The Dissertation Committee for Nathan Michael LeClear Certifies that this is the approved version of the following Dissertation:

Evolution of *Jatropha*: Phylogenetics, Biogeography, and Phylogeography

Committee:

Beryl Simpson, Co-Supervisor

Craig Linder, Co-Supervisor

David Cannatella

Shalene Jha

James Mauseth

Stan Roux

Evolution of *Jatropha*: Phylogenetics, Biogeography, and Phylogeography

by

Nathan Michael LeClear

Dissertation
Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2019

Dedication

To Emily J. Lott. Team *Jatropha* forever and ever.

Acknowledgements

I wish to thank my advisers, Beryl Simpson and Randy Linder, for their patient guidance and encouragement, and for keeping the doors open for me as I wandered. Next, I would like to thank my dissertation committee: Jim Mauseth for helping me to not lose the forest for the topologies, and for recognizing me as a phytophilic educator with a nomination for a teaching award; Shalene Jha for adopting me as a lab member and giving me opportunities to participate in conservation and outreach efforts; Dave Cannatella for instruction on technical aspects pertinent to each chapter and for the best gumbo; and Stan Roux for inviting me to represent the graduate students for the Plant Research Institute, and for keeping an eye on the larger context of a specialized project.

I am deeply indebted to Bijan Dehgan for insightful conversations, for his foundational work on the biology and systematics of *Jatropha*, and for providing material for many of the species used in Chapter 1. I would like to thank George Yatzkievytch, Tom Wendt, and Lindsay Woodruff at the Billie L. Turner Plant Resources Center at The University of Texas at Austin for assistance in arranging collection trips and facilitating loans of specimens. My thanks also goes to Kyle Gervers, who began assisting me in the lab as an undergraduate, who generated much of the sequence data for Chapter 2, and who is now pursuing his PhD. I am grateful to George Montgomery for access to the Arizona Sonora Desert Museum, and for inviting me to his home during collection trips around Tucson. Thank you to Jim Folsom with the Huntington Botanical Gardens for material from living specimens and for an invitation to present to the Cactus and Succulent Society of America.

My deep gratitude goes to Sylvia and Nancy Salas for accompanying me in the field, introducing me to the Commissario in Totolapan, and having lots of patience with my Spanish. Also I am grateful to Luis Cruz, Mary Garcia, and Gonzalo García at SERBO for their assistance in the field and sharing with me the beauty of Oaxaca. Thank you to David Gernandt for permission to sample specimens at MEXU, and to Rosalinda Medina Lemos for a copy of the Euphorbiaceae fasciculo of the Flora del Valle de Tehuacán-Cuicatlán. Collections in the field and herbarium in Patzcuaro, Michoacán were made possible by Yocupitzia Ramírez Amezcua, Victor Steinmann, Catalina Ruiz Domínguezto, and Sergio Zamudio. Thank you to Socorro Gonzalez Elizondo, Jorge Tena, Lizeth Raucho, and Fernando Colin at the herbarium at CIIDIR in Durango for accompanying me in the field and for inviting me to give my first presentation in Spanish. I owe much to David Delgado for helping me in the field in Sonora, and also to José Sánchez Escalante at the herbarium at DICTUS-UNISON.

I am grateful for the wonderful community of graduate students in the Plant Biology program and across the University of Texas at Austin. I learned much from my lab mates in the Simpson and Linder labs: Deise Gonçalves, Amalia Diaz, Edgardo Ortiz, Oscar Vargas, Wei Xiao, Sandra Branham, and Ginny Morrison. Thank you also to Juan Palacio for teaching me lab techniques, sharing a can of sandy, cold mondongo in Cap Rock State Park, and for a brilliant field trip to Mexico. Friday seminars on plant systematics with the Jansen and Theriot labs were invaluable times for presenting new ideas, often further explored on napkins at the Plasmodesma. A non-exhaustive unordered set of graduate students to whom I owe a great deal for both theoretical and applied support follows: Emily Rees, Sarah Cusser, Craig Milroy, Megan O’Connell, Kim Ballare, Bikash Shrestha, Hannah Harrison, and so many more. You all helped make it happen in a big way. An especially big thanks to Nathaniel “Pickles” Pope for banjo

therapy, living room statistics lectures, and collaborative goat roast and crawfish boil experiments.

I would like to thank my parents, Michael and Nancy, for always giving me the green light to pursue my curiosity. The roots of my love for plants stretch back to a garden in Kansas where you showed me how to plant seeds. Thank you to my three wonderful sisters, Rachael, Jami, and McKenna, and to all of my family for your love and support; it would take another dissertation to describe all you have given me.

I would like to thank Cynthia “Bella” Sleight for constant encouragement and singing, adventures and costumes, and for dancing and tacos. I can’t say thank you enough to my Austin family at Dragonfly House: to Thomas Atkinson for assistance in the field, discussions on Mexican biogeography, and excellent jicama salad; to Will Atkinson, for making room for “the other kid” and welcoming me in, thanks brother; and finally, to Emily Lott for many adventures in the Steam-Powered Botinator together and a life-long friendship.

Support for research and travel were provided in part by funding from The University of Texas Plant Biology Program, The University of Texas Graduate School, The University of Texas College of Natural Sciences, The University of Texas Department of Integrative Biology, American Society of Plant Taxonomists, and the Cactus and Succulent Society of America.

Abstract

Evolution of *Jatropha*: Phylogenetics, Biogeography, and Phylogeography

Nathan Michael LeClear, PhD
The University of Texas at Austin, 2019

Supervisors: Beryl B. Simpson & Craig R. Linder

The genus *Jatropha* (Euphorbiaceae) is comprised of approximately 180 species of flowering plants adapted to arid and semi arid climates in tropical regions around the world. As the evolutionary relationships within this group have not been tested using molecular phylogenetic approaches, I used a combination of traditional molecular markers and genomic data generated from restriction site-associated DNA sequencing (RADseq) for this purpose. Using this phylogeny, I answered questions of trends in morphological evolution in the neotropical species of *Jatropha*, and explored potential causes of missing data in RADseq datasets. I reconstructed the biogeographic history of *Jatropha* at the global scale to test hypotheses of vicariance and dispersal pertaining to pantropical disjunct groups. I also investigated the impact of tectonic events on diversification of *Jatropha* within Mesoamerica, and tested standing hypotheses about geographic structure in lineages endemic to seasonally dry tropical forests. Using *J. cardiophylla* as a model, I looked for evidence of Pleistocene refugia in the interior Sonoran Desert using RADseq and ecological niche modeling. *Jatropha* was found to be monophyletic based upon traditional markers, and relationships within *Jatropha* were resolved with RADseq data. Evolutionary analyses indicated the ancestor of *Jatropha* was a shrub bearing a tricarpellate fruit. *Jatropha* originated in the Neotropics and arrived

in Africa via at least two long distance dispersal events. Tectonic events in Mesoamerica impacted diversification of *Jatropha* through both vicariance events and by preventing dispersal between areas. Mixed evidence was found in support of the hypothesis that seasonally dry tropical forest lineages are dispersal limited and primarily experience *in situ* diversification. Genetic analysis of *J. cardiophylla* showed that this species consists of two genetically distinct, yet geographically overlapping lineages. Ages of coalescence for each lineage predate the Last Glacial Maximum. Niche modeling did not identify obvious Pleistocene refugia, but somewhat aligned with spatial patterns of genetic variation. It appears that *J. cardiophylla* has responded to multiple rounds of climate change in the Sonoran Desert, and that different lineages may have responded differently. The sum of this work represents a significant contribution to our understanding of the evolution of *Jatropha* at multiple scales.

Table of Contents

List of Tables……...……xiii

List of Figures……...xv

Chapter 1: Phylogenetics of the Neotropical *Jatropha* (Euphorbiaceae)*:* Using RADseq to Identify Historic Introgression and Elucidate Patterns of Morphological Evolution.....1

Introduction.....1

Materials and Methods.....5

Circumscription of Jatropha.....5

Library preparation and sequencing for phylogenetic analysis of Jatropha *using RAD-seq*.....7

RADseq data processing.....7

Phylogenetic reconstruction of Jatropha *using RADseq data*.....9

Morphological evolution of Jatropha.....10

Patterns of missing data.....11

Topological hypothesis testing.....12

Introgression analyses.....12

Results.....13

Circumscription of Jatropha.....13

RADseq data processing.....14

Phylogenetic reconstruction of Jatropha *using RADseq*.....15

Morphological evolution in Jatropha.....17

Phylogenetic patterns of missing data.....18

Topological hypothesis testing and introgression analysis.....19

Discussion.....20

Circumscription of Jatropha *using standard markers*.....20

Phylogenetics of Jatropha *using RADseq data*.....20

Morphological evolution in Jatropha.....22

Causes of missing data in RADseq datasets, topological incongruence, and introgression.....*23*

Conclusions.....26

Chapter 2: Biogeography of the Neotropical *Jatropha:* Intercontinental disjunctions and Mesoamerican diversification within the Seasonally Dry Tropical Forest.....66

Introduction.....66

Intercontinental disjunction.....66

Diversification in the Neotropical Seasonally Dry Tropical Forests.....68

Methods.....70

Estimating divergence times for Jatropha.....70

Assembly of Locality Data and Defining Biogeographic Areas.....72

Ancestral Area Reconstruction and Estimating Biogeographic Events.....74

Testing for Geographic Structure in the Phylogeny of Jatropha.....75

Results.....76

Divergence dating within Euphorbiaceae and Jatropha.....76

Biogeographic Reconstruction of Jatropha.....77

Intercontinental disjunctions.....78

Mesoamerica.....80

Geographic Structure in the Seasonally Dry Tropical Forests of Mesoamerica.....83

Discussion.....83

Vicariance versus long distance dispersal in shaping intercontinental Biogeography for Jatropha.....83

The Evolution of Jatropha *in the Seasonally Dry Forests of Mesoamerica* 86

Conclusions.....88

Chapter 3: Phylogeography of the Heart-leafed Dragon's Blood (*Jatropha cardiophylla*-Euphorbiaceae), an Endemic Shrub from the Interior Sonoran Desert 117

Introduction.....117

Methods.....121

Data collection and processing……...121

Phylogenetic analyses.....123

Genetic structure and isolation by distance.....124

Genetic diversity and demography.....125

Ecological niche modeling.....126

Results.....127

RADseq.....127

Phylogenetic analysis and population clustering.....128

Genetic structure and isolation by distance.....128

Genetic diversity and demography.....130

Ecological Niche Modeling.....131

Discussion.....132

Conclusions.....135

Literature Cited…….…….....……..……..…..161

List of Tables

Table 1-1: Subgenera, sections, and subsections of *Jatropha* recognized in this study.....29

Table 1-2: Proposed ancestral and derived states for three traits traditionally used to define infrageneric groups in *Jatropha*.....30

Table 1-3:Accession information for sequence data used in phylogenetic analysis.....31

Table 1-4: PCR primers used to generate sequences data.....33

Table 1-5: Clades recovered from phylogenetic analyses of RADseq datasets.....52

Table 1-6: Major clades recovered from phylogenetic analyses of RADseq datasets.....53

Table 1-7:Summary of the disagreements among datasets that produce strongly supported alternative topologies.....55

Table 1-8: Number of transitions between three morphological traits reconstructed for *Jatropha* using stochastic character mapping.....57

Table 1-9: Phylogenetic generalized least squares results…….62

Table 1-10: Results from Shimodaira Hasagawa tests comparing alternative phylogenies inferred from different RADseq datasets.....64

Table 1-11: Results from ABBA-BABA tests between members of clades C1 and C4.…65

Table 2-1: Overview of biogeographic hypotheses.....94

Table 2-2: Top) Fossils and secondary calibration ages used for dating analyses.....95

Table 2-3: Dispersal matrices for global time-stratified analysis.....96

Table 2-4: Dispersal matrices for Mesoamerican time-stratified analysis.....97

Table 2-5: Results from divergence analyses in BEAST and cross-validation for Euphorbiaceae (top) and *Jatropha* (bottom).....99

Table 2-6: Parameter estimates for biogeographic reconstruction of *Jatropha*.....101

Table 2-7: Intercontinental dispersal events by area for Jatropha.....103

Table 2-8: Summary biogeographic events inferred from biogeographic stochastic mapping (BSM) using the top scoring models from BioGeoBEARS.....105

Table 2-9: Vicariance events inferred for *Jatropha* subg. *Curcas*.....112

Table 2-10: Dispersal events inferred for *Jatropha.* subg. *Curcas*.....113

Table 2-11: Results from phylogenetic community structure analysis.....115

Table 3-1: Details of RADseq datasets of *Jatropha cardiophylla* 139

Table 3-2: Nei's F_{ST} distances between Groups, geographic divisions using Group II individuals only, and clusters identified by DAPC for Group II only.....145

Table 3-3: Environmental variables used to construct the ecological niche model of *Jatropha cardiophylla*…..……159

List of Figures

Figure 1-1: Morphological variation in *Jatropha*.....28

Figure 1-2: Parameters file for assembly of RADseq datasets in *ipyrad*.....34

Figure 1-3: Branching strategy used to generate datasets in *ipyrad*.....35

Figure 1-4: Model showing how gene flow produces alternative shared SNP patterns.…36

Figure 1-5: Maximum likelihood phylogeny for low copy nuclear marker EMB2765.....37

Figure 1-6: Maximum likelihood phylogeny for chloroplast marker *rbcL*. 38

Figure 1-7: Maximum likelihood phylogeny of nuclear ribosomal markers *ITS1*, *ITS2*, and *5.8S rRNA*.....39

Figure 1-8: Number of raw reads and recovered loci assembled in *ipyrad* (85min4).....40

Figure 1-9: Characteristics of 12 RADseq data sets assembled from *ipyrad* using 3 similarity clustering thresholds and 4 levels for minimum taxa.....41

Figure 1-10: Recovered loci in RADseq datasets using forward reads only and paired-end reads.....42

Figure 1-11: Missing data in RADseq datasets using forward reads only and paired-end reads.....43

Figure 1-12: Parsimony informative characters recovered in RADseq datasets using forward reads only and paired-end reads.....44

Figure 1-13: Average bootstrap scores across all branches of best scoring maximum likelihood trees using forward reads only and paired-end reads.....45

Figure 1-14: Recovered loci per sample in RADseq datasets using a subset of taxa from *J.* subg. *Curcas* and the full dataset of *Jatropha*.....46

Figure 1-15: Parsimony informative characters in RADseq datasets using a subset of taxa from *J.* subg. *Curcas* and the full dataset of Jatropha.....47

Figure 1-16: Missing data in RADseq datasets using a subset of taxa from *J.* subg. *Curcas* and the full dataset of Jatropha.....48
Figure 1-17: Maximum likelihood phylogeny of concatenated RADseq supermatrix.....49
Figure 1-18: Comparison of consensus topologies from phylogenetic analyses of all 12 datasets using maximum likelihood and coalescence methods.....50
Figure 1-19: Majority rule consensus tree from coalescence analysis.....51
Figure 1-20: Maximum likelihood phylogeny of 85min4 alignment of forward-reads from all samples.....54
Figure 1-21: Ancestral reconstruction of growth habit in *Jatropha*.....56
Figure 1-22: Ancestral state reconstruction of growth form in *Jatropha* using alternative coding of intermediate forms.....57
Figure 1-23: Ancestral reconstruction of carpel number in *Jatropha*.....59
Figure 1-24: Ancestral reconstruction of anther number in *Jatropha*.....60
Figure 1-25: Phylogenetic generalized least squares results.....61
Figure 1-26: Proportion of RAD loci shared across individuals (dataset = 85min4 forward-reads).....63
Figure 2-1: Positions of continental landmasses at present and two times in the past.....91
Figure 2-2: Major mountain ranges, tectonic events, and bioregions used for biogeographic reconstruction for *Jatropha* subg. *Curcas* in Mesoamerica. .92
Figure 2-3: Distribution of *Jatropha* 93
Figure 2-4: Dated BEAST chronogram for Euphorbiaceae from analysis of the chloroplast marker *rbcL*.....98
Figure 2-5: Dated BEAST chronogram of *Jatropha* from analysis of RADseq dataset..100
Figure 2-6: Biogeographic reconstruction of *Jatropha* (model=DEC+j unconstrained). 102

Figure 2-7 Validation of biogeographic stochastic mapping (BSM) the global scale (model=DEC+j).....104
Figure 2-8 Biogeographic reconstruction of *Jatropha* (model=DEC+j + w time-stratified) 106
Figure 2-9 Biogeographic reconstruction of *Jatropha* (model = DIVALIKE+j unconstrained).....107
Figure 2-10: Summary of biogeographic events estimated for *Jatropha* subg. *Curcas*.. 108
Figure 2-11: Biogeographic reconstruction of *Jatropha* subg. *Curcas* (model=DIVALIKE+j+w).....109
Figure 2-12 Biogeographic reconstruction of *Jatropha* subg. *Curcas* (model = DIVALIKE+j time stratified).....110
Figure 2-13 Biogeographic reconstruction of *Jatropha* subg. *Curcas* (model = DEC+j time stratified).....111
Figure 2-14: Correlation of geographic and genetic distance among Mesoamerican *Jatropha* from all habitat types.....114
Figure 2-15: Correlation of geographic and genetic distance among Mesoamerican *Jatropha* from areas of seasonally dry tropical forest.....116
Figure 3-1: Distribution of the major warm deserts of North America, subdivisions of the Sonoran Desert and distribution of *Jatropha cardiophylla*.....138
Figure 3-2: Maximum likelihood phylogeny of *Jatropha cardiophylla* from full dataset 139
Figure 3-3: Geographic distribution of Group I and Group II.....140
Figure 3-4: Phylogenetic networks from SplitsTree for the full dataset and Group II individuals only.....141

Figure 3-5: Scatter plot from discriminant analysis of principal components (DAPC) for the full dataset, excluding *Jatropha vernicosa* (k = 3).....142
Figure 3-6: Scatter plot from discriminant analysis of principal components (DAPC) for the full dataset, excluding *Jatropha vernicosa* (k = 4).....143
Figure 3-7: Scatter plot from discriminant analysis of principal components (DAPC) for the full dataset, excluding *Jatropha vernicosa* (k = 7).....144
Figure 3-8: Scatter plot from discriminant analysis of principal components (DAPC) within Group II (k = 3).....146
Figure 3-9: Scatter plot from discriminant analysis of principal components (DAPC) within Group II (k = 4) with pairwise Nei's F_{ST} for identified clusters.....147
Figure 3-10: Scatter plot from discriminant analysis of principal components (DAPC) within Group II (k = 5) with pairwise Nei's F_{ST} for identified clusters.....148
Figure 3-11: Scatter plot from discriminant analysis of principal components (DAPC) within Group II (k = 6) with pairwise Nei's F_{ST} for identified clusters.....149
Figure 3-12: A - Contingency table of the number of shared individuals between clusters from Group II identified by discriminant analysis of principal components with k=3 and SplitsTree network analysis and the geographic distribution of clusters identified by DAPC.....150
Figure 3-13: Geographic distribution of clusters from Group II identified by discriminant analysis of principal components (k=3).....151
Figure 3-14: Interpolated mapping of the lagged scores of first principal component from spatial PCA for full dataset.....152
Figure 3-15: Interpolated mapping of the lagged scores of first principal component from spatial PCA for Group II dataset.....153

Figure 3-16: Relationship between geographic and genetic distance for *Jatropha cardiophylla* based on the full data excluding *J. vernicosa*.....154
Figure 3-17: Relationship between geographic and genetic distance for the 19 individuals of *Jatropha cardiophylla* of Group I.....155
Figure 3-18: Relationship between geographic and genetic distance for the 285 individuals of *Jatropha cardiophylla* of Group II.....156
Figure 3-19: Bayesian skyline plots for Groups I and II.....157
Figure 3-20: Bayesian skyline plots for geographic areas using Group II individuals only 158
Figure 3-21: Distribution of suitable habitat for *Jatropha cardiophylla* at present date, Last Glacial Maximum, and Last Interglacial.....160

Chapter 1: Phylogenetics of the Neotropical *Jatropha* (Euphorbiaceae)*:* Using RADseq to Identify Historic Introgression and Elucidate Patterns of Morphological Evolution

INTRODUCTION

The genus *Jatropha* L., Euphorbiaceae, is comprised of ~180 species of woody and herbaceous perennials adapted to tropical arid and semi-arid climates (Dehgan and Schutzman, 1994; Govaerts and Frodin, 2000). Within Euphorbiaceae, *Jatropha* is distinguished by its non-milky latex, flowers with well-developed corollas and calyces, pollen grains lacking apertures, and seeds that often bear an oil-rich structure called an elaiosome (Radcliffe-Smith and Esser, 2001). Diversity within *Jatropha* is concentrated in Mexico, the Caribbean, Central and South America, Africa, the Arabian peninsula, and India. *Jatropha* is most widely known for the economically important species, *Jatropha curcas* L., which is cultivated for the production of biodiesel from its oil rich seeds (Fairless, 2007). The genus also displays a high degree of variation in growth form and floral morphology, making it a good candidate for studying patterns of morphological evolution (Dehgan 2012; Fig. 1-1). Growth form is particularly variable, ranging from trees and shrubs exhibiting varying degrees of succulence, to rhizome producing multi-stemmed shrubs, to geophytes that perennially die back to large underground woody organs.

In spite of considerable work on the systematics of the genus, *Jatropha* remains a taxonomically challenging group. The last comprehensive treatment of *Jatropha,* 40 years ago, (Dehgan and Webster, 1979), was based on morphological and anatomical characters and recognized two subgenera (*Jatropha* and *Curcas*), 10 sections, and 10 subsections (Table 1-1). They distinguished the two subgenera primarily by the

degree of fusion of the petals, and distinguished sections, in part, by carpel number, anther number, and growth form (Fig. 1-1). A third subgenus, *Manihotoides*, was created by Radcliffe-Smith (1997) for *J. mahafalensis* Jum & H.Perrier, the only species from Madagascar (Radcliffe-Smith, 1997), although this species has alternatively been assigned to the genus *Givotia* by some taxonomists (Dehgan and Webster, 1979). Based upon the of Dehgan and Webster (1979) treatment, it has been hypothesized that the ancestral form of *Jatropha* was most similar to the extant species *J. curcas* (Dehgan and Schutzman, 1994). Shrubby and geophytic species were hypothesized thought to be more derived, and in Mesoamerica the reduction in form was accompanied by a reduction in carpel number in *J.* subg. *Curcas*, whereas in South America there was a reduction in anther number (Table 1-2).

The position of *Jatropha* within Euphorbiaceae and confirming the monophyly of the genus are important. Phylogenetic analysis of two plastid regions found *Joannesia* to be sister to *Jatropha*, and both genera were placed in a clade sister to a clade including, among other genera, the large genus *Croton* L.in the subfamily Crotonoideae (Wurdack et al., 2005). A second phylogenetic study (Tokuoka, 2007) using three plastid regions and the 18S rDNA supported the affinity of *Jatropha* to *Croton*, however no sample of *Joannesia* was included. Both studies also included only a single species of *Jatropha* (*J. integerrima* Jacq.), and so neither the generic boundaries, nor the taxonomic system of Dehgan and Webster (1979), have been tested using molecular phylogenetics.

In preliminary work for this article, we analyzed individual nuclear and chloroplast regions. We found these to have inadequate variation for resolving relationships within *Jatropha*. The issue of inadequate data from sequencing traditional molecular markers for phylogenetics analysis of closely related species is not unique to

Jatropha (Wessinger et al, 2016). The problem of inadequate data can be compounded by the commonly observed phenomenon of discordance among gene trees, stemming from incomplete lineage sorting and/or interspecific hybridization (Maddison, 1997; Maddison and Knowles, 2006; Vargas et al., 2017). Interspecific hybridization has been demonstrated to occur within *Jatropha*, even among species presumed to be distantly related (Dehgan, 1984).

Jatropha is, therefore, an excellent candidate for using next-generation sequencing methods that can yield large amounts of variable sequence data in evolutionarily shallow groups (McCormack et al., 2013). Several methods are available that are able to sequence large numbers of homologous markers with minimal sequencing effort and cost. In particular, restriction site associated DNA sequencing (RADseq) has proven to be a cost-effective way to collect data from thousands of genetic regions in non-model organisms (Baird, 2008). RADseq has been successfully employed to resolve taxonomically challenging groups of closely related species, such as sedges (*Carex* L.) and oaks (*Quercus* L.) (Escudero et al., 2014; Hipp et al., 2014). In addition, methods for detecting historical introgression between species, a major cause for gene tree discordance, have been developed for use with RADseq data, allowing deeper exploration of the evolutionary processes underlying recalcitrant groups (Eaton and Ree, 2013).

Using RADseq for phylogeny building is not without drawbacks. Virtually every empirical study using RADseq for species level phylogenetics have reported levels of missing data in excess of 50% (Eaton et al., 2017). Missing data in RADseq datasets primarily result from either locus dropout or inadequate sequencing effort, although these are not mutually exclusive. Locus drop out is an evolutionary process in which mutations either alter a cut site, thereby preventing the endonuclease from cutting, or produce a new

cut site close enough to the original that the resulting fragment is smaller than the target size (Andrews et al., 2016). In either case the affected fragment is not sequenced, and because mutations are heritable, any such change in the restriction site of the ancestor of two or more species is inherited by all descendants. Such changes result in the affected locus not being sequenced , and is expected to produce a non-random pattern of missing data. Locus dropout is expected to increase as a function of mutation rate, size of cut site, and phylogenetic distance between samples, suggesting a limit to the phylogenetic depth among samples to which RADseq can be applied. Simulations have estimated this limit to be 60 mya, although empirical studies have attempted to resolve relationships in groups that diverged as much as 80 Mya (Rubin et al., 2012; Herrera and Shank, 2016).

The second cause of missing data, uneven sequencing depth of fragments, arises from the stochastic nature of high-throughput sequencing. More common fragments will be sequenced to a greater degree than rare ones, and PCR steps during library preparation can introduce bias for GC rich fragments and skew the frequency of these fragments to be greater (Aird et al., 2011; Schweyen et al., 2014). Software packages commonly used for RADseq assembly filter loci with inadequate depth of coverage, which results in missing data for many loci in samples that received low read coverage (Catchen et al., 2013; Eaton, 2014). In contrast to locus-dropout, this type of missing data is expected to be randomly distributed across a phylogeny, and to decrease as a function of increased sequencing effort.

The relative importance of these types of missing data on phylogeny reconstruction has recently been explored. A meta-analysis of all available RADseq datasets used for species-level phylogeny construction found that inadequate sequencing effort and phylogenetic distance, singly, and in combination, accounted for a significant amount of variation in missing data among data sets (Eaton et al., 2017). Even so,

simulations have shown that high amounts of missing data have less impact on the accuracy of phylogeny estimation than the reduction in size of, and bias introduced to, datasets by selective exclusion of fast evolving markers found only in a few taxa (Huang and Knowles, 2014).

A third challenge facing RADseq assemblies, particularly for non-model organisms, is establishing the orthology of loci, which is done primarily by adjusting a similarity threshold for clustering sequences. As no hard and fast rules exist for determining proper similarity cut-offs, it is necessary to explore the effects of using different clustering thresholds when assembling RADseq datasets for phylogenetics (Harvey et al., 2015).

The objectives of this study were to: (1) circumscribe *Jatropha* using traditional molecular markers; (2) generate a well-resolved phylogeny of *Jatropha*, emphasizing the neotropical *J.* subg. *Curcas*, using RADseq; (3) map evolutionary trends in growth form and floral morphology in *Jatropha;* (4) determine the probable sources of missing data in RADseq datasets generated for *Jatropha*; and (5) test for historic introgression among species of *Jatropha* as a possible explanation for observed topological incongruities among datasets.

MATERIALS AND METHODS

Circumscription of Jatropha

To circumscribe the boundaries of *Jatropha*, we sampled 57 individuals (56 *Jatropha* species and 1 *Joannesia* species) for sequencing of nuclear and plastid markers. We also used available sequences of the same markers from Genbank for species from subfamily Crotonoideae *sensu* Wurdack et al. (2005) (Table 1-3). Leaf

samples were collected from field sites, living collections, and herbarium specimens and then stored in silica. DNA was extracted with DNeasy Plant Mini Kits (QIAGEN, Valencia, CA), with the initial lysis step performed at 65° C for 10-20 minutes, followed by the addition 25 mL of Proteinase K and overnight incubation at 45° C. Extracted DNA was quantified using a Qubit broad-range kit (Life Technologies, Carlsbad, CA) and visually inspected for quality in 1% agarose gels.

Three markers were amplified: chloroplast *rbcL*, exon 9 of the low copy nuclear gene *EMB2765*, and nuclear ribosomal internal transcribed spacers ITS I and II with the 5.8s rRNA subunit. PCR amplification was done on a Thermo Fisher C1000 Thermal Cycler, (primer sequences: Table S2). PCR products were cleaned using ExoSap and cycle-sequenced with the same primers used for amplification using the Big Dye Terminator kit v 3.1 (Applied Biosystems, Foster City, CA) at The University of Texas at Austin Core Facility.

Sequences were imported into Geneious v. 7.1.9 and aligned using MAFFT (Katoh et al., 2002; Kearse et al., 2012). Separate Maximum likelihood (ML) and Bayesian analyses were performed for each marker, with the best model of substitution determined using jModelTest2 (Darriba, et al., 2012). Maximum likelihood analyses were performed using RAxML v. 8.2.9 HPC2-work flow using GTR+Γ (Stamatakis, 2014). To find the highest scoring tree we performed 100 independent searches. Branch support on the highest scoring tree was assessed by 1000 non-parametric bootstraps. Bayesian analyses were implemented in MrBayes v. 3.2.6 with GTR+Γ (Ronquist et al., 2012). For each analysis we ran two independent MCMC chains for 10 million generations and sampled tree and parameter distributions every 10,000 generations. Both ML and Bayesian analyses were run through the CIPRES Science Research Gateway (Miller, 2010).

Library preparation and sequencing for phylogenetic analysis of **Jatropha** *using RAD-seq*

The protocol of Peterson et al. (2012) was used to prepare 77 libraries for double-digest RADseq: 70 samples of *Jatropha* (61 species including nine duplicated species), and seven outgroup species (one *Joannesia* Pers., five *Croton* L., and one *Manihot* Mill.). We digested 200-600 ng of genomic DNA with *sphI* and *ecoRI* restriction endonucleases (NEB, Ipswich, MA). These enzymes were chosen based on the fragment length distributions of digests of *J. cinerea* and *J. macrocarpa*. Distributions were measured with a Bioanalyzer (Agilent, Santa Clara, CA). Digested DNA was ligated to adapters containing unique in-line barcodes for multiplexing, then pooled and size selected (target: 410 bp, range: 370-450 bp) using a Pippin Prep (Sage Science, Beverly, MA). Libraries were sequenced using two lanes on the Illumina HiSeq 4000 platform with a target of 1.5 million 2x150 bp paired-end reads per sample. All libraries were prepared and sequenced by the Genome Sequencing and Analysis Facility at The University of Texas at Austin.

RADseq data processing

Quality scores of the raw Illumina reads were assessed with the program FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). Inspection of fastq files revealed that reverse reads had consistently lower quality scores than forward reads. Restriction site overhangs were also very low quality, and were trimmed using the ‘reformat.sh’ script from the software package BBTools v36.30 (Bushnell, 2016). Reads were demultiplexed using the program deML v1 (Renaud et al., 2015) allowing for up to

two mismatches in barcodes. Assembly of demultiplexed reads into orthologous groups (hereafter loci) was done using the software package iPyrad v0.7.21 (Eaton and Overcast, 2016) on the Lonestar5 cluster of the Texas Advanced Computing Center at The University of Texas at Austin (http://www.tacc.utexas.edu) (Assembly parameters: Fig. 1-2).

A branching strategy was used to generate datasets that varied by two parameters: percent similarity (%Sim) and minimum number of taxa (minTaxa) (Fig 1-3). The %Sim is the threshold for clustering reads within and across samples, and minTaxa is the minimum number of taxa in which a locus is present to be included in a dataset. When %Sim is too low, some paralogs will be incorrectly aligned as orthologs, but when set too high, some variable orthologs will be split up as separate loci (Harvey, et al., 2015). To explore the effects of changing the clustering threshold on missing data and phylogeny estimation, %Sim was set at 85%, 88%, or 90%. MinTaxa affects the number of loci in a dataset and the amount of missing data therein. Increasing minTaxa reduces the proportion of missing data, but also decreases the total number of loci and phylogenetically informative sites recovered. To explore its impacts on patterns of missing data and phylogeny reconstruction the minTaxa parameter was set at 4, 8, 16, or 32. A total of 12 datasets was assembled using these parameter combinations. To refer to datasets assembled with different combinations of these parameters we give the %Sim followed by the minTaxa, (e.g. 85min4 refers to the datasets assembled at 85% similarity with a minimum of four taxa for a locus to be included).

Once it was determined that *Jatropha* subg. *Curcas* was consistently recovered as a stable clade, we generated 12 additional datasets for all samples from *J.* subg. *Curcas* and eight closely related species using the same parameter combinations described above. These datasets are referred to below as the *Curcas* datasets. By

excluding species more distantly related to *J.* subg. *Curcas* we produced larger datasets, with less missing data, that enabled us resolve relationships within the subgenus.

Phylogenetic reconstruction of **Jatropha** *using RADseq data*

To determine if species for which replicated samples were included were recovered as monophyletic, and to identify potential rogue taxa, ML phylogenetic analyses were performed using 6 datasets assembled from both paired-end and forward-read-only RADseq datasets and containing all 77 samples: 85min4, 88min4, 90min4, 90min8, 90min16, and 90min32. These datasets were chosen because they spanned the range of values for parameters used for assembly.

Ten independent searches using ML analysis were performed using RAxML v. 8.2.9 (GTR+Γ) (Stamatakis, 2014) HPC-2 Workflow on the CIPRES research gateway (Miller et al., 2012) to find the highest scoring tree. Branch support was assessed by performing 500 non-parametric bootstraps. *Croton* was not recovered as monophyletic (consequently neither was *Jatropha*) in analyses using paired-end datasets. Analyses of forward read datasets recovered *Croton* as monophyletic, but the inclusion of the outgroup samples resulted in *Joannesia* and *Manihot* both falling within *Jatropha*. This was in contrast to our findings from analysis of all Sanger markers (presented below), and likely resulted from the large amounts of missing data in the outgroup samples (Fig. 1-8B) These findings, along with the low number of reads for most of the outgroup samples (Fig. 1-8A), indicated that *Jatropha*, *Croton*, and *Manihot* are too phylogenetically distant for informative use of RADseq data. Therefore the *Croton* and *Manihot* samples were removed from subsequent analyses, along with four species of *Jatropha* (*J. dhofarica* Radcl.-Sm., *J. tehuantepecana* J.Jiménez Ram. & A. Campos, *J.*

contrerasii J.Jiménez Ram. & Mart. Gord., and *J. pelargoniifolia* Courbon) which had very low numbers of recovered loci (Fig. 1-8B). *Joannesia princeps* Vell. also had a low number of recovered loci, but was retained as the best outgroup to root trees. We also removed one sample from all species for which we had included duplicates or multiple varieties after determining each pair to be monophyletic.

Using the reduced taxon set of 56 species, which we refer to as the *Jatropha* datasets, phylogenetic analyses were conducted on datasets of forward-reads spanning all parameter combinations (12 in total) using RAxML, with settings as described above, and a quartet-based method that models a multi-species coalescent process implemented in Tetrad v. 0.4.4 (Chifman and Kubatko 2014). Aligned loci were concatenated into unpartitioned datasets for maximum likelihood analyses and alignments of single nucleotide polymorphisms (SNPs) were used for coalescent analyses. All possible quartets were assessed in the coalescent analyses, and branch support was assessed with 500 non-parametric bootstraps.

Morphological evolution of Jatropha

We used stochastic character mapping to reconstruct the ancestral states of growth habit, carpel number, and anther number for *Jatropha* onto the best ML phylogeny inferred from the 85min4 dataset (Huelsenbeck et al., 2003). Character data were compiled from plants in the field, living collections, herbarium specimens, and taken from the literature when species could not be directly observed. For species of *Jatropha* that could be classified as either small trees or large shrubs, two alternative codings for habit were used: one that considered intermediate forms as trees, and the other as shrubs. Geophytic and rhizomatous habits were treated as separate states. Carpel

and anther states were coded as the observed number of each structure. For each character three transition rate models were compared (equal rates, symmetrical reversed rates, and all-rates-different) with the likelihood scores calculated using the R package *ape* (Paradis et al., 2004) and compared via ANOVA. Using the best fit transition model, 100 independent stochastic character maps were run, and the ancestral state probabilities were summarized and plotted onto the starting phylogeny using the R package *phytools* (Revell, 2012).

Patterns of missing data

To determine if patterns of missing data in RADseq assemblies were the result of loci-dropout or unequal sequencing depth among samples we used a quartet-based modification of the phylogenetic generalized least squares (PGLS) analysis (Grafen, 1989; Eaton et al., 2017). We tested for a significant association between the number of shared loci among taxa with either the phylogenetic distance between taxa (dist) or the number of raw reads recovered for taxa (reads). Datasets assembled at the four levels of minTaxa (at 85% similarity) were analyzed along with their respective highest scoring ML trees. Phylogenetic distance was measured as the sum of standardized GTR+Γ branch lengths for all quartets extracted from the starting tree. The number of shared loci between samples and the raw number of reads, log-median transformed, were taken from iPyrad output from the assembly of each dataset and standardized. Regressions and confidence intervals were calculated using 100 sub-samples of 200 quartets using Python and R scripts available at (https://github.com/dereneaton/RADmissing/blob/master/emp_and_sims_nb_pgls.ipynb).

Topological hypothesis testing

Shimodaira-Hasegawa (SH) tests (Shimodaira and Hasegawa, 1999) were performed to test the significance of strongly supported (>80% BS) topological incongruities found when different datasets were analyzed. We performed the SH tests in IQTree v1.5.4 (Nguyen el at., 2015), allowing the program to calculate a starting tree, and the performed one thousand RELL bootstrap replicates to generate a distribution of site-specific likelihood scores for the starting tree to which the differing topologies were compared (Kishino et al.,1990).

Introgression analyses

To test if the observed significant topological incongruities observed were the result of hybridization we calculated Patterson's D statistics using the ABBA-BABA test (Durand et al., 2011), implemented in iPyrad v0.7.21 (https://github.com/dereneaton/ipyrad/blob/master/tests/cookbook-abba-baba.ipynb). This test assumes that, given a sequence alignment and a rooted quartet, discordant polymorphic site patterns can be a product of either introgression or incomplete lineage sorting. Incomplete lineage sorting should cause equal frequencies of the two possible discordant patterns (ABBA and BABA), whereas gene flow (introgression) should cause one pattern to predominate (Fig. 1-4). Tests for introgression were conducted with the 85m4 dataset because it supported the alternative placement of *J. elbae* (in clade C2 rather than C4; see Results) and provided the greatest number of loci for making comparisons.

Patterson’s D was calculated for all quartets that included *Jatropha elbae* as a possible species having introgressed with *J. jaimejimenezii* V.W.Steinm., rooting

quartets with *J. gaumeri* Greenm. Quartets were generated from a starting tree in which *J. elbae* J.Jiménez Ram. was in clade C-4 (see Fig. 1-17), the position supported by the majority of phylogenetic analyses and most in accordance with traditional taxonomy.

The four taxon ABBA-BABA test identifies pairs of taxa that have experienced past gene flow but cannot determine the direction of this gene flow. To determine the direction of gene flow, we calculated partitioned D-statistics with a five taxon extension of the ABBA-BABA test after identifying pairs of potentially introgressing taxa with the four taxon test (Eaton and Ree, 2013; Eaton et al., 2015). We used the same starting topology, dataset, and root as described for the four taxon test. All D-statistics were standardized to generate z-scores to assess significance. Z-scores were adjusted for multiple comparisons with a Bonferroni correction, using a conservative cutoff of a corrected $\alpha = 0.01$.

