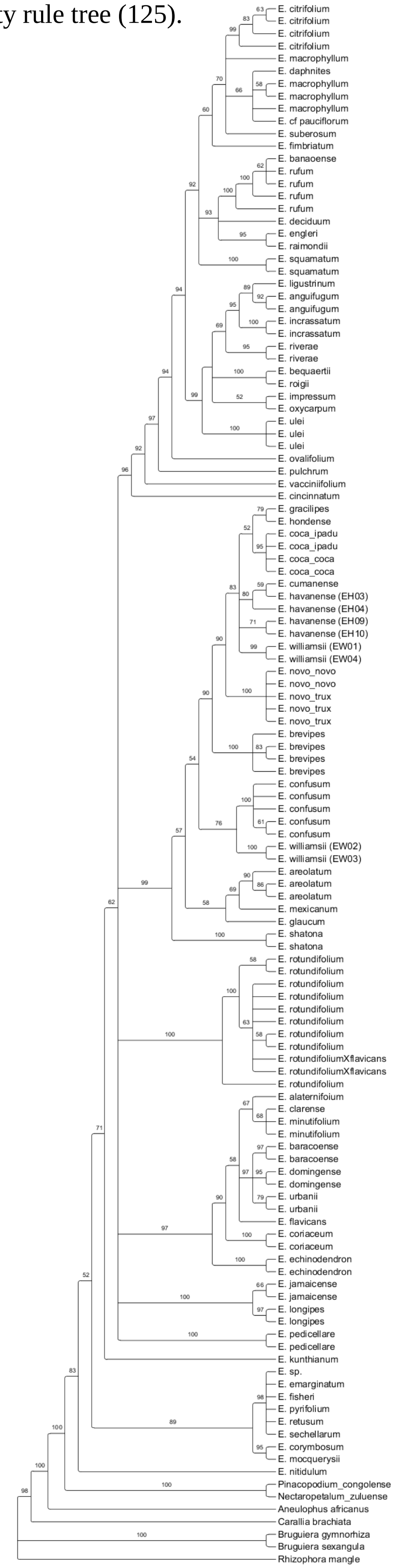
1. foliar sclereid type: 0 sclerosed cells of the spongy mesophyll, 1 dilated tracheids in a terminal position on veinlets, 2 ramified fibers derived from the vein and veinlet fiber sheaths, 3 absent.
2. epidermal cell shape (mature leaves): 0 polygonal, 1 convoluted, 2 intermediate.
3. epidermal papillae: 0 absent, 1 sparse or incipient, 2 abundant and prominent.
4. leaf duration: 0 evergreen, 1 deciduous, 2 semievergreen.
5. leaf secondary venation type : 0 brochiododromous, 1 campto – reticulodromous, 2 campto - cladodromous, 3 campto – eucamptodromous, 4 crasped-semi, 5 crasped-mixed, 6 actinodromous.
6. leaf tertiary vein pattern: 0 ramified-transverse, 1 ramified-admedial, 2 ramified-exmedial, 3 reticulate-random4 reticulate-orthogonal.
7. leaf marginal venation type:0 fimbriate, 1 incomplete, 2 looped.
8. leaf areole development: 0 imperfect, 1 well-developed.
9. leaf areole shape: 0 irregular, 1 quadrangular, 2 pentagonal, 3 polygonal.
10. freely ending veinlets branching pattern: 0 none, 1 simple-linear, 2 simple-curved, 3 branched-once, 4 branched-twice, 5 branched-three times.
11. flower sexuality: 0 bisexual, 1 unisexual.
12. stipule duration: 0 persistant, 1 deciduous.
13. stipule striations (sclerified vascular bundles): 0 absent, 1 present.
14. stipule margin: 0 entire, 1 fimbriate, 2 denticulate.
15. number of stipule setae: 0 zero, 1 one, 2 two.
16. texture of setae: 0 entire, 1 hairy.
17. stipule apex shape: 0 acute, 1 mucronate, 2 emarginate, 3 obtuse, 4 acuminate, 5 bisected.
18. style fusion: 0 free, 1 connate.
19. endocarp shape: 0 terete, 1 6-sulcate, 2 3-sulcate, 3 4-sulcate.
20. phloem fiber pattern: 0 absent, 1 discrete bundles, 2 continuous bands, 3 tangentially interrupted bands.
21. leaf structural type: 0 ramified fibro-sclereids present, 1 ramified fibro-sclereids absent.
22. habitat: 0 xeric, 1 semi-xeric, 2 mesic.
23. growth rings: 0 absent, 1 faint but inconspicuous, 2 very prominent, but irregular.
24. wood crystal type: 0 calcium oxalate, 1 silica grains, 2 absent.
25. wood cell type with crystal: 0 absent, 1 ray, 2 parenchyma.
26. schulz's sections: 0 pogonophorum, 1 macrocalyx, 2 rhabdophyllum, 3 leptogramme, 4 erythroxylum, 5 archerythroxylum, 6 megalophyllum, 7 mastigophorum, 8 microphyllum, 9 melanocladus, 10 gonocladus, 11 sethia, 12 lagynocarpus, 13 coelocarpus, 14 eurysepalum, 15 venelia, 16 pachylobus, 17 schistophyllum, 18 oxystigma.
27. hygrine: 0 absent, 1 present.
28. methyl ecogine: 0 absent, 1 present.
29. cocaine: 0 absent, 1 present.
30. cinnamolycocaine: 0 absent, 1 present.

Appendix 3. Trees from all analyses, with support values, and no collapsed clades. Maximum parsimony majority rule trees have jackknife values above the branches. Maximum likelihood trees are the best tree with bootstrap values above the branches. Number of accessions for that analysis follows the title of each tree.

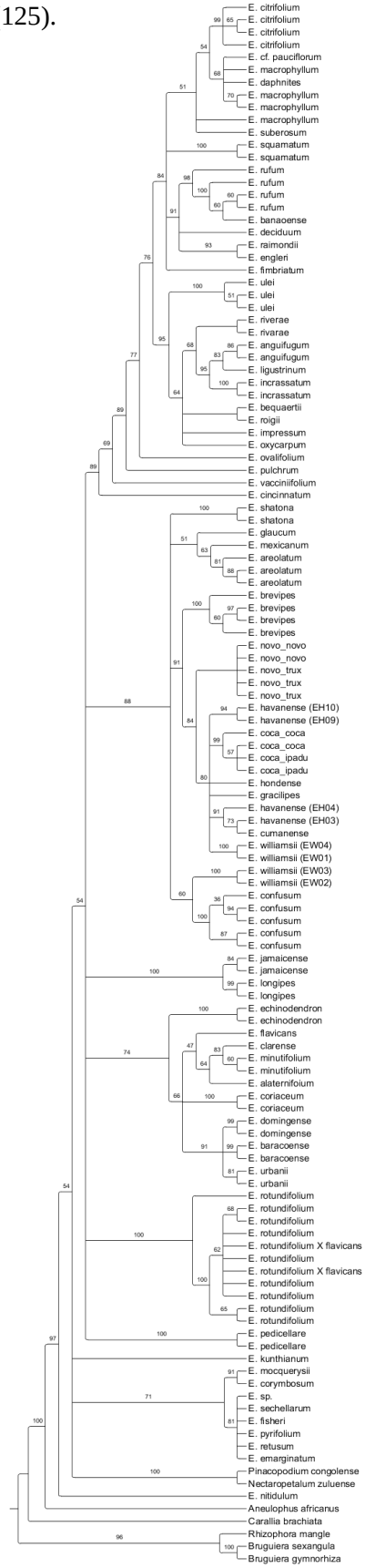
ITS maximum parsimony strict consensus (125).



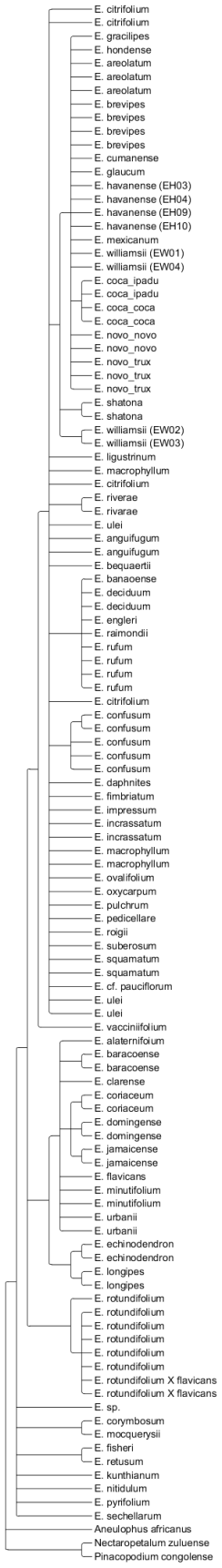
ITS maximum parsimony majority rule tree (125).