RESULTS

Circumscription of Jatropha

Maximum likelihood and Bayesian reconstructions for each of the three markers inferred *Jatropha* as a monophyletic group. Two of three markers (EMB2765, BS = 97/BPP = 0.91 and rbcL, BS = 72/0.87) supported *Joannesia* as sister to *Jatropha* (Figs. 1-5 and 1-6), whereas the nuclear ITS1-5.8SrRNA-ITS2 region inferred *Jatropha* to be sister to a clade including *Joannesia*, *Croton*, and three other genera (BS = 67, Fig. 1-7). Based upon the strong support from EMB2765 reconstructions and the agreement of the rbcL ML reconstructions, *Joannesia* was used as the outgroup for *Jatropha* in subsequent analyses of RADseq datasets.

RADseq data processing

Illumina sequencing of 77 samples generated ca. 108 million paired-end reads, 98.6% of which were high quality. An average of $1.35x10^6$ reads/sample (SD = $9.6x10^5$) was recovered but coverage varied widely across samples (Fig. 4A) (range: $5.1x10^4$ – $4.0x10^6$ reads/sample). Samples with high read coverage tended to have high numbers of recovered loci in final alignments, but this was not true in all cases (compare Figs. 1-8 and 1-8B).

Datasets produced from different parameter combinations (%Sim and minTaxa), read-type (paired-end and forward only), and taxon sub-sampling (all *Jatropha* or only subg. *Curcas*) showed several similar trends. The total number of recovered loci and the percentage of missing data both increased with increasing %Sim for all but the highest level of minTaxa, a trend that was more pronounced for datasets assembled from paired-end reads (Figs. 1-9A and B, 1-10,and 1-11). Together these findings suggest that more stringent %Sim values may have over-split variable loci during assembly. The number of parsimony informative sites was consistently higher in datasets assembled from lower clustering thresholds, and nearly doubled when paired-end reads were used (Figs. 1-9C and 1-12).

Average bootstrap values, computed across all nodes of the best ML trees from all datasets, were consistently above 90 for all forward-read datasets (Fig. 1-9D), but were increasingly unstable for trees inferred from paired-end data sets as %Sim was increased (Fig. 1-13). In particular, support values were lowest for the tree inferred from the 90min4 paired-end dataset, the parameter combination most likely to include over-split loci (Fig. 1-13). Due to the higher levels of missing data, greater chances of over-

splitting loci with some parameter settings, and greater variability in bootstrap support for trees inferred from paired-end reads, we present the phylogenies produced from forward-read datasets.

Using a subsample of closely related taxa to generate additional datasets increased the amount of recovered data. The *Curcas* datasets had nearly double the average number of loci per sample (range *Curcas*: 780-2,606; range full: 330-1,563), more parsimony informative sites (range *Curcas*: 8,311-53,349; range full: 5,635-49,805), and less missing data (range *Curcas*: 9.7-65.8; range full: 27.8-82.4) compared to the full *Jatropha* datasets (Figs. 1-14 – 1-16).

Phylogenetic reconstruction of Jatropha *using RADseq*

Maximum likelihood and coalescent-based analyses of all *Jatropha* datasets supported similar backbone topologies in nearly all cases (Figs. 1-17 – 1-19). The major clades identified from this study, and their species, are given in Table 1-5, and ranges in bootstrap support for the composition of these clades and their placement in the tree given in Table 1-6. With the exception of *J. gaumeri*, *J.* subg. *Curcas*, as currently circumscribed, was recovered as monophyletic (BS=100 in all analyses). *Jatropha gaumeri*, a species from Yucatan, Mexico, was consistently recovered in a clade of Caribbean species ascribed to *J.* subg. *Jatropha* sect. *Polymorphae*, which we refer to hereafter as the Caribbean clade. The remainder of *J.* subg. *Jatropha* was recovered as a grade, in which the arrangement of taxa depended on whether or not samples of *Croton* were included as outgroups. The largest difference was that the African clade of *J.* subg. *Jatropha* was recovered as monophyletic and sister to *J.* subg. *Curcas* + the Caribbean

clade without the *Croton* samples, but was recovered as a basal grade to the rest of *Jatropha* when *Croton* samples were included (Figs. 1-17 and 1-20).

Within *Jatropha*. subg. *Curcas*, four clades (hereafter C-1 through 4) were strongly supported by analyses of nearly all datasets and reconstruction methods (Fig. 1-17, Table 1-6). Clade C-4 was always recovered as sister to the remainder of *J.* subg. *Curcas*, but there was some disagreement among analyses about the relationship among C-1, C-2, and C-3. Five of twelve *Jatropha* datasets analyzed with ML placed C-2 as sister to C-1 with moderate to high support, and this increased to half of the datasets with higher support when the *Curcas* datasets were analyzed, (BS range = 66-100 for full datasets; 88-100 for *Curcas* datasets; Table 1-6). Datasets that supported the sister relationship of C-2 and C-1 had the least amount of missing data (minTaxa = 16 and 32). Seven of twelve ML analyses of *Jatropha* datasets, half of the *Curcas* datasets, and all coalescent analyses recovered C-3 as sister to C-2 (BS range = 32-97 for full dataset; 40-95 for *Curcas*).

Relationships within C-4 changed across datasets and were generally poorly resolved. A major topological disagreement among data sets was the placement of *J. elbae* (Fig. 7, Table S3). Maximum likelihood analyses resolved *J. elbae* as part of either C-2 or C-4, whereas coalescent analyses consistently placed *J. elbae* in C-4 (Fig. 1-18, Table 1-7). We investigated this topological disagreement more closely for significance and evidence of introgression below.

Sister to *Jatropha* subg. *Curcas* was the Caribbean clade + *J. gaumeri*, followed by the five species in the African clade of *J.* subg. *Jatropha*. The next branch was a single North American species, *J. macrorhiza* Benth., belonging to the geographically widespread *J.* subg. *Jatropha* sect. *Peltatae*. The remaining *J.* subg. *Jatropha* species were recovered in a large clade predominately consisting of South

American taxa along with one African and one North American species, *J. mahafalensis* and *J. cathartica* Terán & Berland. respectively. For convenience we refer to this clade as the South American clade of *J.* subg. *Jatropha* (J-SA), and recognize two clades within it: J-SA1 and J-SA2 (Fig. 1-17). J-SA1 corresponded to *J.* subg. *Jatropha* sect. *Jatropha* and J-SA2 included all the species sampled from *J.* sect. *Peltatae*, with the exception of *J. macrorhiza* which was sister to the African clade and *J.* subg. *Curcas* (see above). Also included in J-SA2 were *J. mahafalensis* from the monotypic subgenus *Manihotoides* and *J. martiusii* Baill., from the monotypic section *Martiusiae.* Relationships within J-SA2 were influenced strongly by method of reconstruction. Maximum likelihood analyses supported *J. mahafalensis* as sister to *J. weddelliana* Baill., deeply nested within the clade, whereas coalescent analyses recovered *J. mahafalensis* as sister to the remainder of J-SA2 (Fig. 1-18).

Morphological evolution in Jatropha

Stochastic character mapping (equal-rates transition matrix) reconstructed the shrub habit as the ancestral state for *Jatropha* (Fig. 1-20) regardless of how intermediate forms were coded, but the number of transitions from shrub to tree habit was affected by character coding (Table 1-8). When intermediate forms were coded as trees, six independent transitions from shrub to tree habit were inferred: three in J-SA2, one in the Caribbean clade with a reversion to shrub form, one in C2 with a single reversion, and the sixth along the branch to C1 with two subsequent reversions (Fig. 1-21). When intermediate forms were coded as shrubs, there were still six transitions from shrub to tree but transitions were delayed, eliminating reversions to the shrub habit, and J-SA2 had only one transition (on the branch to *J. mahafalensis*) whereas C1 contained three

(Fig. 1-22). The geophyte habit evolved from the shrub habit independently along the branches leading to *J. cathartica* and *J. macrorhiza*, and the rhizomatous shrub habit evolved from the shrub habit four times: twice in the African clade, once in *J. dioica*, and once in *J. cardiophylla* (Figs. 1-21 and 1-22). The number of transitions to these two habits was unaffected by alternative character coding (Table 5).

The ancestral state of carpel number in *Jatropha* was three (all-rates-different transition matrix), with a transition to two carpels on the branch leading to the ancestor of *J.* subg. *Curcas* (Fig. 1-23). Within *J.* subg. *Curcas* we inferred three reductions to one carpel: twice in C-4 and once in C-3. The ancestor of C-2 reverted to the tricarpellate state, which persisted in all sampled taxa in that clade, and in C-1 we inferred numerous shifts between two and three carpels (Fig. 1-23).

Stamen number in the common ancestor for all *Jatropha* was eight (equal-rates transition matrix; Fig. 1-24). In *J.* subg. *Jatropha* there was a single transition to six in *J. martiusii*, and two shifts from eight to ten, one within the Caribbean clade and one in the branch leading to to *J.* subg. *Curcas*. Within *J.* subg. *Curcas* there was a single shift from ten to five anthers in *J. jaimejimenezii* and from ten to eight in *J. bullockii* E.J.Lott.

Phylogenetic patterns of missing data

Results from phylogenetic generalized least squares indicated that, for all datasets analyzed, phylogenetic distance among taxa explained a significant amount of variation in the number of shared loci, whereas the raw number of reads per sample did not (Fig. 1-25; Table 1-9). The main explanatory factor for missing data in our datasets was phylogenetic distance between taxa. Pairwise comparison of the number of shared loci between species showed a strong degree of partitioning between major clades

identified in our phylogenetic analyses (Fig. 1-26). Therefore it is probable that further sequencing would have primarily increased read depth for loci rather than novel loci discovery.

Topological hypothesis testing and introgression analysis

Shimodaira-Hasegawa tests were performed to compare alternative placement of the problematic *Jatropha elbae* based on analyses of different datasets (Table 1-10). When *J. elbae* was constrained to its alternative placement for a given dataset, a significantly reduced likelihood score resulted (range: $p = 0.001$ to $p < 0.0001$), which was a further indication of hybridization between *J. elbae* and another species. Results of ABBA-BABA tests, however, indicated no significant gene flow between *Jatropha elbae* and *J. jaimejimenezii* (Table 1-11), and that hybridization between these two species was not likely the cause for the topological incongruities observed for *J. elbae.*

The ABBA-BABA tests did however reveal unexpected signals of introgression between *J. elbae* and members of a sub-clade in C1 comprised of *J. alamanii* Müll.Arg, *J. platyphylla* Müll.Arg, *J. ciliata* Sessé, and *J. bartlettii* Wilbur, and additional ABBA-BABA tests showed extensive introgression between this sub-clade of C1 and all of clade C4 (Table 8). This led us to perform five-taxa introgression tests to assess the likely source of this signal. Results indicated that all introgression took place between either *J. alamanii* or *J. ciliata* and a branch containing either *J. oaxacana* and *J. dioica* Sessé or *J. oaxacana* J.Jimènez Ram. & R. Torres and *J. cuneata* Wiggins & Rollins from C4 (Table 9). These results suggest that introgression took place between the ancestor of C4 and *J. alamani* and/or *J. ciliata;* however introgression was not also

detected between C1 species and either *J. elbae* or *J. neopauciflora* Pax, both of which branched off later within C4. This pattern could be explained by *J. ciliata* or *J. alamanii* having crossed with *J. oaxacana*, which subsequently crossed with the ancestor of *J. dioica* and *J. cuneata.* Alternatively, *J. oaxacana* crossed with the widespread species *J. dioica* alone, and *J. dioica* acted as a conduit of gene flow into *J. cuneata.* These scenarios are somewhat complicated by the fact that *J. dioica* and *J. cuneata* are both confirmed tetraploids. Chromosome counts and fine-scale sampling from populations of these species would be necessary to test these hypotheses. Also, the lack of resolution within the C-4 clade makes it difficult to determine which might have been the case, but this evidence of hybridization might be a clue why the relationships in this clade have proven challenging to resolve.

DISCUSSION

Circumscription of Jatropha *using standard markers*

No change to the generic circumscription of *Jatropha* was necessary as a result of our analyses as all markers confirmed its monophyly. This includes the contested species from Madagascar, *J. mahafalensis*, which is strongly supported as being a species of *Jatropha*. Two markers, *EMB2765* and *rbcL*, supported *Joannesia* as sister to *Jatropha,* a result that agrees with the only other molecular phylogenetic analysis of the family that included both *Jatropha* and *Joannesia* (Wurdack et al., 2005). Analysis of ITS weakly supported *Joannesia* as sister to *Croton* rather than to *Jatropha*, but it is probable that ITS was too variable for reconstructing relationships at this phylogenetic depth, as has been found in other studies (Hearn, 2006).

Phylogenetics of Jatropha *using RADseq data*

RADseq data were adequate to resolve infrageneric relationships within *Jatropha* to a degree which traditional markers have failed. Further, this first attempt to resolve relationships within *Jatropha* using molecular data significantly improved our understanding of evolutionary trends within the genus. We were not, however, able to generate sufficient informative data from RADseq to confidently assign the placement of every species of *Jatropha* included in this study, particularly those for which samples were collected from herbarium specimens which had very low numbers of reads and consequently low numbers of recovered loci. Future work should be able to improve inferences of relationships in *Jatropha* using RADseq as more samples can be added to RADseq datasets at a later time as long as the same restriction enzymes and size selection steps are maintained (Hipp et al., 2014). Therefore we are hopeful that using fresh material instead of herbarium specimens and a more complete sample of species, especially the South American, Caribbean, and Old World taxa, will yield a better resolved and supported phylogeny.

We found strong support for the monophyly of *Jatropha* subg. *Curcas* with the single exception of *J. gaumeri*, which fell in the Caribbean clade of *J.* subg. *Jatropha*. The sectional boundaries within *J.* subg. *Curcas* of Dehgan and Webster (1979), however, were not supported, nor were any morphological synapomorphies evident among the characters we investigated that would help to define the major clades identified in our phylogeny.

At the limited level of sampling for *Jatropha* subg. *Jatropha*, the subgenus was not recovered as a monophyletic group, but as a grade comprised of three clades corresponding to three biogeographic regions: South America, Africa, and the Caribbean. *Jatropha mahafalensis* was strongly placed within the South American clade of *J.* subg.

Jatropha, which suggests that *J. mahafalensis* should not be considered a separate subgenus, despite its geographical and morphological uniqueness. It is possible that increased taxon sampling from Africa could change this finding. The sections of Dehgan and Webster (1979) within the South American clade of *J.* subg. *Jatropha* were monophyletic, with the exception of *J. macrorhiza*, and allowing for the inclusion of *J. mahafalensis* within *J.* subg. *Jatropha* sect. *Peltatae.*

Morphological evolution in Jatropha

Previous evolutionary hypotheses proposed that *Jatropha curcas* is the extant species most representative of the ancestral form of the genus because it possesses, among other traits, an arborescent growth habit, 10 monadelphous stamens, and tricarpellate capsules (Dehgan and Webster, 1979; Dehgan, 2012). We found that carpel number is the only trait we observed in *J. curcas* that was shared with the ancestor of the genus, and that *J. curcas* is a derived species rather than basal within the genus.

We inferred the ancestral habit for *Jatropha* to have most likely been a shrub, and that numerous transitions and reversions between shrub and tree habit took place, especially within *J.* subg. *Curcas*. We inferred geophytic and rhizomatous habits to be derived states, which is in agreement with previous authors' assertions that these traits evolved late in the genus, potentially as adaptations to arid or freezing conditions (Dehgan and Webster, 1979). Growth habit in *Jatropha* appears to be labile and the high number of independent origins of geophytism, rhizome formation, and arborescence would make *Jatropha* an interesting group for studying the anatomical basis for the evolution of growth form.

Carpel number, while stable in *Jatropha* subg. *Jatropha*, is a labile trait within *J.* subg. *Curcas* and therefore of little utility for defining sections. We inferred that the basal lineages of *J.* subg. *Curcas* (C3 and C4) underwent reductions from the ancestral state of three carpels. Previous authors have noted that the reduction of carpel number, like the tendency towards geophytic and rhizomatous growth habits, is more common in species occurring in the northern portion of the range in *J.* subg. *Curcas* (Dehgan and Webster, 1979). In contrast, species of *J.* subg. *Jatropha* occurring at higher latitudes in South America do not show a similar trend in reduced carpel number (Dehgan, 2012). This difference could be the result of phylogenetic constraint of carpel number in *J.* subg. *Jatropha,* or perhaps that selection has acted differently on carpel number for species the Northern versus the Southern Hemisphere. Two species of *J.* subg. *Jatropha, J. cathartica* and *J. macrorhiza*, that occur as far north as Texas and Arizona, possibly as a result of long distance dispersal, have retained the tricarpellate state, which would support a phylogenetic constraint hypothesis.

Anther number was nearly uniform across *Jatropha* subg. *Curcas*, in agreement with previous taxonomic work (Dehgan and Webster, 1979). The reduction in anther number from ten to five in *J. jaimejimenezii* was inferred to have been independent from the reduction from ten to eight in *J. bullockii.* Ten anthers would be a morphological synapomorphy for *J.* subg. *Curcas* if *J. gaumeri* were included within the subgenus, however this would necessitate including the Caribbean taxa currently assigned to *J.* subg. *Jatropha.* Anther number also varied little in *J.* subg. *Jatropha,* shifting only a single time from eight to six in *J. martiusii*, but our taxon sampling did not capture the majority of the variation reported for this character in the South American taxa. As a result of this limited sampling, we could not assess the taxonomic utility for anther number in defining sections in *J.* subg. *Jatropha*.

Causes of missing data in RADseq datasets, topological incongruence, and introgression

Our RADseq datasets were highly informative and largely helped to resolve relationships in *Jatropha*. However, we also found a high degree of variation in the number of reads per sample, recovered loci, and parsimony informative sites across datasets assembled with different %Sim and minTaxa parameters. Most of the samples that were removed from analyses due to high amounts of missing data were from herbarium specimens and generally had very low read coverage. Low read number in samples can be caused by DNA degradation, which has been shown to decrease read quality and lower the number of recovered loci in RADseq datasets (Graham et al., 2015). Wessinger et al. (2016) showed that it is possible to use herbarium samples in RADseq phylogenetic studies, but Beck and Semple, (2015) found a negative correlation between age of herbarium specimen and read quality and a positive correlation between specimen age and the amount of missing data. No herbarium samples included in the present work were greater than 20 years old, but it was still evident that these samples received fewer reads on average than samples derived from fresh material. At the very least using herbarium samples would require greater sequencing effort to produce comparable data to what can be acquired from fresh material.

Read depth directly affects the number of recovered loci because assembly programs like iPyrad use a minimum depth of coverage for statistically determining base calls. Previous work has shown that inadequate read depth can be a major cause of missing data in RADseq datasets, in addition to loci-dropout resulting from mutations in restriction sites (Eaton et al., 2017). With the library preparation methods and enzyme

combination used in this study, we found that only phylogenetic distance between tips significantly correlated with missing data.

The underlying concern about the source of missing data in any phylogenetic study is whether these missing data impact our ability to infer relationships correctly. Simulations have demonstrated that excluding loci present in only a few taxa can yield datasets biased towards slowly evolving loci, which reduces the accuracy of maximum likelihood phylogeny estimation (Huang and Knowles, 2014). Although high amounts of missing data might not negatively impact accuracy of phylogeny reconstruction when the missing data are randomly dispersed across the tree, it could impact the ability of ML analysis to consistently recover the correct topology when a small number of samples have much more missing data.

We found systematic differences in the placement of *Jatropha elbae*, a herbarium sample with low read count and recovered loci (101,00 reads and 735 loci compared to the mean of all samples of 1.3 million reads and 1,527 loci), by ML analyses of datasets with varying amounts of missing data. For the complete *Jatropha* taxon sets, ML analyses of datasets with the least amount of missing data (minTaxa = 16 or 32) and all coalescent analyses supported the placement of *J. elbae* in the C-4 clade, which largely corresponds to *J.* subg. *Curcas* sect. *Mozinna* to which *J. elbae* is currently assigned. The taxonomically reduced *Curcas* datasets, which had more parsimony informative sites and fewer missing data inferred the same relationship. The placement of *J. elbae* in C-2 was only seen in ML analysis of datasets with the greatest amounts of missing data.

We closely investigated the potential cause of the topological disagreement among analyses of different RADseq datasets with respect to the placement of *Jatropha elbae.* No signal of past introgression was detected between this species and

any species from clade C2, where *J. elbae* was placed in a number of reconstructions. It seems probable, as our sample of *J. elbae* was taken from herbarium material, that low read number and its production of high levels of missing data was the cause of the observed incongruence. These findings indicate that our strategy of creating reduced-taxon datasets of more closely related species can help to resolve the relationships of species with high amounts of missing data.

Our tests for introgression uncovered evidence for gene flow between members of clades C4 and C1, and further investigation is warranted to determine if the widespread species *J. dioica* has acted as a conduit for gene flow between geographically distant species.

CONCLUSIONS

The circumscription of *Jatropha* and understanding about the relationships therein were improved by using a combination of traditional genetic regions and RADseq data for phylogenetic analysis. With the exception of *J. gaumeri*, *J.* subg. *Curcas* was found to be monophyletic, but the currently recognized sections within this subgenus were not. *Jatropha* subg. *Jatropha* was not recovered as monophyletic, but the sections largely were. Increased taxon sampling, especially of South American and Old World species, would improve our understanding of relationships across the genus as a whole.

Previous authors hypothesized that the ancestor of *Jatropha* was arborescent and had flowers with ten anthers and fruits with three carpels, most similar to *J. curcas*. Our analyses inferred the ancestor of *Jatropha* to most likely have been a shrub with bearing flowers with eight anthers and a fruit with three carpels.

Levels of missing data in RADseq datasets were significantly correlated with phylogenetic distance between species, but not read depth. Although we included samples of species from *Croton* and *Manihot* for RADseq library preparation we were unable to recover sufficient loci shared between these genera and *Jatropha* for phylogenetic analysis to confidently resolve relationships among them, indicating that the phylogenetic distance between these genera exceeds that for which RADseq is useful. This was especially the case when assembling datasets from paired-end reads, which did not successfully recover the genera as monophyletic.

Figure 1-1: Morphological variation in *Jatropha*. Variation in habit (A-C) and fruit and floral traits D-J of *Jatropha:* A-Tree form (*J. stephani*), B-Shrub exhibiting above ground succulence (*J. podagrica*), C-Multibranched rhizomatous shrub (*J. cuneata*), D-Unilocular capsule (*J. cuneata*), E-Bilocular capsule (*J. oaxacana*), F-Trilocular capsule (*J. macrorhiza*), G-Solitary female flower with fused corolla typical of J. subgenus Curcas (*J. stephani*), H-Staminate flowers with free petals typical of *J*. subgenus *Jatropha* (*J. cathartica*), I-Staminate flower (perianth removed) showing 10 biseriate stamens (*J. curcas*; bar = 2mm), J-Staminate flower (perianth removed) showing 8 biseriate stamens (*J. nudicaulis*; bar = 2mm).

Jatropha

Subg. *Jatropha*
- **Sect. *Jatropha***
  - **Subsect. *Adenophorae***
  - Subsect. *Purpurae*
  - Subsect. *Isabellae*
- Sect. *Collenucia**
- **Sect. *Spinosae****
- **Sect. *Tuberosae****
- **Sect. *Peltatae***
  - **Subsect. *Peltatae***
  - **Subsect. *Multifidae***
  - **Subsect. *Macrorhizae***
- **Sect. *Martiusae***
- **Sect. *Polymorphae***
  - **Subsect. *Polymorphae***
  - **Subsect. *Hernandiifoliae***

Subg. *Curcas*
- **Sect. *Curcas***
- **Sect. *Platyphyllae***
  - **Subsect. *Platyphyllae***
  - **Subsect. *Fremontioides***
  - **Subsect. *Gaumeri***
- **Sect. *Loureira***
  - **Subsect. *Loureira***
  - **Subsect. *Canescentes***
  - **Subsect. *Neopauciflorae***
- **Sect. *Mozinna***

Subg. *Manihotoides**

Table 1-1: Subgenera, sections, and subsections of *Jatropha* recognized in this study. Bold names indicate groups that were sampled for genetic sequencing, and asterisks indicate African groups.

Character	Ancestral	Derived
Habit	Arborescent	Shrub or geophyte
Stamens	10	6 or 8
Carpels	3	1 or 2

Table 1-2: Proposed ancestral and derived character states for three traits traditionally used to define infrageneric groups in *Jatropha.*

Taxa	Voucher		GenBank accession information			
			ITS	***EMB2765***	***rbcL***	**NGS**
Astraea lobata	van Ee 486	(WIS)	EU586945	NA	NA	NA
Cladogelonium madagascarensis	Randrianaivo et al. 561	(MO)	NA	NA	AY794868	NA
Cnidoscolus urens	Wurdack D002	(US)	NA	NA	AY794874	NA
Codiaeum variegatum 1		()	JQ898648	NA	NA	NA
Codiaeum variegatum 2	Wurdack D33	(US)	NA	FJ669753	NA	NA
Conceveiba mayanensis	Bell 93-x	(US)	DQ006005	NA	NA	NA
Croton alabamensis	Wurdack D08	(US)	NA	NA	AY788171	NA
Croton alamosanus	Van Devender 2006-1284	(MICH)	NA	HM564240	NA	NA
Croton atroites	Van Ee 537	(WIS)	NA	HQ654593	NA	NA
Croton capitatus	W. R. Carr 36145	(TEX)	NA	NA	NA	653
Croton cupulifer	Rodriguez 1416	(WIS)	EU478063	NA	NA	NA
Croton ekmanii	HAJB 81786	(MICH)	NA	NA	EF405860	NA
Croton glandulosus	W. R. Carr 36137	(TEX)	NA	NA	NA	645
Croton grangerioides	Haevermans 561	(P)	KP878342	NA	NA	NA
Croton guildingii subsp. Tiarensis	Riina 1271	(WIS)	AY971254	NA	NA	NA
Croton lechleri	Riina 1497	(WIS)	NA	HQ654593	NA	NA
Croton lindheimerianus	W. R. Carr 35868	(TEX)	NA	NA	NA	652
Croton lobatus	Steinmann 2024	(RSA)	NA	NA	AY794905	NA
Croton lucidus	Van Ee 378	(WIS)	EF421765	NA	NA	NA
Croton lucidus	Wurdack D117	(US)	NA	NA	AY794909	NA
Croton maestrensis	HAJB 81958	(MICH)	NA	NA	EF405857	NA
Croton michauxii var. ellipticus	Archer 40	(WIS)	NA	HM564296	NA	NA
Croton malvaviscifolius	Mogenses 1042	(DAV)	EU478080	NA	NA	NA
Croton mayarum	Leon 118	(WIS)	EU478038	NA	NA	NA
Croton monanthogynus	W. R. Carr 36280	(TEX)	NA	NA	NA	649
Croton priscus	Riina 1535	(WIS)	NA	HQ654594	NA	NA
Croton sampatik	Riina 1447	(WIS)	NA	NA	EF405859	NA
Croton setiger	Hughey s.n.	(US)	NA	NA	AY794910	NA
Croton stockeri	Forster PIF10580	(BRI)	KP878394	NA	NA	NA
Croton texensis	W. R. Carr 36322	(TEX)	NA	NA	NA	648
Croton trigonocarpus	HAJB 81960	(MICH)	NA	NA	EF405861	NA
Croton urucurana	Riina 1317	(MICH)	NA	HQ654595	NA	NA
Elateriospermum tapos	Soepadmo and Suhaimi s193	(NY)	NA	NA	AY794873	NA
Endospermum chinense		()	KP092925	NA	NA	NA
Glycerodendron amazonicum	Gillespie 4546	(US)	NA	NA	AY794876	NA
Hevea sp	Gillespie 4272	(US)	NA	NA	AY788175	NA
Hevea brasiliensis		()	KJ665775	NA	NA	NA
Jatropha alamanii	14-133	()	NA	NA	NA	8_2
Jatropha andrieuxii	B. Dehgan et al. B86.098	(FLAS)	2_1	2_1	2_1	NA
Jatropha andrieuxii	14-211	()	NA	NA	NA	23_4
Jatropha bartletii	B. Dehgan and B. Schutzman B86.084	(FLAS)	NA	NA	NA	364
Jatropha bullockii	B. Dehgan et al. B86.079	(FLAS)	2_3	2_3	2_3	19_2
Jatropha campestris	Arid lands	()	NA	NA	16_6	16_6
Jatropha canescens	38900	(hunt)	2_4	2_4	2_4	350
Jatropha capensis		()	5_8	5_8	NA	639
Jatropha cardiophylla3		(baboquiviris)	1_1	1_1	NA	NA
Jatropha cardiophylla4		(silverbells)	4_5	4_5	NA	NA
Jatropha cardiophylla1		()	NA	NA	NA	421
Jatropha cardiophylla2		()	NA	NA	NA	424
Jatropha cathartica		()	5_4	5_4	5-4	353
Jatropha chamelensis	B. Dehgan B86.081	(FLAS)	19_13	19_13	19_13	365
Jatropha ciliata	14-209	()	8_5	8_5	NA	23_5
Jatropha cinerea	44641	(hunt)	2_6	4_1	4_1	347
Jatropha clavuligera var. p	B. Dehgan and F. Almira B05.015	(FLAS)	4_2	4_2	4_2	NA
Jatropha clavuligera var. p	Arid lands	()	NA	NA	16-10	16_10
Jatropha conterasii	Hidalgo 1828	(MEXU)	NA	NA	NA	25_3
Jatropha conzattii		()	8_9	9_1	NA	642
Jatropha cordata	Dgh 57701	(hunt)	2_7	2_7	2_7	345
Jatropha costaricensis	G. L. Webster and L. J. Povada 22160	(DAV)	NA	NA	2_8	360
Jatropha cuneata	289(collection number)	(OPCNM)	1_2	1_2	1_2	NA
Jatropha cuneata	15-17	()	NA	NA	NA	22_1
Jatropha curcas		()	1_3	1_3	NA	19_17
Jatropha dehganii	Flores 209	(MEXU)	NA	NA	NA	25_4
Jatropha dhofarica	Arid lands	()	NA	NA	16_13	16_13
Jatropha dioica var. dioica		()	3_2	3_2	3_2	370
Jatropha dioica var. gracilis		()	NA	NA	NA	371

Table 1-3:Accession information for sequence data data used in phylogenetic analysis.

Jatropha elbae	Garcia s.n.	(MEXU)	NA	NA	NA	25_6
Jatropha ellenbeckii	Arid lands	()	NA	NA	16_12	16_12
Jatropha excisa var. veridiflora		()	5_7	5_7	NA	375
Jatropha excisa	B. Dehgan and F. Almira B05.020	(FLAS)	7_1	7_1	NA	NA
Jatropha fremontioides1	14-142	()	8_4	8_4	NA	23_2
Jatropha fremontioides2	14-141	()	NA	NA	NA	388
Jatropha galvanii1	Soto 19184	()	NA	NA	NA	25_7
Jatropha galvanii2	Flas 224859	()	NA	NA	NA	383
Jatropha gaumeri	B. Schutzman and T. Nance S-601A	(FLAS)	2_9	NA	2_9	376
Jatropha gossypiifolia var. jamacensis		()	7_3	7_3	NA	377
Jatropha gossypiifolia var. staphysagrifolia	G. Holstein and W. S. Armbruster 2033	(DAV)	NA	NA	NA	378
Jatropha grossidentata	B. Dehgan and F. Almira B05.004	(FLAS)	NA	7_4	NA	359
Jatropha hernandiifolia var. hernandiifolia	G. L. Webster and E. Miller 8830	(DAV)	7_5	7_5	7_5	366
Jatropha hernandiifolia var. portoricensis	B. Dehgan and Breckon B86.146	(FLAS)	7_6	7_6	NA	367
Jatropha hieronymii	42570	(hunt)	7_7	7_7	NA	346
Jatropha horizontalis	Arid lands	()	5_9	5_9	NA	NA
Jatropha humboldtiana	P. C. Hutchison and J. K. Wright 3610	(DAV)	NA	7_8	7_8	16_11
Jatropha integerima		()	NA	NA	NA	354
Jatropha integerima	G. L. Webster 4662	(DAV)	5_3	5_3	NA	
Jatropha integerima var. tupifolia	B. Dehgan B03.001	(FLAS)	7_9	7_9	NA	
Jatropha integerima var. integerima		()	7_10	7_10	NA	
Jatropha jaimejimenezii	Rojas 60	(MEXU)	NA	NA	NA	25_8
Jatropha macrantha2	53222	(hunt)	NA	NA	NA	348
Jatropha macrantha1	H. Luther s.n.	(FLAS)	4_3	4_3	4_3	NA
Jatropha macrocarpa	B. Dehgan and F. Almira B05.009	(FLAS)	4_4	4_4	NA	NA
Jatropha macrorhiza		()	NA	NA	NA	351
Jatropha mahafalensis		()	3_1	3_1	3_1	19_16
Jatropha malacophylla	Mine #?	(UA arboretum)	NA	7_12	NA	
Jatropha malacophylla	14-149	()	8_8	NA	NA	22_4
Jatropha martiusii	B. Dehgan and G. L. Webster B86.346	(FLAS)	NA	NA	NA	357
Jatropha mcvaughii	77317	(hunt)	2_10	2_10	2_10	349
Jatropha molissima	B. Dehgan and G. L. Webster B86.318	(DAV)	NA	NA	NA	361
Jatropha moranii	B. Dehgan B74.052	(DAV)	2_11	NA	NA	NA
Jatropha moranii	41253	(hunt)	3_4	3_4	3_4	355
Jatropha multifida		()	NA	5_2	NA	352
Jatropha mutabilis	B. Dehgan and G. L. Webster B76.312	(DAV)	7_15	7_15	7_15	362
Jatropha neopauciflora	14-212	()	NA	NA	NA	23_3
Jatropha nudicaulis	H. Luther s.n.	(FLAS)	5_10	5_10	NA	373
Jatropha oaxacana	14-154	()	NA	NA	NA	641
Jatropha pereziae	E. J. Lott 1146	(FLAS)	19_3	2_12	2_12	356
Jatropha platyphylla	42566	(hunt)	NA	NA	NA	344
Jatropha pelargonifolia	Arid lands	()	NA	NA	16_3	16_3
Jatropha podagrica		()	5_1	5_1	NA	638
Jatropha pseudocurcas1	Salas 5239	(MEXU)	NA	NA	NA	380
Jatropha pseudocurcas2	Ismael 18649	(MEXU)	NA	NA	NA	382
Jatropha spathulata	28210	(DBG)	7_16	7_16	NA	NA
Jatropha stephani1	Mine #?	()	NA	NA	NA	368
Jatropha stephani2	Mine #?	()	NA	NA	NA	369
Jatropha standleyi	B. Dehgan B86.058	(FLAS)	2_13	2_13	NA	NA
Jatropha sympetala1	14-139	()	NA	NA	NA	640
Jatropha sympetala2	MacDougal/King H446/1391	(?)	2_14	2_14	2_14	358
Jatropha tehuantepecana	Salinas 8207	(MEXU)	NA	NA	NA	379
Jatropha unicostata	Arid lands	()	NA	NA	16_8	16_8
Jatropha varifolia	Arid lands	()	NA	NA	16_4	16_4
Jatropha vernicosa		()	3_3	3_3	3_3	19_18
Jatropha wedelliana	Webster et al. 25316	(DAV)	7_18	19_15	NA	363
Joannesia princeps	unvouchered	(UC Berkley s.n.)	19_1	19_1	19_1	19_1
Joannesia princeps	Chase 1262	(K)	NA	NA	AJ418808	NA
Manihot esculenta	unvouchered	(NA)	JQ743203	NA	NA	NA
Manihot grahamii	Wurdack s.n.	(US)		FJ669761		NA
Manihot grahamii		()		NA	AY794875	NA
Manihot sp.		()			NA	NA
Micrandra inundata	Berry 6350	(MO)	NA	NA	AY794877	NA
Ophellantha spinosa	Breedlove 46994	(NY)	AY971263	NA	NA	NA
Sandwithia guyenensis	Ek et al. 906	(NY)	NA	NA	AY794904	NA
Sagotia racemosa	Smith 253	(US)	AY971264	NA	AY794903	NA
Suregada	Rakotomalaza et al. 1292	(MO)	NA	NA	AY788189	NA
Suregada eucleoides	D. Harder et al. 1569	(MO)	DQ006007	NA	NA	NA
Suregada glomerata	Chase 1272	(K)	NA	FJ669770	NA	NA
Tetrorchidium macrophyllum	Bell et al. 93-204	(US)	NA	FJ669771	NA	NA

Table 1-3 (continued).

Primer	Sequence (5’ to 3’)	gene	genomic compartment	Source
1F	ATGAGTTGTAGGGAGGGACT	rbcL	chloroplast	Wurdack, 2002
1460R	TCCTTTTAGTAAAAGATTGGGCCGAG	rbcL	chloroplast	Wurdack, 2002
EMB 2765 ex 9 F	TAAATGCTCTTCCWCTTTATCCAAATGA	Exon 9 EMB2765	nucleus	Wurdack and Davis, 2009
EMB 2765 ex 9 R	TTGGTCCAYTGTGCWGCAGAAGGRT	Exon 9 EMB2765	nucleus	Wurdack and Davis, 2009
ITS1	GTCCACTGAACCTTATCATTTAG	ITS1 - 5.8S RNA – ITS2	nucleus	Urbatsch et al., 2000
ITS4	TCCTCCGCTTATTGATATGC	ITS1 - 5.8S RNA – ITS2	nucleus	White et al., 1990

Table 1-4: PCR primers used to generate sequences data.

```
------- ipyrad params file v0.7.21)--------------------
kept                    ## [0] [assembly_name]: Used to name output directories for assembly steps
./                      ## [1] [project_dir]: Project dir (made in curdir if not present)
                        ## [2] [raw_fastq_path]: Location of raw non-demultiplexed fastq files
                        ## [3] [barcodes_path]: Location of barcodes file
./reads/*_R1_.fq.gz     ## [4] [sorted_fastq_path]: Location of demultiplexed/sorted fastq files
denovo                  ## [5] [assembly_method]: Assembly method
                        ## [6] [reference_sequence]: Location of reference sequence file
ddrad                   ## [7] [datatype]: Datatype (see docs): rad, gbs, ddrad, etc.
CATCG, AATTC            ## [8] [restriction_overhang]: Restriction overhang (cut1,) or (cut1, cut2)
7                       ## [9] [max_low_qual_bases]: Max low quality base calls (Q<20) in a read
26                      ## [10] [phred_Qscore_offset]: phred Q score offset
6                       ## [11] [mindepth_statistical]: Min depth for statistical base calling
6                       ## [12] [mindepth_majrule]: Min depth for majority-rule base calling
99999999                ## [13] [maxdepth]: Max cluster depth within samples
0.85/0.88/0.90          ## [14] [clust_threshold]: Clustering threshold for de novo assembly
0                       ## [15] [max_barcode_mismatch]: Allowable mismatches in barcodes
2                       ## [16] [filter_adapters]: Filter for adapters/primers (1 or 2=stricter)
35                      ## [17] [filter_min_trim_len]: Min length of reads after adapter trim
4                       ## [18] [max_alleles_consens]: Max alleles per site in consensus sequences
14, 14                  ## [19] [max_Ns_consens]: Max N's (uncalled bases) in consensus (R1, R2)
16, 16                  ## [20] [max_Hs_consens]: Max Hs (heterozygotes) in consensus (R1, R2)
4/8/16/32               ## [21] [min_samples_locus]: Min # samples per locus for output
140                     ## [22] [max_SNPs_locus]: Max # SNPs per locus (R1, R2)
140                     ## [23] [max_Indels_locus]: Max # of indels per locus (R1, R2)
1.0                     ## [24] [max_shared_Hs_locus]: Max # heterozygous sites per locus (R1, R2)
0, 0, 0, 0              ## [25] [trim_reads]: Trim raw read edges (R1>, <R1, R2>, <R2) (see docs)
0, 0, 0, 0              ## [26] [trim_loci]: Trim locus edges (see docs) (R1>, <R1, R2>, <R2)
*                       ## [27] [output_formats]: Output formats (see docs)
                        ## [28] [pop_assign_file]: Path to population assignment file
```

Figure 1-2: Parameters file for assembly of RADseq datasets in iPyrad.

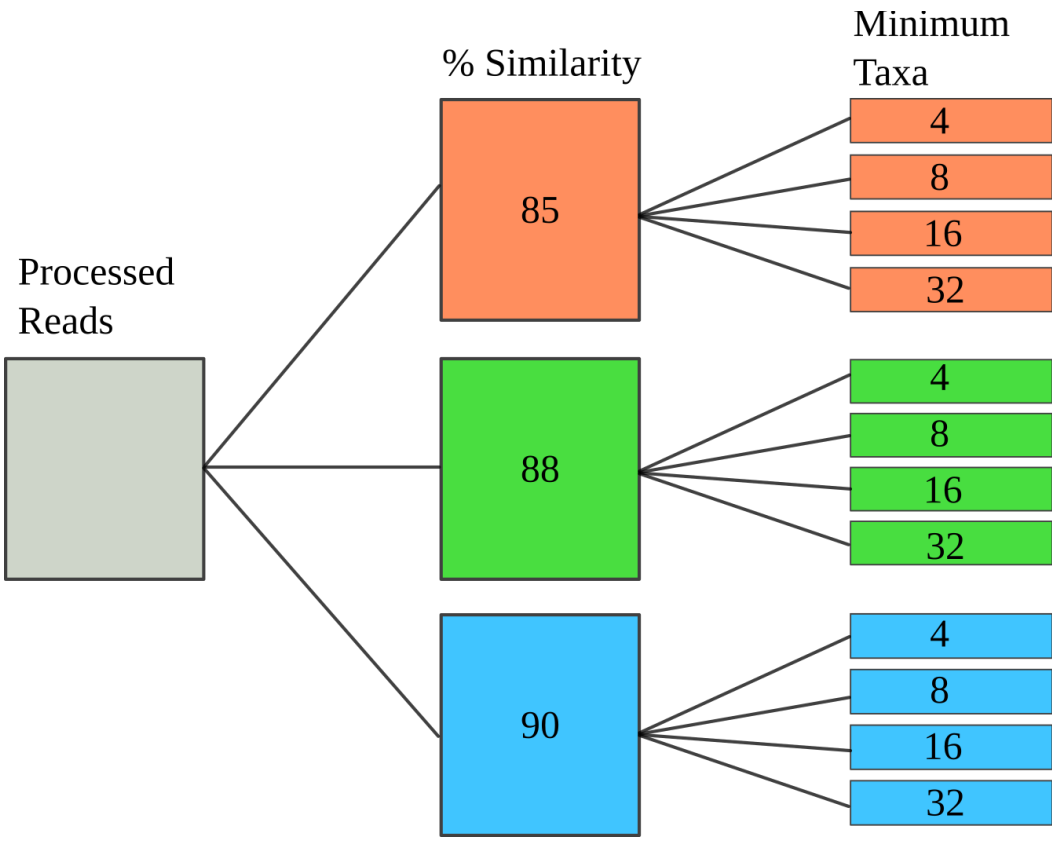


Figure 1-3: Branching strategy used to generate datasets with different assembly parameters in iPyrad.

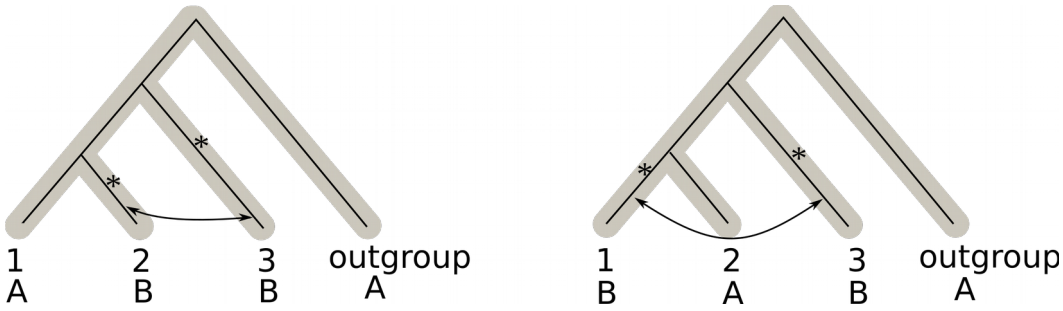


Figure 1-4: Two quartets showing how gene flow between species 1 and 3 or 2 and 3 could produce alternative shared SNP patterns from what would be expected without gene flow. Letters at the bottom of the figure indicate the state for a sampled SNP: ‘A’ = ancestral state and ‘B’ = derived. Looking at the quartet on the left and starting at the root with an ancestral state ‘A’, the derived trait ‘B’ arises via mutation (asterisk) in one of the branches leading to either species 2 or 3. Subsequent gene flow (arrow) leads to the derived state being shared by both species. The quartet on the right shows the same process for species 1 and 3.

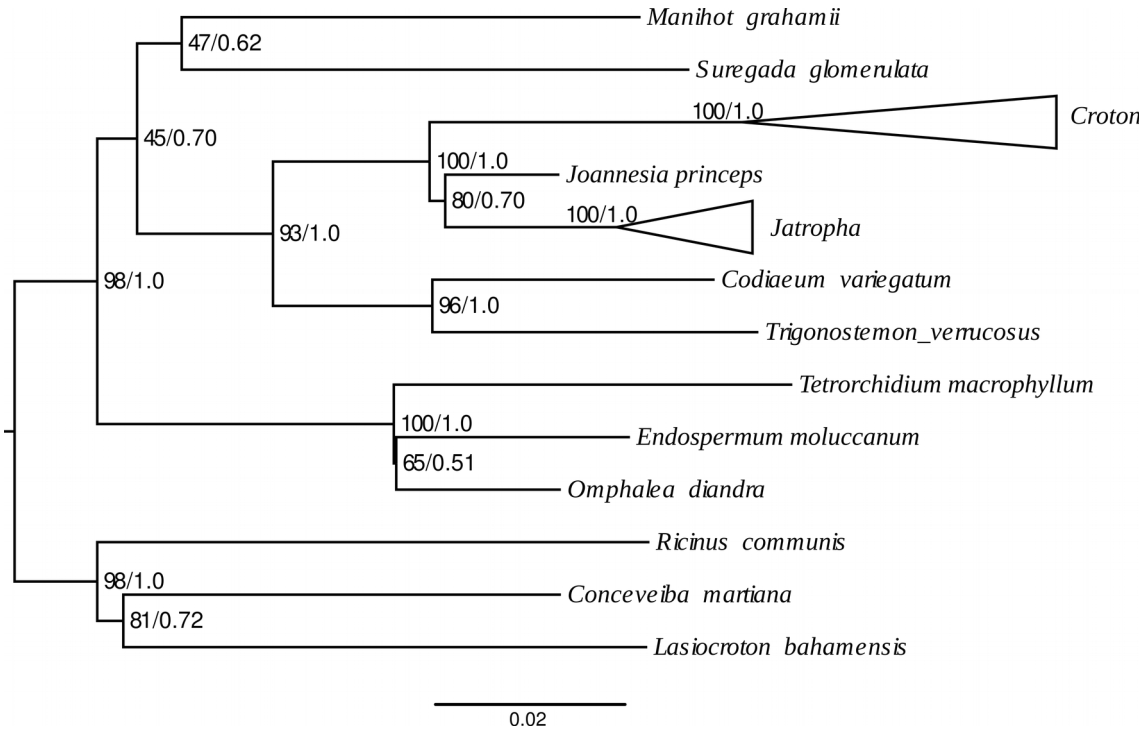


Figure 1-5: Highest scoring maximum likelihood phylogeny from RAxML analysis of the low copy nuclear marker EMB2765. Values on branches are bootstrap scores followed by Bayesian posterior probabilities.

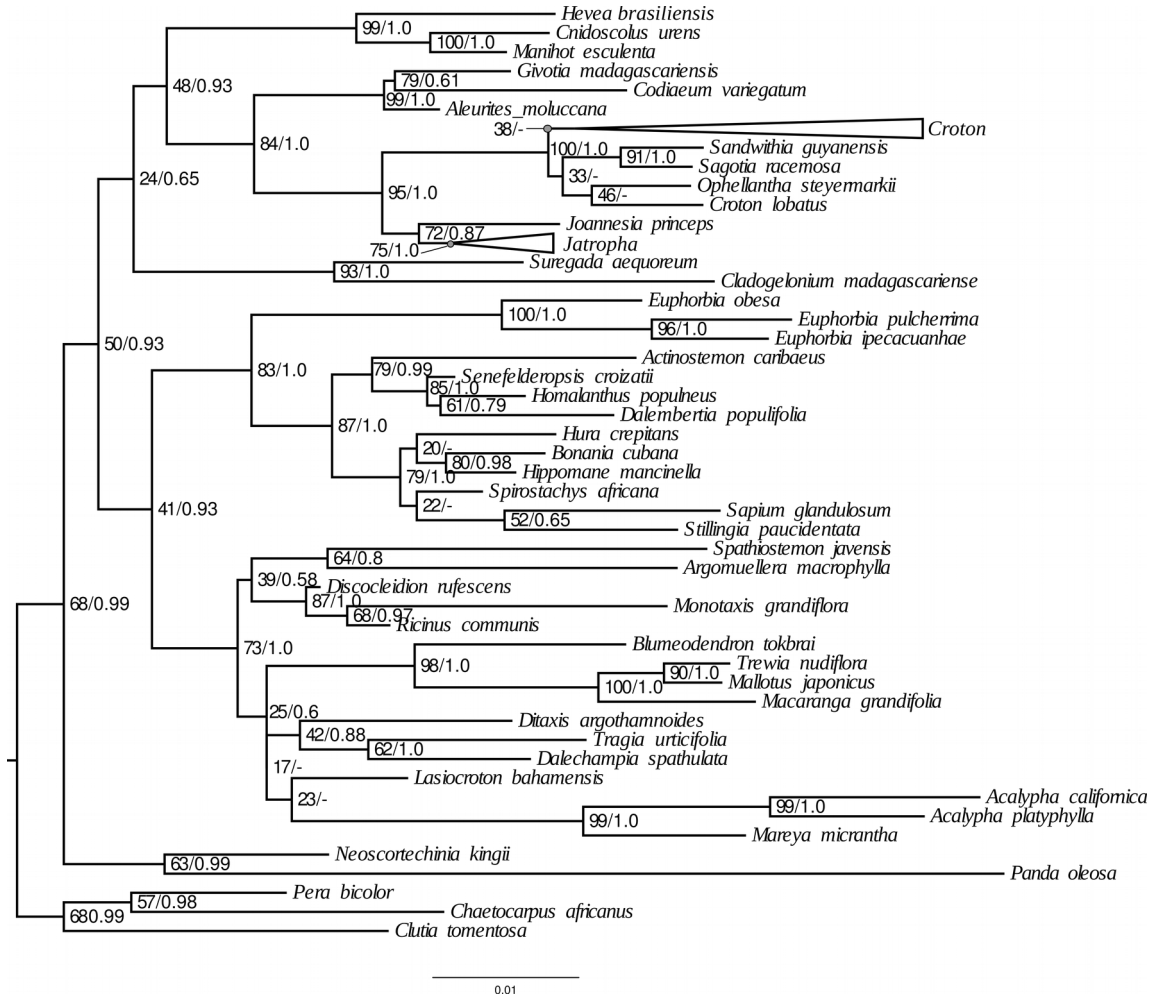


Figure 1-6: Best tree resulting from maximum likelihood analysis of chloroplast marker *rbcL*. Support values at nodes are non-parametric bootstrap scores (n=500) followed by Bayesian posterior probabilities. A "-" indicates a clade not recovered in Bayesian analysis

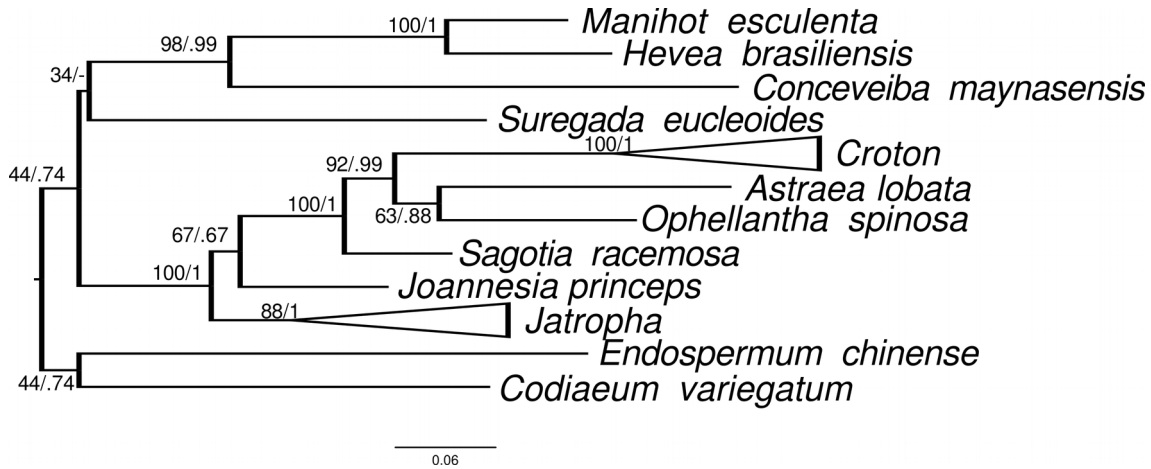


Figure 1-7: Best tree resulting from maximum likelihood analysis of nuclear ribosomal markers *ITS1*, *ITS2*, and *5.8S rRNA*. Support values at nodes are non-parametric bootstrap scores followed by Bayesian posterior probabilities. A "-" indicates a clade not recovered in Bayesian analysis

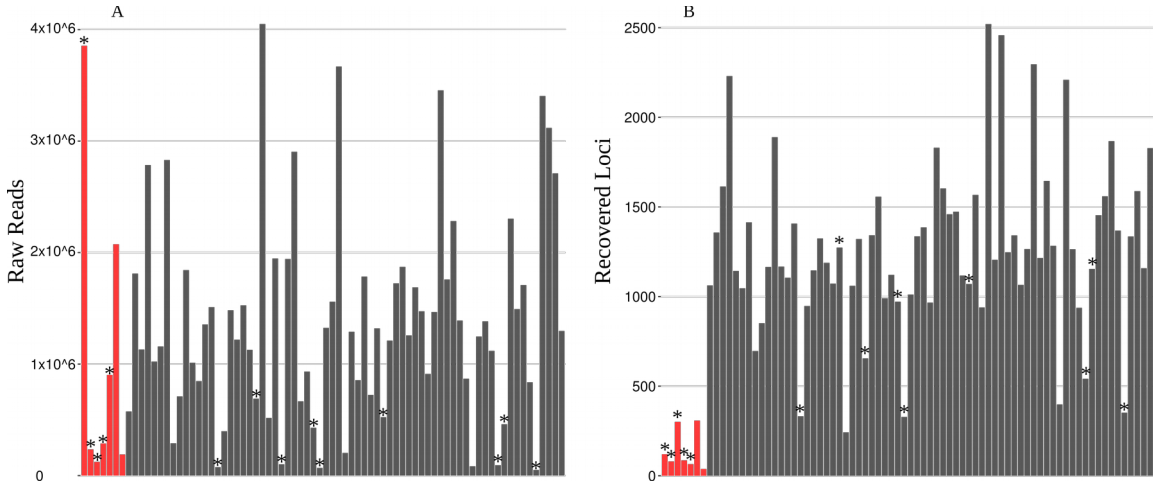


Figure 1-8: Histograms showing: A) the number of raw Illumina reads across all samples and B) the number of loci assembled in *ipyrad* (85min4). Outgroup taxa are shown in red and herbarium samples have an asterisks above the bar.

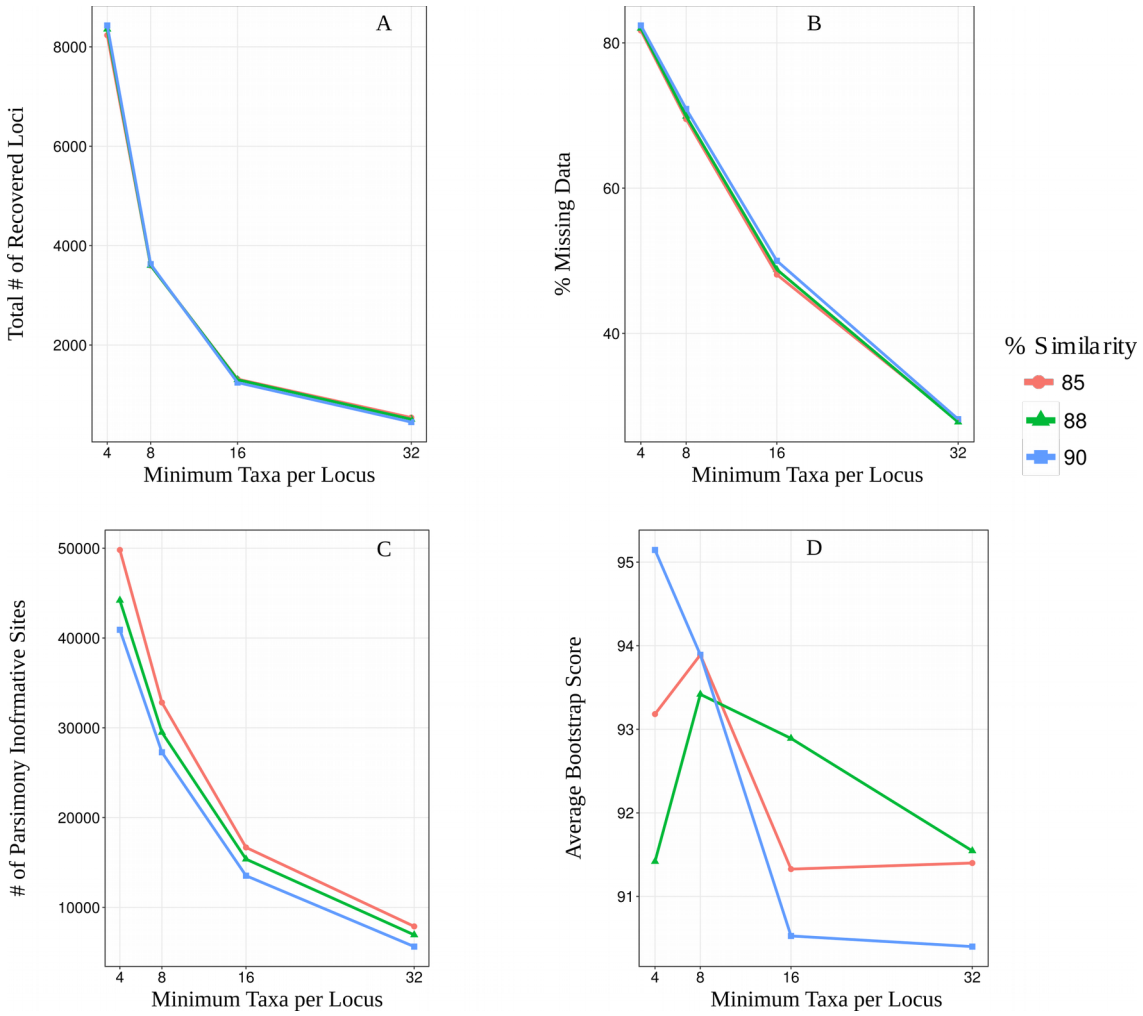


Figure 1-9: Characteristics of 12 RADseq data sets assembled from forward reads in *ipyrad* using 3 similarity clustering thresholds and 4 levels for minimum taxa: A-number of recovered loci, B-number of parsimony informative characters, C)-percent missing data, and D-average bootstrap scores (500 replicates) across branches of the highest scoring maximum likelihood phylogeny inferred for each dataset in RAxML.

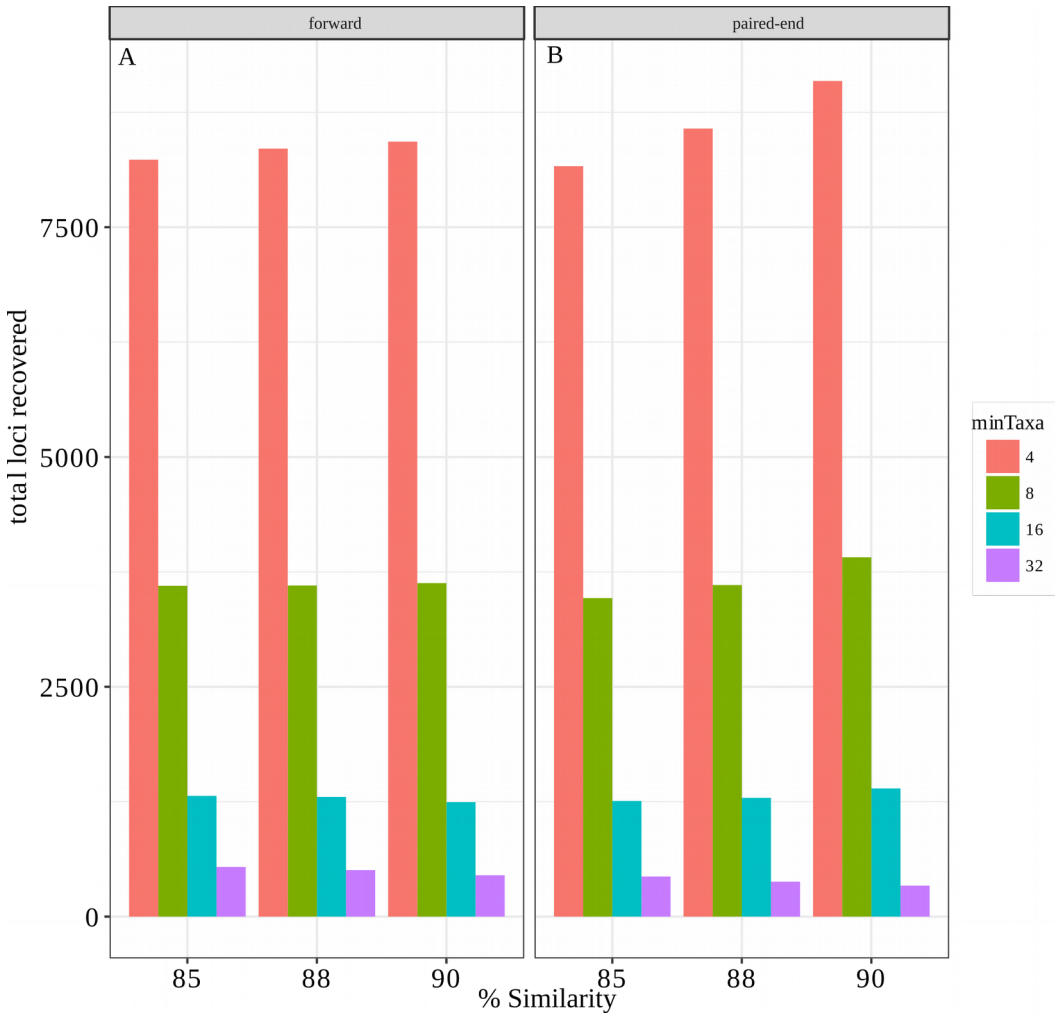


Figure 1-10: Total recovered loci in RADseq assemblies using: A) forward reads only and B) Paired-end reads

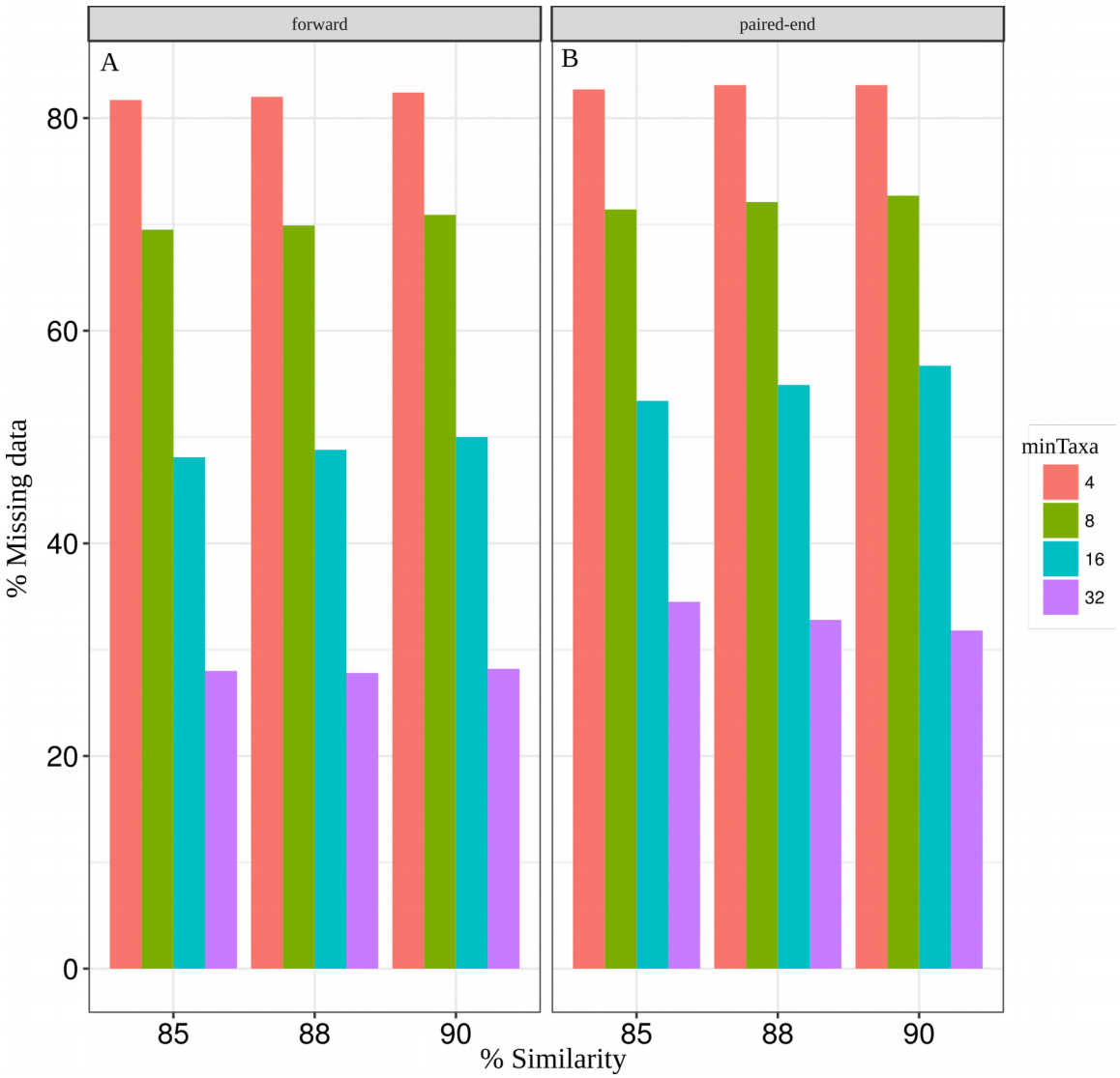


Figure 1-11: Percentage of missing data in RADseq assemblies using: A) forward reads only and B) paired-end reads

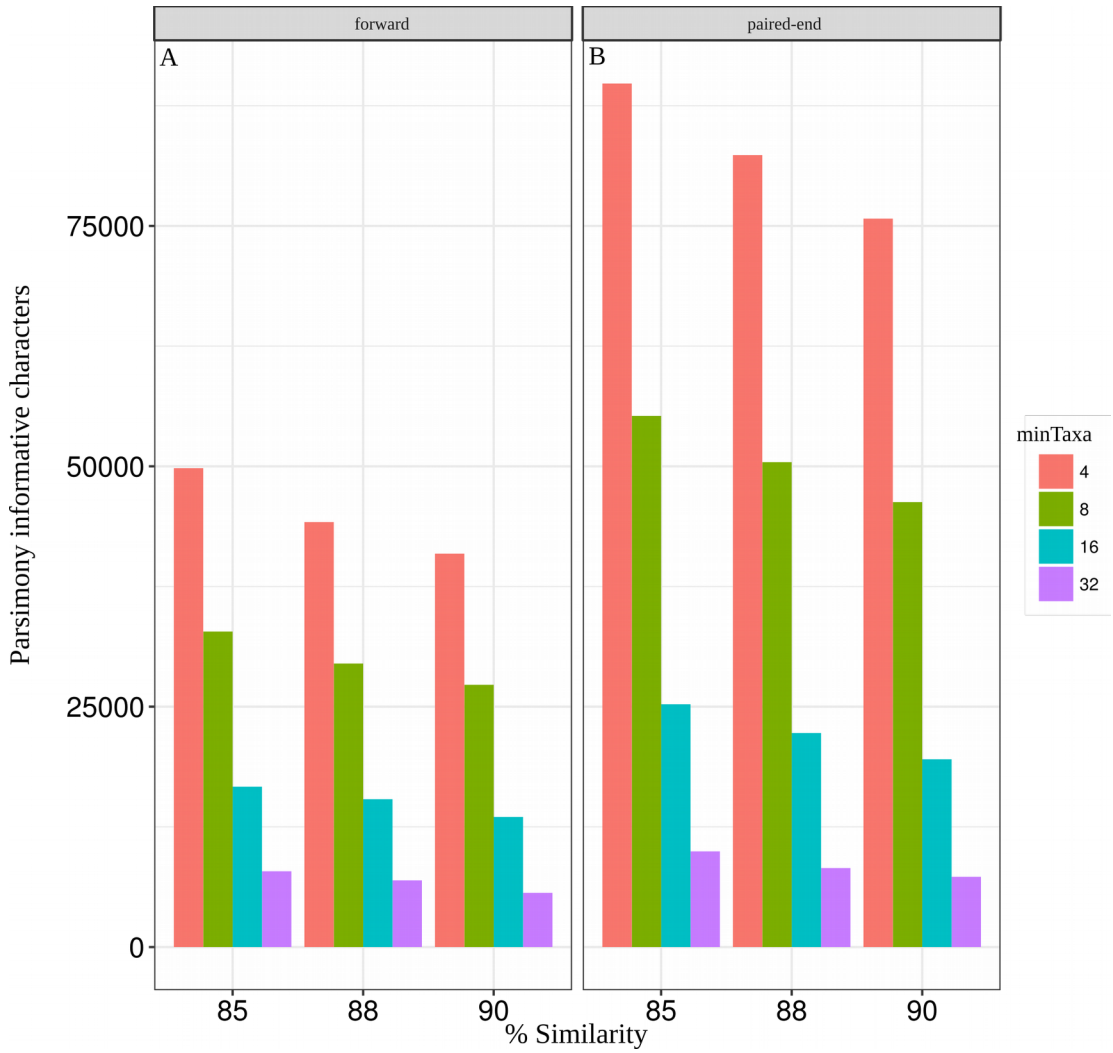


Figure 1-12: Total parsimony informative characters recovered in RADseq assemblies using: A) forward reads only and B) paired-end reads

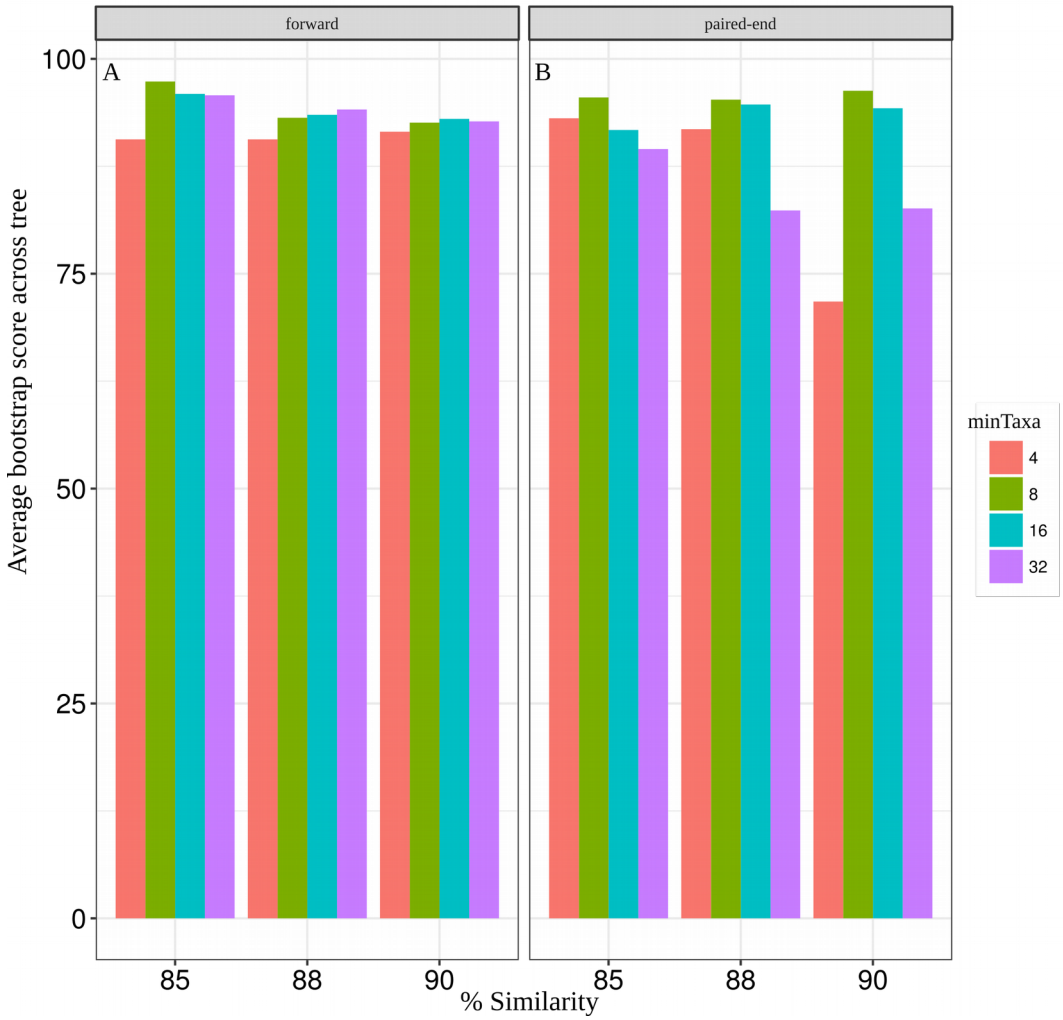


Figure 1-13: Average bootstrap scores across all branches of best scoring maximum likelihood trees from RAxML for RADseq assemblies produced using: A) forward reads only and B) paired-end reads

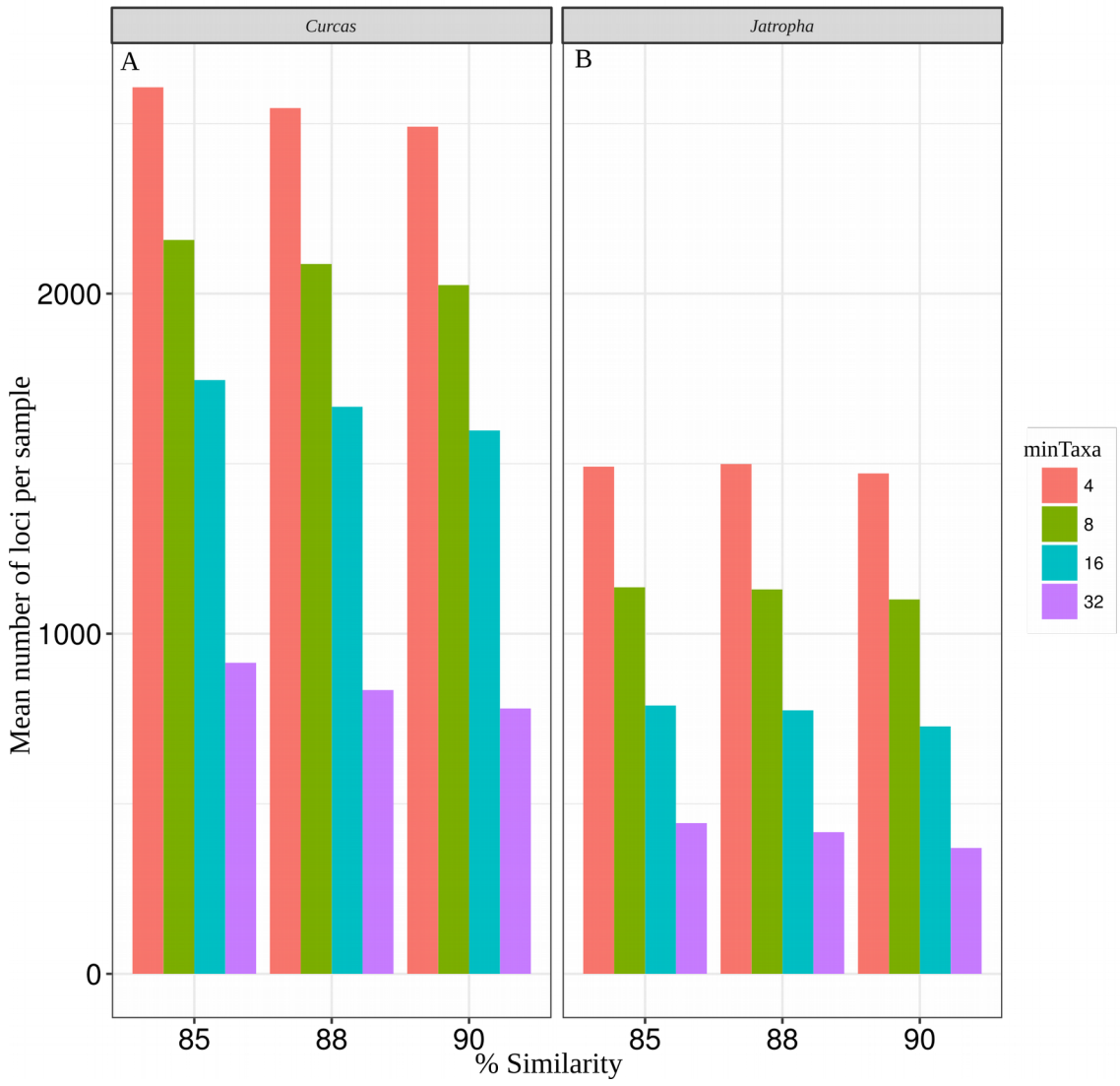


Figure 1-14: Mean number of recovered loci per sample in RADseq assemblies using: A) a subset of taxa from *J.* subg. *Curcas* and B) the full dataset of *Jatropha*