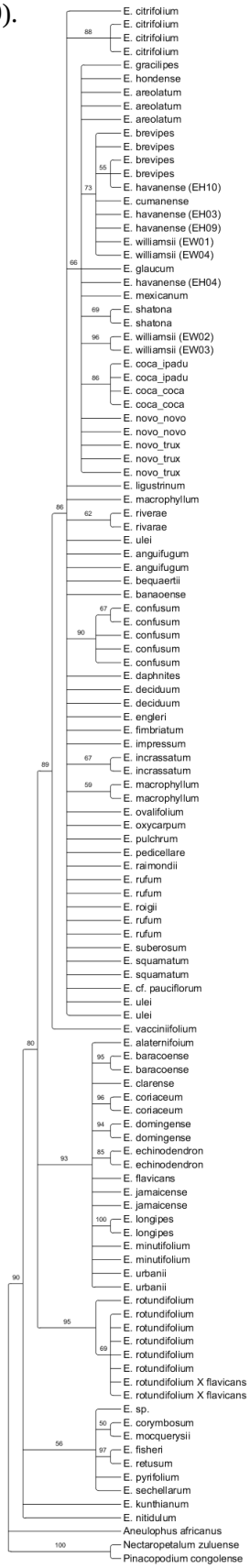
ITS maximum likelihood tree (125).



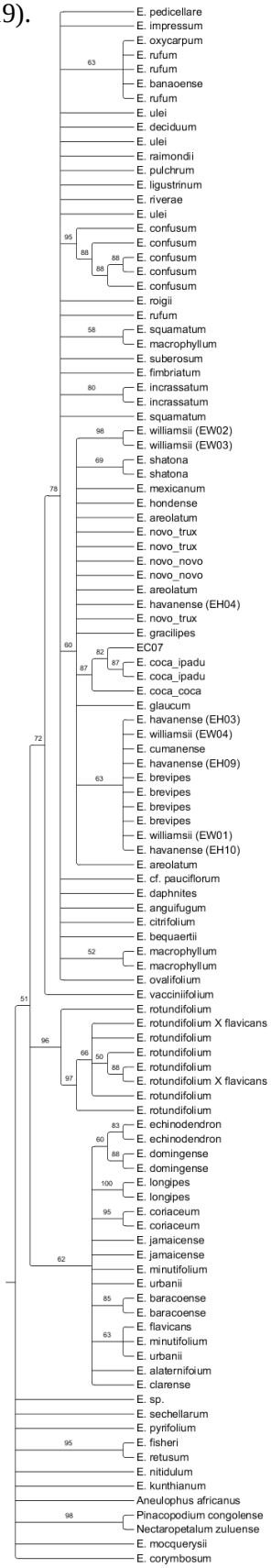
rpl32-trnL maximum parsimony strict consensus (119).



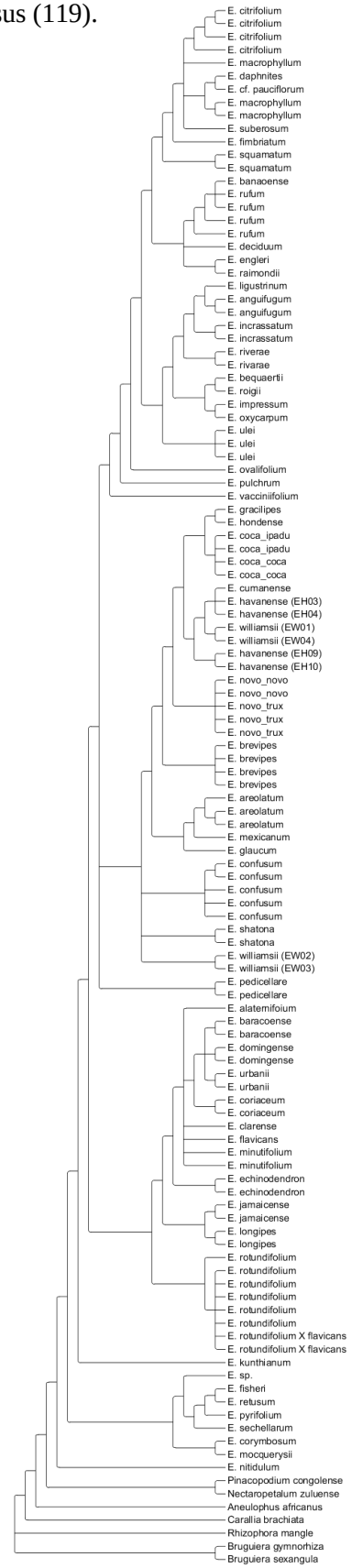
rpl32-trnL maximum parsimony majority rule tree (119).



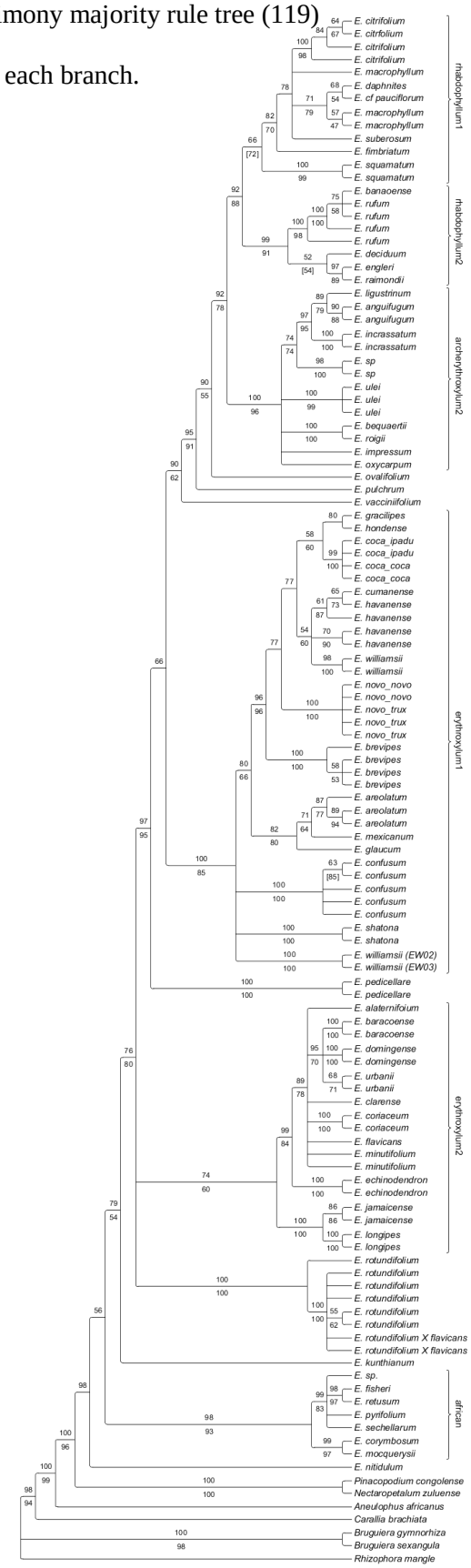
rpl32-trnL maximum likelihood tree (119).



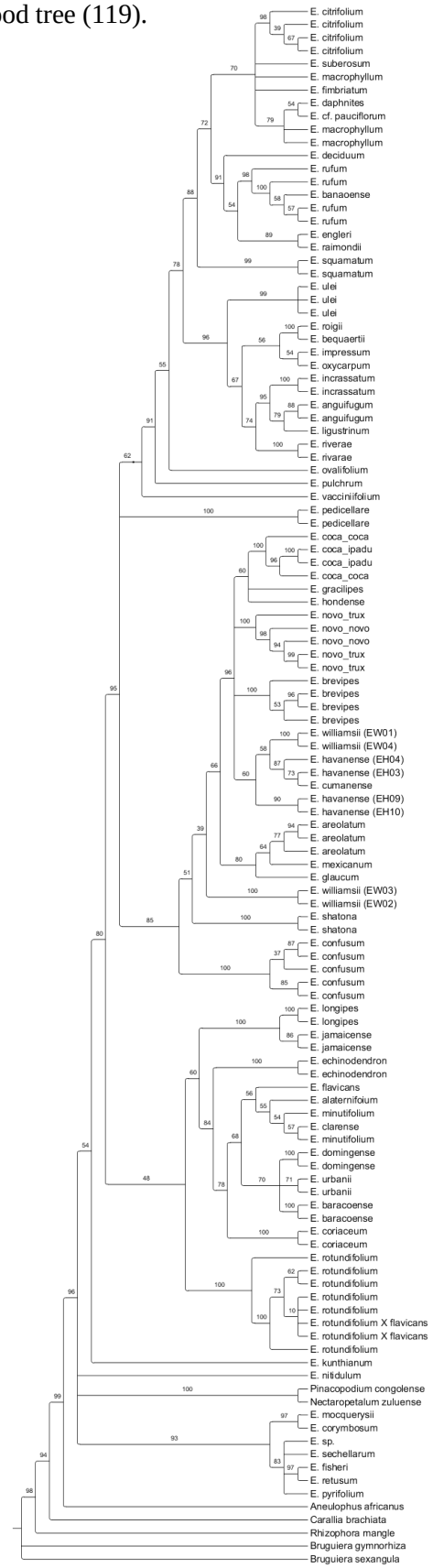
all molecular maximum parsimony strict consensus (119).