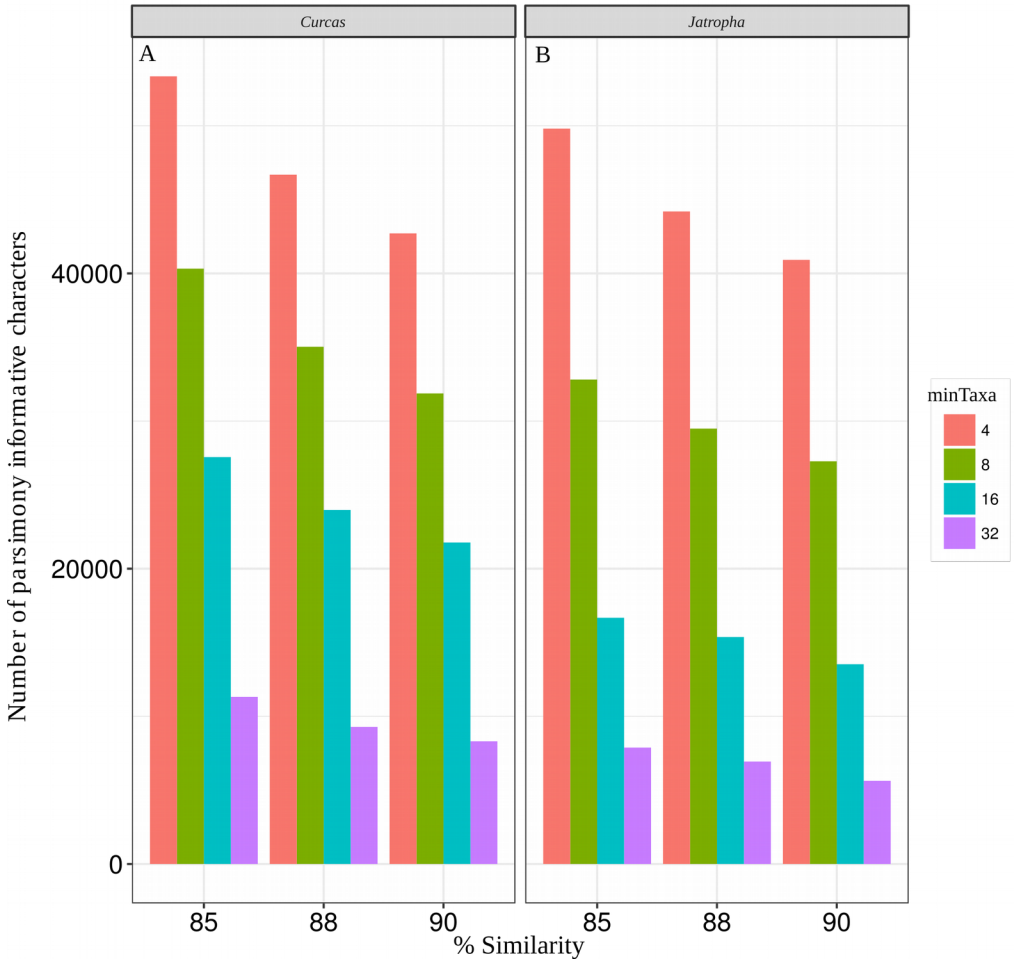


Figure 1-15: Number of parsimony informative characters in RADseq assemblies using: A) a subset of taxa from *J.* subg. *Curcas* and B) the full dataset of *Jatropha*

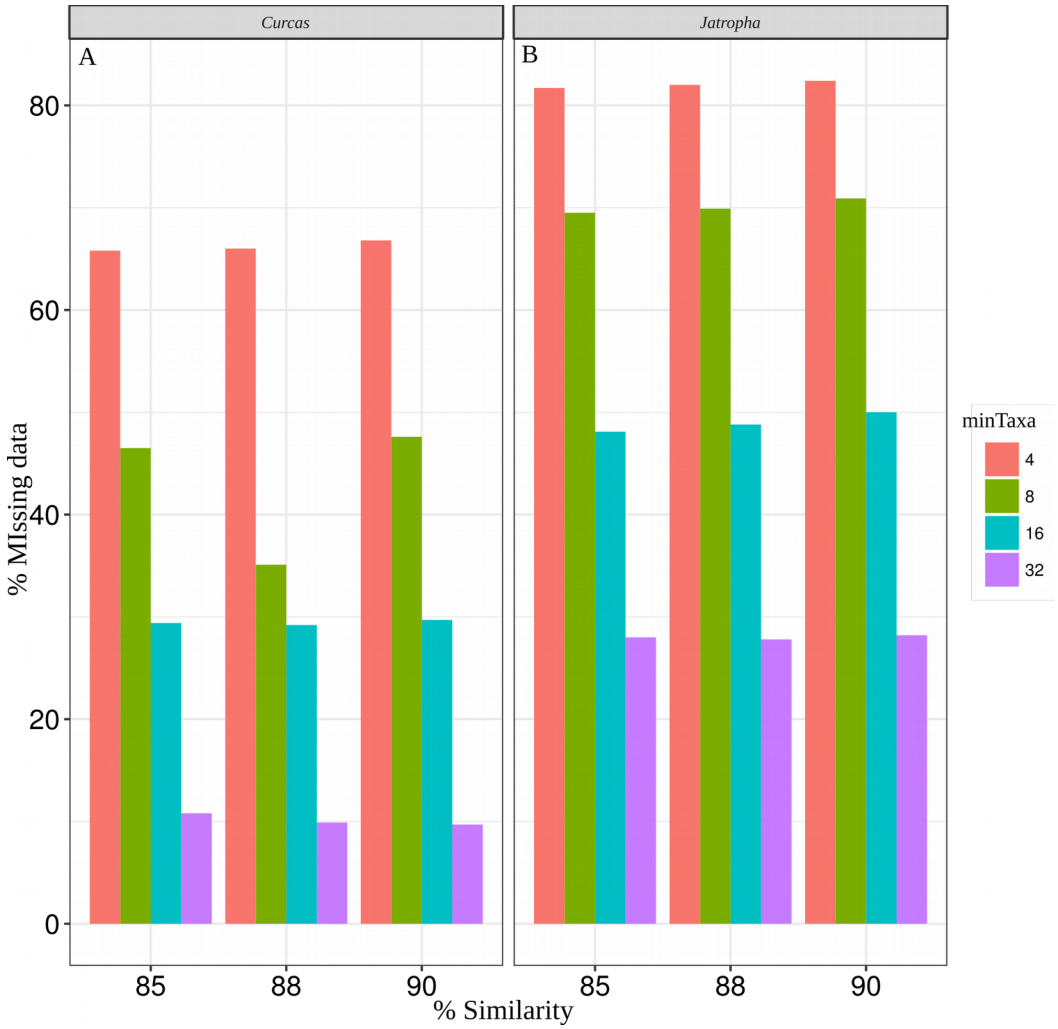


Figure 1-16: Percentage of missing data in RADseq assemblies using: A) a subset of taxa from *J.* subg. *Curcas* and B) the full dataset of *Jatropha*

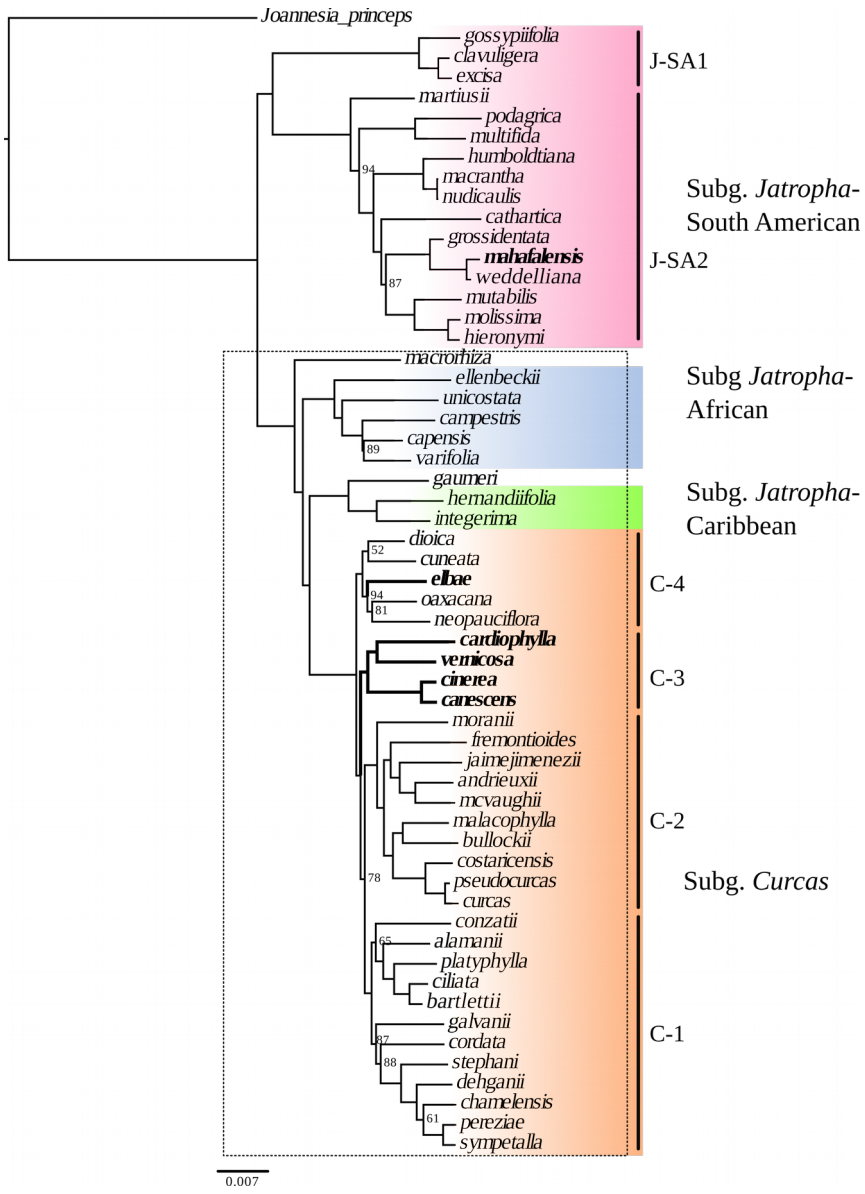


Figure 1-17: Best tree from RAxML analysis of concatenated RADseq supermatrix (85min16). Support values shown on branches are non-parametric bootstrap scores < 100. Colored boxes and labels mark clades discussed in the text, and the dashed box indicates the species used to assemble the *Curcas* datasets. Bold species/clades were recovered in alternative positions in analyses of other datasets, and were tested for introgression.

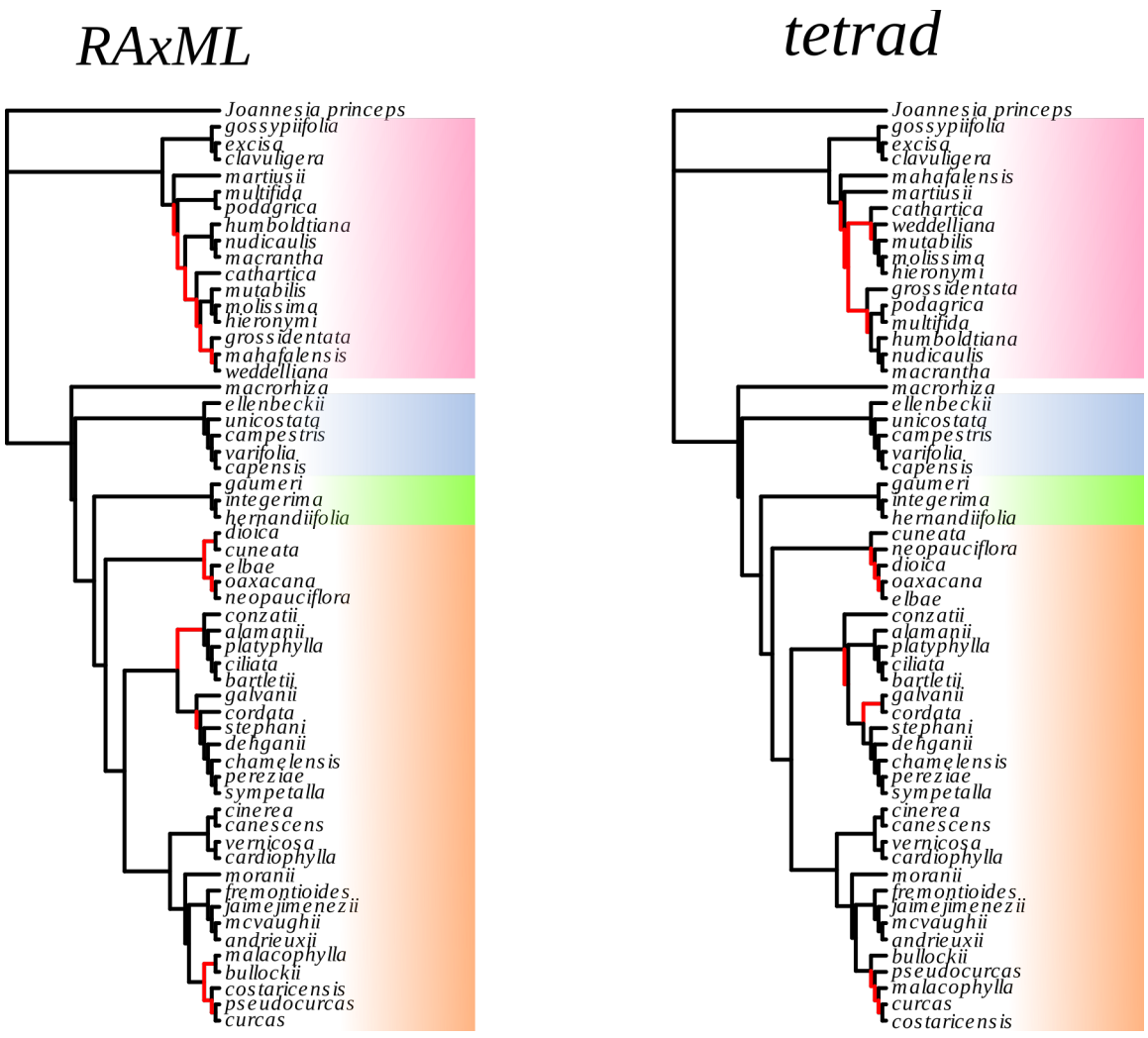


Figure 1-18: Comparison of consensus topologies from phylogenetic analyses of all 12 datasets assembled of forward reads using maximum likelihood (left) and coalescence (right) methods. Clades in conflict are shown in red. Colors show major clades discussed in the text (subclades not shown).

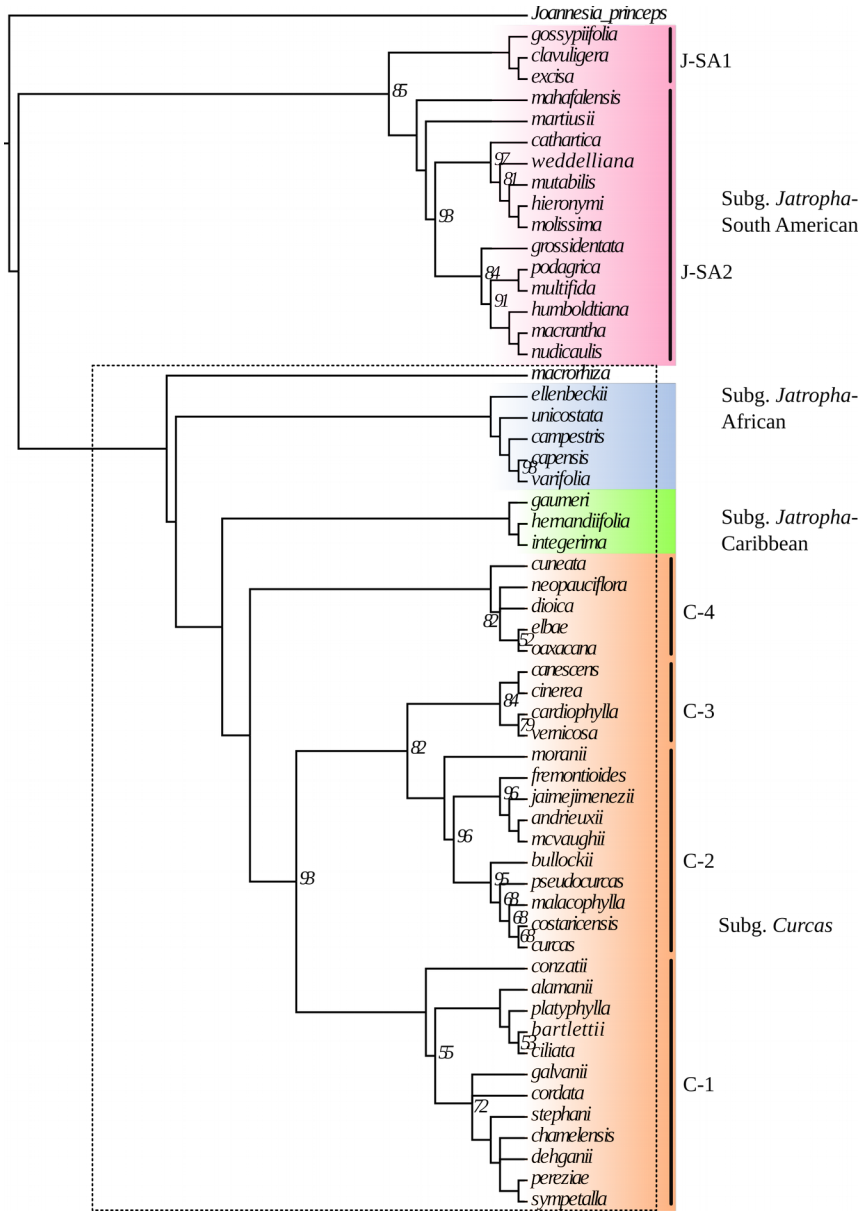


Figure 1-19: Majority rule consensus topology from coalescence analysis of the 85%SimminTaxa16 forward read alignment in Tetrad. Support values are bootstrap scores (500 replicates) and colored boxes and names along right-hand side correspond to major clades discussed in the text. Collapsed branches received BS values <50, whereas bifurcating branches with no score received BS = 100.

Clade	Species	Previous taxonomic assignment: Section, Subsection (if applicable)
C1	*J. conzatii*	Loureira, Loureira
	J. alamanii	Platyphyllae, Platyphyllae
	J. platyphylla	Platyphyllae, Platyphyllae
	J. ciliata	Loureira, Loureira
	J. bartletii	Platyphyllae, Platyphyllae
	J. galvanii	Loureira, Loureira
	J. cordata	Loureira, Loureira
	J. stephani	Platyphyllae, Platyphyllae
	J. dehganii	Loureira, Loureira
	J. chamelensis	Platyphyllae, Platyphyllae
	J. pereziae	Platyphyllae, Platyphyllae
	J. sympetalla	Loureira, Loureira
C2	*J. moranii*	Platyphyllae, Fremontioides
	J. fremontioides	Platyphyllae, Fremontioides
	J. jaimejimenezii	Platyphyllae, Fremontioides
	J. andrieuzii	Platyphyllae, Platyphyllae
	J. mcvaughii	Curcas
	J. malacophylla	Curcas
	J. bullockii	Platyphyllae, Fremontioides
	J. costaricensis	Platyphyllae, Platyphyllae
	J. pseudocurcas	Platyphyllae, Platyphyllae
	J. curcas	Curcas
C3	*J. cardiophylla*	Mozinna
	J. vernicosa	Loureira, Loureira
	J. cinerea	Loureira, Canescentes
	J. canescens	Loureira, Canescentes
C4	*J. dioica*	Mozinna
	J. cuneata	Mozinna
	J. elbae	Mozinna
	J. oaxacana	Loureira, Canescentes
	J. neopauciflora	Loureira, Canescentes
J-CB	*J. gaumeri*	Platyphyllae, Gaumeri
	J. hernandiifolia	Polymorphae, Polymorphae
	J. integerima	Polymorphae, Polymorphae
J-AF	*J. ellenbeckii*	Spinosae
	J. unicostata	Tuberosae, Tuberosae
	J. campestris	Tuberosae, Tuberosae
	J. capensis	Tuberosae, Capensis
	J. varifolia	Tuberosae, Capensis
J-SA1	*J. gossypiifolia*	Jatropha, Jatropha
	J. clavuligera	Jatropha, Jatropha
	J. excisa	Jatropha, Jatropha
J-SA2	*J. martiusii*	Martiusae
	J. podatriga	Peltatae, Multifidae
	J. multifida	Peltatae, Multifidae
	J. humboldtiana	Peltatae, Peltatae
	J. macrantha	Peltatae, Peltatae
	J. nudicaulis	Peltatae, Peltatae
	J. cathartica	Peltatae, Multifidae
	J. grossidentata	Peltatae, Peltatae
	J. mahafalensis	Subgenus Manihotoides
	J. weddelliana	Peltatae, Peltatae
	J. mutabilis	Peltatae, Peltatae
	J. molissima	Peltatae, Peltatae
	J. hieronymi	Peltatae, Peltatae

Table 1-5: List of clades recovered from phylogenetic analyses of RADseq datasets with the previous taxonomic assignments of species within each clade. Letters in the left hand columns signify subgenera (J = *Jatropha* and C = *Curcas*) and geographic region (CB = Caribbean, AF = African, and SA = South American).

	ML – full	ML – *Curcas*	Tetrad – full	Tetrad – *Curcas*
Jatropha subg. *Curcas*	100 (100%)	100 (100%)	100 (100%)	100 (100%)
C1-content	97-100 (100%)	93-100 (100%)	64-100 (100%)	100 (100%)
C2-content	66-100 (100%)	100 (100%)	100 (100%)	100 (100%)
C3-content	76-100 (100%)	94-100 (100%)	67-88 (100%)	63-100 (100%)
C4-content	66-100(75%)	94-100 (100%)	98-100 (100%)	89-100 (100%)
C3(C2,C1)	66-100 (41.6%)	88-100 (50%)	0 (0%)	0 (0%)
C1(C2,C3)	32-97 (58.3%)	40-96 (50%)	65-82 (100%)	60-90 (100%)
Caribbean-content	100 (100%)	-	100 (100%)	-
Caribbean-position	100 (100%)	-	99-100 (100%)	-
African-content	100 (100%)	-	100 (100%)	-
African-position	80-100 (100%)	-	100 (100%)	-
J-SA1-content	100 (100%)	-	100 (100%)	-
J-SA2-content	100 (100%)	-	99-100 (100%)	-
Remainder (J-SA1, J-SA2)	97-100 (100%)	-	83-91 (66.7%)	-
J-SA1(J-SA2, Remainder)	0 (0%)	-	<50 (33.3%)	-

Table 1-6: Summary of the major clades recovered from phylogenetic analyses of RADseq datasets and the range of bootstrap scores for each clade (and percentage of all datasets supporting that clade's position) depending on the dataset and reconstruction methods used. The first two columns show results from ML analysis in RAxML on the full dataset of *Jatropha* samples and the taxonomically reduced set of just *J.* subg. *Curcas* and closely related species. The second two columns are the same but analyzed via coalescent methods in Tetrad.

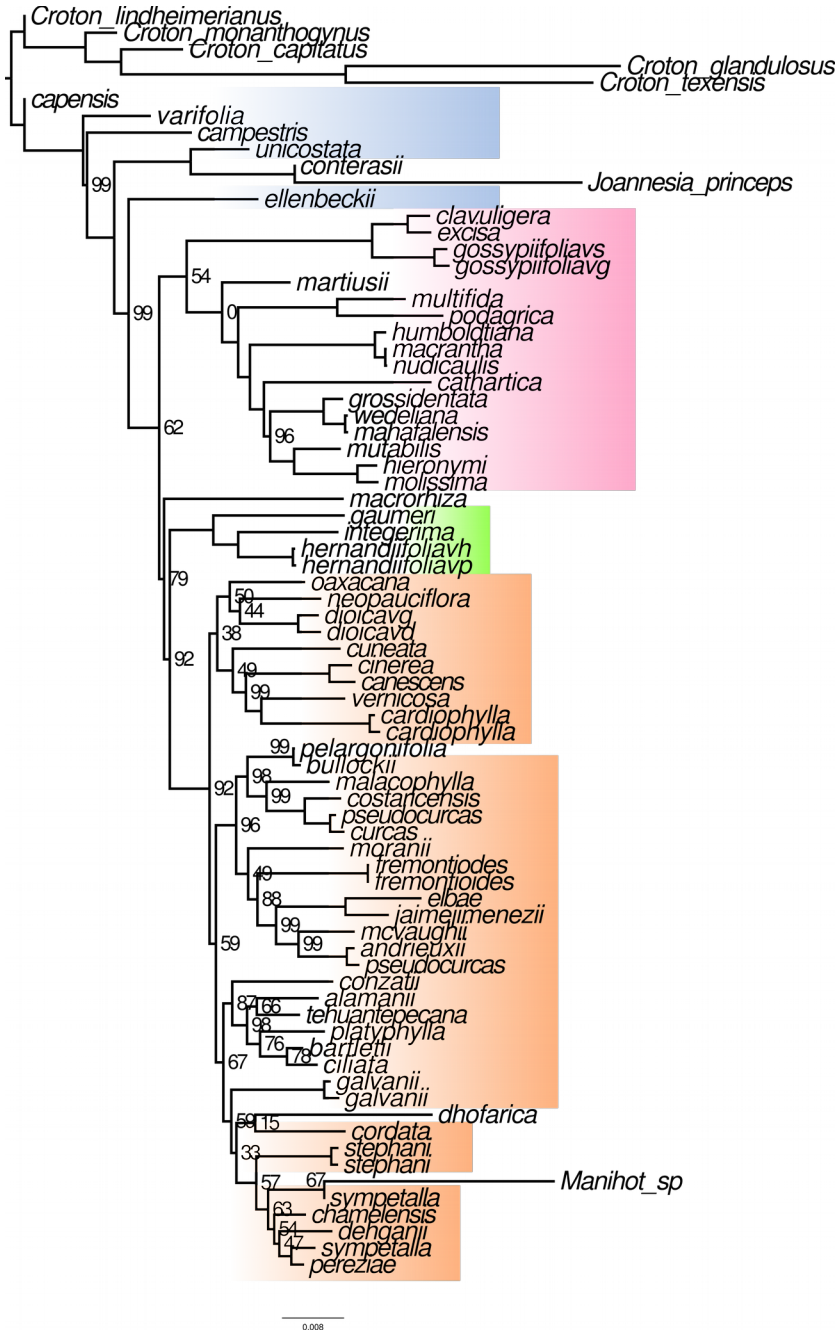


Figure 1-20: Best tree from ML analysis of 85min4 alignment of forward-reads from all samples. Colors denote major clades discussed in text. Values on branches are BS scores (100 replicates); BS = 100 not shown.

Clade	Topology	Method	Full-forward	Curcas-forward	Full-paired-end
J. elbae	T1-C4	RAxML	All-others	All	all m16 & m32
	T2-C2	RAxML	85m4,85m8,88m4	None	all m4 & m8
	T1-C4	Tetrad	All	All	
	T2-C2	Tetrad	None	None	
Sister to Curcas	T1-J.macrorhiza	RAxML	90m4	NA	85m4,85m8,88m4,90m4
	T2-African	RAxML	All-others	NA	all others
	T1-J.macrorhiza	Tetrad	None	NA	
	T2-African	Tetrad	All	NA	
J. mahafalensis	T1-Deeply nested	RAxML	All	NA	all others
	T2-Basal	RAxML	None	NA	85m32,88m32,90m32
	T1-Deeply nested	Tetrad	None	NA	
	T2-Basal	Tetrad	All	NA	
C4 relationships	(dio,cun,(oax,neo))	RAxML	85m4,88m4		
	(cun,(dio,(neo,oax)))	RAxML	85m8		
	(cun,dio),(elb,(oax,neo	RAxML	85m16,85m32,88m8,16,32, 90m32		
	cun+others	RAxML	90m4,90m8,90m16		
	cun+others	Tetrad	85m4,85m32, 88m*,90m*		
	cun+elb,oax+others	Tetrad	85m8,85m16		

Table 1-7:Summary of the disagreements among datasets that produce strongly supported alternative topologies.

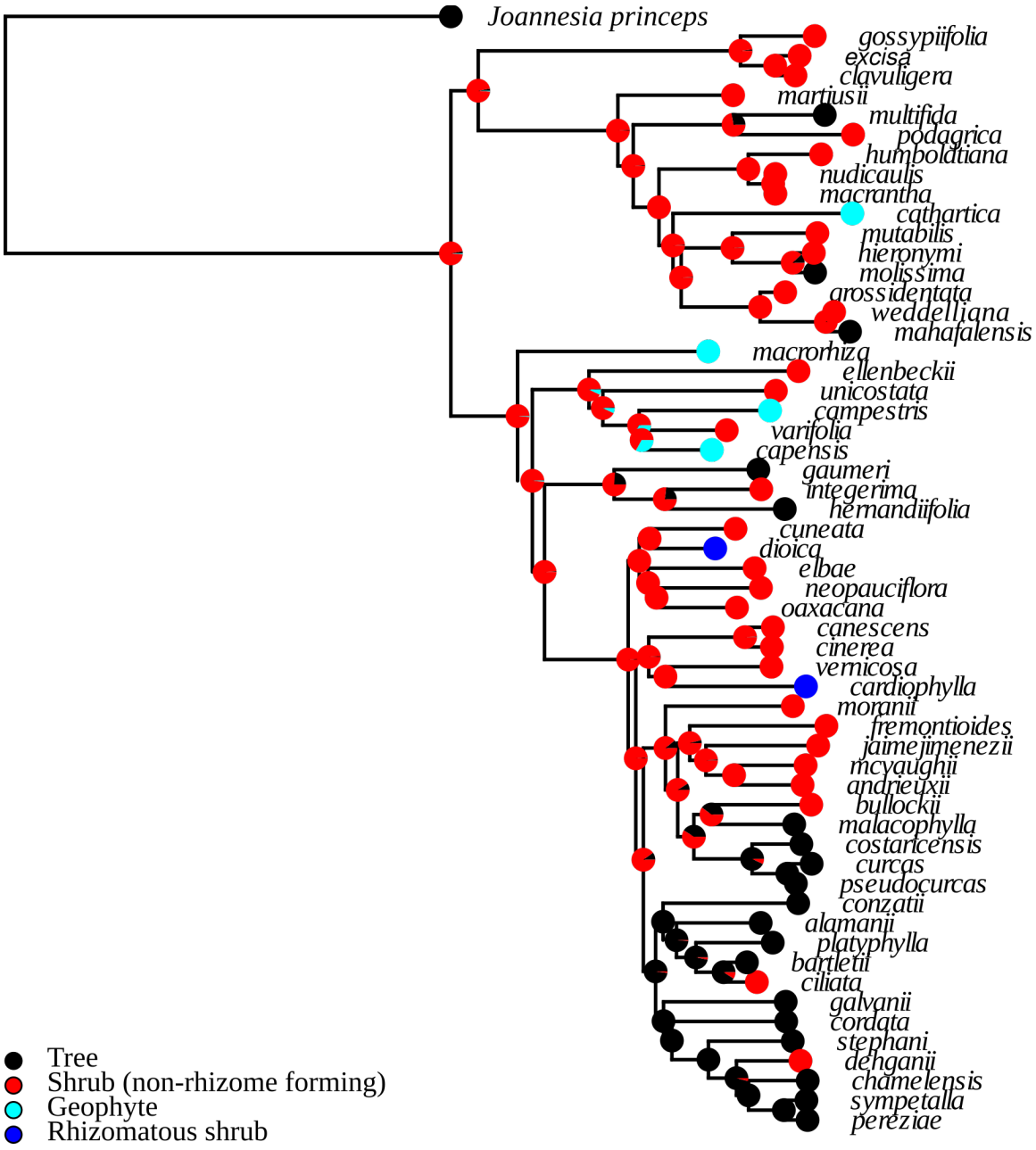


Figure 1-21: Ancestral reconstruction of growth habit (intermediate forms between trees and shrubs coded as trees) in *Jatropha* using stochastic character mapping. Ancestral states are the average posterior probabilities summarized from 100 stochastic maps.

Character	Transition Matrix					
		Shrub → tree	Tree → shrub	Shrub → geophyte	Tree → geophyte	
Habit 1	Symetrical	8	2	6	0	
Habit 2	Equal rates	6	1	6	0	
Carpels	All rates different	3 → 2	2 → 1	2 → 3	3 → 1	
		6	5(4)	1	0	
Anthers	Equal rates	8→6	8 → 10	10 → 8	10 → 5	10 → 6
		1	2	1	1	1

Table 1-8: Number of transitions between character states of three morphological traits reconstructed for *Jatropha* using stochastic character mapping. Two habit codings were used where intermediate forms were coded as trees (habit 1) or shrubs (habit 2).

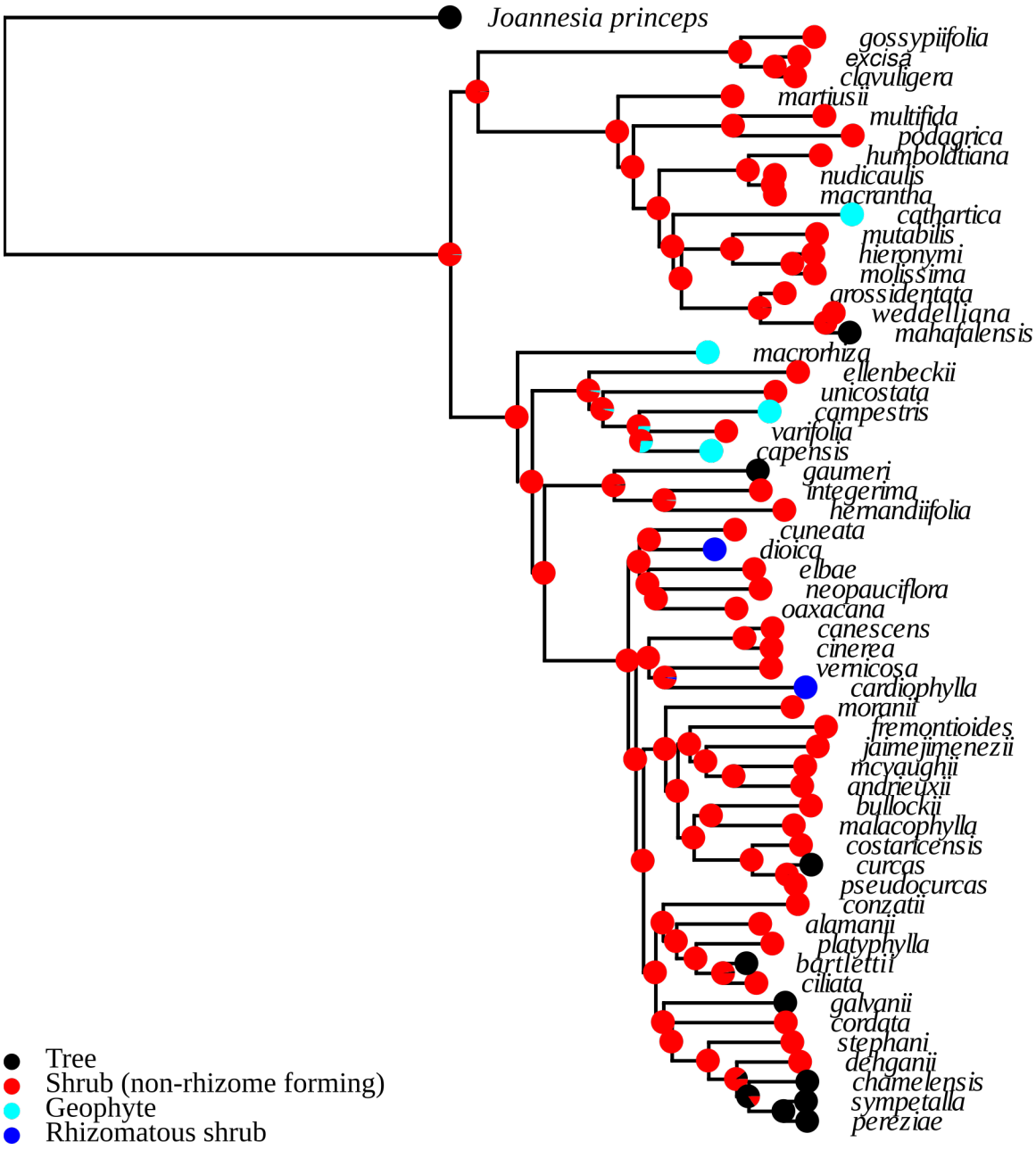


Figure 1-22: Ancestral state reconstruction of growth form in *Jatropha* using stochastic character mapping on the best scoring maximum likelihood tree from RAxML using the 85%Sim-minTaxa16 RADseq dataset.

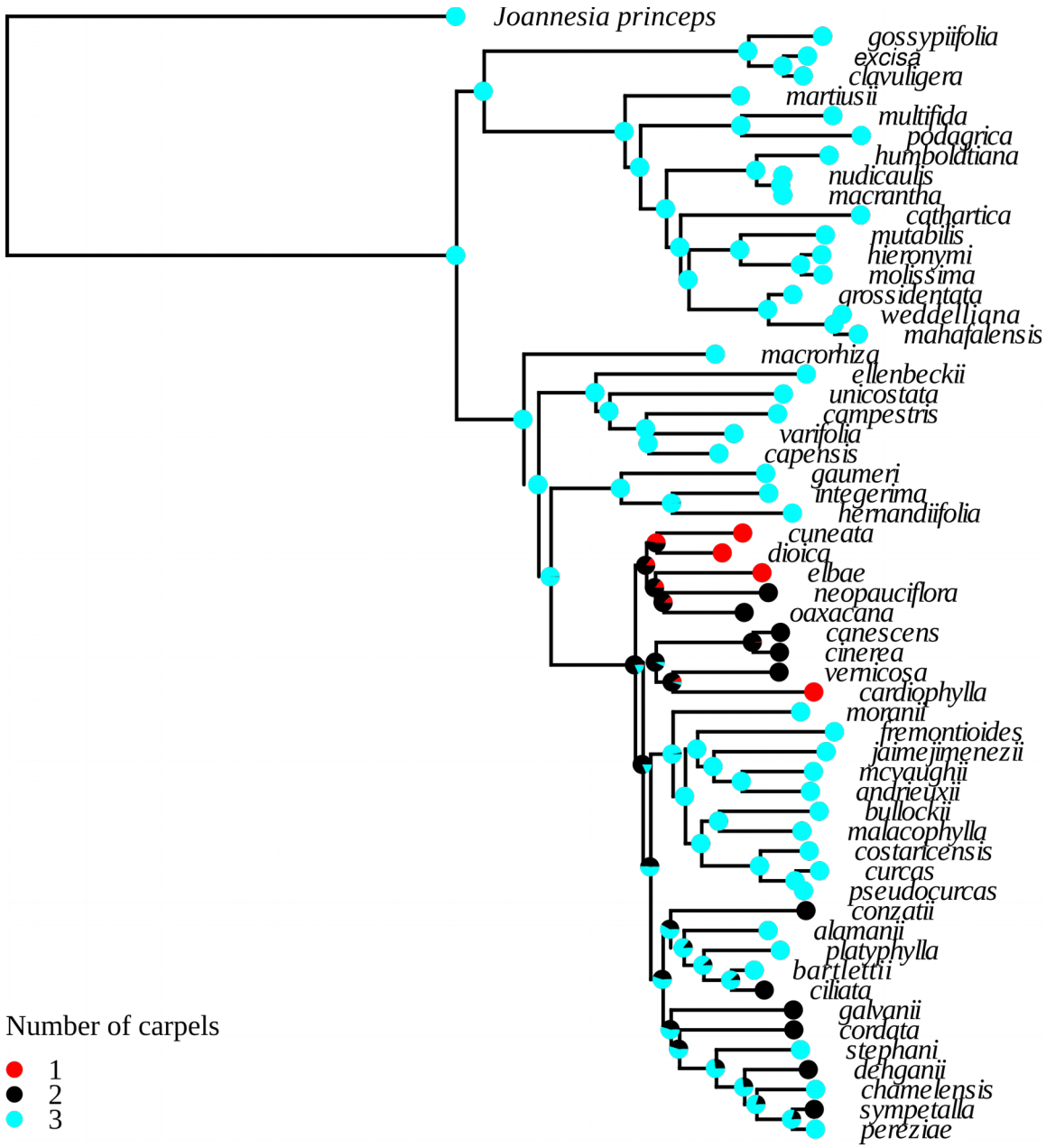


Figure 1-23: Ancestral reconstruction of carpel number in *Jatropha* using stochastic character mapping and the best scoring maximum likelihood tree from RAxML produced from the 85%Sim-minTaxa16 RADseq dataset.

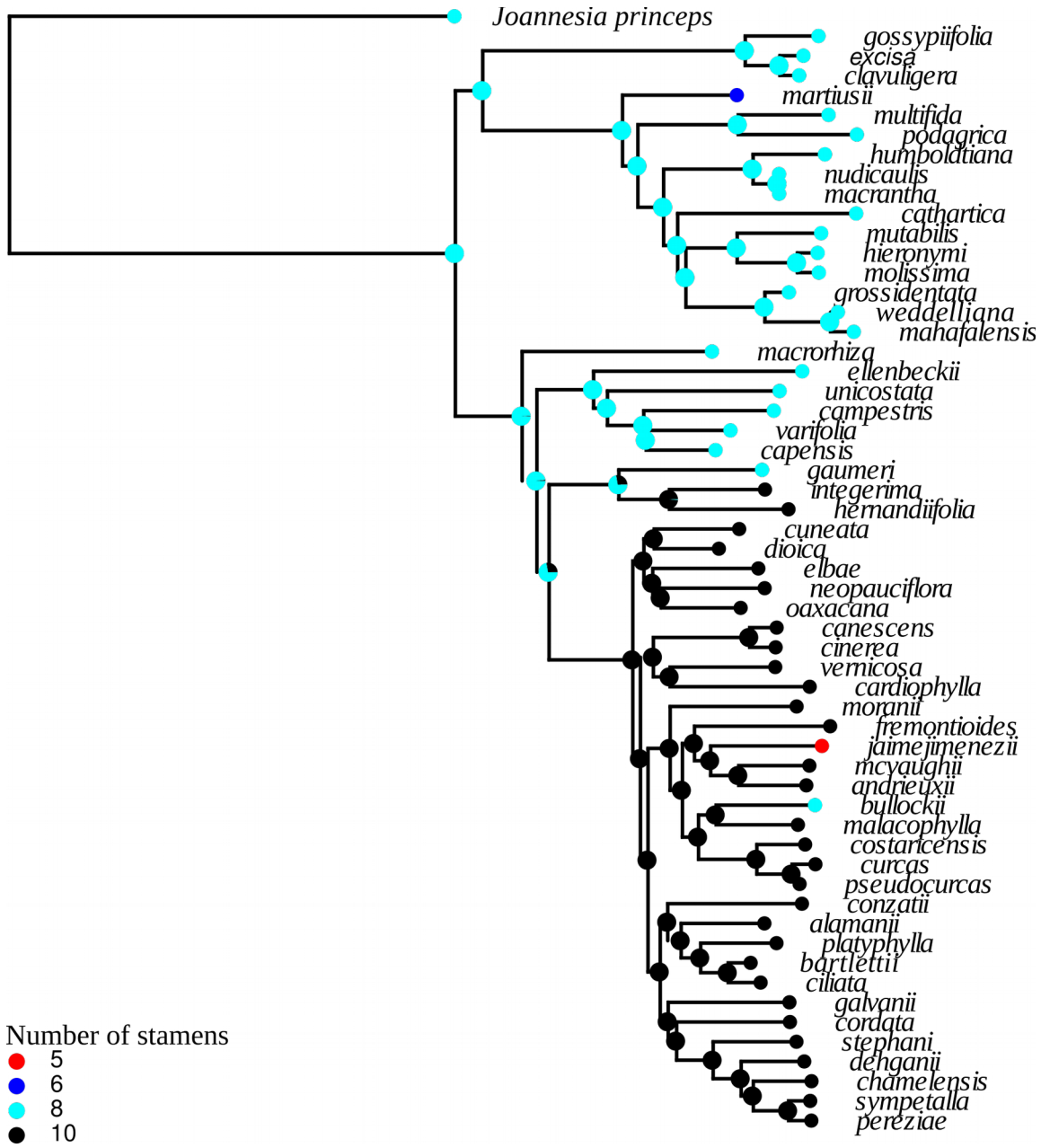


Figure1-24: Ancestral reconstruction for the number of anthers in *Jatropha* using stochastic character mapping and the best scoring maximum likelihood RAxML tree using the 85%Sim-minTaxa16 RADseq dataset.

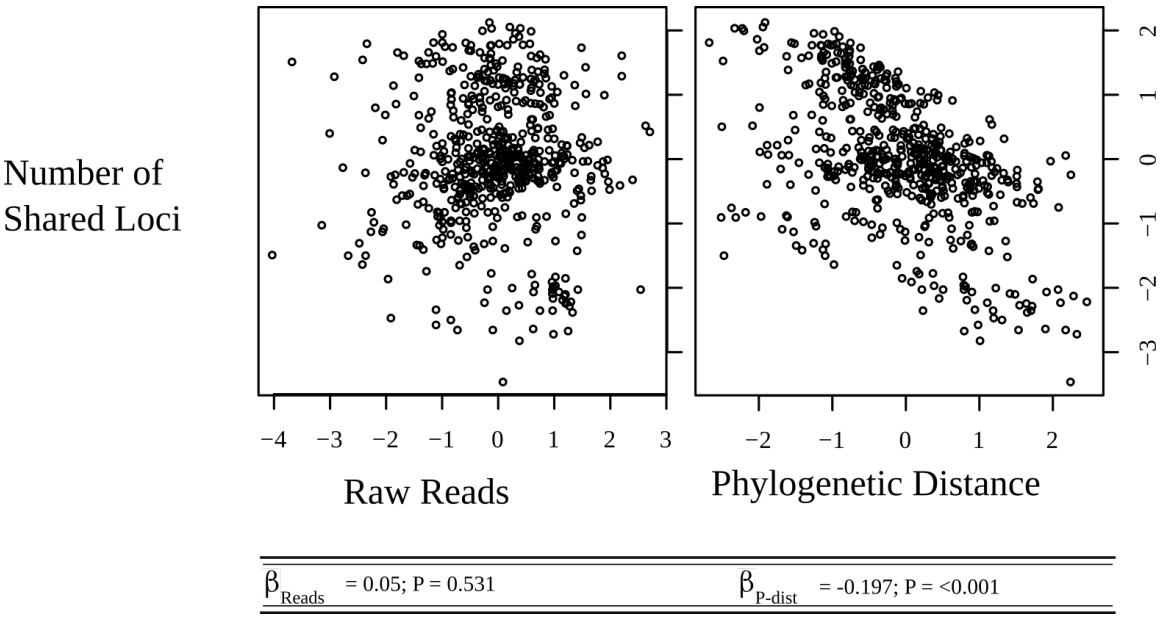


Figure 1-25: Results from fitting phylogenetic regressions between log median number of raw reads per sample (Reads) and phylogenetic distance (P-dist) to predict the loci shared among taxa. Regression coefficient is the mean of 100 replicate subsamples of 200 quartets using a 56-tip tree of *Jatropha*.

Dataset	-lnL	Pagel's λ	Reads - β	Reads - p	Dist - β	Dist - p
85min4	215.17	0.999	0.005	0.531	-0.197	< 0.001
85min8	217.28	0.999	0.0004	0.562	-0.193	< 0.001
85min16	218.24	0.999	0.0007	0.527	-0.203	< 0.001
85min32	234.46	0.999	-0.006	0.568	-0.175	< 0.001

Table 1-9: Results from phylogenetic generalized least squares analysis showing the relationship between missing data with number of reads (Reads B significance Reads P) and phylogenetic distance (Phy B significance Phy P).

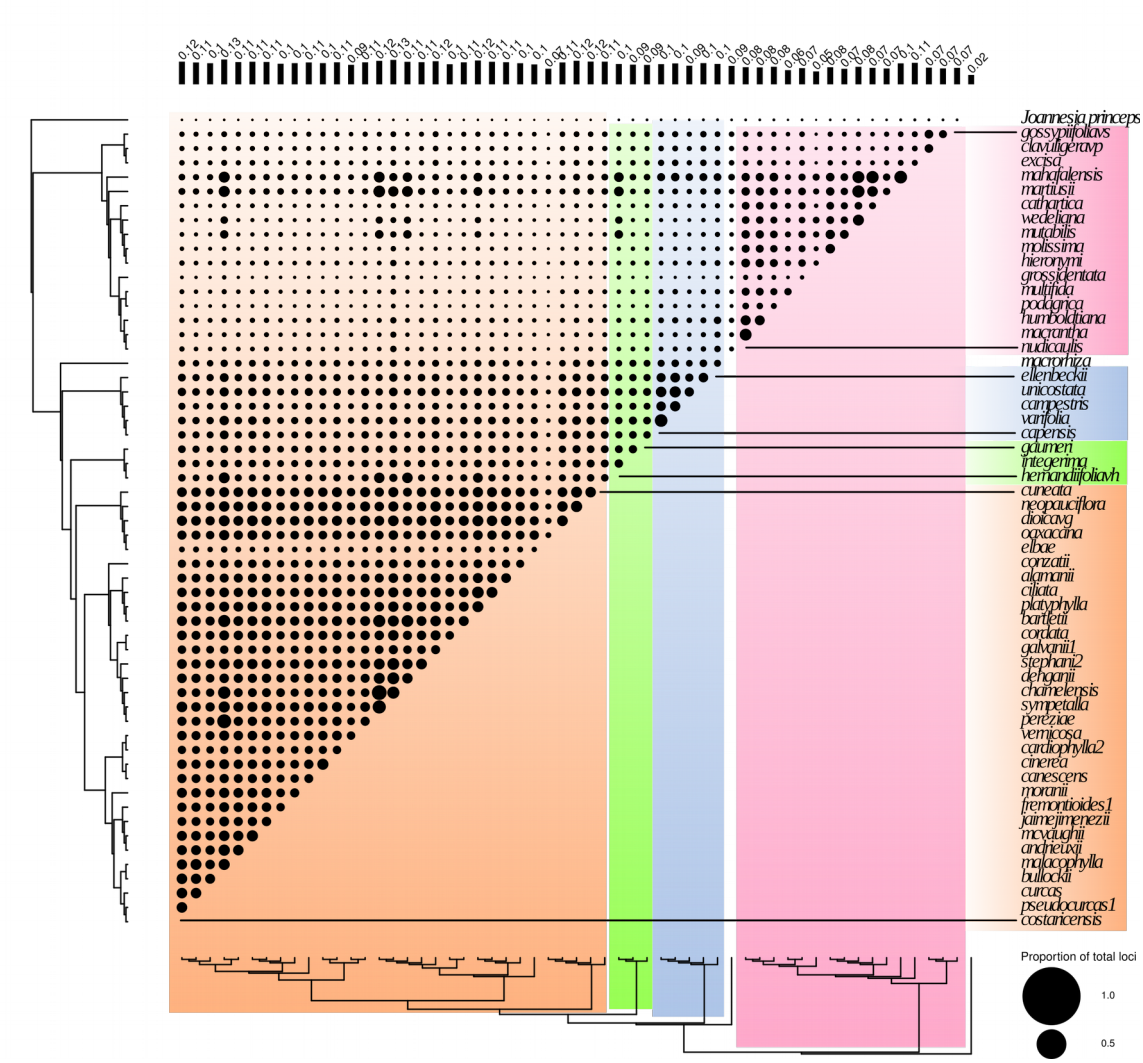


Figure 1-26: Proportion of RAD loci shared across individuals (dataset = 85min4 forward-reads). Circle size in the matrix is proportional to the number of shared loci between two species (scale in lower right-hand corner), and colors correspond to the major clades discussed in the text. Bars along top show the proportion of all loci in the alignment present for each species.

Phylogenetic Hypothesis		Topology	Data set	Δ - lnL	p-value
Position of *J. elbae* in subgenus *Curcas*	T_1	C4	85min8	413	0.001
	T_2	**C2**		0	0.99
	T_1	**C4**	85min16	0	1
	T_2	C2		1838	<0.0001
Position of C-2 in subgenus *Curcas*	T_1	**C1+C2**	85min16	0	1
	T_2	C2+C3		39.4	0.225
	T_1	C1+C2	85min32	57.8	0.046
	T_2	**C2+C3**		0	1

Table 1-10: Results from Shimodaira Hasagawa tests comparing alternative phylogenies inferred from different RADseq datasets. Dataset abbreviations are the assembly parameter combinations used (e.g. 85min8 is an 85% similarity threshold for clustering 8 minimum taxa in which a locus must be present. Bold entries in the topology column denote the position that taxon was recovered in on the best ML tree from RAxML for that dataset.

Test	P1	P2	P3	Outgroup	Z
1.1	Fre	Ala	Elb	Gau	*3.49*
1.2	Fre	Pla	Elb	Gau	*3.07*
1.3	Fre	Cil	Elb	Gau	3.04
2.1	Jai	Cil	Oax	Gau	**5.48**
2.2	Jai	Ala+Pla+Bar+Cil	Oax	Gau	**4.65**
2.3	Jai	Cil	Neo+Oax	Gau	**4.64**
2.4	Jai	Bar+Cil	Oax	Gau	**4.41**
2.5	Jai	Cil	Elb+Neo+Oax	Gau	**4.34**
2.6	Jai	Pla+Bar+Cil	Oax	Gau	*4.10*
2.7	Jai	Bar+Cil	Neo+Oax	Gau	*3.98*

Table 1-11: Results from ABBA-BABA tests for introgression between members of clades C1 and C4 (in columns P2 and P3 respectively). Z scores are standardized from D-statistics and shown for the ABBA arrangement of taxa indicating introgression between P2 and P3; bold values indicate tests that were significant at a Bonferonni corrected $\alpha = 0.01$. Species names abbreviations are: Fre = *J. fremontioides* Standl., Jai = *J. jaimejimenezii*, Ala = *J. alamanii*, Pla = *J. platyphylla*, Cil = *J. ciliata*, Bar = *J. bartlettii*, Elb = *J. elbae*, Oax = *J. oaxacana*, Neo = *J. neopauciflora.*

Chapter 2: Biogeography of the Neotropical *Jatropha:* Intercontinental disjunctions and Mesoamerican diversification within the Seasonally Dry Tropical Forest

INTRODUCTION

Intercontinental disjunction

Intercontinental disjunction is a biogeographic phenomenon in which a lineage occurs across geographically different continents, and can result from vicariance, long distance dispersal (LDD), or a combination of the two. The primary method for testing vicariance and LDD based hypotheses explaining disjunction patterns compares time calibrated phylogenies and biogeographic reconstruction with hypotheses about past connections between areas. As the means for generating molecular data for phylogeny reconstruction have now become widely available, along with a number of statistical frameworks for modeling evolutionary histories, the application of these methods for testing competing hypotheses for various intercontinental disjunction patterns have become increasingly prevalent.

Pantropical disjunction is an specific intercontinental distribution in which a lineage occurs in continental tropical regions separated by the Atlantic and/or Pacific Oceans. One vicariance hypothesis explaining pantropical disjunctions is the breakup of the Gondwanan supercontinent into present day South America, Africa, Antarctica, Australia, and India (Raven and Axelrod, 1974). Expected divergence times for groups originally occurring in both Africa and South America have been attributed to the breakup of Gondwana ca 105 million years ago (Mya) (McLoughlin, 2001). Few dated molecular phylogenies of pantropical groups have found divergence times in accordance

with the Gondwanan vicariance hypothesis (but see Chanderbali et al., 2001). The North Atlantic Land Bridge (NALB) hypothesis is an alternative vicariance explanation for pantropical disjunction which invokes migration along boreal land connections that existed during the global thermal optimum of the middle Eocene approximately 45 Mya (Tiffney, 2000, 1985; Morley, 2003; Fig. 2-1). Some estimated divergence times of families Burseraceae, Vitaceae, and Malpighicaeae support this pattern (Davis et al., 2002; Weeks et al., 2005, Nie et al., 2012). Estimated branching times between neotropical and African tropical lineages less than 40 Mya support LDD as the cause, as has been documented within Sapotaceae and Simaroubaceae (Bartish et al., 2011; Clayton et al., 2009).

A strictly neotropical intercontinental disjunction that has long interested and challenged biogeographers is between tropical areas of North and South America, and is also related to the establishment of the Isthmus of Panama (IP) (Gutiérrez-García and Vázquez-Domínguez, 2013; Marshall et al., 1982). The date for final emergence of the IP remains highly contested A widely accepted time, associated with the sudden appearance of mammal fossils in the record indicating movement between the continents, is ~3.5 Mya (Coates et al., 2004; Jaramillo et al., 2017; Montes et al., 2015; O'Dea et al., 2016), but recent geologic evidence corroborated by phylogenies of both terrestrial and marine organisms has suggested an earlier age of 10-15 Mya (Bacon et al., 2015; Montes et al., 2015). Divergence dates for disjunct clades separated by the IP older than 15 Mya can be taken to support LDD as the underlying cause of the distribution, whereas splits younger than 3.5 Mya can be considered probable cases of direct migration across the IP. The mechanism of disjunction for divergence times between 3.5 and 15 Mya may or may not be taken as support for a Miocene age of the Isthmus of Panama, depending on the context and biology of the organisms in question.

A third biogeographic pattern involving disjunction between landmasses on different continental plates pertains to the relative roles of vicariance versus LDD in shaping the Caribbean biota and the continental sources for Caribbean lineages. As with pantropical disjunction, acceptance of plate tectonic theory prompted many biogeographers to invoke vicariance to explain the origin of the Caribbean biota, postulating that the Antilles are effectively land rafts dislodged from continents the moved west to east to their current positions (Rosen, 1975). Alternatively, another explanation is a proposed contiguous land span extending from northeastern South America to the Greater Antilles along the now submerged Aves Ridge (GAAR) during the Oligocene (35-32 Mya) allowed migration as far as present day Cuba (Iturralde-Venent, 2006; Iturralde-Vinent and MacPhee, 1999). The GAARlandia hypothesis, as it is known, has been challenged and supported by both biological and geological evidence (Ali, 2012; Hedges, 2006; Nieto-Blázquez et al., 2017; van Ee et al., 2008). The growing consensus from dated phylogenies is that long distance dispersal to the Caribbean, primarily from the New World continental regions, followed by *in situ* diversification has largely shaped the Caribbean biota (Cervantes et al., 2016; Chakrabarty 2006; Matos-Maraví et al., 2014). Work is ongoing to determine the relative contributions of taxa from North and South America.

Diversification in the Neotropical Seasonally Dry Tropical Forests

The seasonally dry tropical forest (SDTF) is a unique biome that has received increased attention in the last 30 years due to its recognition as a biodiversity hot-spot that has been relatively understudied compared to other neotropical forest types (Janzen, 1988; Myers et al., 2000). Seasonally dry tropical forests are found in frost free

zones where rainfall is adequate to sustain trees as the dominant growth form, but strong seasonality of rainfall results in long periods (typically in excess of six months) of dormancy when foliage is lost (Mooney et al., 1995). Much of the global SDTF habitat is highly fragmented, a factor that has been hypothesized to have influenced the evolution of endemic plant lineages (Schrire et al., 2009).

With a growing number of dated molecular phylogenies, an emerging narrative is that SDTF patches have been historically isolated and ecologically stable, and that plant lineages found in these patches are dispersal limited and diversify within patches (Lavin et al., 2004; Pennington et al., 2000; Pennington et al., 2009). This model of evolution for SDTF specialists predicts that phylogenies for these groups will display significant phylogenetic structure, phylogenetic niche conservatism, and branching times leading to terminal taxa that will be older than lineages from adjacent younger biomes such as wet tropical forests (Lavin, 2006). These predictions have been verified with empirical studies conducted primarily in South America (but see also Schrire et al., 2009). More studies in other regions are needed to confirm if the model applies to SDTF groups more generally. To date, *Bursera* Jacq. ex L. is the only Mesoamerican SDTF group studied using molecular phylogenetic dating to test these predictions explicitly (DeNova et al., 2012). Although *Bursera* exhibits phylogenetic geographic structure and niche conservatism, species from the SDTF are younger on average than tropical rainforest species, contrary to the observed patterns for many South American lineages (DeNova et al., 2012, Pennington et al., 2009).

The largest tracts of dry forest in Mesoamerica are found along the Pacific slopes of the Sierra Madre Occidental and Sierra Madre del Sur (Trejo and Dirzon, 2002) (Fig. 2-2). The Pacific side of the Sierra Madre Occidental is protected from Arctic cold fronts by the mountains, allowing for the persistence of tropical communities as far north

as Alamos, Sonora, Mexico ~27° N. From the early Miocene through to the present, rounds of volcanism in the Trans-Mexican Volcanic Belt (TVB) sequentially dissected Mexico latitudinally at 19-20° N (Ferrari et al., 2012). This belt has been demonstrated to have played a large role in the diversification of taxa in the Mexican highlands (Bryson and Riddle, 2012; Bryson et al., 2012; 2012; Mastretta-Yanes et al., 2015). If the SDTF is an old and stable biome in Mesoamerica, with taxa that have persisted for millions of years, then uplift of the TVB may have resulted in speciation via vicariance, which would be reflected in divergence dates between northern and southern lineages.

The global distribution of *Jatropha* makes it a useful group for testing hypotheses about mechanisms underlying intercontinental disjunction patterns, and the close association of the genus with the seasonally dry tropical forest on multiple continents is ideal for studying the evolutionary process within a unique biome (Fig. 2-3). The objectives of this study were to create a time calibrated phylogeny for *Jatropha* to answer the following questions: (1) are the observed intercontinental disjunctions of *Jatropha* the product of vicariance, long distance dispersal, or a combination of both; (2) what role did tectonic forces play in the evolution of *Jatropha* subg. *Curcas* in Mesoamerica, and (3) is there phylogenetic geographic structure in the lineages of *Jatropha* most closely associated with seasonally dry forests in Mesoamerica (Table 2-1)?

METHODS

Estimating divergence times for Jatropha

We estimated the crown age for *Jatropha* by first conducting a family wide analysis, and then using this age as a secondary calibration point to date nodes within *Jatropha*. The crown age of *Jatropha* was estimated with sequence data from the

plastid marker *rbcL* for 61 species of Euphorbiaceae, 12 of which were species of *Jatropha*. Samples were included from the three largest subfamilies (Euphorbioideae, Acalyphoideae, and Crotonoideae) and several basal genera following Wurdack et al., (2005). Two fossil calibration points were used: *Acalyphapolenites* Sun M.R., a pollen fossil similar to modern *Acalypha* L. (subfamily Acalyphoideae) from the early Eocene, and *Crepetocarpon perkinsii* Berry, a fossilized fruit belonging to tribe Hippomaneae (subfamily Euphorbioideae) from the middle Eocene (Sun et al., 1989; Dilcher and Manchester, 1988). We also used secondary calibration points for the genus *Croton* (subfamily Crotonoideae) and the crown of Euphorbiaceae (Table 2-2). Two previous divergence analyses using *Crepetocarpon perkinsii* as a calibration point differed in how the age constraint was applied. Van Ee et al., (2008) assumed a close affinity between the fossil and the genus *Hippomane* and therefore constrained the node leading to *Hippomane* L. and *Bonania* A.Rich, whereas Xi et al., (2012) constrained the age of the tribe Hippomaneae at the most recent common ancestor of *Hippomane* and *Actinostemon* Mart. ex Klotzsch. We estimated divergence times using both calibration schemes to compare the outcomes. A cross-validation of divergence time estimates for the crown age of *Jatropha* was performed by removing each individual calibration point and rerunning the analysis.

Sequence alignment was performed with MAFFT (Katoh et al., 2002), and divergence times were estimated in BEAST2 v2.5.2 (Bouckaert et al., 2014) using GTR+Γ model of substitution. We used an uncorrelated log-normal relaxed clock and the calibrated yule model as the tree prior in order to place prior distributions on age calibrated nodes (Drummond et al., 2006 ; Heled and Drummond, 2015; Table 2-2). Control files with parameters for analyses were generated in BEAUti v2.5.2 (Bouckaert et al., 2014). We ran two independent MCMC chains for 20,000,000 generations each,

and sampled from the posterior distribution every 10,000 steps. BEAST log files were examined in Tracer v1.6.0 (Rambaut et al., 2014) to assess MCMC convergence, and tree files were combined using LogCombiner v2.5.2 (Bouckaert et al., 2014) with a 20% burn-in. Median node heights with 95% highest probability densities (HPD) were written to the maximum clade credibility tree with TreeAnnotator v2.5.2 (Bouckaert et al., 2014). Visualization of trees with geologic time scales was done using R scripts adapted from the BEAST2 divergence time estimation tutorial (Barido-Sottani et al., 2018).

Divergence times within *Jatropha* were estimated using the age distribution for the crown of *Jatropha* resulting from the family level analysis and a fixed topology for *Jatropha* from the phylogenetic analysis of Chapter 1. Genomic data were produced from double-digest restriction site-associated sequencing (RADseq) for 56 species of *Jatropha* and *Joannesia princeps* (Peterson et al., 2012). The dataset was assembled using ipyrad v0.7.21 (Eaton and Overcast, 2016) with a similarity clustering threshold of 85% and minimum number of taxa of 55. Parameters for the BEAST run were the same as for the family analysis, and results from BEAST summarized as described above.

Assembly of Locality Data and Defining Biogeographic Areas

A database of georeferenced occurrence data was compiled for all species of *Jatropha* used in the divergence analysis. Sources for locality data included the Global Biodiversity Information Facility (https://www.gbif.org/; accessed September 15, 2018), the SEINet Portal Network (http://swbiodiversity.org/seinet/; accessed May 25, 2018), CONABIO (https://www.gob.mx/conabio; accessed May 28, 2018), and personal observations from the field. Records were screened manually for duplicate entries, and

then plotted using QGIS v2.8.6 (QGIS Development Team, 2014) to identify observations within urban centers or bodies of water. Because our taxon sampling was biased toward the Neotropics (primarily within Mesoamerican *J.* subg. *Curcas*) our findings may have underestimated the total number of vicariance and dispersal events between the Neotropics and Paleotropics. The widely cultivated species *J. curcas* was coded as a Mesoamerican species for the global biogeographic analyses, but was excluded from the Mesoamerican analysis due to its widespread cultivation. Likewise, *J. gossypiifolia* L. var. *staphysagrifolia* (P. Miller) Müll.Arg. and *J. gossypiifolia* var. *gossypiifolia* were coded as Mesoamerican and South American, respectively, for the global analysis, but were excluded from Mesoamerican analyses because they are widely naturalized and occur in disturbed areas throughout the tropics.

We considered four bioregions for the inter-continental biogeographic analysis: Mesoamerica, South America, the Caribbean islands, and Africa (Table2-1; Fig. 2-1).

For biogeographic reconstruction of *Jatropha* subg. *Curcas* within Mesoamerica we chose ten bioregions from the Commission for Environmental Cooperation (1997) (Table 2-1; Fig 2-2). The dry forests and desert regions of the Baja peninsula were treated as a single biogeographic unit, with a date of 5.5 Mya for the split of the peninsula from mainland Mexico (Carreno and Helenes, 2002). To test if the TVB has acted as a significant biogeographic barrier for *Jatropha* we divided the roughly contiguous stretch of SDTF along the Pacific slope of the Sierra Madre Occidental and Sierra Madre del Sur into the Sinaloan (northern) and Guerreran (southern) regions. We used the age estimates for four major periods of volcanism associated with of uplift of the TVB: middle to late-Miocene arc (19-10 Mya), Silicic (7.5-5 Mya), bimodal (5-3 Mya),

and late Pliocene-Quarternary arc (3.5 Mya-present) (Ferrari et al., 2012; García-Palomo et al., 2002).

Ancestral Area Reconstruction and Estimating Biogeographic Events

Ancestral area reconstructions were performed using a maximum likelihood framework for comparison of parametric models with the R package *BioGeoBEARS* v0.2.1 (Matzke, 2013). We compared three base models: one designed to approximate the DEC model of Lagrange (Ree and Smith, 2008), one that approximated DIVA (Ronquist, 1997), and one that approximated BAYArea (Landis et al., 2013). For these three "unconstrained" models two free parameters, dispersal to an adjacent area without simultaneous cladogenesis (*d*) and extinction from an area (*e*), were estimated. For the next set of models we estimated a third free parameter (*j*) for the rate of jump dispersal at the time of cladogenesis (Matzke, 2014). Explicit temporospatial constraints incorporating prior knowledge of geography and tectonic history were introduced into the models as dispersal matrices with rates reflecting relative probabilities of movement between areas (Table 2-3 – 2-4). These models were labeled "time-stratified" models. We then added, a scaling parameter *w,* which exponentiated the elements of the dispersal matrices following Dupin et al., (2017). The combination of all parameters yielded 18 unique models. The 6 variations of each base model were unconstrained, unconstrained+j, time-stratified, time-stratified+j, time-stratified+w, time-stratified+j+w. Model selection was done using AICc, a modification of the Akaike information criterion corrected for small sample sizes (Burnham and Anderson 2002).

Biogeographic reconstructions were done for *Jatropha* at the global scale and for *Jatropha* subg. *Curcas* within Mesoamerica. For the global analysis, ancestral

states at nodes were allowed to include up to two bioregions, with no combination of bioregions excluded. For the Mesoamerican biogeographic analysis, up to three bioregions were allowed at each node. Using the best fit model, or models, we determined the number of cladogenic biogeographic events (dispersal, vicariance, sympatric speciation) using a modified form of Bayesian stochastic mapping called biogeographic stochastic mapping (BSM) implemented in BioGeoBEARS (Huelsenbeck et al., 2003; Dupin et al., 2017). Using the dated chronogram, compiled species distribution data, and parameters from the best model(s) we performed 1000 realizations (independent mappings) of the ancestral states. Validation of biogeographic stochastic mapping was done by comparing the mean state probabilities calculated for each node by stochastic mapping to the state probabilities estimated from maximum likelihood inference using linear regression and the R package *phytools* v0.6.0 (Revell, 2012).

Testing for Geographic Structure in the Phylogeny of Jatropha

To test predictions of geographic structure for species from the seasonally dry tropical forest we performed distance regression and phylogenetic community analyses of species of *Jatropha* occurring in Mesoamerica. First, Mantel tests for correlation of log transformed Euclidean distance (based on latitude and longitude of collected specimens) vs the genetic distance (GTR+Γ ML branch lengths from the RADseq tree from Chapter 1). Genetic distances were extracted using the R package *ape* v5.2 (Paradis et al., 2004) and Mantel tests were performed (9999 repetitions) with the R package *ade4* v.1.7.13 (Chessel et al., 2004). *S*econd, we used the software package PHYLOCOM v4.2 to quantify phylogenetic community structure (Webb et al., 2008). Using the RADseq ML phylogeny and the bioregions of Mesoamerica defined in the

biogeographic analysis we calculated the net related index and nearest taxa index for each clade. Positive values for these indices indicate species are geographically aggregated, whereas negative values indicate over-dispersion. Calculated indices were compared to distributions from 10,000 randomly generated geographic states for the same tree to test for significance.

RESULTS

Divergence dating within Euphorbiaceae and Jatropha

The recovered topology of the chronogram from BEAST strongly supported the monophyly of the three subfamilies (Euphorbioideae, Acalyphoideae, and Crotonoideae) with the exception of the placement of *Neoscortechinia kingii* Pax & K.Hoffm. as part of Crotonoideae (Fig. 2-4), contrary to previous studies which placed *N. kingii* as sister to the remainder of Euphorbiaceae (Wurdack et al., 2005). Relationships among the subfamilies were poorly resolved with only a 20% posterior probability for Euphorbioideae being sister to Acalyphoideae and Crotonoideae (Fig 2-4). *Jatropha* was recovered as monophyletic with high support (BPP = 1.0), and *Joannesia princeps* was well supported as being sister to *Jatropha* (BPP = 1.0) (Fig. 2-4). Divergence analysis using four calibration points recovered a median age for the split between *Jatropha* and *Joannesia* of 40.5 Mya (58.3-21.3 95% highest probability density (HPD)), and estimated a median crown age of *Jatropha* of 24.6 Mya (39.6-11.5 95% HPD) (Fig. 2-4). Cross-validation showed significant changes in the estimated ages for calibration nodes of Euphorbiaceae when they were not constrained (shown in bold in Table 2-4). The median crown and stem ages of *Jatropha* estimated in cross-validation runs, however, were not significantly different (remaining within the 95% HPD) from the ages recovered in the 4

calibration analysis. We therefore used the age estimates from the 4 calibration point analysis for subsequent divergence dating within *Jatropha*.

Results from divergence analysis within *Jatropha* recovered a slightly younger crown age for the genus than was found in the family analysis (22.2 Mya, 31.5-13.8 95% HPD (Fig. 2-5; Table 2-5). The observed heterogeneity in substitution rates across branches (ucld.stdev of the combined runs = 0.328) indicated that molecular evolution in *Jatropha* has not been strictly clock-like. Ages for major clades within *Jatropha* (see Chapter 1 – Fig. 6 for clade names) are shown in Table 2-5. We estimated an early Miocene origin (19.1 Mya, 27.5-11.6 95% HPD) origin for the crown of the South American clade of *J.* subg. *Jatropha* and a late Miocene origin (6.4 Mya, 9.4-3.8 95% HPD) for *J.* subg. *Curcas* (Fig. 2-5; Table 2-5). The stem ages for the African and Caribbean clades were 11.0 Mya (15.9-6.2 95% HPD) and 10.0 Mya (11.0-3.9 95% HPD) respectively.

Biogeographic Reconstruction of Jatropha

Biogeographic model testing yielded high AICc scores for more than one model for both global and Mesoamerican analyses (Table 2-6), and biogeographic stochastic mapping was performed for all best scoring models (weighted AICc > 0.1). Reconstructed biogeographic histories are presented for the best scoring model at each scale, and differences among the inferred histories from other high-scoring models are discussed.

Intercontinental disjunctions

Top scoring models indicated jump-dispersal as the primary cause of range shifts between continents for *Jatropha*, and that a single vicariance event, at most, occurred between Mesoamerica and South America (Fig. 2-6; Table 2-7). Maximum likelihood reconstruction of *Jatropha* found three models to be similar with respect to AICc scores: DEC+j unconstrained, DIVA+j unconstrained, and DEC+j+w time-stratified (Table 2-6). For the DEC+j+w time-stratified model, the optimized parameter *w* was less than 1 ($w = 0.25$), indicating a better fit of the model when values in the dispersal matrix were more dispersed (Table 2-6). There was highly significant positive correlation between mean state probabilities for ancestral areas inferred at nodes from BSM and those calculated from the maximum likelihood for all models (Fig. 2-7; Table 2-6). Similar mean numbers of dispersal, cladogenic, and vicariance events were inferred by biogeographic stochastic mapping for each model, although DIVA estimated slightly more founder events and fewer vicariance events than either DEC model (Table 2-8).

Divergence dating and biogeographic reconstruction indicated that the pantropical disjunction of *Jatropha* resulted from long distance dispersal events in the Neogene and Quaternary. The DEC models indicated that the most recent common ancestor (MRCA) of *Jatropha* was most likely distributed in both South America and Mesoamerica, whereas the DIVA model recovered an ancestral area restricted to South America to be more probable(Fig. 2-6, 2-8 – 2-9). The different reconstruction explains the lack of inferred vicariance events by the DIVA model. All three models inferred similar states at internal nodes after the MRCA of *Jatropha*, and inferred jump-dispersal events at the same locations in the tree (Fig. 2-6). Two trans-Atlantic jump dispersals were inferred, one from Mesoamerica, giving rise to the African clade of *J.* subg. *Jatropha*, and one from South America to Madagascar along the branch leading to *J.*

mahafalensis Jum. & H. Perrier. Since the estimated crown age of *Jatropha* was ~22 Mya (upper limit of the 95% HPD = 31.5 Mya), we rejected both Gondwanan breakup and North Atlantic land bridge migration hypotheses and accepted the alternative hypothesis of long distance dispersal as the cause for pantropical disjunction in *Jatropha*.

Multiple range shifts from South America to Mesoamerica were detected, with some alternatively inferred as vicariance due to the splitting of an ancestor distributed across both continents (Fig. 2-6). The earliest shift occurred 22.2 Mya (31.5-13.8 Mya 95% HPD) at the crown node of *Jatropha* on the branch leading to the MRCA of *J. macrorhiza,* the Caribbean clade, the African clade, and *J.* subg. *Curcas*. The inferred probability of the crown node only occurring in South America, as opposed to also occurring in Mesoamerica, however, was only around 25%, therefore there was strong support (~75%) for the disjunction observed at this node to have resulted from vicariance instead (see Discussion). The second shift from South America to Mesoamerica occurred on the branch leading to the MRCA of *J. podagrica* Hook. and *J. multifida* L. 9.4 Mya (13.7-9.4 Mya 95%HPD), with a subsequent shift to the Caribbean for *J. multifida.* The third shift occurred on the branch leading to *J. cathartica* 7.6 Mya (11.1-4.3 Mya 95% HPD) (Fig. 2-6). The median estimated age of the range shift to Mesoamerica for the clade including *J. macrorhiza* was prior to the hypothesized Miocene age (10-15 Mya) for the Isthmus of Panama, which indicates long distance dispersal to be the cause of the distribution. The younger limit of the 95% HPD, however, was 13.8 Mya, and thus direct migration can't be entirely ruled out for this clade. The estimated median ages and 95% HPD of the range shifts observed for *J. cathartica* and the MRCA of *J. podagrica* and *J. multifida* were all between the Miocene and traditional Pliocene ages proposed for the Isthmus of Panama, leaving the underlying mechanism for these cases of disjunction unresolved.

Two separate dispersal events to the Caribbean were detected: one along the branch leading to *Jatropha multifida* at 5.6 Mya (8.9-2.9Mya 95% HPD) and the other on the branch leading to the clade with *J. integerrima* Jacq. and *J. hernandiifolia* Vent. At 7.2 Mya (11.0-3.9 95% HPD) (Fig. 2-6). In both cases the dispersal took place after the hypothesized time for the postulated existence of the GAARlandia land span from South America. *Jatropha multifida* was inferred to have arrived in the Caribbean from Central America, while the ancestor of *J. integerrima* and *J. hernandiifolia* likely arrived from the Yucatan Peninsula (Fig. 2-6; Table 2-8). *Jatropha integerrima* and *J. hernandiifolia* belong to *J.* subg. *Jatropha* sect. *Polymorphae*, a group of eight species that comprise the only substantial radiation of *Jatropha* in the Caribbean.

Mesoamerica

Three time-stratified models scored highly using AICc in biogeographic reconstructions of *Jatropha* subg. *Curcas*: DIVALIKE+ j+w, DIVALIKE+j, and DEC+j (Table 2-6). In contrast to the global analysis, the *w* parameter was estimated as greater than 1 (1.53 in the DIVALIKE+j+w model, the highest scoring model), indicating a better fit using more aggregated rates in the dispersal matrices and suggesting higher levels of dispersal between areas. Regression analysis showed BSM and ML estimates of ancestral state probabilities at nodes were highly congruent for all three best scoring models models (Table 2-6). All three models inferred similar numbers of biogeographic events, but the two DIVALIKE models inferred more anagenetic dispersals whereas DEC+*j* inferred more sympatric speciations (Fig. 2-10, Table 2-6). These differences were not the result of the *w* parameter since the two DIVALIKE models (with *w* and without *w*) estimated very similar numbers of all types of biogeographic events (Fig 2-11, Table 2-7).

This indicated that, at least for DIVALIKE models, results were robust to changes in the dispersal matrix, whereas model choice did impact the types of inferred events.