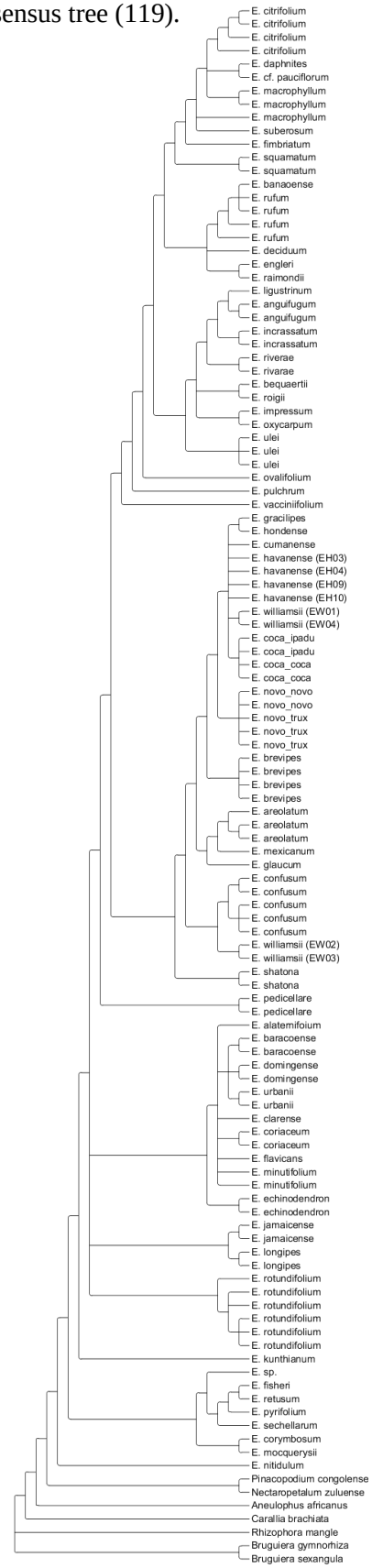
all molecular maximum parsimony majority rule tree (119) with likelihood values below each branch.



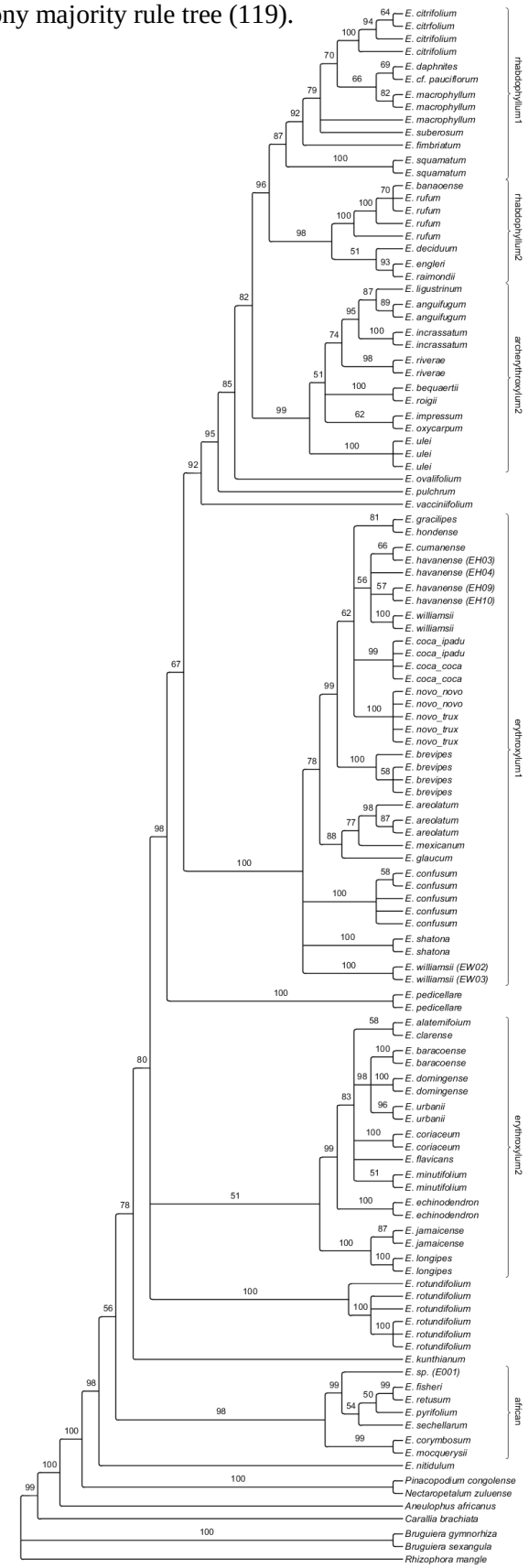
all molecular maximum likelihood tree (119).



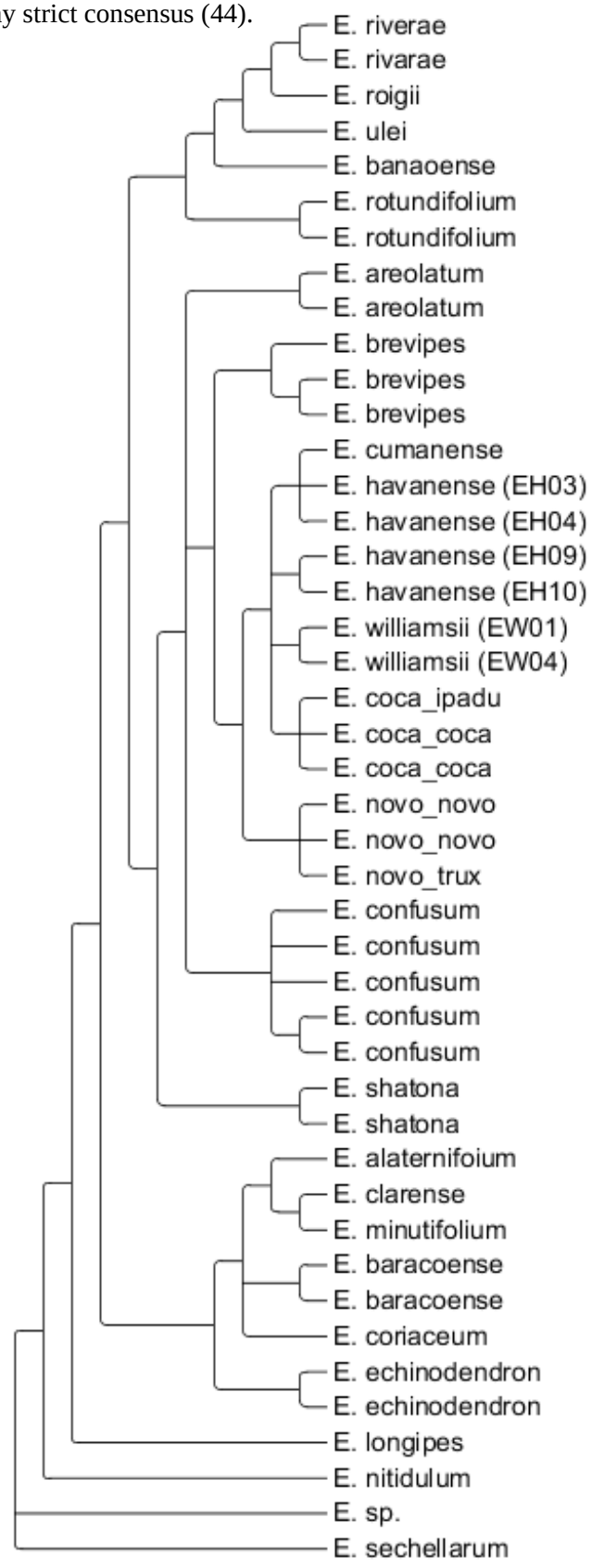
all characters maximum parsimony strict consensus tree (119).



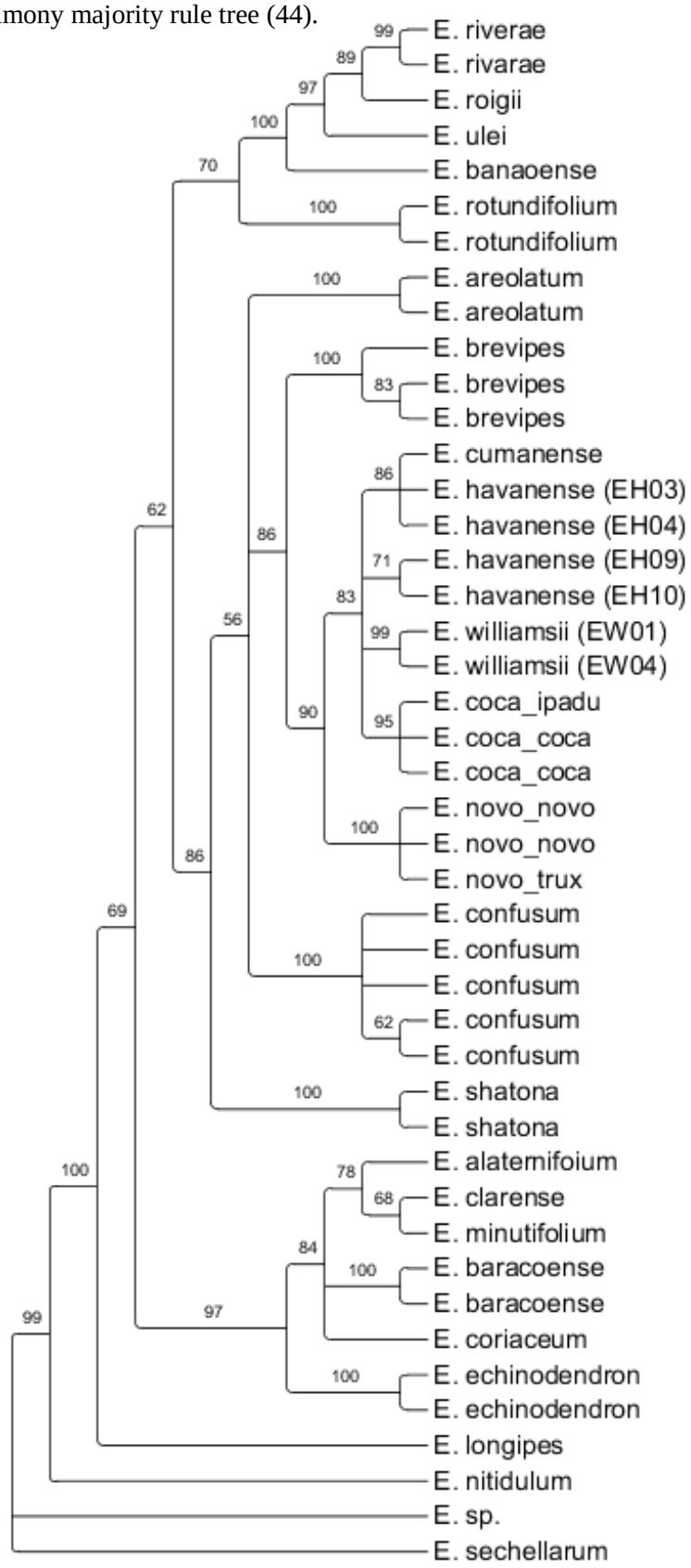
all characters maximum parsimony majority rule tree (119).



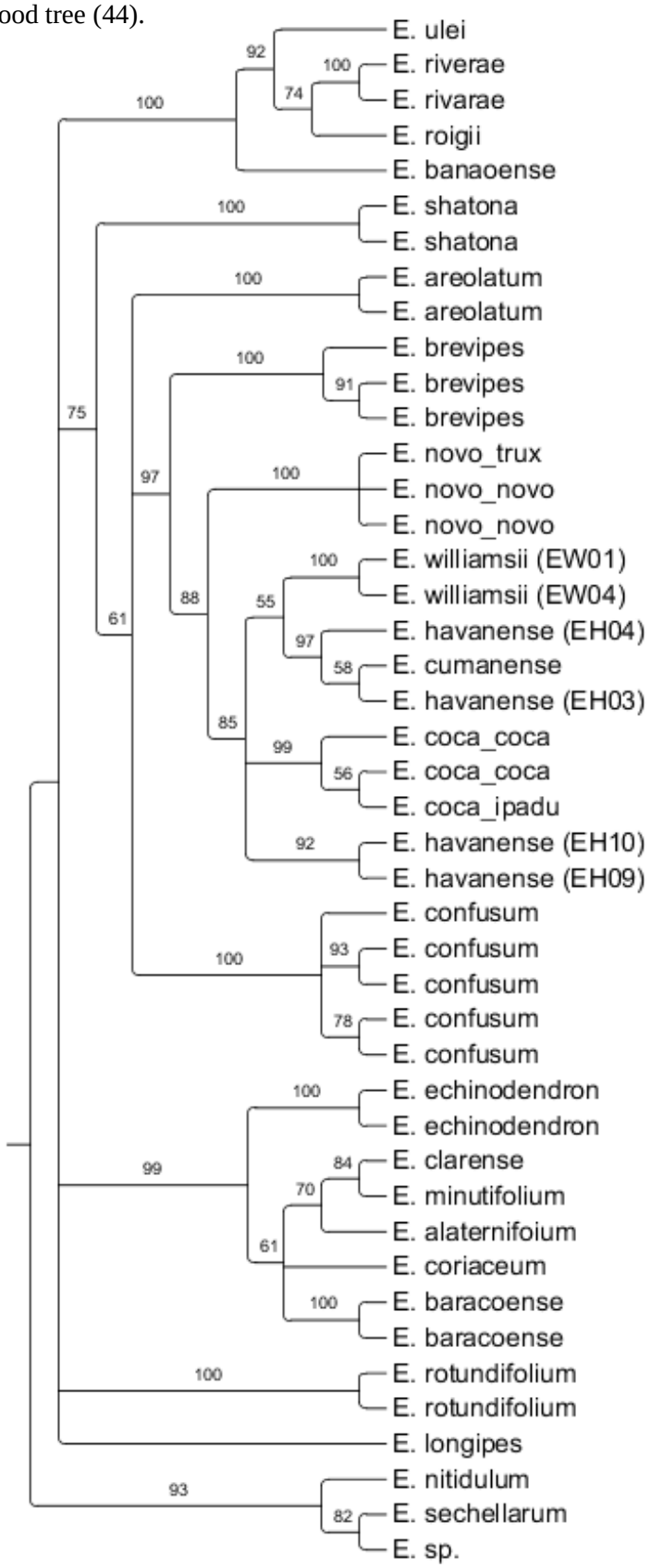
ITS maximum parsimony strict consensus (44).



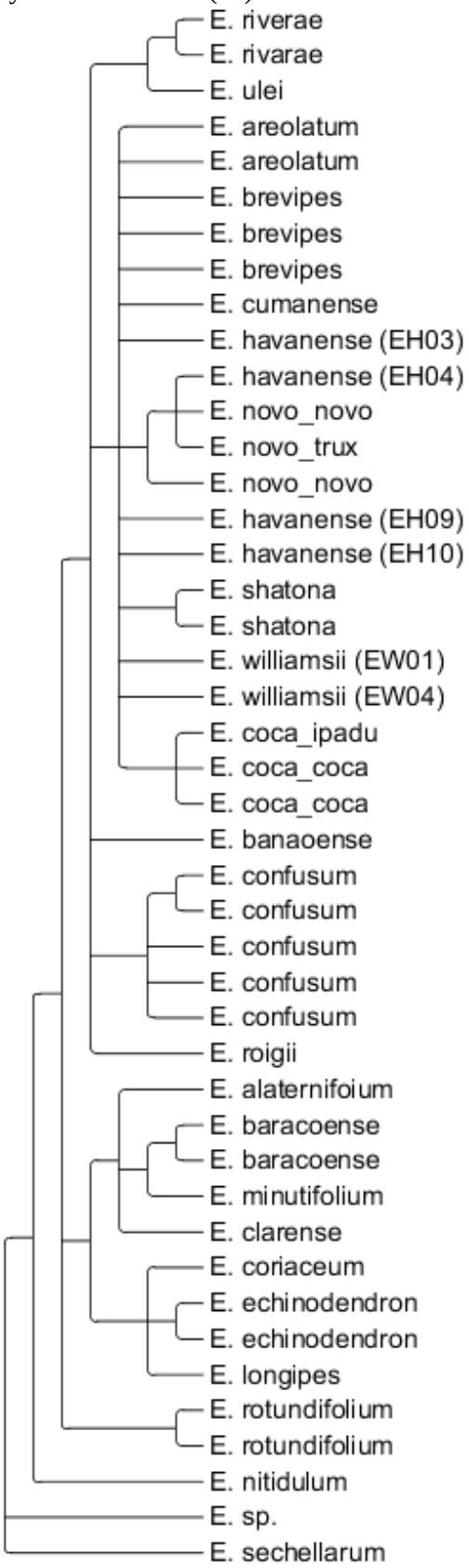
ITS maximum parsimony majority rule tree (44).