The high concordance in the ancestral area reconstructions of internal nodes across models resulted in a single, albeit complicated, biogeographic history of the subgenus (Fig. 2-9). All models inferred the ancestral area of the MRCA of *Jatropha* subg. *Curcas* to have been the Sonoran Desert and the Yucatan peninsula, with The DEC+j model additionally including the Balsas Depression (Fig. 2-11 – 2-13). Clade C-4 originated in the Sonoran Desert, and then diverged into a northern lineage and a southern lineage in which divergences occurred progressively farther southeast: first in the Balsas Depression (*J. elbae*), then in the Tehuacan valley (*J. neopauciflora*), and then in dry valleys of Oaxaca (*J. oaxacana*). Clade C-3 originated and diversified within northwest Mexico, whereas the MRCA of C-2 occurred in Baja California and then subsequently formed a lineage widely distributed across southern Mexico (Fig. 2-11). Clade C-1 originated in southern Mexico in the Guerreran dry forests, where it has largely remained and diversified, also spreading into the Balsas Depression (Fig. 2-11).

Biogeographic stochastic mapping favored dispersal over vicariance as the major force behind diversification of *Jatropha* in association with tectonic events in Mesoamerica. The Sonoran Desert species *J. cardiophylla* (Torr.) Müll.Arg. and Baja Californian *J. vernicosa* Brandegee diverged around the time of the formation of the Gulf of California (Fig. 2-11). This divergence was inferred as vicariance in about 14% of BSM realizations, and otherwise as dispersal between areas (Table 2-11). Also, the MRCA of clade C-2, which occurred in Baja California, gave rise to *J. moranii* Dehgan and G.L.Webster in Baja California and the remainder of clade C-2, which was distributed throughout the Guerreran SDTF (Fig. 2-11). Despite this divergence being synchronous with the splitting of the Peninsula from the mainland, it was inferred as

vicariance in only 10% of BSM realizations (Table 2-11). Marginal support was also found for vicariance as the explanation for the divergence between *J. cuneata* and *J. dioica* from the Sonoran and Chihuahuan Deserts respectively (Fig. 2-11). More vicariance events were inferred south of the TVB than to the north, primarily between the Guerrera seasonal dry forest and the dry forests of the Balsas Depression: first at the node leading to *J. deghanii* J.Jiménez Ram. from the Balsas Depression and three species from the Guerreran SDTF in the late Pliocene, and second between *J. ciliata* from the Balsas and *J. bartlettii* from the Guerreran SDTF in the Pleistocene (Fig. 2-11).

Numerous founder events (jump dispersals with cladogenesis) and anagenic range expansions were inferred from biogeographic reconstruction for *Jatropha* subg. *Curcas* (Fig. 2-11; Table 2-7). Founder events or expansions across the TVB were detected in all clades except C-3, and the Guerreran SDTF was the single area most associated with shifts across the TVB (Table 2-10). Southward founder events from northwestern Mexico across the TVB in the late Miocene-early Pliocene were inferred to have occurred twice into the Guerreran SDTF and once into the Balsas Depression (Fig. 2-11). Four Pliocene-Pleistocene northward range expansions into the Sinaloan SDTF were inferred at nodes leading to terminal branches: three from the Guerreran SDTF (*J. platyphylla J. mcvaughii* Dehgan & G.L.Webster, and *J. malacophylla* Standl.) and once from the Balsas Depression (*J. cordata* Müll.Arg) (Fig. 2-11). Jump dispersal events between areas of seasonally dry forest south of the TVB were common throughout the Pliocene and Pleistocene, (Table 2-10).

Geographic Structure in the Seasonally Dry Tropical Forests of Mesoamerica

Tests for geographic structure in the among species of *Jatropha* in the Mesoamerican seasonally dry forests yielded mixed results. The Mantel test for all species of *J.* subg. *Curcas* showed a marginally significant positive correlation between geographic and genetic distances (R^2 = 0.15, p = 0.093, number of replicates = 9,999; Fig. 2-14), with a similar trend when only species from seasonally dry forests were analyzed (R^2 = 0.17, p = 0.087, number of replicates = 9,999; Fig. 2-15). Phylogenetic community analysis indicated that both the net relatedness index and nearest taxon index were significant for most regions of seasonally dry tropical forest in Mesoamerica, but not for desert regions (Table 2-11). The areas of dry tropical forest that did not exhibit significant phylogenetic structure were the Isthmus of Tehuantepec and Central America, likely because of small sample size (N=2 in each both areas). The conflict from the results of our analyses do not strongly support the conclusion that dispersal limitation has driven localized diversification for lineages of *Jatropha* associated with seasonally dry tropical forests in Mesoamerica.

DISCUSSION

Vicariance versus long distance dispersal in shaping intercontinental Biogeography for Jatropha

Estimated divergence times between Old World and New World species are sufficiently later than hypothesized land connections between the Neotropics and Paleotropics, leading us to favor long distance dispersal over vicariance to account for the pantropical distribution in *Jatropha* (McLoughlin, 2001; Tiffney 2000; 1985). Other biogeographic studies of flowering plant groups with pantropical disjunctions resulting

from long distance dispersal have supported either trans-Atlantic or trans-Pacific introductions (Bartish et al., 2011; Clayton et al., 2009; Michalak et al., 2010). *Jatropha* occurs in India and the Arabian Peninsula in addition to Africa, and it is possible that the introduction to the Old World was from either the east or west. Increased taxon sampling of African and Asian lineages is the necessary to address this question.

Biogeographic reconstructions indicated that the Mesoamerica-South America disjunction observed for *Jatropha* was the result of both vicariance and long distance dispersal events. Ancestral area reconstruction strongly favored the most recent common ancestor (MRCA) of *Jatropha* as being distributed across both South and Mesoamerica, and that a vicariance event led to the two descending lineages that were eventually restricted to single continents where they are presently found. The MRCA of *Jatropha* and *Joannesia* was also inferred to have been distributed across both continents, which is surprising given the presumed isolation of South America from North America for several tens of millions of years prior to this vicariance event. One possible scenario is that sometime prior to the time of the MRCA of *Jatropha* and *Joannesia* there was long distance dispersal between the continents, possibly by island-hopping along the Proto-Antilles island arc between North and South America during the Paleogene (Rosen, 1975). Another possibility is that the ancestral state reconstruction lacked adequate outgroup sampling to determine the state at the root of the tree with high certainty, and was thus overly influenced by the heavy sampling of Mesoamerican taxa. In either case, it is clear from the timing of this earliest divergence in *Jatropha* that the Isthmus of Panama was not involved.

In contrast, three unambiguous shifts from South America to Mesoamerica potentially facilitated by the Isthmus of Panama were also detected. Two of these shifts, *Jatropha cathartica* and *J. podagrica*, are concordant with a Middle Miocene estimate for

the Isthmus, whereas *J. gossypiifolia* var. *staphysagrifolia* arrived in Mesoamerica after the Pliocene estimate for the Isthmus. The timing for the movement of *J. cathartica* and *J. podagrica* could be interpreted as supporting evidence for an earlier established Isthmus of Panama, but given the abundance of intercontinental dispersal detected for other species of *Jatropha* it seems unnecessary to invoke an early connection between North and South America to explain this particular disjunction pattern.

Biogeographic analysis favored long distance dispersal as the mechanism by which *Jatropha* arrived in the Caribbean, which is concordant with other studies showing that the Caribbean flora is comprised of elements that have colonized from all over the world (Nieto-Blázquez et al., 2017). Two independent dispersal events from Mesoamerica in the Late Miocene were inferred, one of which led to a subsequent radiation in the Greater Antilles (assuming monophyly of *J.* sub. *Jatropha.* sect. *Polymorphae*). Our findings are in alignment with those of Cervantes et al., (2016) who inferred Mesoamerica as the source of multiple Caribbean lineages of Euphorbiaceae subfamily Acalyphoideae. Greater sampling of species in the Greater Antilles would help to establish the role of inter-island dispersal and vicariance from changes in sea level.

The mechanism(s) by which *Jatropha* has undergone so many independent intercontinental dispersals is(are) not immediately clear. Neither of *Jatropha*'s means of dispersal—explosive dehiscence (Dehgan, 2012) and ant dispersal due to an oil-rich elaiosome (Leal et al., 2007)--can account for movement at the intercontinental scale. Standley (1923) observed that the seeds of *J. gossypiifolia*, which possess elaiosomes, "are eaten by doves and domestic fowls." All instances of intercontinental dispersal inferred in our analyses involved lineages from *J.* subg. *Jatropha*, possessing of an elaiosome so birds are the most likely candidates for zoochorous intercontinental dispersal in *Jatropha.* Alternatively, colonization of the Caribbean from Mesoamerica has

been attributed to ocean currents moving from west to east through the Central American Seaway during the Miocene (Nieto-Blázquez et al., 2017; Sepulchre, et al., 2013). The Caribbean species *J. multifida* is an exception in that it lacks an elaiosome, and may therefore have required ocean currents for its successful colonization, rather than birds.

The Evolution of **Jatropha** *in the Seasonally Dry Forests of Mesoamerica*

Biogeographic reconstruction inferred the ancestral distribution for the MRCA of *Jatropha* subg. *Curcas* as northwest Mexico and the dry forests of the Yucatan Peninsula. The latter because of the phylogenetic position of *J. gaumeri*, the only species endemic to Yucatan and the sister species of *J. integerrima* and *J. hernandiifolia* from the Caribbean. Taking a broader phylogenetic perspective, a possible explanation for this geographic pattern could be that the MRCA of the clade containing *J. macrorhiza,* the African and Caribbean lineages, and *J.* subg. *Curcas* dispersed to Mesoamerica and was widely distributed from northwest Mexico to the Yucatan. Subsequent fragmentation during the Miocene, likely caused by uplift of the Sierra Madre Occidental and Central Plateau, and a long distance dispersal to Africa, resulted in the four observed lineages (Mastretta-Yanes et al., 2015).

The scale and scope of the present study give a rough idea of the timing for divergence between species of *Jatropha* on the Peninsula and continental Mexico. Diversification of major lineages in *Jatropha* subg. *Curcas* occurred in the Baja California-Sonoran Desert and Guerreran SDTF-Balsas Depression regions four to six million years ago at the Miocene-Pliocene boundary. We found support for allopatric speciation due to vicariance resulting from at least three tectonic events: the formation of Gulf of California, uplift of the Sierra Madre Occidental, and uplift of the Trans-Mexican

Volcanic Belt (TVB). The initial uplift of the TVB in the Middle Miocene predated diversification within of *Jatropha* in Mesoamerica. However, the second phase of volcanism (7.5-3 Mya) centered in the eastern TVB was synchronous with the northern and southern lineages of clade C-4, which are separated by the TVB (García-Palomo et al., 2002). This phase of uplift of the TVB has been inferred as vicariance in other plant lineages (e.g., Gándara and Sosa, 2014). Vicariance detected at the node leading to *J. dioica* and *J. cuneata* in the Chihuahuan and Sonoran Deserts respectively is explained by uplift in the northern Sierra Madre Occidental, as has been observed in other taxa (Bell et al., 2010; Mastretta-Yanes et al., 2015; Provost et al., 2018). Formation of the Gulf of California split *J. vernicosa* from *J. cardiophylla* during the early Pliocene, and possibly split *J. moranii* from the rest of C-2 around the same time. The timing of events associated with the formation of the Baja Peninsula and dates for its last connection to mainland Mexico are extensively debated in the literature (see Hafner and Riddle, 2011). Finer scale sampling of *J. vernicosa* and *J. cardiophylla*, as well as the sister species *J. cinerea* which is distributed on both sides of the Gulf of California, would help elucidate the precise timing of movement of *Jatropha* onto the Baja Peninsula.

Dispersal across the Trans-Mexican Volcanic Belt happened in two bursts. We inferred dispersals during the Miocene from north to south once each within clades C-1, C-2, and C-4, with the dispersal in C-4 alternatively inferred as a vicariance event related to the uplift of the TVB in some BSM iterations. In the Pliocene *J. cordata, J. platyphylla, J. mcvaughii,* and *J. malacophylla* independently expanded their ranges north from the Guerreran into the Sinaloan coastal forest. Most dispersal between areas of seasonally dry tropical forest south of the TVB (i.e., between the Guerreran SDTF and Balsas Depression, Isthmus of Tehuantepec, and Tehuacan Valley). This is reasonable as

restrictions to movement between these areas are small relative to crossing the TVB or Sierra Madre Occidental.

We found mixed support for the predictions of Pennington et al., (2009) about patterns of diversification in lineages specialized to the seasonally dry tropical forest. Distance regressions for species of *Jatropha* from the seasonally dry tropical forests of Mesoamerica were marginally significant, whereas phylogenetic community analysis indicated significant phylogenetic structure for dry forest regions with more than two samples. Preliminary phylogenetic community and distance regression analyses of *Jatropha* in South America shows similar findings, but increased taxon sampling is needed to improve the statistical strength for comparison to findings for Mesoamerica. *Jatropha* is well represented in the SDTF on three continents, and a larger scale analysis of other groups using the approach presented here would help determine whether the predictions about plant diversification in this biome are more generally applicable across continents and clades.

CONCLUSIONS

Jatropha diversified in the dry tropics during the Oligocene and Miocene, a time of global aridification. As with many pantropical groups of flowering plants, application of molecular data for estimating divergence times shows long distance dispersal to be the primary force behind intercontinental disjunctions. The African lineage of *J.* subg. *Jatropha* was inferred to have dispersed from Mesoamerica, whereas *J. mahafalensis* arrived in Madagascar from South America. The possible mechanisms for long distance dispersal in *Jatropha* zoochory by birds and ocean currents for Caribbean species *J. multifida*, but more work is needed to demonstrate the viability of these

methods. It is unclear whether *Jatropha* arrived in Africa via trans-Atlantic or trans-Pacific dispersal or both, and sampling of Arabian and Indian species would be of great help in answering this question.

Estimated times for two dispersal events from South America to Mesoamerica for *Jatropha* are concordant with a Miocene establishment of the Isthmus of Panama, however as long distance dispersal cannot be ruled out as an alternative explanation for these range shifts we cannot conclude that our findings strongly support an earlier date for land connection between the continents. Two independent introductions of *Jatropha* to the Greater Antilles from Mesoamerica were inferred as long distance dispersal events, one of which resulted in a radiation.

We found mixed support for the hypothesis that lineages of plants specialized to seasonally dry tropical forests, a unique and threatened habitat, exhibit phylogenetic signatures of dispersal limitation. Phylogenetic community analysis of *Jatropha* in Mesoamerica indicated that seasonally dry tropical forest specific lineages are geographically structured, whereas desert lineages are not. The trend from distance regression methods in agreement, but the relationship between genetic and geographic distance between species was only marginally significant. Increased taxon sampling of South American and Old World lineages would be beneficial for intercontinental comparisons of this pattern.

Ancestral state reconstruction inferred the most recent common ancestor for *J.* sub. *Curcas* to have been in northwest Mexico and the Yucatan Peninsula, and major clades originated in northwestern Mexico and the Guerreran seasonally dry tropical forests along the Pacific slope of the Sierra Madre Occidental and Sierra Madre del Sur. The formation of the Gulf of California, and uplift of the northern Sierra Madre Occidental and the Trans-Mexican Volcanic Belt (TVB) all impacted diversification of *J.*

subg. *Curcas*. Uplift of the TVB largely predated diversification in *J.* subg. *Curcas*, but likely created a barrier between the Central Plateau and the Balsas Depression in the early Pliocene and restricted dispersal of *Jatropha* between northern and southern Mexico during this time. The barrier of the TVB was not complete, however, as several cases of northward range expansions along the Pacific slopes of the Sierra Madre Occidental were inferred.

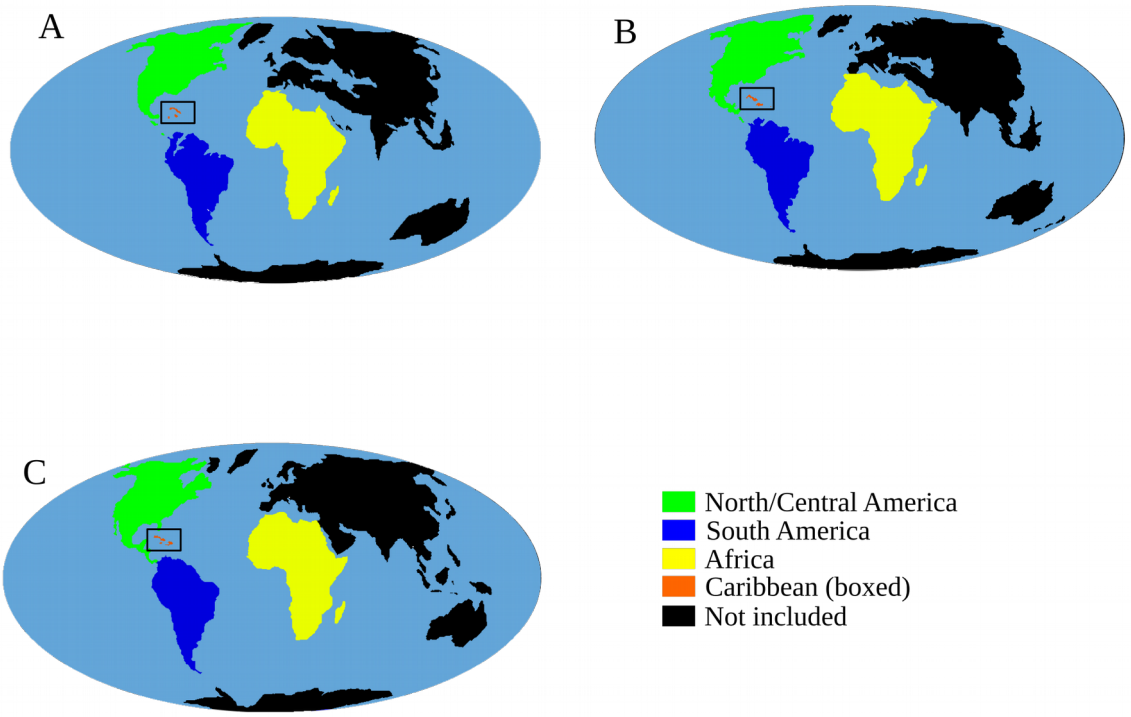


Figure 2-1: Positions of continental landmasses at A) Oligocene (36 mya), B) Miocene (20 mya), and C) Present.

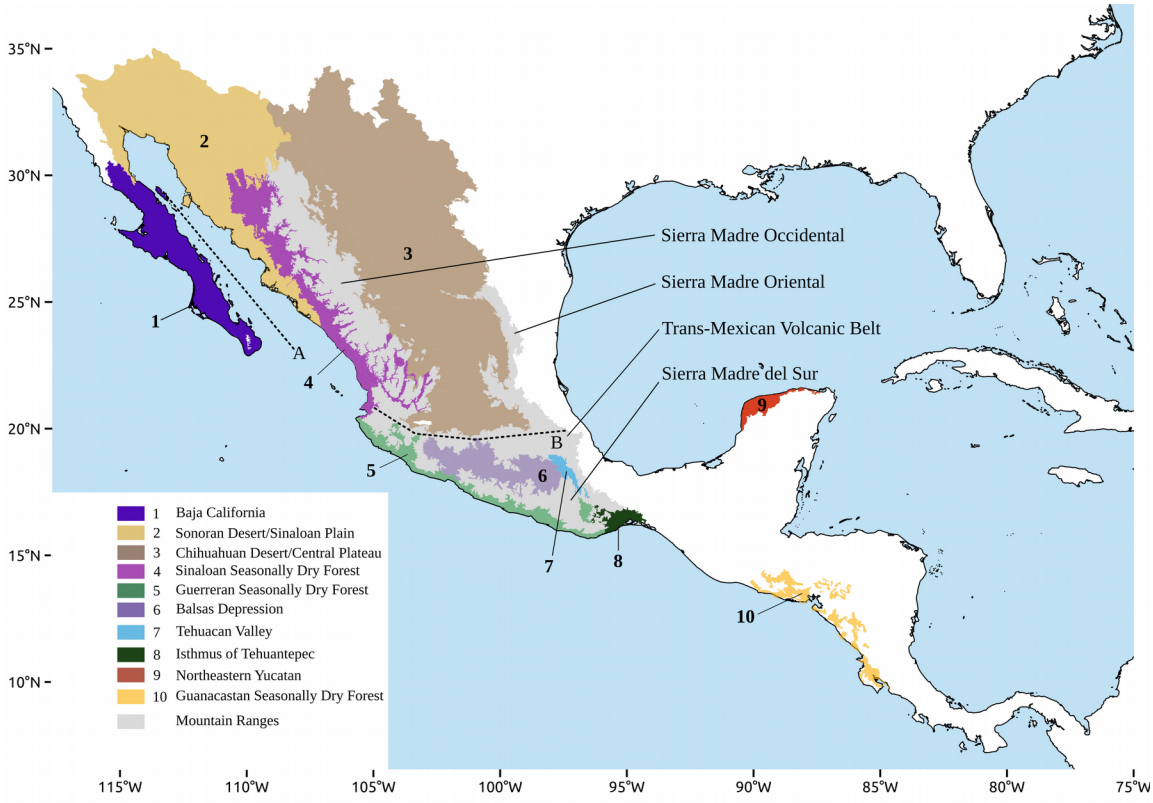


Figure 2-2: Map showing major mountain ranges and bioregions (indicated by numbers in the legend) used for biogeographic reconstruction for *Jatropha* subg. *Curcas* in Mesoamerica. Dotted lines show present day locations for geologic events encoded in dispersal models: A) the formation of the Gulf of California and B) uplift of the Trans-Mexican Volcanic Belt

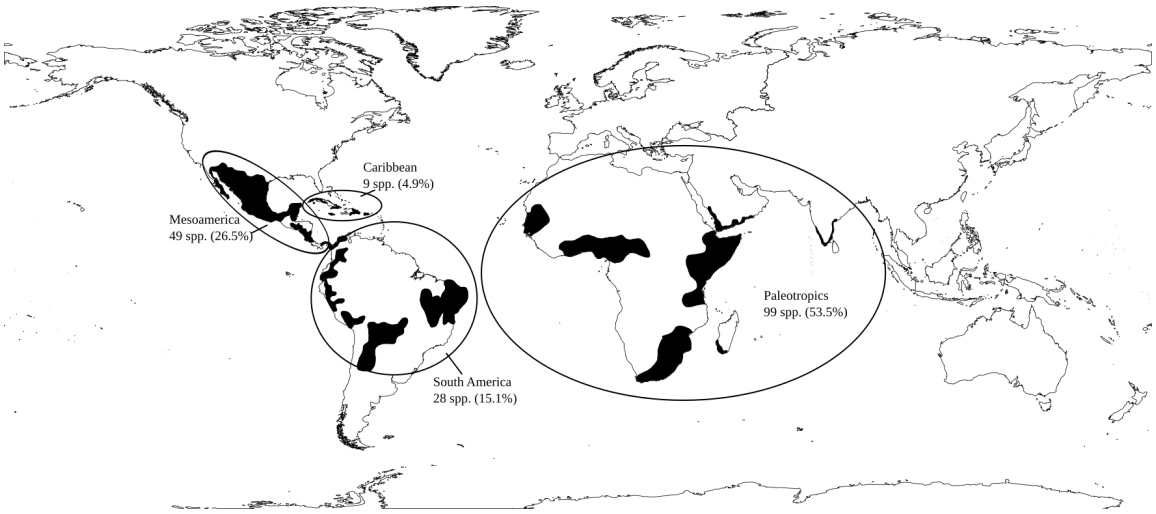


Figure 2-3: Distribution of *Jatropha* based upon all available georeferenced records from GBIF (accessed September 15, 2018) excluding widely cultivated or weedy species (i.e. *J. curcas*, *J. gossypiifolia*, and *J. multifida*).

Scale Bioregions	Events: *Dates*	Source
Intercontinental		
North/Central America	Breakup of Gondwana: *105 mya*	McLoughlin, 2001
South America	North Atlantic Land Bridge: *45 mya*	Tiffney, 2000
Caribbean	GAARlandia: *35-32 mya*	Iturralde-Venent, 2006
Africa	Miocene Isthmus of Panama: *10-15 mya*	Montes et al, 2015
	Pliocene Isthmus of Panama: *3.5 mya*	O'dea et al., 2016
Mesoamerica		
Baja California	Uplift of Trans-Mexican Volcanic Belt	Gómez et al., 1996
Sonoran Desert	-Middle to late-Miocene arc: *19-10mya*	
Chihuahuan Desert	-Silicic: *7.5-5mya*	
Sinaloan SDTF	-Bimodal: *5-3 mya*	
Guerreran SDTF	-Late Pliocene-Quarternary arc: *3.5mya-present*	
Balsas Depression	Gulf of California formation: *5.5 mya*	Carreno and Helenes, 2002
Tehuacan Valley		
Isthmus of Tehuantepec		
Northeastern Yucatan		
Guanacastan SDTF		

Table 2-1: Overview of biogeographic analyses. The first column indicates the scale or area of focus for an analysis along with the bioregions used. Bioregions marked with the same symbol were combined in analyses that reduced the number of bioregions. Specific events that were investigated for association with diversification in *Jatropha* are given with sources for dates in columns two and three.

Euphorbiaceae

Node	Taxon/Calibration	Age constraint and distribution in mya	Source
1	*Crepetocarpon*	exponential (μ = 1.0, offset = 40)	Dilcher and Manchester, 1988
2	*Acalyphaepollenites*	exponential (μ = 1.0, offset = 61)	Sun et al., 1989
3	*Croton* (crown)	normal (μ = 38.4, σ = 0.8)	van Ee et al., 2008
4	Euphorbiaceae (crown)	normal (μ = 89.9, σ = 5)	Xi, 2012

Jatropha

1	Stem	normal (μ = 40.5, σ = 9.0)
2	Crown	normal (μ = 24.6, σ = 6.0)

Table 2-2: Top) Fossil (nodes 1 and 2) and secondary calibration (nodes 3 and 4) ages used for dating analyses of Euphorbiaceae. Bottom) Priors used for divergence dating within *Jatropha*.

3.5-P	South America	North America	Africa	Caribbean
South America	-	1	0.2	0.8
North America		-	0.1	0.8
Africa			-	0.1
Caribbean				-

20-3.5	South America	North America	Africa	Caribbean
South America	-	0.8	0.2	0.8
North America		-	0.1	0.8
Africa			-	0.1
Caribbean				-

41-20	South America	North America	Africa	Caribbean
South America	-	0.5	0.3	0.8
North America		-	0.2	0.7
Africa			-	0.2
Caribbean				-

Table 2-3: Dispersal matrices for global time-stratified analysis in BioGeoBEARS. Time ranges (millions of years) for which each matrix was implemented are encircled to the left of the matrix. P = Present.

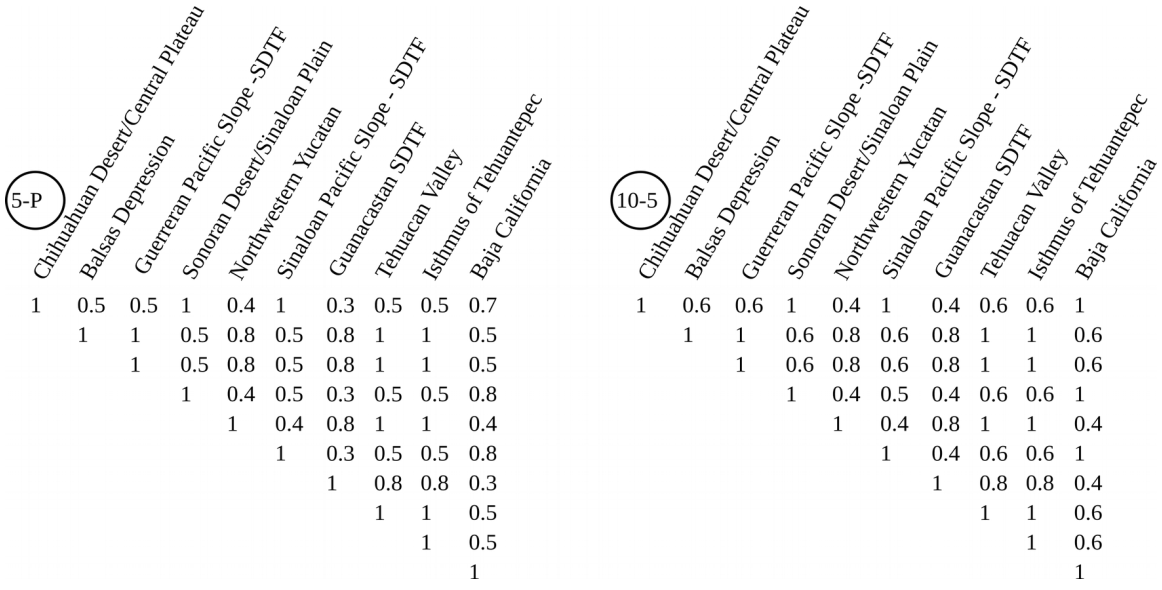


Table 2-4: Dispersal matrices for Mesoamerican time-stratified analysis in BioGeoBEARS. Time ranges (millions of years) that each matrix was enforced are encircled to the left of the matrix. P = Present.

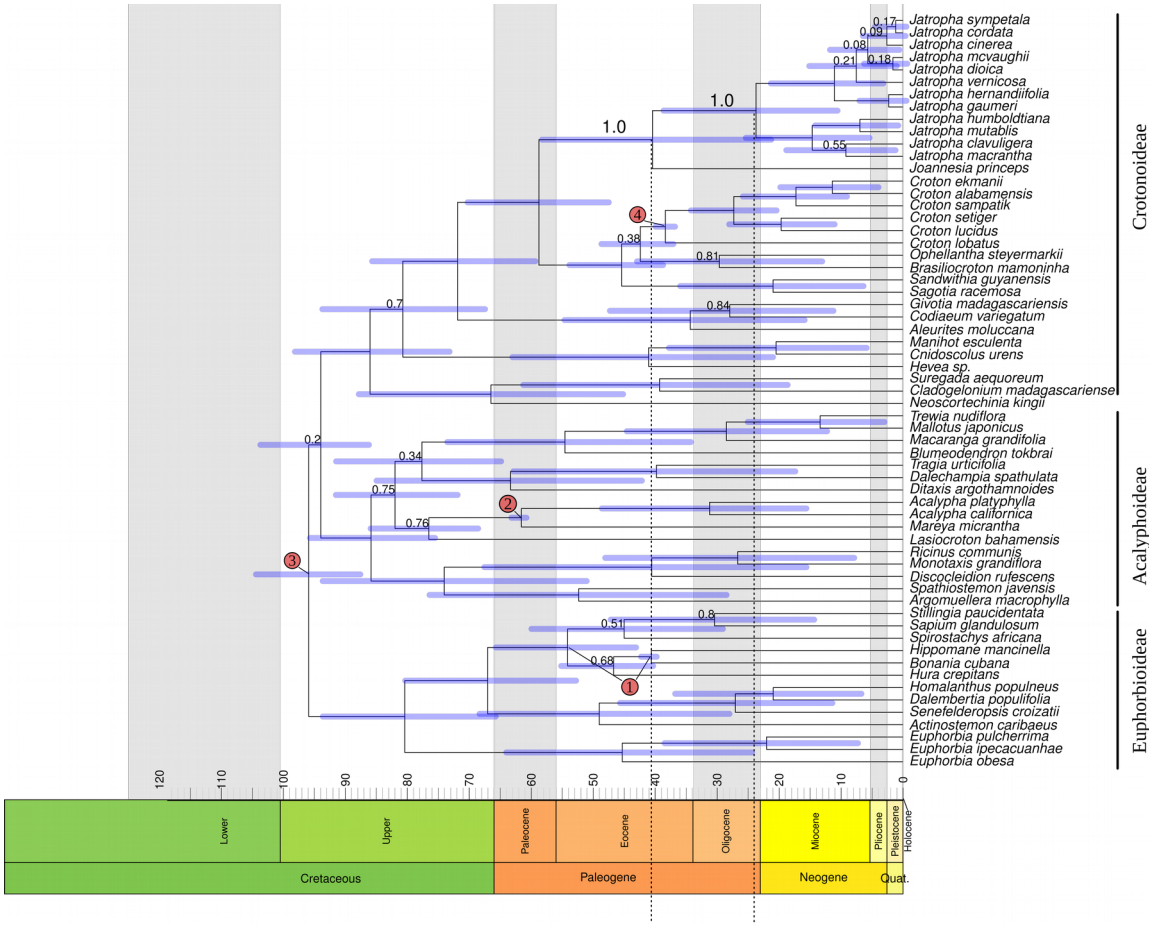


Figure 2-4: Dated BEAST chronogram for Euphorbiaceae from analysis of the chloroplast marker *rbcL*. Values on branches are Bayesian posterior probabilities (nodes without values BPP > 0.90). Circled numbers show calibration points described in Table 2. Blue bars are 95% confidence intervals of the median node ages. Dotted lines mark the stem and crown ages of *Jatropha* used in subsequent analysis.

Euphorbiaceae

	4 calibrations	No *Acalypha*	No *Crepetocarpon*	No *Croton*
Jatropha - stem	40.5 (58.3-21.3)	38.9 (57.3-20.4)	38.9 (57.1-21)	37.5 (62-16.7)
Jatropha - crown	24.6 (39.6-11.5)	23.4 (38.1-11.6)	23.1 (37.3-10.8)	22.7 (42.3-9.5)
Croton - crown	38.3 (39.9-36.7)	38.3 (39.9-36.7)	38.3 (39.8-36.8)	**25.4 (40.7-13.6)**
Acalypha - stem	61.8 (63.4-61)	**38.2 (53.7-22.6)**	61.8 (63.4-61)	61.8 (63.4-61)
Crepetocarpon	40.8 (42.5-40)	40.8 (42.5-40)	**19.9 (34.9-6.7)**	40.8 (42.5-40)
Euphorbiaceae - crown	93.1 (102.1-84.2)	93.1 (102.1-84.2)	94.9(104.5-86.1)	94.9(104.4-86.1)

Jatropha

Stem	40.1 (55.1-25.9)
Crown	22.2 (31.5-13.8)
S. Am	19.1 (27.5-11.6)
Africa	11 (15.9-6.2)
Caribb.	10 (11-3.85)
Curcas	6.4 (9.4-3.77)
C-1	6.11 (8.14-3.22)
C-2	6.02 (8.71-3.43)
C-3	6.02 (8.71-3.43)
C-4	6.4 (9.4-3.77)

Table 2-5: Results from divergence analyses in BEAST and cross-validation for Euphorbiaceae (top) and *Jatropha* (bottom). All values are median ages (millions of years) with the 95% highest probability densities (HPD) in parentheses. Bold ages denote nodes that were estimated in a cross-validation analysis, but were constrained in the full analysis.

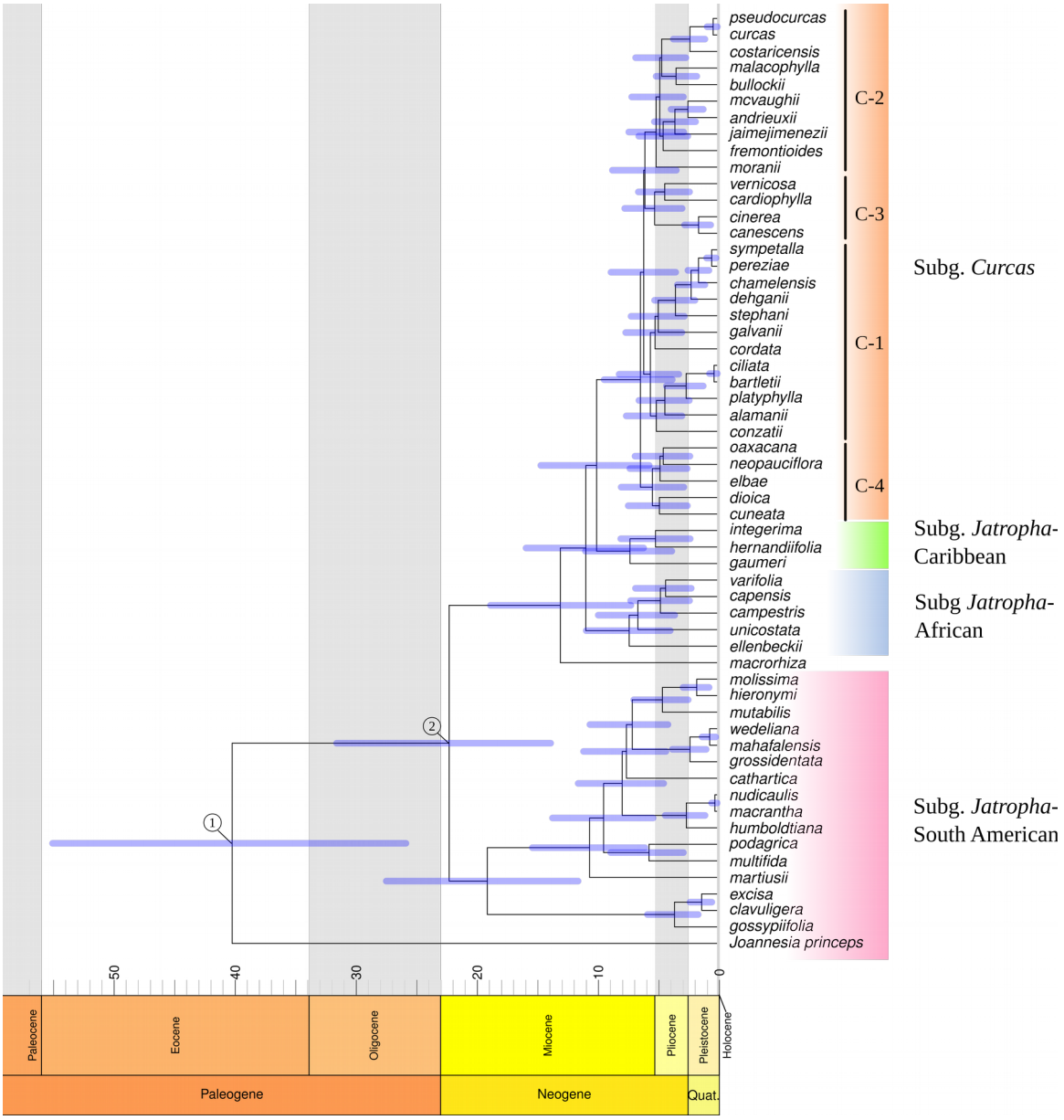


Figure 2-5: Dated BEAST Chronogram of *Jatropha* from analysis of RADseq dataset using a constrained topology from Chapter 1. Circled numbers are calibrated nodes based on the family wide divergence analysis (see Figure 4 and Table 4). Blue bars are 95% HPD intervals of median node ages.

	free parameters	*d*	*e*	*j*	*w*	*log likelihood*	*AIC*	*weighted AIC*	*validation R2 - intercept- slope*
A: Global									
DEC+*j* Unconstrained	3	1.0e-12	1.0e-12	0.025	1.00	-35.30	76.60	0.31	**1 - 0 - 0.999**
DIVALIKE+*j* Unconstrained	3	1.0e-12	1.0e-12	0.028	1.00	-36.04	78.08	0.15	**1 - 0 - 0.999**
DEC+*j*+*w* Time-stratified	4	1.0e-12	1.0e-12	0.033	0.25	-35.17	78.33	0.13	**1 - 0 - 0.999**
B: Mesoamerica									
DIVALIKE+*j*+*w* Time-stratified	4	0.019	1.0e-12	0.102	1.53	-93.91	195.8	0.42	**0.87 - 0 - 0.93**
DIVALIKE+*j* Time-stratified	3	0.016	1.0e-12	0.083	1.00	-94.99	196.0	0.38	**0.87 - 0 - 0.93**
DEC+*j* Time-stratified	3	0.014	1.0e-12	0.098	1.00	-96.12	198.2	0.13	**0.84 - 0 - 0.94**

Table 2-6: Parameter estimates for biogeographic reconstruction of *Jatropha* at (A) the global scale and (B) for *J.* subg. *Curcas* in Mesoamerica. Parameters are: dispersal (*d*), extinction (*e*), jump-dispersal (*j*), and the scaling parameter for dispersal matrices (*w*) from each of the best scoring models from BioGeoBEARS. Weighted AICc is the probability of being the best model among all those tested. The last column show regression values from validating biogeographic stochastic mapping.

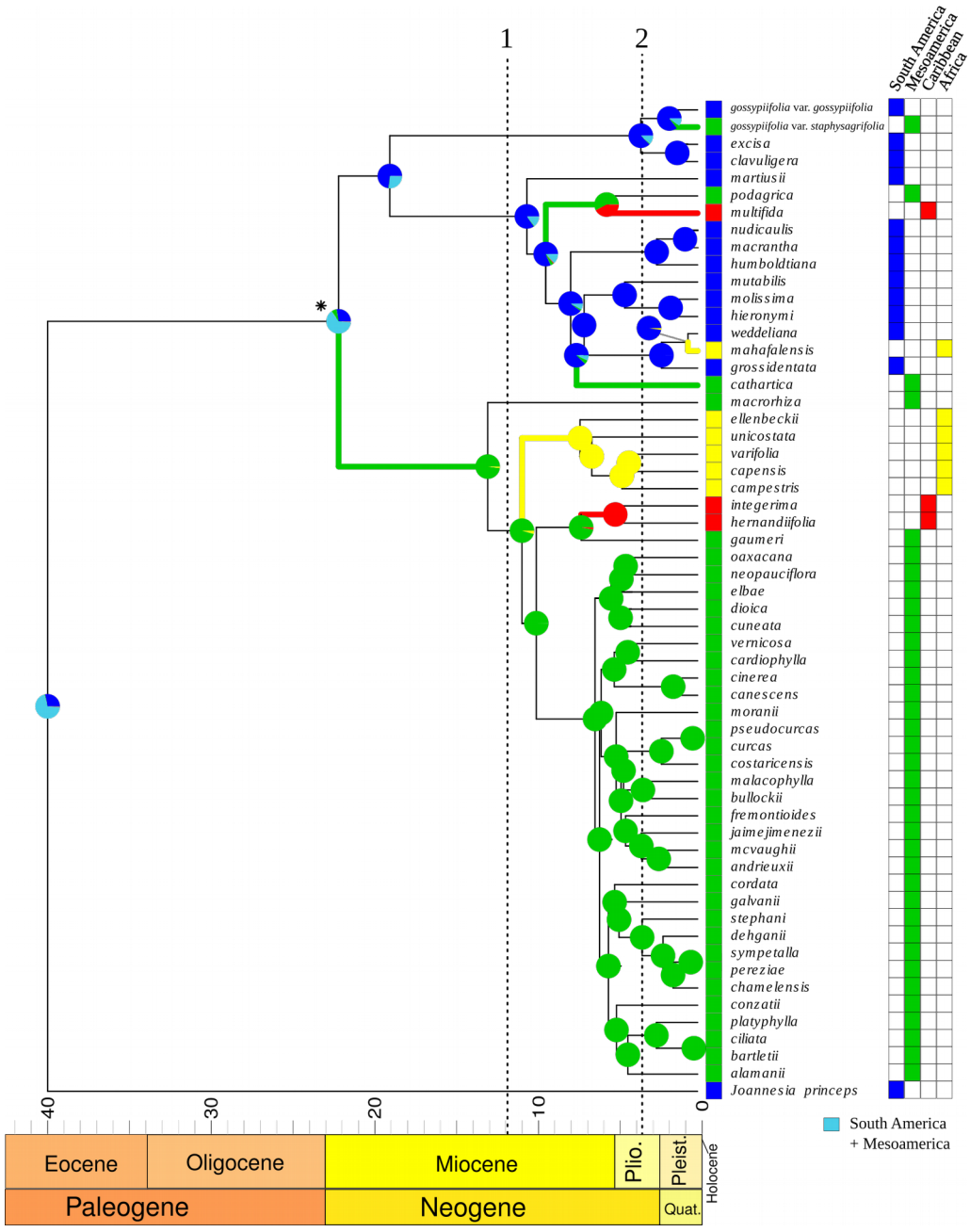


Figure 2-6: Biogeographic reconstruction of *Jatropha* (model = DEC+j unconstrained). Pie charts show the probabilities of different ancestral areas. Colored branches indicate intercontinental dispersal events, and the asterisks was inferred as a vicariance event. Dotted lines show: 1) the Middle Miocene and 2) Pliocene ages of the Isthmus of Panama.

From \ To	South America	Mesoamerica	Africa	Caribbean
South America	-	2.45 (0.91)	1 (0.063)	0.42 (0.49)
Mesoamerica	0.23 (0.6)	-	0.97 (0.16)	1.58 (0.5)
Africa	0.036 (0.2)	0.054 (0.3)	-	0 (0)
Caribbean	0.006 (0.077)	0.45 (0.54)	0.002 (0.045)	-

Table 2-7: Mean estimated intercontinental dispersal events by area for *Jatropha*. Determined from biogeographic stochastic mapping in BioGeoBEARS (standard deviations in parentheses). Warmer colors indicate higher frequencies of dispersal from (rows-names) source areas to (column-names) destination areas.

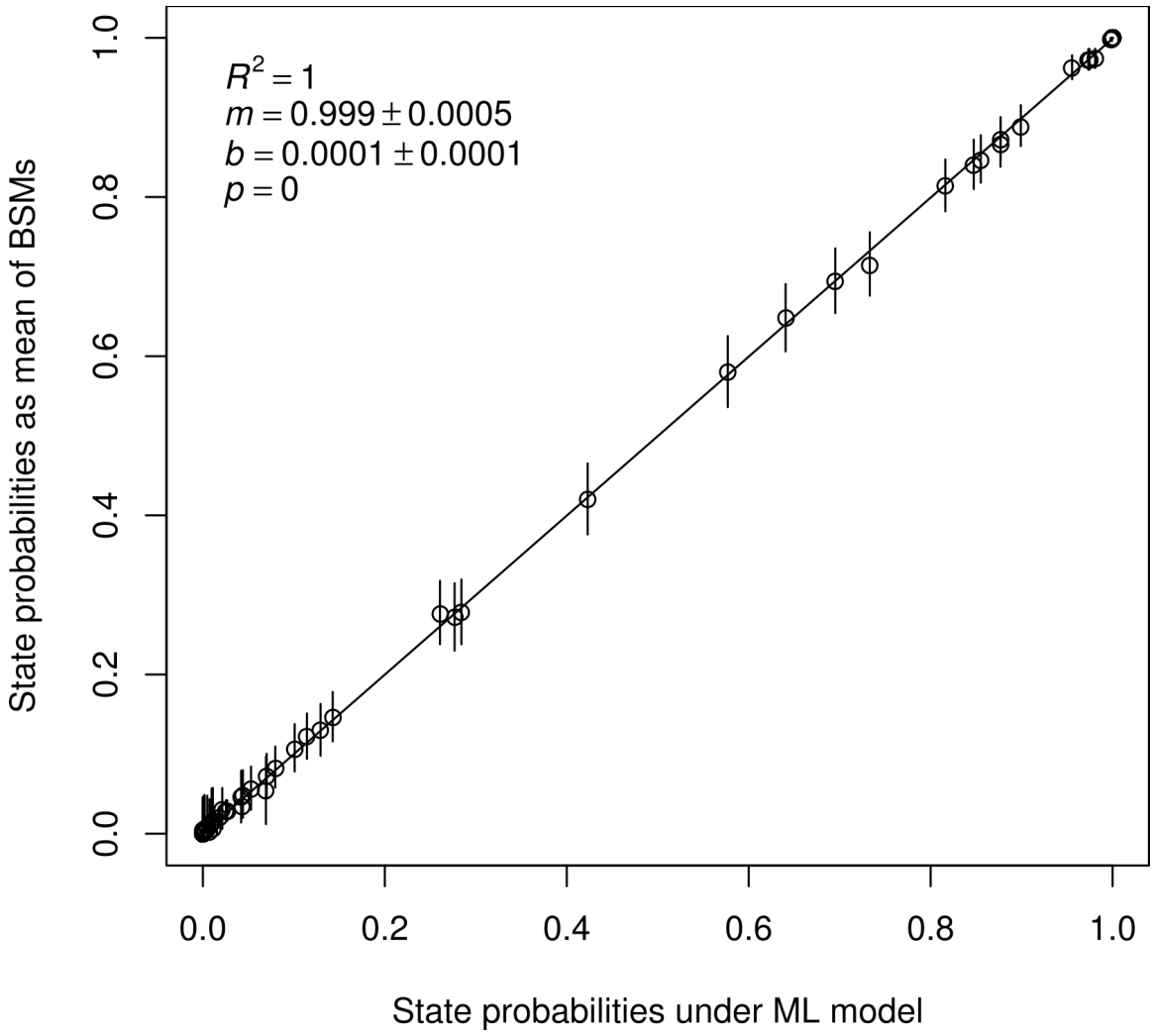


Figure 2-7 Linear regression for validation of biogeographic stochastic mapping (BSM) of ancestral areas at nodes for global scale biogeographic reconstruction of J*atropha* using the DEC+j model.

Scale Model	founder events	anagenetic dispersal	vicariance	narrow sympatry
Global				
DEC+j unconstrained	7.21 (1.01)	0 (0)	0.72 (0.45)	47.46 (1.32)
DIVALIKE+j unconstrained	8.3 (0.54)	0 (0)	0.38 (0.48)	48.32 (0.72)
DEC+j+w TS	7.21 (1.01)	0 (0)	0.72 (0.45)	47.46 (1.32)
Mesoamerica				
DIVALIKE+j+w TS	14.14 (1.93)	13.32 (1.34)	3.94 (1.43)	10.92 (1.5)
DIVALIKE+j	13.78 (1.88)	13.38 (1.37)	4.02 (1.49)	11.19 (1.48)
DEC+j TS	13.51 (1.97)	0 (0)	2.83 (1.23)	10.11 (1.72)

Table 2-8: Summary of the number of biogeographic events inferred from biogeographic stochastic mapping (BSM) using the top scoring models from BioGeoBEARS. Values are the mean numbers of events estimated from 500 realizations of BSM (standard deviation in parentheses).

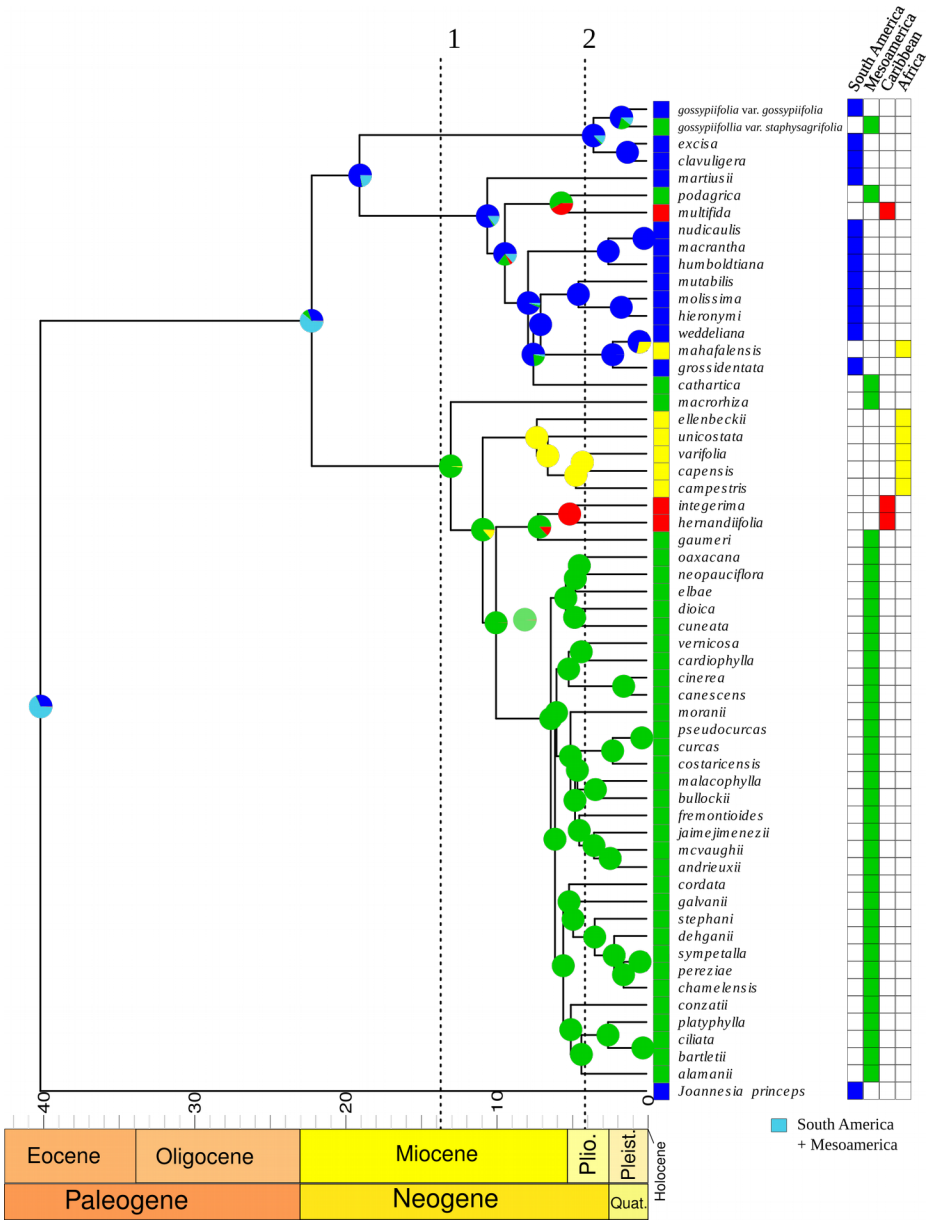


Figure 2-8 Biogeographic reconstruction of *Jatropha* (model = DEC+j + w time-stratified). Pie charts show the probabilities of different ancestral areas. Colored branches indicate intercontinental dispersal events, and the asterisks was inferred as a vicariance event. Dotted lines show: 1) the Middle Miocene and 2) Pliocene ages of the Isthmus of Panama.

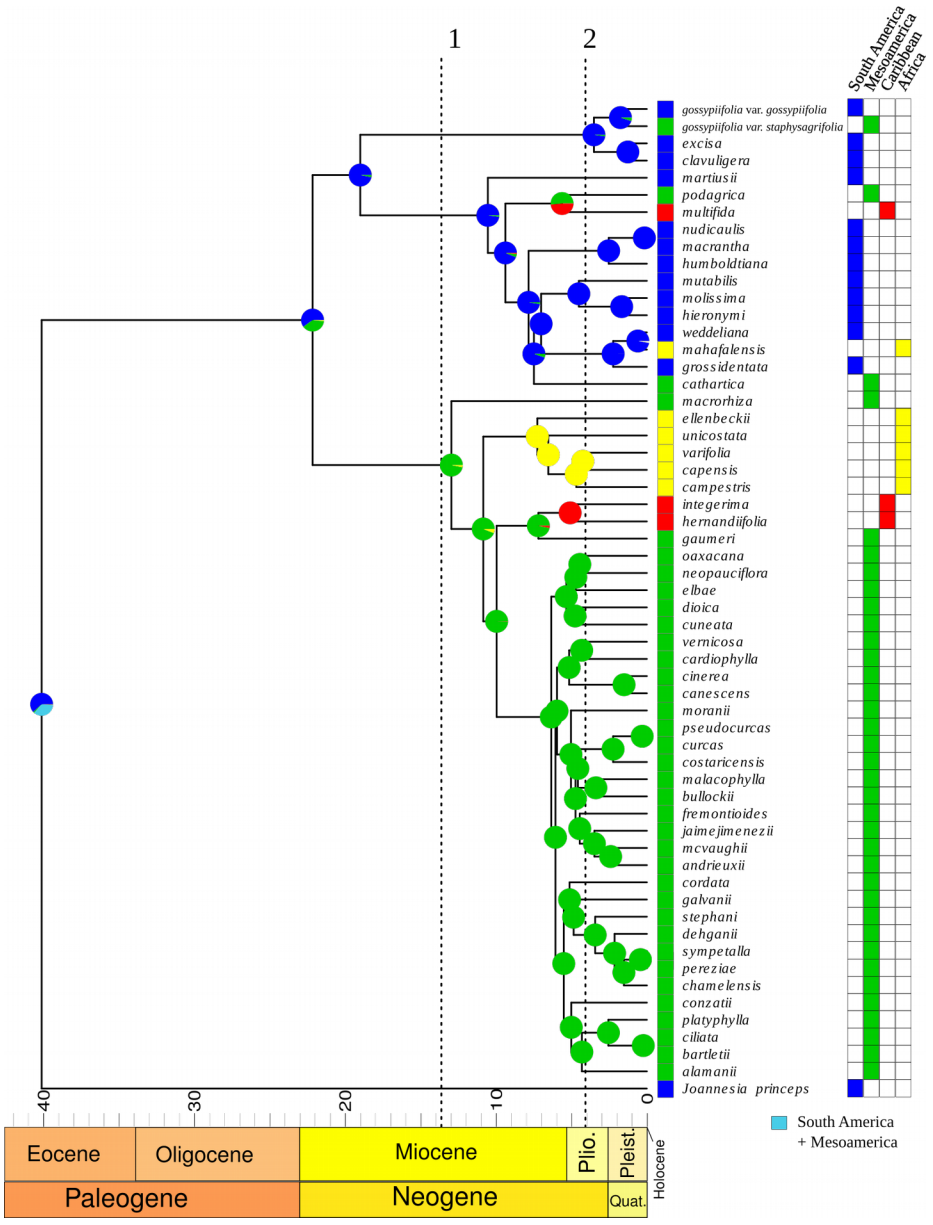


Figure 2-9 Biogeographic reconstruction of *Jatropha* (model = DIVALIKE+j unconstrained). Pie charts show the probabilities of different ancestral areas. Colored branches indicate intercontinental dispersal events, and the asterisks was inferred as a vicariance event. Dotted lines show: 1) the Miocene and 2) Pliocene ages of the Isthmus of Panama.

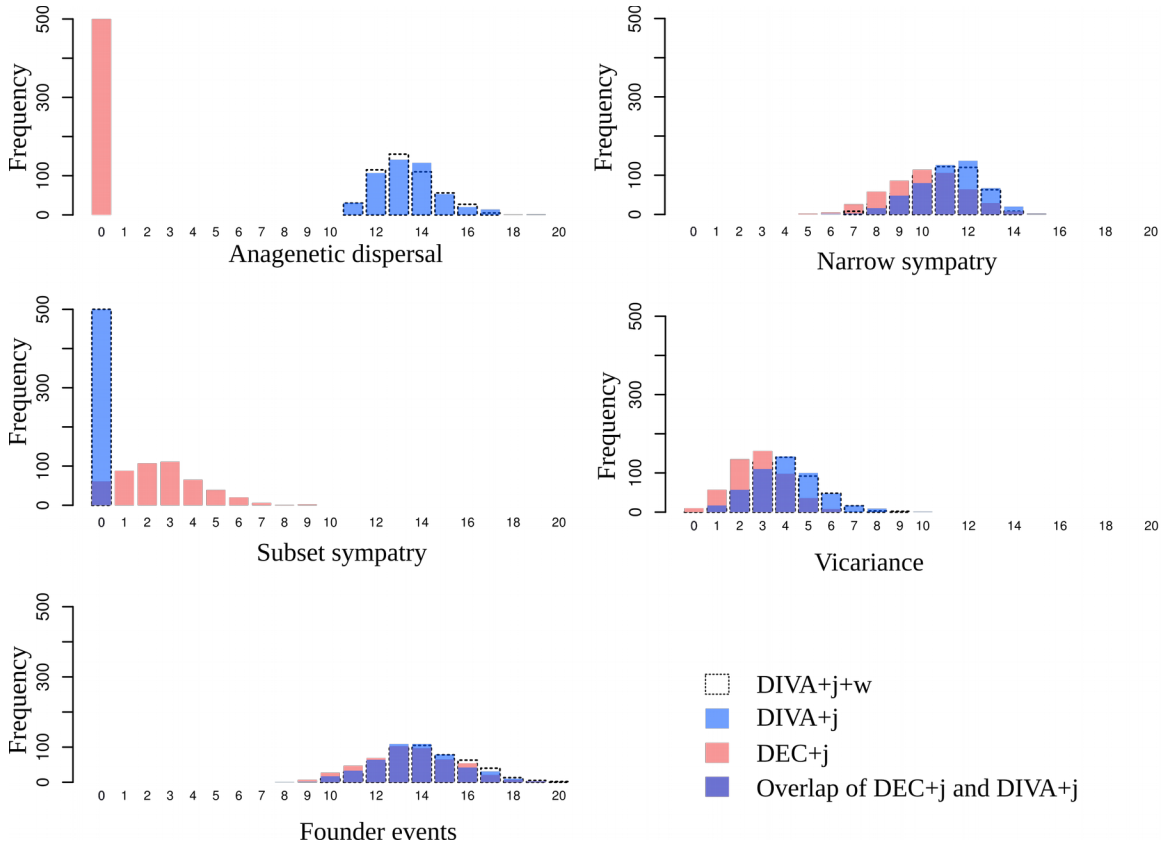


Figure 2-10: Histograms summarizing the number of biogeographic events estimated from 500 realizations of biogeographic stochastic mapping using the top three scoring models from the global biogeographic analysis of the Mesoamerican clade *Jatropha.* subg. *Curcas.*

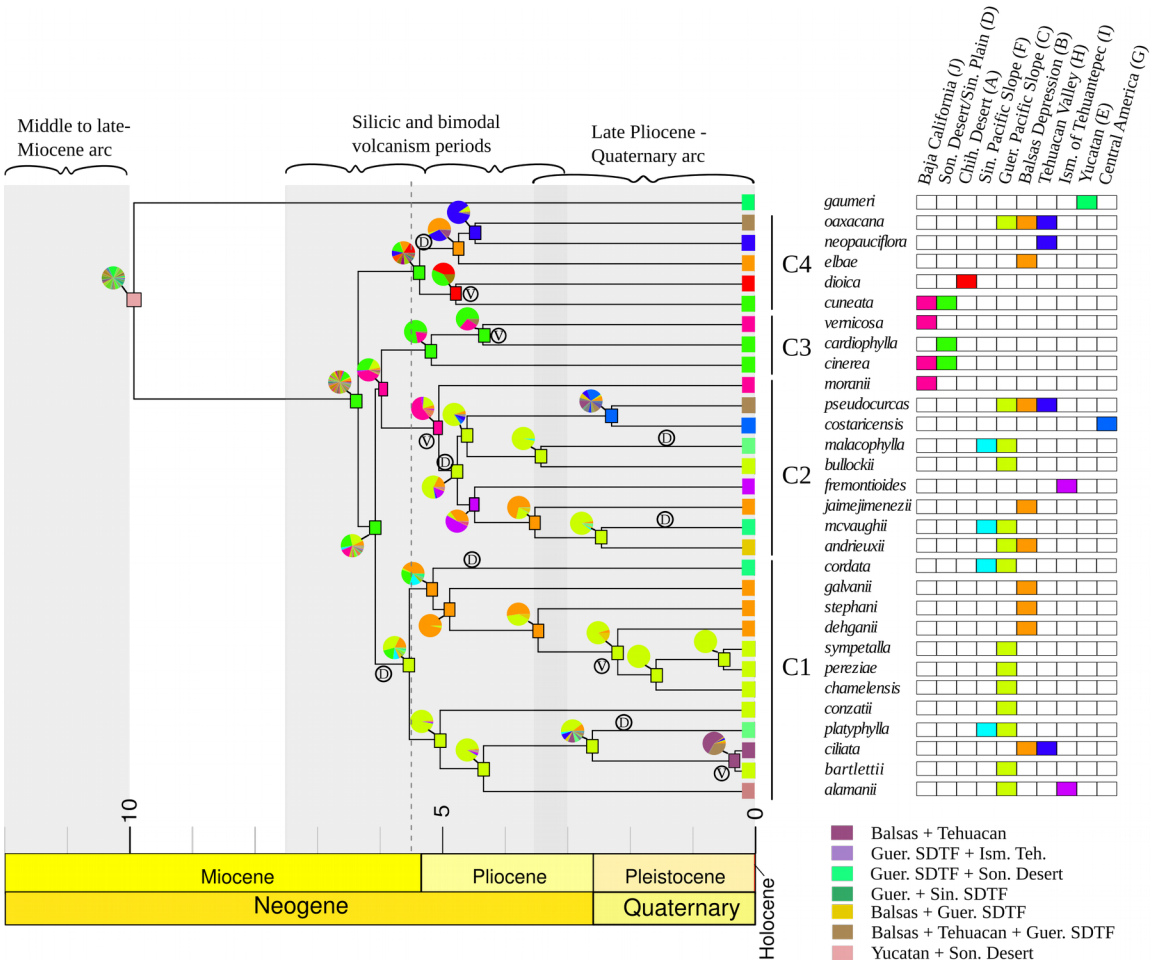


Figure 2-11: Biogeographic reconstruction of *Jatropha* subg. *Curcas* (model = DIVALIKE+j+w). Pie charts show the probabilities of different ancestral areas and vicariance and dispersal events are marked by circled ‘V’ and ‘D’ respectively. Gray boxes show major periods of volcanism in the Trans-Mexican Volcanic Belt and dotted line indicates formation of the Gulf of California.

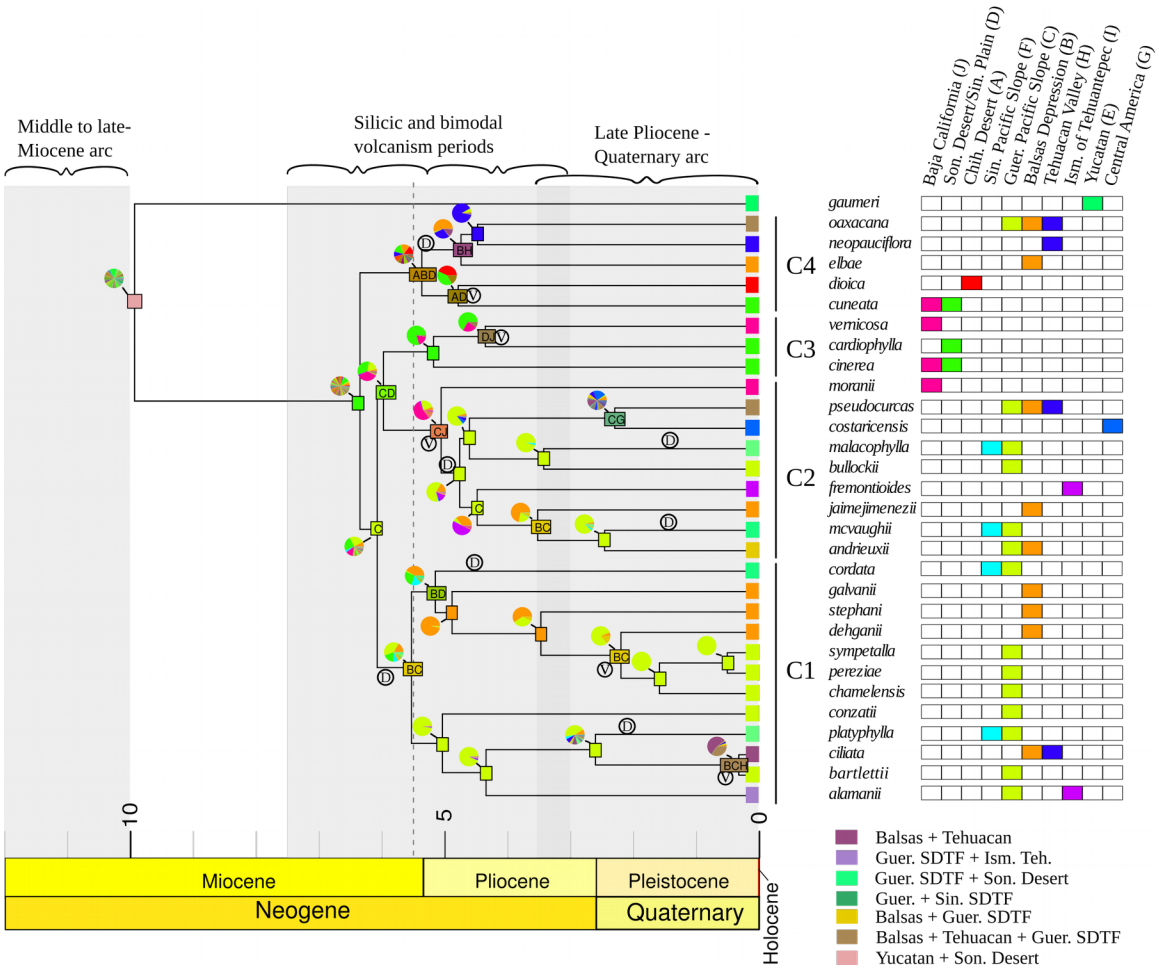


Figure 2-12 Biogeographic reconstruction of *Jatropha* subg. *Curcas* (model = DIVALIKE+j time stratified). Pie charts show the probabilities of different ancestral areas and vicariance and dispersal events are marked by circled ‘V’ and ‘D’ respectively. Gray boxes show major periods of volcanism in the Trans-Mexican Volcanic Belt and dotted line indicates formation of the Gulf of California.

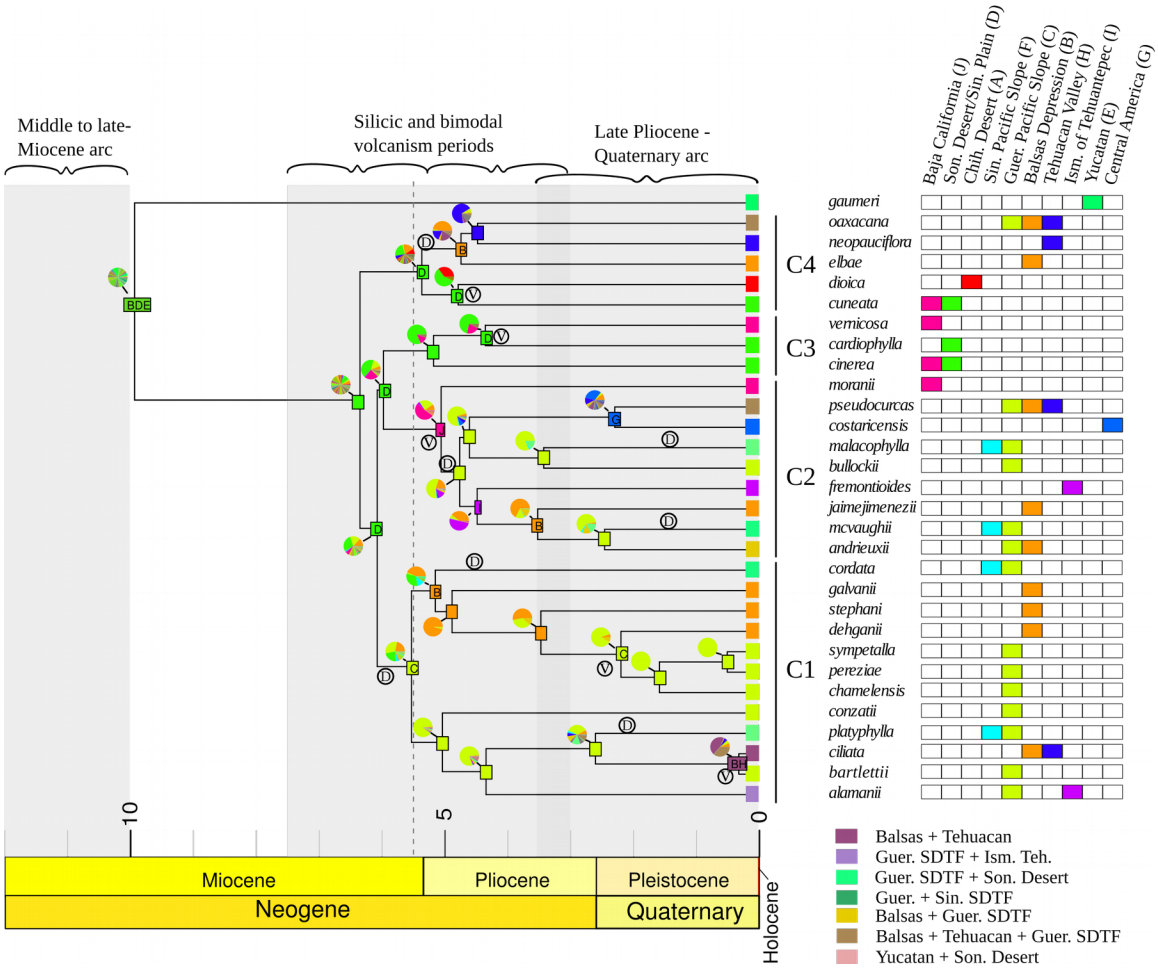


Figure 2-13 Biogeographic reconstruction of *Jatropha* subg. *Curcas* (model = DEC+j time stratified). Pie charts show the probabilities of different ancestral areas and vicariance and dispersal events are marked by circled ‘V’ and ‘D’ respectively. Gray boxes show major periods of volcanism in the Trans-Mexican Volcanic Belt and dotted line indicates formation of the Gulf of California.

	BC->B,C	BCH->C,BH	AD->A,D	DJ->D,J	BD->B,D
% present in BSM	38.6	39.4	18.2	14.4	13.6

Table 2-9: Top five most frequent vicariance events inferred from biogeographic stochastic mapping (BSM) in for *Jatropha* subg. *Curcas*. Area codes: (A) Chihuahuan Desert, (B) Balsas Depression, (C) Guerreran SDTF, (D) Sonoran Desert, (E) Yucatan, (H) Tehuacan Valley, and (J) Baja California.

From \ To	chihuahuan desert	balsas depression	guerreran sdtf	sonoran desert	yucatan	sinaloan sdtf	central america	tehuacan valley	isthmus of tehuantepec	baja california
chihuahuan desert	-	0.15 (0.37)	0.028 (0.17)	0.52 (0.55)	0.01 (0.1)	0.008 (0.089)	0 (0)	0.084 (0.28)	0 (0)	0.032 (0.18)
balsas depression	0.1 (0.3)	-	3.05 (1.44)	0.41 (0.58)	0.052 (0.22)	0.35 (0.55)	0.23 (0.42)	1.59 (1.05)	0.3 (0.47)	0.074 (0.27)
guerraran sdtf	0.018 (0.13)	3.31 (1.46)	-	0.32 (0.59)	0.03 (0.17)	2.92 (0.48)	0.55 (0.5)	1.38 (0.9)	1.48 (0.55)	0.32 (0.49)
sonoran desert	0.49 (0.5)	0.49 (0.62)	0.36 (0.58)	-	0.016 (0.13)	0.57 (0.5)	0.002 (0.045)	0.078 (0.27)	0.026 (0.16)	1.05 (0.68)
yucatan	0.002 (0.045)	0.004 (0.063)	0 (0)	0.002 (0.045)	-	0 (0)	0 (0)	0.006 (0.077)	0 (0)	0 (0)
sinaloan sdtf	0.012 (0.11)	0.22 (0.44)	0.39 (0.63)	0.44 (0.52)	0.004 (0.063)	-	0 (0)	0.026 (0.17)	0.006 (0.077)	0.046 (0.21)
central america	0 (0)	0.056 (0.24)	0.092 (0.35)	0 (0)	0 (0)	0 (0)	-	0.032 (0.18)	0.004 (0.063)	0.006 (0.077)
tehuacan valley	0.066 (0.25)	1.54 (1.06)	1.09 (0.96)	0.11 (0.34)	0.03 (0.17)	0.028 (0.17)	0.19 (0.4)	-	0.016 (0.13)	0.016 (0.14)
isthmus of tehuantepec	0.002 (0.045)	0.28 (0.49)	0.45 (0.72)	0.026 (0.17)	0.008 (0.089)	0.006 (0.077)	0.016 (0.13)	0.022 (0.15)	-	0.052 (0.22)
baja california	0.03 (0.17)	0.19 (0.43)	0.45 (0.56)	0.84 (0.94)	0.014 (0.12)	0.048 (0.21)	0.006 (0.077)	0.022 (0.15)	0.14 (0.35)	-

Table 2-10: Mean estimated intercontinental dispersal events by area from biogeographic stochastic mapping for the Mesoamerican clade *Jatropha*. subg. *Curcas* in BioGeoBEARS (standard deviations in parentheses). Warmer colors indicate higher frequencies of dispersal from (rows-names) source areas to (column-names) destination areas.

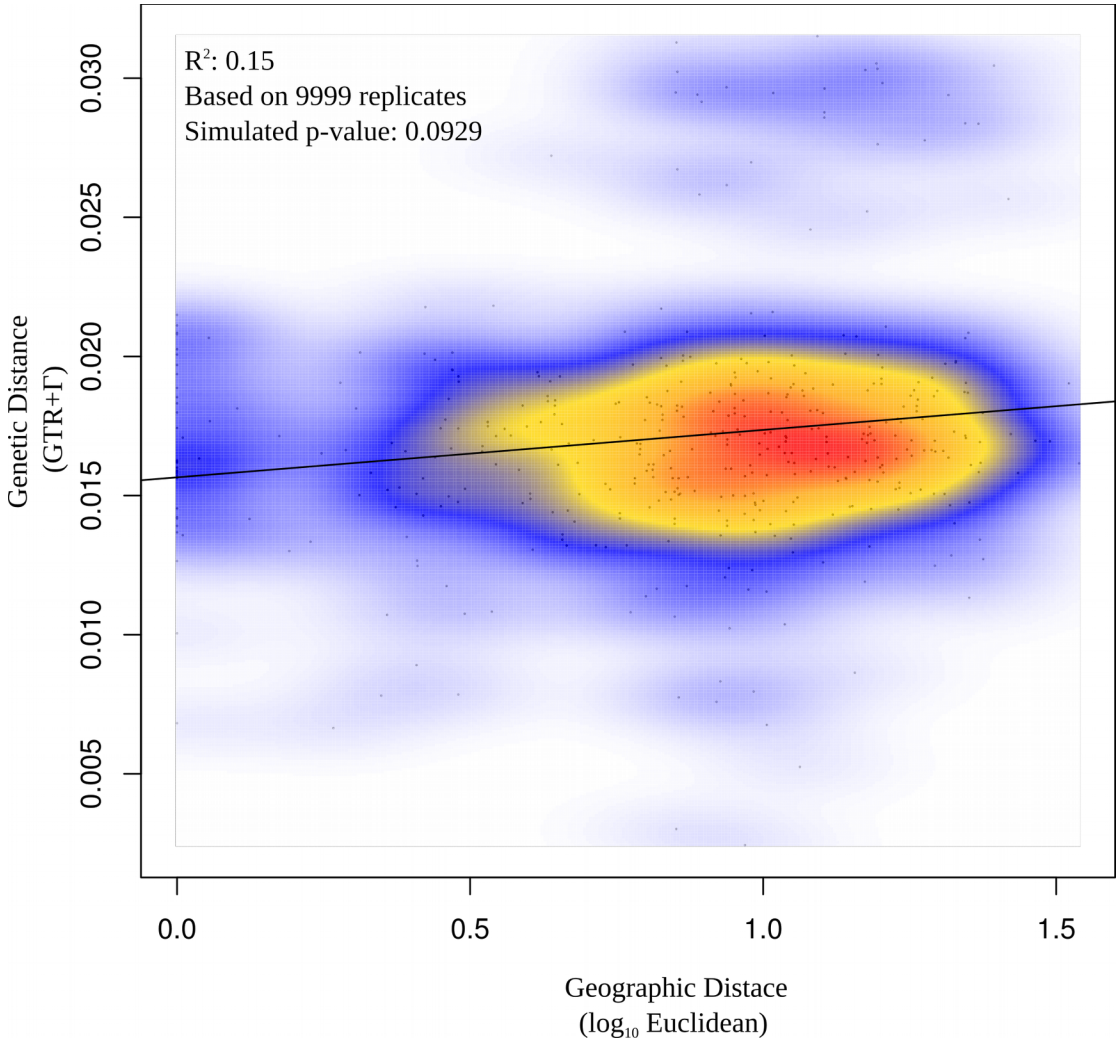


Figure 2-14: Two-dimensional kernel density plot for the correlation of geographic and genetic distance among Mesoamerican *Jatropha* from all habitat types. Warmer colors indicate a higher density of points. Results from Mantel test in upper left.