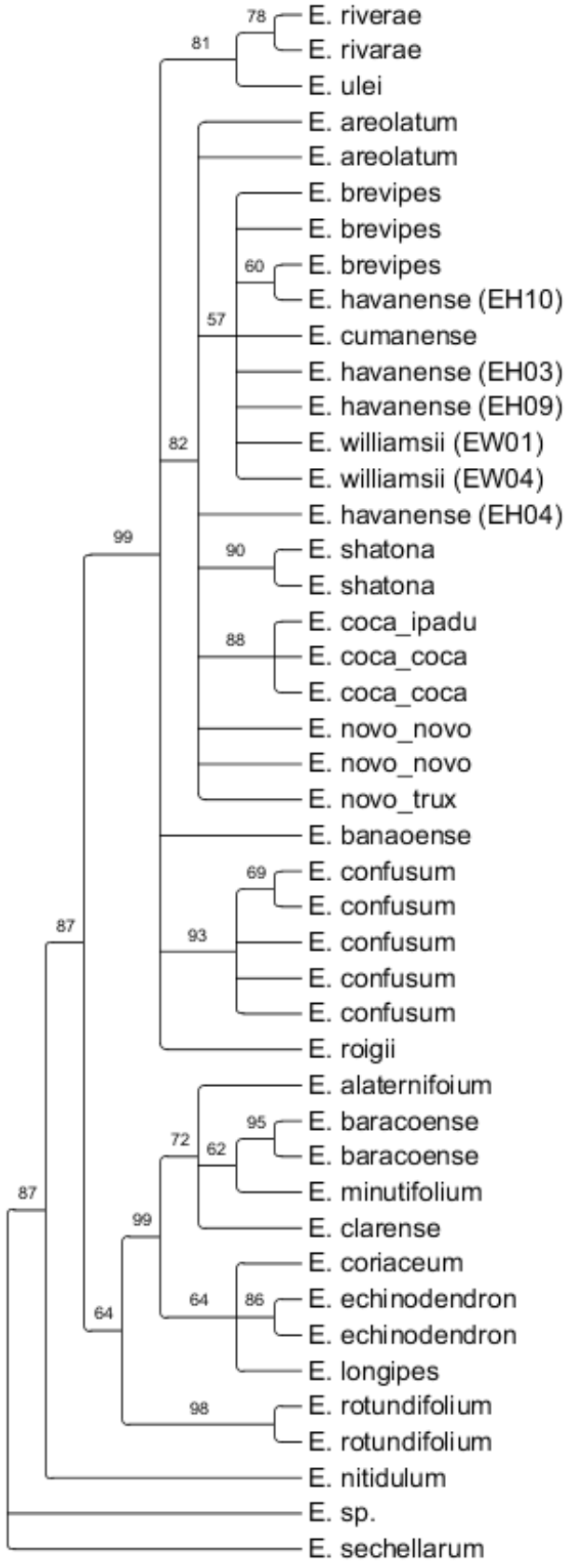
ITS maximum likelihood tree (44).



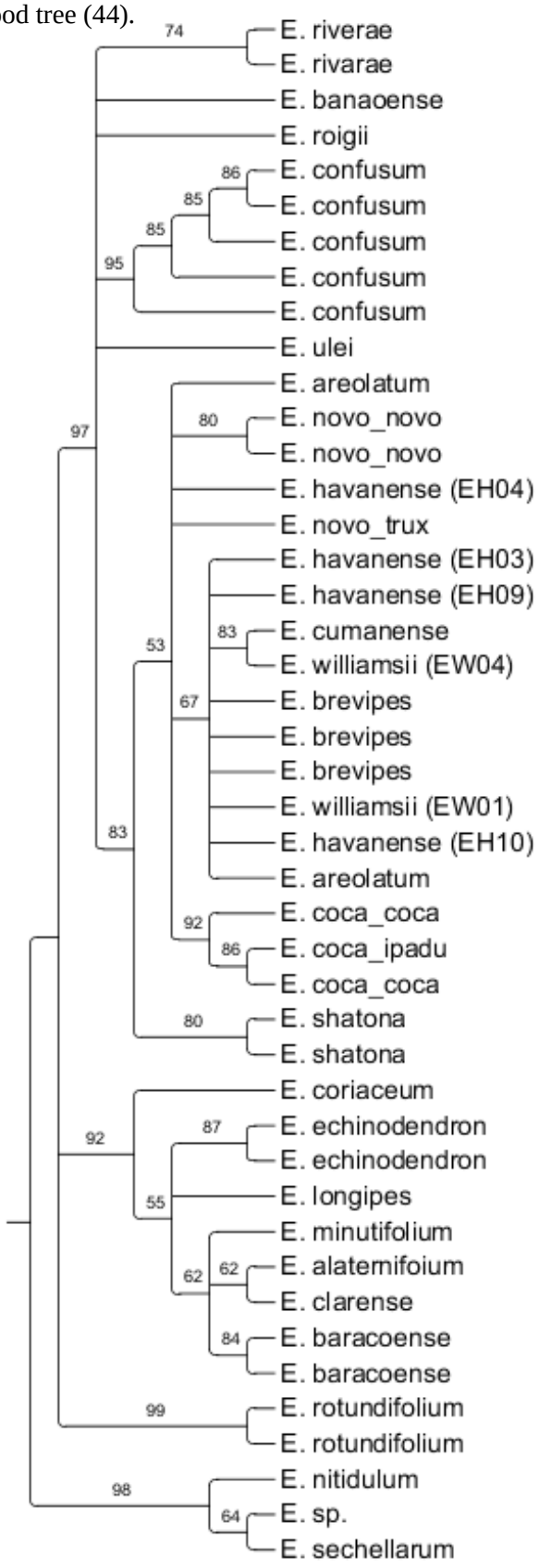
rpl32-trnL maximum parsimony strict consensus (44).



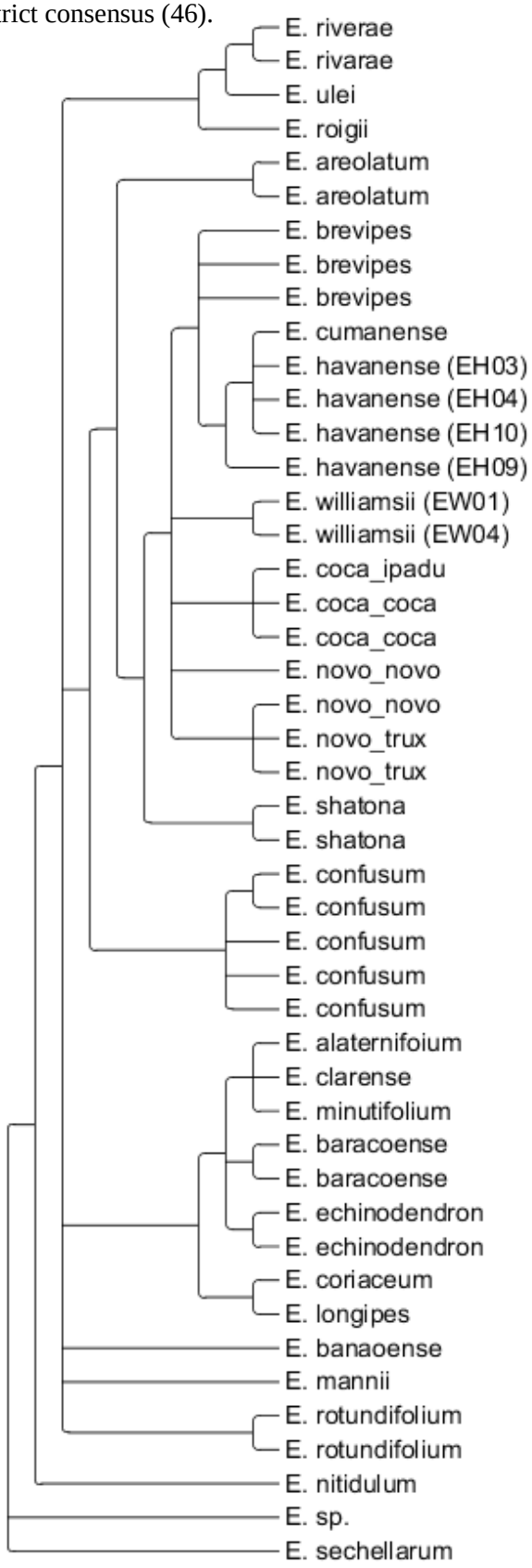
rpl32-trnL maximum parsimony majority rule tree (44).



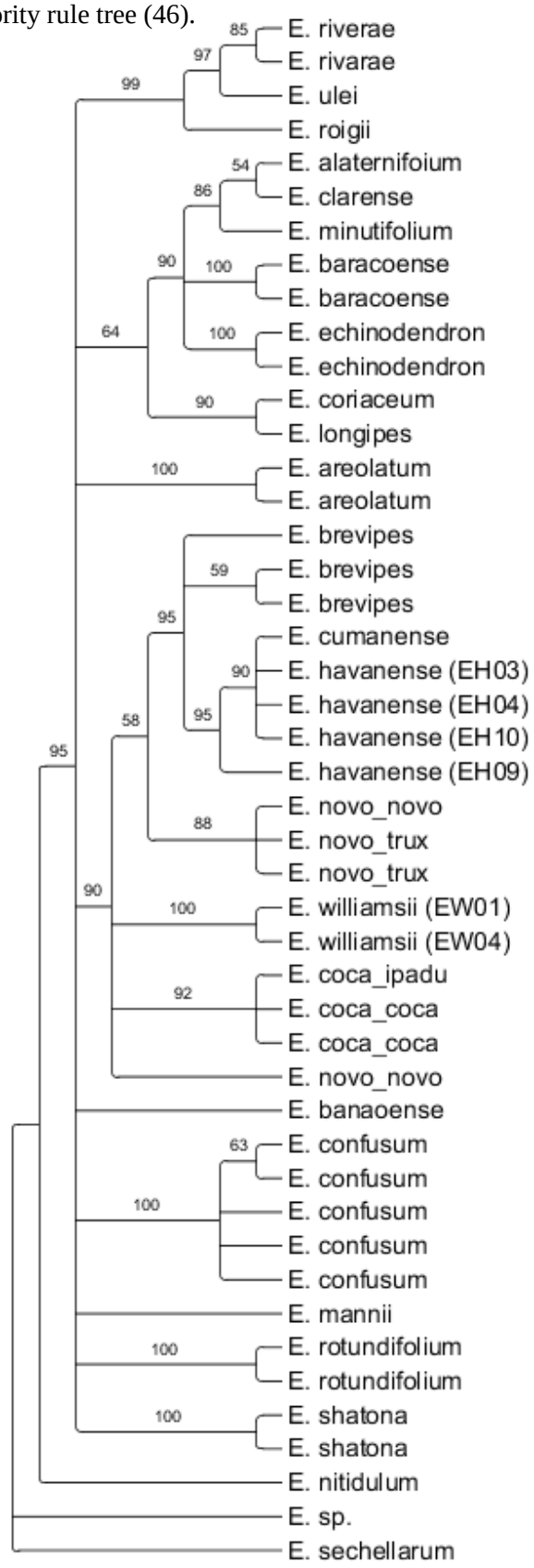
rpl32-trnL maximum likelihood tree (44).



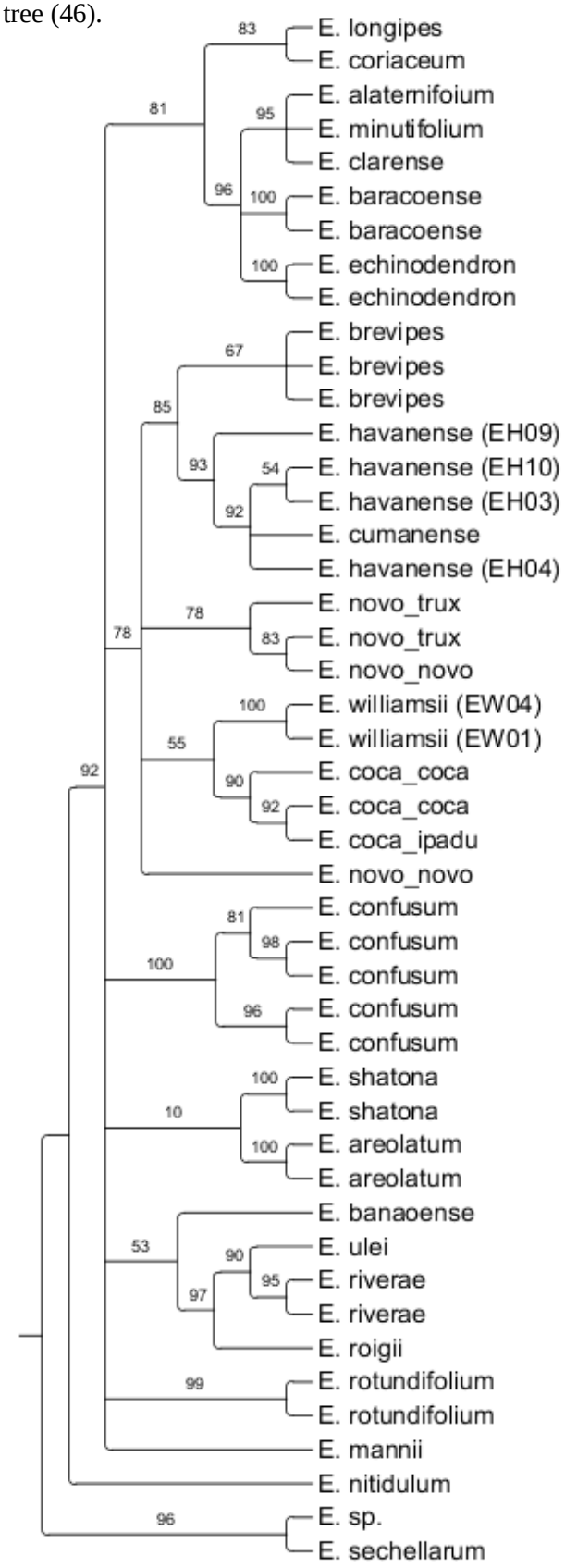
idh maximum parsimony strict consensus (46).



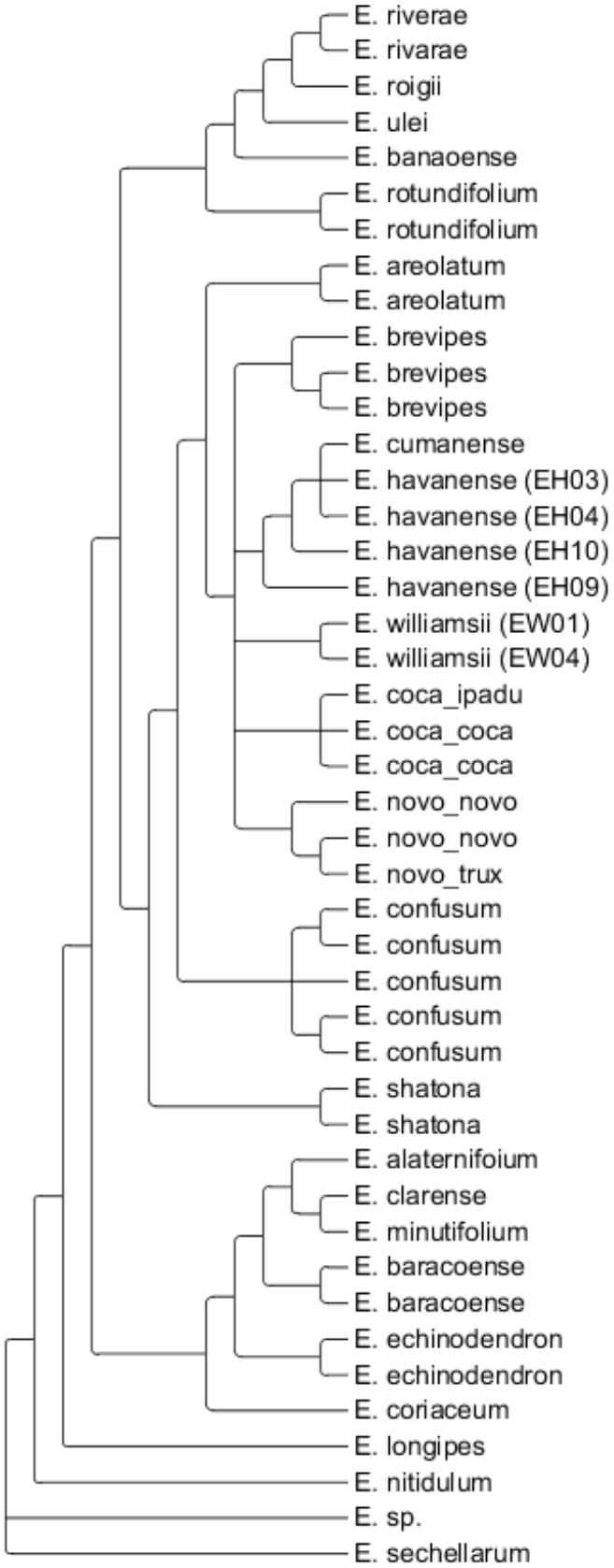
idh maximum parsimony majority rule tree (46).



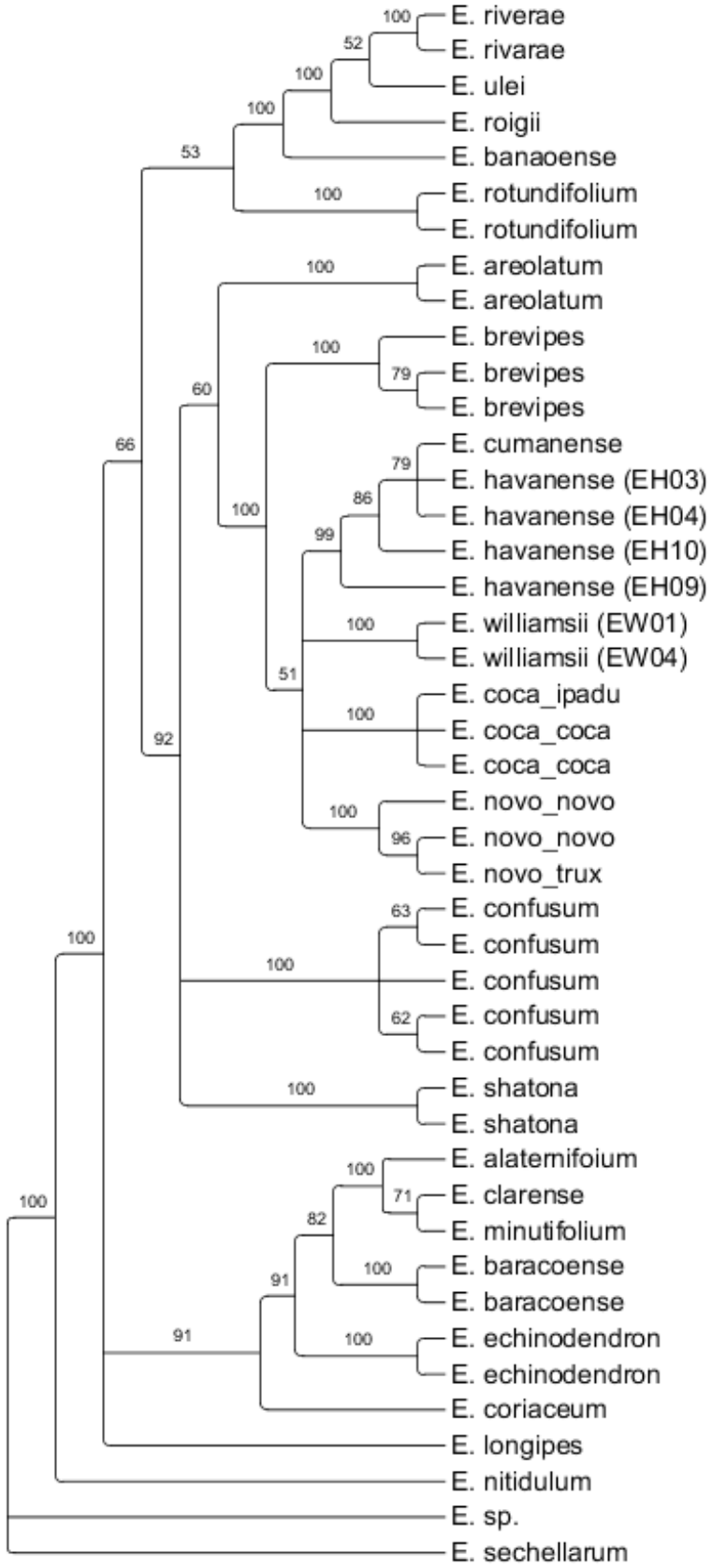
idh maximum likelihood tree (46).



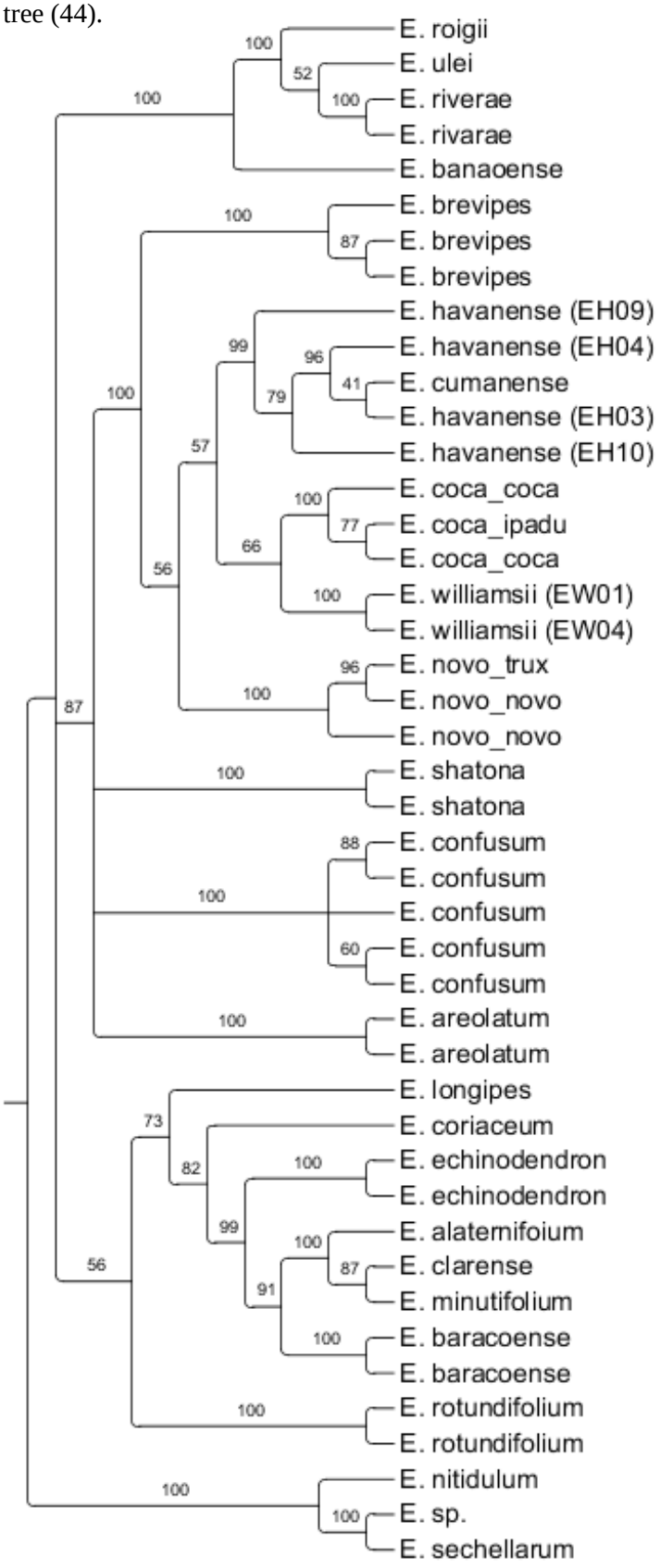
all nuclear loci maximum parsimony strict consensus tree (44).



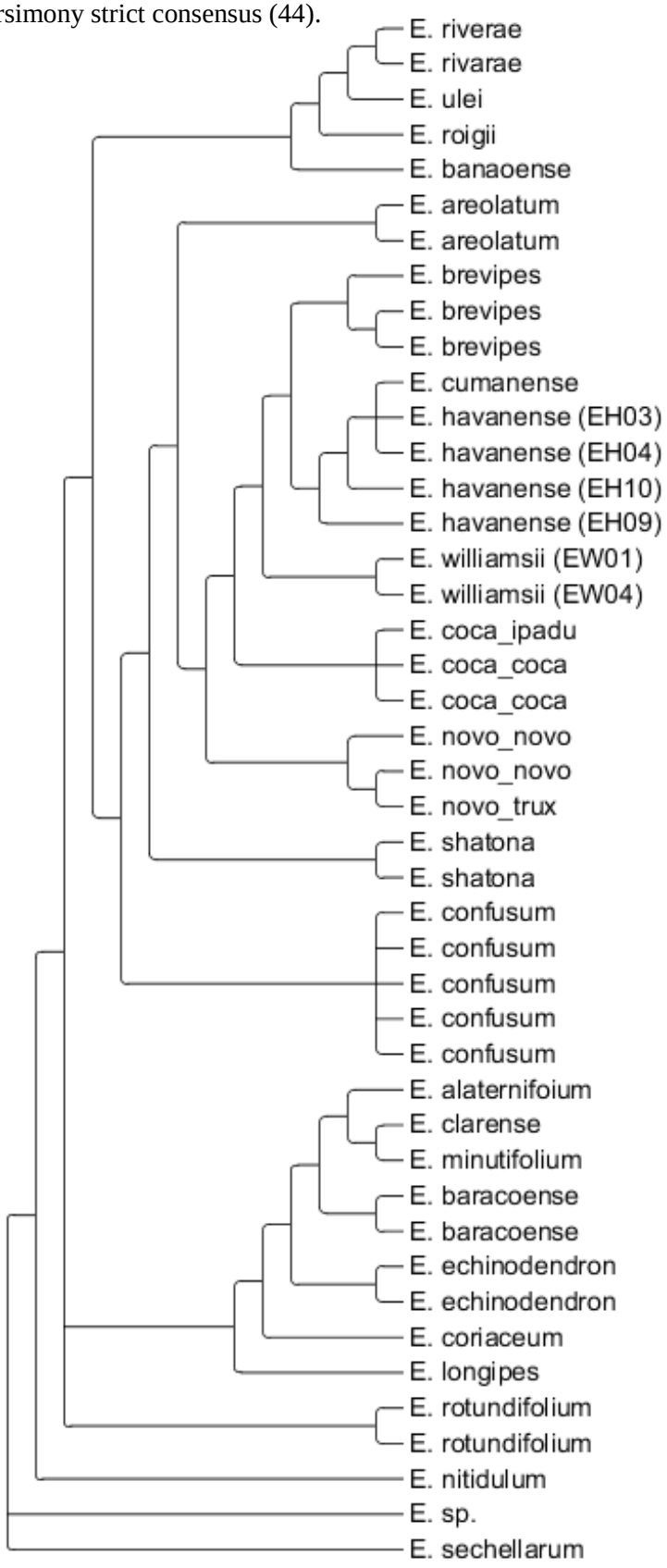
all nuclear maximum parsimony majority rule tree (44).



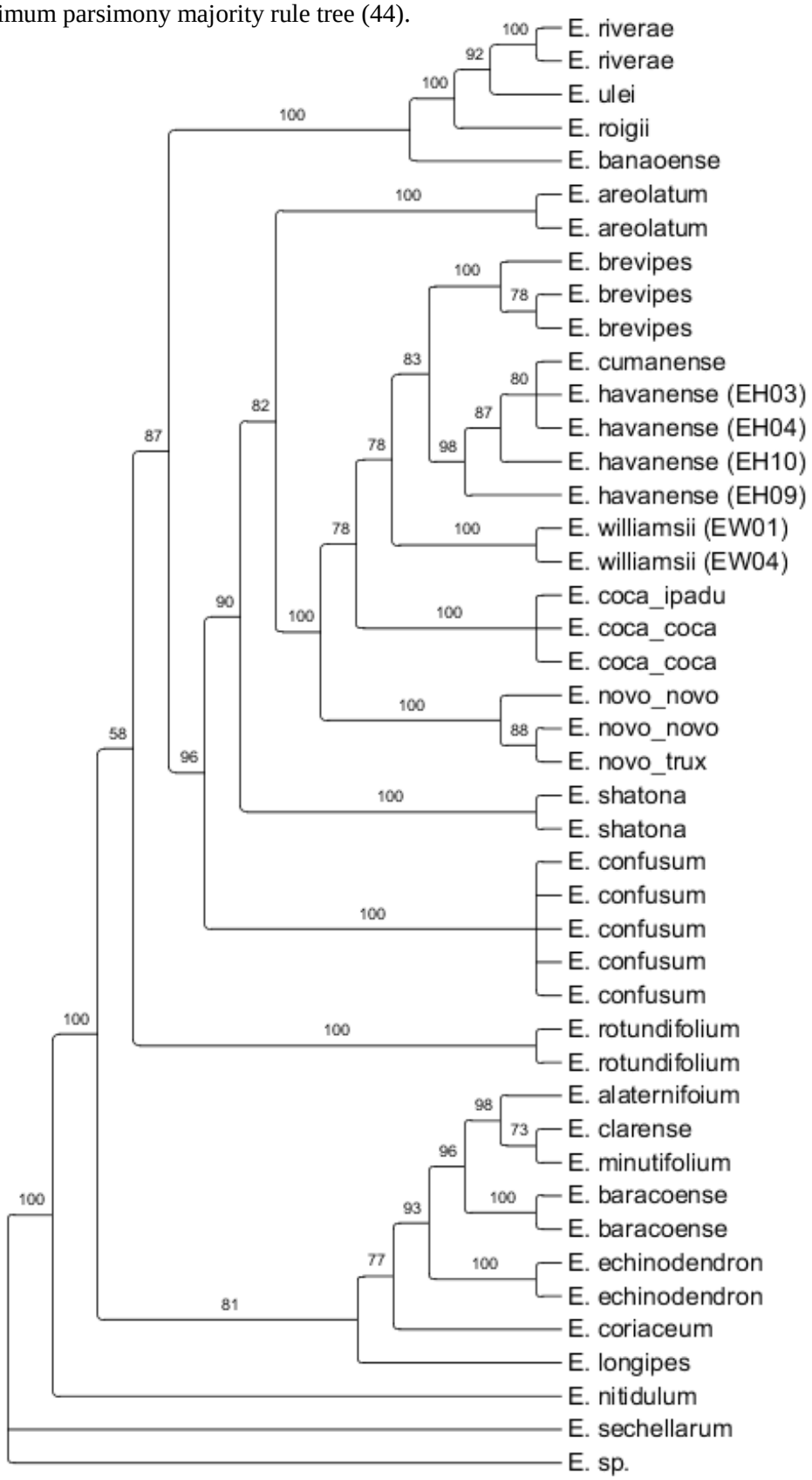
all nuclear likelihood tree (44).



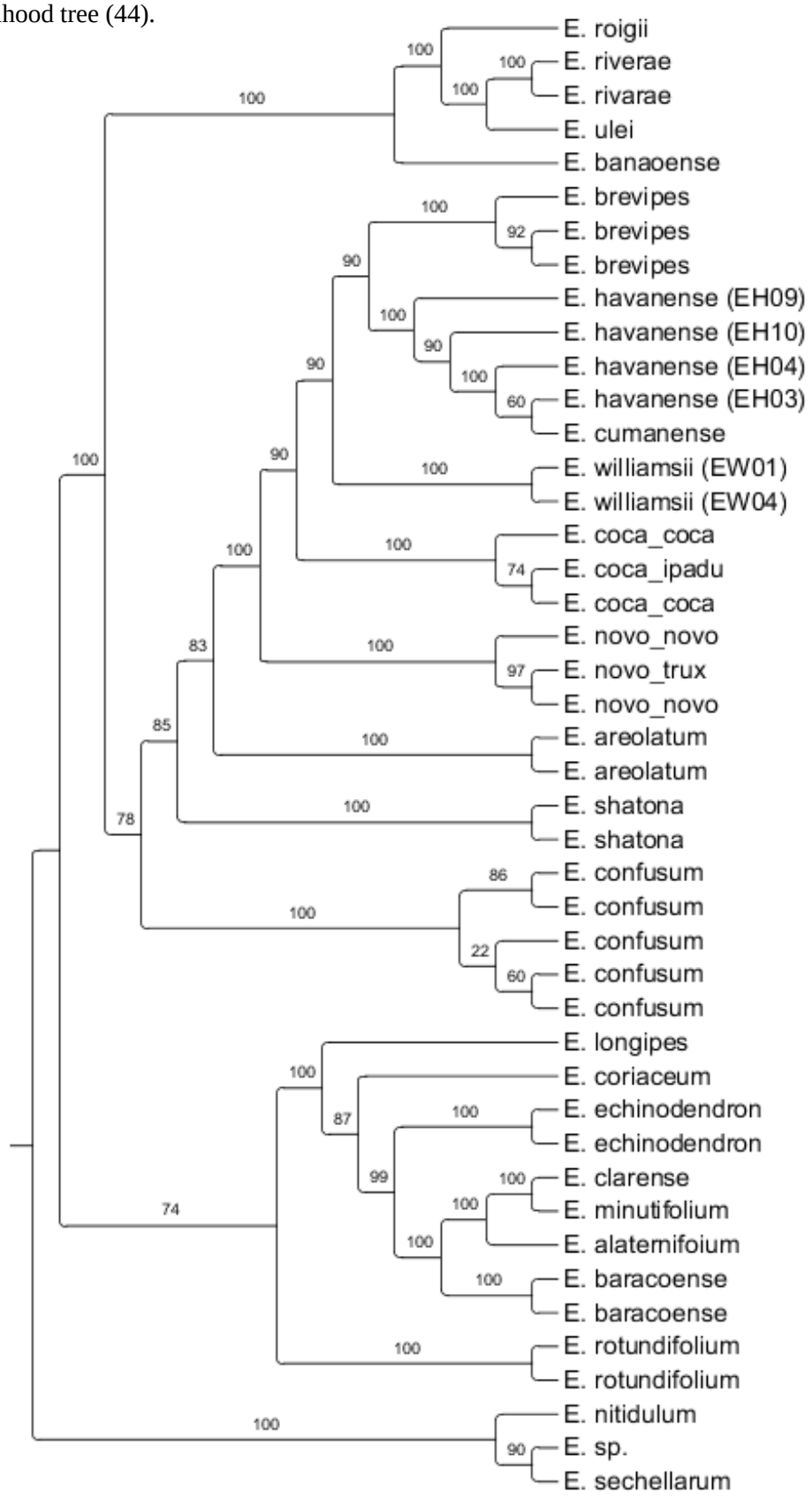
all loci maximum parsimony strict consensus (44).



all loci maximum parsimony majority rule tree (44).



all loci likelihood tree (44).



Appendix 4. List of Caribbean *Erythroxylum* species by island, adopted from Oviedo (2002) with refinements based on field work with R. Oviedo Prieto.

Species	*Notes on distribution*
Bahamas	
E. areolatum	Widespread
E. confusum	Widespread
E. reticulatum	Endemic
E. rotundifolium	Widespread
	4 species; 1 endemic; 3 widespread
Cuba	
E. alaternifolium	Endemic
E. areolatum	Widespread
E. armatum	Endemic
E. banaoense	Endemic
E. baracoense	Endemic
E. brevipes	Three islands, Cuba, Hispañiola, Puerto Rico
E. clarense	Endemic
E. confusum	Widespread
E. coriaceum	Endemic
E. dumosum	Endemic
E. echinodendron	Endemic
E. flavicans	Endemic
E. havanense	Widespread
E. lineolatum	Endemic
E. longipes	Endemic
E. minutifolium	Endemic
E. mogotense	Endemic
E. pedicellare	Endemic
E. roigii	Endemic
E. rotundifolium	Widespread
E. spinescens	Endemic
	21 species; 16 endemic; 1 endemic to three islands; 4 widespread
Hispañiola	
E. areolatum	Widespread
E. barahonense	Endemic
E. brevipes	Three islands, Cuba, Hispañiola, Puerto Rico
E. domingense	Endemic
E. rotundifolium	Widespread
E. rufum	Widespread
E. urbanii	Two islands, Hispañiola and Puerto Rico
?E. williamsii	Endemic
	8 species; 3 endemic; 1 endemic to three islands, 1 endemic to two islands; 3 widespread
Puerto Rico	
E. areolatum	Widespread
E. brevipes	Three islands, Cuba, Hispañiola, Puerto Rico
E. rotundifolium	Widespread
E. rufum	Widespread
E. urbanii	Two islands, Hispañiola and Puerto Rico
	5 species; 1 endemic to three islands, 1 endemic to two islands; 3 widespread

Species	***Notes on distribution***
Jamaica	
E. areolatum	Widespread
E. confusum	
E. incrassatum	Endemic
E. jamaicense	Endemic
E. rotundifolium	Widespread
	5 species; 2 endemic; 3 widespread
Lesser Antilles	
E. havanense	Widespread
E. squamatum	Widespread
	2 species both widespread
Cayman Islands	
E. areolatum	Widespread
E. confusum	Widespread
E. rotundifolium	Widespread
	3 species all widespread