Areas	# taxa	NRI	Quantile	NTI	Quantile
Sonoran Desert	5	1.79	9282	0.91	8197
Chihuahuan Desert	2	-1.04	2339	-1.04	2339
Baja California*	2	1.12	8254	1.12	8257
Balsas Depression	9	**3.71**	**9982**	**1.88**	**9640**
Sinaloan SDTF	4	**2.11**	**9581**	1.49	9293
Guerreran SDTF	15	**5.43**	**10000**	**3.28**	**9997**
Central America	2	-1.30	554	-1.30	559
Tehuacan Valley	3	**2.02**	**9908**	**1.78**	**9808**
Isthmus of Tehuantepec	2	1.04	7666	1.04	7666

Table 2-11: Results from phylogenetic community structure analysis. Bold numbers for Net Relatedness Index (NRI) and Nearest Taxon Index (NTI) indicate significant positive geographic structure at the $\alpha = 0.05$ level (repetitions = 10,000). Underlined areas are regions of seasonally dry tropical forest. *desert and dry forest regions of Baja California were combined.

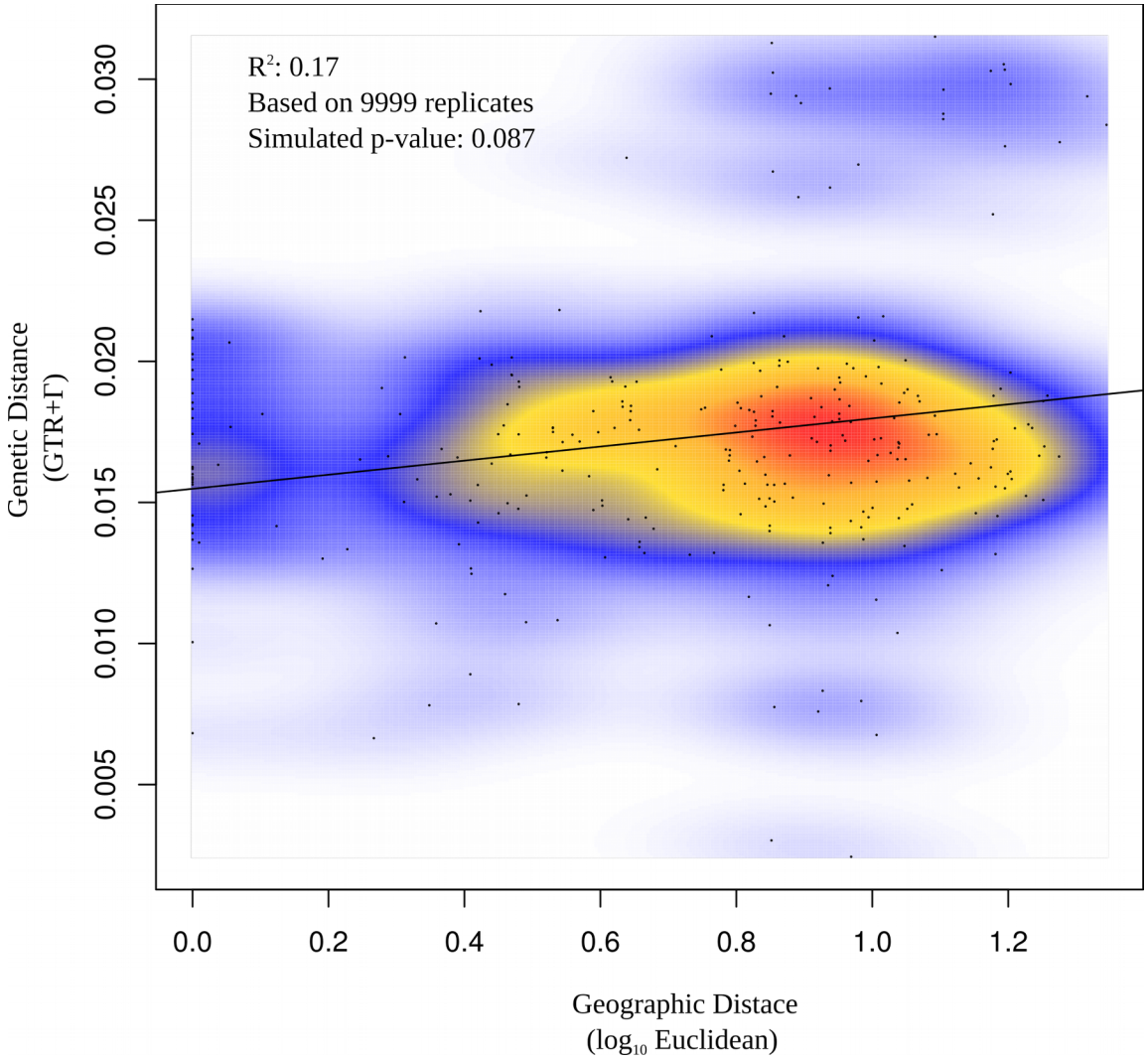


Figure 2-15: Two-dimensional kernel density plot for the correlation of geographic and genetic distance among Mesoamerican *Jatropha* from areas of seasonally dry tropical forest. Warmer colors indicate a higher density of points. Results from Mantel test in upper left.

Chapter 3: Phylogeography of the Heart-leafed Dragon's Blood (*Jatropha cardiophylla*-Euphorbiaceae), an Endemic Shrub from the Interior Sonoran Desert

INTRODUCTION

Phylogeography, the study of infraspecific lineages across time and space, plays an increasingly large role in facilitating our understanding of biological evolution, the response of species to environmental change, and the dynamics of community assemblage (Avise, 2000). In particular, the response of organisms to past climate change has been the focus of many studies, in some cases leading to the identification of refugia that potentially harbored species during the adverse climate conditions of the Pleistocene. As means for molecular data acquisition and statistical analyses have improved, phylogeographic methods have been applied studying an increasingly wide set of organisms (Linder, 2013). Used with ecological niche modeling (ENM), phylogeographic analysis can be an effective way of identifying Pleistocene refugia (Richards et al., 2007; Waltari et al., 2007).

In temperate regions of the Northern hemisphere, Pleistocene glacial cycles have been shown to have impacted many species distributions, in some cases resulting in range fragmentation and population bottlenecks, followed by range expansion and secondary contact (Gugger et al., 2011; Schönswetter et al., 2005). Statistical frameworks have been developed to detect genetic signatures in modern populations resulting from these demographic processes (Excoffier and Heckel, 2006; Fu, 1997; Holsinger and Weir, 2009; Rosenberg and Nordborg, 2002). While glaciation did not directly displace populations in the tropical and sub-tropical low elevation regions of North America, the associated shifts in climate patterns, especially temperature and

rainfall patterns, did impact species distributions (Metcalfe, 2006; Van Devender et al., 1994). Prior to the Pleistocene, tectonic events in the late Miocene and Pliocene, such as the formation of the Gulf of California, the Bouse Embayment, and mountain building have been shown to contribute to population differentiation and speciation in tropical North America (Gándara and Sosa, 2014; Hafner and Riddle, 2011; Provost et al., 2018). Establishing the relative importance of these forces in shaping the genetic structure of modern populations is an ongoing goal of phylogeographers.

The warm arid zone of North America, divided into the Chihuahuan, Sonoran, Peninsular, and Mojave deserts, is a good system for tracing warm temperate plant population responses to recent climate change (Hafner and Riddle, 2011; Fig. 3-1A). Many species native to these desert regions are sensitive to changes in temperature and/or precipitation, and the interaction between elevation and latitude produces a variety of fragmented patches of suitable habitat (Shreve, 1922), which could change during glacial and interglacial periods. Geologically, much of the region has been and is still , having experienced considerable change in topology and climate during the Neogene and Quaternary Periods (Hafner and Riddle, 2011; Metcalfe, 2006).

Phylogeographic studies of regional and widespread desert plant and animal species have identified several common patterns. First, tectonic forces during the Miocene and Pliocene caused speciation of desert taxa via vicariance (Bell et al., 2010; Graham et al., 2015; Jaeger et al., 2005; Myers et al., 2017). Second, cycles of Pleistocene glaciation associated climate change caused distribution fragmentation into refugia causing population bottlenecks, which in turn led to lineage diversification and spatial genetic structuring, followed by rapid expansion during interglacial periods (Fehlberg and Fehlberg, 2017; González-Trujillo et al., 2016; Loera et al., 2017; Myers et al., 2018; Rebernig et al., 2010; Scheinvar et al., 2016; Vásquez-Cruz and Sosa, 2016).

Among the regional North American deserts, the Sonoran has remained relatively understudied from a phylogeographic standpoint. It is situated in northwestern Mexico and southwestern United States between the Sierra Madre Occidental and Gulf of California, extending south from the Colorado Plateau until transitioning to the Sinaloan seasonally dry forest (Fig 3-1A). Sonoran Desert plant communities are largely determined by rainfall amount and periodicity, which averages between 3 to 16 inches annually split between heavy Summer monsoons and light winter precipitation (Shreve and Wiggins, 1964). In their classic work on the flora of the Sonoran Desert Shreve and Wiggins (1964) designated six sub-regions based on vegetation and physiography: the Lower Colorado Valley, Arizona Uplands, Plains of Sonora, Foothills of Sonora, and Central Gulf Coast, and Baja California (now considered a separate biogeographic unit, the Peninsular desert) (Hafner and Riddle, 2011). Modern classification systems largely agree with Shreve and Wiggins (1964), with the exception that portions of the Foothills of Sonora are treated as a part of the Sinaloan dry forest (Brown et al., 2007; Commission for Environmental Cooperation, 1997; Fig. 3-1B).

Few phylogeographic studies have been conducted with plant species endemic to the interior region of the Sonoran Desert (i.e., the Arizona Uplands, Plains of Sonora, and the Foothills of Sonora). As a result, little is known about how plants in these communities were affected by Pleistocene climate change. Phylogeographic studies of plants in the Chihuahuan and Peninsular deserts identified glacial refugia in both the southern and northern portions of modern population distributions, indicating that some, but not all, species responded to climate change simply by moving south (Scheinvar et al., 2016; Vásquez-Cruz and Sosa, 2016). The Lower Colorado Valley of the Sonoran Desert has also been identified as a Pleistocene glacial refugium for widely distributed desert species (Castoe et al., 2007; Fehlberg and Fehlberg, 2017; Jaeger et al., 2005;

Rebernig et al., 2010). Conditions in present day northern Mexico and southwestern United States were cooler and wetter during the last glacial maximum (18,000-22,000 ypb), and large pluvial lakes persisted through much of the area (Metcalfe, 2006). Fossil evidence indicates that the Sonoran Desert underwent considerable change with post-Pleistocene aridification, and that contemporary communities are as young as 4,000-8,000 ybp (Betancourt and Van Devender, 1990; Van Devender et al., 1994). Areas now dominated by xeric scrubland were largely piñon-juniper woodland during the Pleistocene, although some desert community elements may have been little affected and co-occurred with these woodlands (Anderson et al., 1995; Axelrod, 1979).

To identify potential Pleistocene refugia in the interior Sonoran Desert, we conducted phylogeographic analysis and ecological niche modeling (ENM) of the Heart-leafed Dragon's Blood, *Jatropha cardiophylla* (Torr) Muell.Arg. *Jatropha cardiophylla* is endemic to, and widely distributed throughout, the interior Sonoran desert and the adjacent Sinaloan dry forest and Madrean Sky Islands (Fig. 3-1B). Fossilized remains collected from packrat middens in the Tucson mountains suggest *J. cardiophylla* occurred at the northern edge of its present distribution by at least 12,000 kya, a possible indication of a northern Pleistocene refugium (Van Devender, 1973). We used restriction site associated DNA sequencing (RADseq) to generate large numbers of single nucleotide polymorphisms (SNPs) for 318 individuals collected from across the latitudinal extent of its distribution (32.5° N to 26.7° N) (Fig. 3-1B).

To examine the potential influence of Pleistocene climatic fluctuations on the distribution of *Jatropha* cardiophylla, we addressed the following questions: 1) does *J. cardiophylla* exhibit the genetic signature of population expansion from Pleistocene refugia, 2) if so, are these refugia identifiable by ecological niche modeling, and 3) is there genetic/phylogeographic structure to modern populations? If the distribution of *J.*

cardiophylla were restricted to one or more refugia during the Last Glacial Maximum (LGM) (refugial hypothesis) we predict genetic variation will be highest in populations within these regions, and significant genetic signatures of rapid expansion will be seen in populations outside of putative refugia, but not within. If expansion occurred from a single refugium, we predict a strong isolation by distance (IBD) pattern, whereas expansion from two or more refugia might not, depending on the locations of the refugia and their contributions to expansion. We also predict that ecological niche modeling will indicate fragmentation of suitable habitat during the Last Glacial Maximum. Alternatively, if *J. cardiophylla* persisted more or less in its current distribution through the last glacial cycle then we predicted that populations throughout the range would show high levels of genetic variation and no detectable signal of rapid population expansion (persistence hypothesis). In this case, ecological niche modeling should indicate a stable distribution of suitable habitat during the Last Glacial Maximum. An isolation by distance pattern may be detectable depending on how populations are structured.

METHODS

Data collection and processing

Leaf tissue was collected from 318 individuals of *Jatropha cardiophylla* and dried in silica gel. Collections spanned the latitudinal range of the species, with localities in the Arizona Uplands, Lower Colorado Valley, Plains of Sonora, Foothills of Sonora/Sinaloan Dry Forest, and one locality on the Central Gulf Coast (Fig. 3-1B). One to six individuals were collected per locality, taking care to make sure that all samples were from discrete plants and not from potential clusters of clones. DNA was extracted using DNeasy Plant Mini Kits (QIAGEN, Valencia, CA) with the initial lysis step

performed at 65° C for 10-20 minutes, followed by the addition 25 mL of Proteinase K and overnight incubation at 45° C. Extracted DNA was quantified using a Qubit broad-range kit (Life Technologies, Carlsbad, CA) and visually inspected for quality in 1% agarose gels.

Double-digest RADseq libraries were prepared using the protocol of Peterson et al. (2012) using 200-600 ng of genomic DNA digested with *sphI* and *ecoRI* restriction endonucleases (NEB, Ipswich, MA). Digested DNA was ligated to adapters containing unique in-line barcodes for multiplexing, then pooled and size selected (target: 410 bp, range: 370-450 bp) using a Pippin Prep (Sage Science, Beverly, MA). Libraries were prepared and sequenced on two lanes of Illumina HiSeq 4000 (targeting 1 million 150 bp paired-end reads per sample) at the The University of Texas at Austin Genome Sequencing and Analysis Facility.

Quality scores of the raw Illumina reads were assessed with the program FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). Restriction site overhangs were trimmed using the 'reformat.sh' script from the software package BBTools v36.30 (Bushnell, 2016). Reads were demultiplexed using the program deML v1 (Renaud et al., 2015) allowing for up to two mismatches in barcodes. Assembly of demultiplexed reads into orthologous groups (hereafter loci) was done using the software package ipyrad v0.7.29 (Eaton and Overcast, 2016) on the Lonestar5 cluster of the Texas Advanced Computing Center at The University of Texas at Austin (http://www.tacc.utexas.edu). Reads generated from Chapter 1 for the sister species, *Jatropha vernicosa* Brandegee, were included for rooting phylogenetic trees.

Reads were assembled into datasets using 1) a similarity threshold of 85% for clustering within and across individuals, 2) a Phred score of 33 as a quality threshold for reads, and 3) allowing a maximum of seven low quality bases per read, and a

minimum read depth of 10 for statistical base calling. Loci were filtered from final datasets if they had more than eight uncalled bases, two alleles, twelve heterozygous sites, six indels, forty-five SNPs, or were missing from more than 10% of individuals. Fourteen of 318 samples were excluded due excessively low numbers of recovered loci, likely resulting from insufficient read depth. The reads of the remaining samples were reassembled into the "full" dataset, containing 304 individuals of *J. cardiophylla* and one individual of *J. vernicosa*). Datasets were assembled for each of four geographic groups: Arizona (n=15), northern Sonora (n=152), central Sonora (n=73), and southern Sonora (n = 64). Finally, datasets were assembled for two distinct genetic groups (Groups 1 & II, see below) identified in clustering analyses, using the parameters as described above.

Phylogenetic analyses

Evolutionary relationships among individuals were reconstructed using maximum likelihood in RAxML v8.2.4 after concatenating all loci(Stamatakis, 2014). We analyzed the full dataset, with the GTR+Γ model of substitution, conducting ten independent searches for the highest scoring tree, and assessing branch support with 500 bootstrap replicates. Analyses were performed on the CIPRES Science Gateway (Miller, et al., 2010). Bootstrap values were summarized onto the best tree using IQTree v1.5.4 (Nguyen et al., 2005). Since a bifurcating tree can be a misleading representation of intraspecific evolutionary history, we also conducted phylogenetic network analysis of the full dataset using SplitsTree v4.14.8 (Huson and Bryant, 2006). The network was constructed with a neighbor joining algorithm based on the raw pairwise distance between samples. Since two highly divergent clusters were identified (Groups I and II:

Figs. 3-2 and 3-4), another network was constructed with just the 285 individuals of Group II to better resolve relationships within it.

Genetic structure and isolation by distance

Several tests for genetic structure were performed using all samples and solely Group II. First, discriminant analysis of principal components (DAPC) was performed using the R package *adegenet* v2.1.1 (Jombart, 2008; Jombart and Ahmed, 2011). Output from ipyrad was filtered to allow only biallelic loci using vcftools v0.1.16 (Danecek et al., 2011) and then imported into R and converted into 'genind' format using the package *vcfR* v1.8.0 (Knaus and Grunwald, 2016). DAPC requires the number of groups/populations be defined *a priori*, so the k-means clustering algorithm was used with the Bayesian information criterion to determine the optimal number of clusters. We ran DAPC with a range of k-values and mapped the results to determine if the identified clusters differed in a biologically meaningful sense. To avoid over parameterization, only the first 30 principal components were included in the discriminant analysis, but sensitivity analyses were conducted using 50 and 80 PCAs to determine the impact of this parameter on group assignment. A contingency table was constructed to compare the composition of clusters identified from SplitsTree and discriminant analysis of principal components. Second, pairwise F_{ST} scores were calculated between Groups I and II, between geographic areas using Group II individuals only, and between clusters identified by DAPC within Group II (Nei, 1987). Pairwise F_{ST} calculations were computed using the R package *hierfstat* v0.4.22. Third, we performed spatial analysis of principal components (sPCA) with the R package *adespatial* v0.3.4 (Dray et al., 2018). Spatial PCA is a good way of visualizing spatial patterns of genetic variation by

combining ordination methods with Moran's index of spatial autocorrelation (Jombart et al., 2008; Moran, 1950). Because sampling locations were spatially aggregated, the neighborhood by distance method was used to create a connection network (distance range: 0.1-1°).

We also tested for isolation by distance (IBD) in the full dataset and both Group I and Group II datasets using a Mantel test with 1000 replicates in *adegenet* v2.1.1. Two dimensional kernel density plots were used to determine if significant patterns of IBD were due to clinal variation or the existence of distinct patches.

Genetic diversity and demography

Mean pairwise sequence divergence was calculated for all datasets using R package *pegas* v0.11 (Paradis, 2010), and observed levels of heterozygosity were calculated using *hierfstat* v0.4.22 (Goudet and Jombart, 2015). To test for population expansion we calculated Tajima's D for the full, Group I, and Group II datasets using the R package *strataG* v2.0.2 (Archer et al., 2017; Tajima, 1989). We additionally calculated Tajima's D for each geographic region, first using all individuals and then only Group II individuals.

Changes in effective population size (N_e) over time were estimated by modeling the coalescent process using the Bayesian skyline tree prior implemented in BEAST v2.5.2 (Bouckaert et al., 2014; Drummond et al., 2005). Effective population sizes were estimated, allowing five time periods, using a rate of $4.54x10^{-9}$ substitutions/site/year, based upon a published rate of synonymous substitutions for Euphorbiaceae (De La Torre et al., 2017). Two independent MCMC iterations were run for 100 million generations each, and the software package Tracer v1.6.0 was used to

assess convergence of chains and summarize results into skyline plots using a 10% burn-in (Rambaut et al., 2014). Effective population sizes were estimated for the same datasets as used in calculating Tajima's D.

Times to coalescence for the most recent common ancestor (MRCA) for both Groups I and II and for each geographic region using only Group II individuals were estimated using BEAST v2.5.2. Two coalescent tree priors were used: one modeling constant population size and one modeling exponential population growth. The substitution rate used was the same as in the Bayesian skyline analyses above. Two independent MCMC iterations were run for 10 million generations each and assessed for convergence in Tracer v1.6.0 (Rambaut et al., 2014). Tree files were combined using LogCombiner v2.5.2 (Bouckaert et al., 2014) with a 10% burn-in. Median node heights and 95% highest probability densities (HPD) were written to the maximum clade credibility trees with TreeAnnotator v2.5.2 (Bouckaert et al., 2014) and visualized with Figtree v1.4.2 (http://tree/bio.ed.ac.uk/software/figtree/).

Ecological niche modeling

As an additional means of identifying putative glacial refugia, we used the software package Maxent v3.4.1 (Phillips et al., 2006) to construct an ecological niche model for *Jatropha cardiophylla* based on its present distribution. A database of georeferenced occurrence records of *J. cardiophylla* was compiled from the SEINet Portal Network (http://swbiodiversity.org/seinet/; accessed May 25, 2018), the Global Biodiversity Information Facility (https://www.gbif.org/; accessed September 15, 2018) and direct observations made in the field. After removing duplicates, the database contained 764 records. The niche model was constructed using these occurrence records

and 19 environmental variables from Worldclim v1.4 downloaded at 30 arc-second resolution (Hijmans et al., 2005). Environmental raster layers were cropped to include only the study area using QGIS v2.8 (QGIS Development Team, 2016) and then assessed for correlation between variables using the package *raster* in R (R Core Team, 2018). For environmental variables that were highly correlated with each other (Pearson's $r > 0.85$) only one of the correlated variables was retained, leaving 10 environmental variables (Table 3-3).

To predict how suitable habitat for *Jatropha cardiophylla* changed over the last glacial cycle we projected the contemporary niche model onto modeled climate conditions from the Last Glacial Maximum (LGM) (22,000 ybp) and the Last Interglacial (LIG) (120-140,000 ybp). We used the Models for Interdisciplinary Research on Climate (MIROC) model for the LGM climates (2.5 arc-minutes resolution) and the Community Climate System Model (CCSM) (30 arc-second resolution) for the LIG (Otto-Bliesner et al., 2006). The same 10 environmental variables used for the present day distribution model were used for the past distribution models.

RESULTS

RADseq

Illumina sequencing of 318 samples produced 1.03×10^8 reads. The average number of reads per sample was 6.38×10^5 (range: 5.35×10^4 – 5.44×10^6). The average number of reads for the 14 excluded samples was 6.37×10^4, an order of magnitude lower. The number of recovered loci and SNPs for datasets assembled from the 304 remaining samples plus *Jatropha vernicosa*, and subsets thereof (i.e., geographic and genetic clusters), are reported in Table 3-1.

Phylogenetic analysis and population clustering

Maximum likelihood analysis of the full dataset recovered two highly supported clades: Group I and Group II (BS = 94 and 99 respectively; Fig. 3-2). Group I was comprised of 19 individuals collected from southern Arizona and central to southern Sonora, whereas the 285 individuals from Group II were widely distributed across the entire sampled area (Fig. 3-3). Relationships among individuals within Group I were mostly well resolved, but there was poor resolution within Group II, which also displayed relatively lower levels of nucleotide diversity (π = 0.0213 and 0.0007 for Groups I and II respectively) (Fig. 3-2; Table 3-1).

Network analysis also identified Groups I and II and revealed a similar pattern of contrasting genetic variation within each group (Fig. 3-4A). Network analysis of the Group II dataset somewhat improved the resolution among individuals (Fig. 3-4B). Clusters in the SplitsTree network were circumscribed in accordance with the clusters identified in DAPC (see below).

Genetic structure and isolation by distance

The lowest BIC score for k-means selection for the full dataset indicated 23 clusters as optimal for DAPC (Fig. 3-5). The largest discriminant function in all analyses across k values separated Group I from an increasingly subdivided Group II (Figs. 3-6 – 3-7). Nei's F_{ST} between Groups I and II was 0.47, indicating very little gene flow between groups (Table 3-2). Within Group II k-means selection indicated six as the optimal number of clusters (Fig. 3-8). DAPC (k = 3 to 6) consistently recovered two

clusters (D-1 and D-3), but a third cluster (D-2) was increasingly subdivided into geographically overlapping clusters for when k > 3 (Fig. 3-8 – 3-11). Since pairwise F_{ST} scores for the k > 3 subdivided clusters (range: 0.039-0.118) revealed little to moderate differentiation (F_{ST}) we treated them as one cluser concordant with the k = 3 D-2 cluster (Fig. 3-8 – 3-11). Cluster D-1 contained nearly all samples from cluster S-1 from SplitsTree and was restricted to a small area in southern Sonora (Figs. 3-12 – 3-13). Cluster D-2 was distributed across almost the entirety of the sampled area, only slightly overlapping with D-1, and contained most individuals from cluster S-2 from SplitsTree (Figs. 3-12 – 3-13). Cluster D-3 consisted of three highly divergent samples relative to those of D-1 and D-2, which were widely distributed across the Plains of Sonora (Figs. 3-8 and 3-13).

Spatial principal components analysis of the full dataset (excluding *Jatropha vernicosa*) showed little genetic variation across northern Sonora and southern Arizona (Fig. 3-14). A transition from positive to negative spatial autocorrelation in genetic variation was seen beginning in central Sonora and increasing into south Sonora. The shift from genetic homogeneity to heterogeneity also was detected for Group II alone, beginning farther south, indicating that the variation seen in central Sonora was entirely due to individuals from Group I (Fig. 3-15).

Tests for isolation by distance were significant for the full dataset (R^2 = 0.135, $p < 0.001$) and for Group II alone (R^2 = 0.454, $p < 0.001$), but a non-significant for Group I alone (R^2 = -0.061, p = 0.66). The kernel density plot for all samples shows that the majority of the comparisons reflect a genetic cline with a smaller number of samples clustering separately (Fig. 3-16). Graphing Groups I and II separately shows that Group II individuals alone exhibited the clinal pattern (Figs. 3-17 – 3-18).

Genetic diversity and demography

Genetic divergence among all individuals was low ($\pi = 0.0029$), but varied largely between Group I and Group II ($\pi = 0.213$ and 0.0007 respectively; Table 3-1). Arizona and central Sonora showed the highest levels of genetic diversity ($\pi = 0.019$ and 0.06 respectively), but when only comparing Group II individuals the greatest diversity was in southern Sonora (Table 3-1). The observed heterozygosity was higher in Group II ($H_{obs}= 0.036$) than in Group I ($H_{obs}=0.021$). There was pronounced spatial variation in heterozygosity within Group II with individuals from Arizona being much higher ($H_{obs}= 0.186$) than any other area (H_{obs} range = 0.066-0.077).

Tajima's D was significant for the full dataset ($D_{full} = -2.75$, $p < 0.001$), indicating recent population expansion. When Groups I and II were analyzed separately, however, only Group II showed a significant recent expansion ($D_I = -1.85$, $p = 0.051$; $D_{II} = -2.58$, $p < 0.01$). Looking across the sampled area, Group II had significant Tajima's D values for each geographic area (Table 3-1).

Bayesian skyline plots showed that Group I experienced a substantial increase in effective population size (N_e) approximately 1mya, but has remained stable since the mid-Pliocene, whereas Group II has been expanding steadily throughout the Pleistocene (Fig. 3-19). The coalescent times to the most recent common ancestors (MRCAs) for Group I and Group II were 4.50 Mya (3.95-5.00 Mya 95% HPD) and 0.33 Mya (0.30-0.37 Mya 95% HPD) respectively. Estimated ages of the MRCAs of Group II for each geographic region were all during the LIG (range for median ages: 160-380 kya), with the time to coalescence significantly later in Arizona than all other regions (Table 3-1).

Bayesian skyline plots of Group II individuals from each geographic region revealed pronounced spatial differences in changes in N_e (Fig. 12A-D). Southern

and northern Sonora experienced rapid periods of population expansion during the LIG, with estimated dates of expansion being earlier in southern Sonora (~ 100 kya) than northern Sonora (~50 kya) (Fig. 3-20). In contrast, central Sonora and Arizona remained relatively stable over the same time period. These two areas are where Groups I and II co-occur, and so we tested if gene flow between groups could account for the different demographic pattern. Genetic differentiation (F_{ST}) between Groups I and II within Arizona and central Sonora were 0.46 and 0.45 respectively, similar to the F_{ST} between all of Groups I and II, indicating no increased gene flow between groups within these areas.

Ecological Niche Modeling

The area under the curve (AUC) score for the niche model constructed based on the current distribution of *Jatropha cardiophylla* was 0.915, indicating that the model predicted distribution for this species significantly better than a random model. The two most important environmental variables for predicting occurrence of *J. cardiophylla* were the mean temperature in the coldest quarter and total annual precipitation, which accounted for nearly 60% of the contribution to the model together (Table 3-3). Projection of the contemporary niche model onto a map shows a nearly contiguous distribution with some patchy breaks in the north (Fig. 3-21A). These gaps might reflect a dearth of collections from border regions and the Tohono O'odham reservation rather than a true break in the distribution. The model also predicted a low likelihood of occurrence along the central gulf coast of Sonora where *J. cardiophylla* occurs sparsely along with *J. cinerea* Müll.Arg. On the western margin of the distribution the model showed fragmentation where *J. cardiophylla* is restricted to lower elevations valleys in the foothills of the eastern slopes of the Sierra Madre Occidental.

Paleodistribution modeling indicated that the distribution of *Jatropha cardiophylla* remained stable throughout the most recent cycle of glaciation (Fig. 3-21B-C). As at the present, the central Plains of Sonora and the adjacent foothills of the Sierra Madre Occidental, and the Arizona Uplands region of southern Arizona all had the highest probability for the occurrence of suitable habitat during the Last Glacial Maximum (LGM) and Last Interglacial (LIG). During the LGM there was an expansion of suitable habitat from the central Plains of Sonora region to the north and south (Fig. 3-21B). The LIG distribution model was nearly identical to the current distribution of *J. cardiophylla*, with slightly higher predictions for suitable habitat in northern Sonora and a westward shift in southern Arizona (Fig. 3-21C). Overall, paleodistribution modeling of *J. cardiophylla* for the LGM and LIG reflected increased connectivity between areas, rather than fragmentation, of suitable habitat compared with present conditions.

DISCUSSION

Phylogeographic and genetic clustering analyses of RADseq data showed that *Jatropha cardiophylla* consists of two genetically distinct lineages: Groups I and II. Group I is divided in distribution between southern Arizona and central Sonora and genetically highly variable relative to Group II, which is widely distributed across the study area. The degree of differentiation between Groups I and II is striking given their sympatric distributions, and indicative of a strong barrier to gene flow between the groups. The nature of the barrier to gene flow is not currently known, but one possibility is that Groups differ in ploidy. Tetraploids have been documented for *J. cuneata* and *J. dioica* from the Sonoran and Chihuahuan Deserts respectively, but polyploidy has not yet been reported for *J. cardiophylla* (Dehgan and Webster, 1979). Variation in ploidy has

been linked with significant spatial genetic structure in *Melampodium leucanthum* Torr. & A. Gray and *Larrea tridentata* Cav., both widespread desert plant species (Laport et al., 2012; Rebernig et al., 2010). Polyploidy might also explain the lower number of recovered loci for samples in Group I relative to Group II (45.5 and 69.0 loci/individual respectively), if reads mapped to duplicated regions were clustered together and subsequently removed by filtering for paralogs during assembly. Chromosome counts of individuals from southern Arizona and central Sonora would be necessary to verify this.

We found support for refugial scenarios on two temporal scales: first that Arizona and central Sonora might each represent former semi-refugia predating the Pleistocene, and second that central-southern Sonora was a refugium for Group II during the Pleistocene. Spatial variation in genetic diversity was observed for *Jatropha cardiophylla* in analyses of the full dataset. The geographic regions with the highest genetic diversities (mean pairwise distance per site) were Arizona and central Sonora (using datasets containing both Groups I and II individuals together). Ecological niche modeling also strongly predicted the occurrence of suitable habitat in Arizona and central Sonora throughout the LGM and LIG. The observed elevated diversity in Arizona and central Sonora was, however, largely driven by Group I individuals. When Group II was analyzed separately, however, marginally higher levels of genetic diversity were observed in southern Sonora. This spatial pattern was supported by spatial PCA, which revealed a genetic cline between central and southern Sonora in analyses of both the full and Group II only datasets. Ecological niche models from the Pleistocene indicated more suitable habitat in southern Sonora than at present. The significant tests for population expansion and strong pattern of isolation by distance with a single cline together support a single refugium scenario for Group II.

Evidence in favor of a central Sonoran refugium for Group II includes the significant expansions observed in the skyline plots for the southern and northern Sonora, whereas central Sonora has remained stable over the last 200,000 years. The order of expansion events in the south and north matched the order of the ecological niche modeling predictions of increased suitable habitat in these regions. It is important to note that the absolute dating of demographic events from genetic data is dependent on, among other factors, the substitution rate. This is a challenge for *denovo* assembled RADseq datasets of non-model organisms, where the identity of the markers is rarely known, and therefore published substitution rates must often be used as a best approximation.

The conflicting spatial patterns of genetic variation between Groups I and II are not particularly surprising in light of the vastly different estimated coalescence times for each group. The most recent common ancestor (MRCA) of Group I dated back to the mid-Pliocene (4.49 mya), whereas the MRCA of Group II dated back to the late Pleistocene (330 kya). It is possible that Group II experienced a bottleneck prior to the last glaciation cycle, and was restricted to a refugium in southern Sonora, reflected by the elevated genetic diversity observed there. Group I likely experienced an earlier population bottleneck, perhaps in response to an earlier glaciation cycle. Genetic signals for temporally separate population retractions to different refugia have been observed for *Agave lecheguilla* in the Chihuahuan Desert (Scheinvar et al., 2016). Alternatively, if the two groups are in fact different species then each may have responded differently to the pressure of climate change. Group I may have persisted in central Sonora and Arizona while Group II was restricted further south. The present limited distribution of Group I and lack of significant signal for expansion both suggest the possibility of differential response to recent climate change between groups. Modeling response to past climate

change separately with each group would be one way of testing this, however the small sample size of individuals in Group I precludes this at present.

CONCLUSIONS

RADseq data clearly indicated that *Jatropha cardiophylla* is comprised of two genetically distinct groups, which may represent cryptic species, or perhaps incompatible ploidy races, given the near complete sympatry of Group I with Group II. The nature of the barrier to gene flow between groups is an unresolved question that merits further research, perhaps a survey for polyploid individuals. We found support for temporally distinct refugial scenarios for each group, with implications of differential response to past climate fluctuations across the area. Ecological niche modeling did not identify clearly isolated Pleistocene refugia, as have have been documented in other species. However, regions with the highest predicted climatic suitability over the time period considered did coincide with the regions of highest genetic variability, which strongly overlapped with the distribution of Group I. Low heterozygosity, restricted spatial distribution, a lack of significant signal for expansion, and a mid-Pliocene age for the MRCA point to the possibility that Group I has undergone, or is currently undergoing, a bottle neck. Group II, on the other hand, has undergone a recent rapid population increase. This expansion either predates the most recent glaciation cycle, or molecular substitution rates are faster for RADseq markers in *J. cardiophylla* than those estimated for family Euphorbiaceae that were used for analyses. Evidence from analyses for just Group II supported two alternate refugial scenarios. Spatial patterns of genetic diversity supported a southern Sonora refugium, whereas coalescent models of effective population sizes over time favored a central Sonoran refugium. In either case, results from tests for

isolation by distance supported a single refugium scenario. We showed that Pleistocene climate change appears to have been the major driving force shaping patterns of genetic structure for Group II, and potentially also from Group I. The age of Group I does leave open a greater window of possible alternative forces, potentially tectonic events or earlier climate change, as has been documented for other desert taxa with lineage divergence dates in the Neogene, however there are no obvious landscape features the immediately suggest tectonically induced vicariance between groups or between patches of Group I individuals. We recommend surveying for variation in ploidy level and increased sampling from eastern populations situated higher in the valleys of the Sierra Madre Occidental as possible next steps for addressing some of the outstanding questions pertaining to the genetic patterns identified in this study.

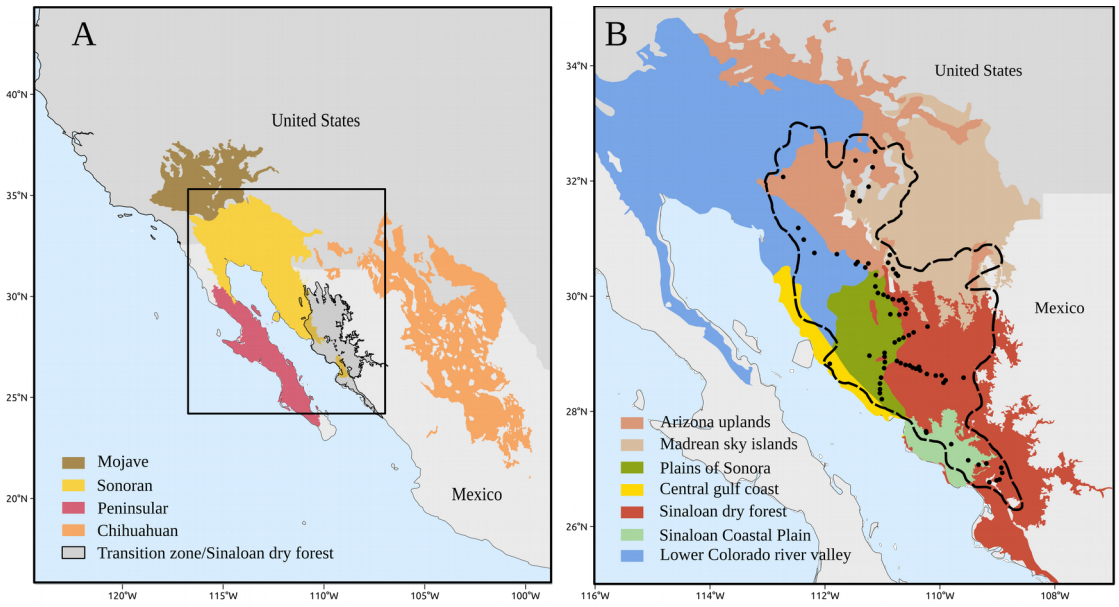


Figure 3-1: A) Distribution of the major warm deserts of North America. B) Distribution of *Jatropha cardiophylla* in the subdivisions of the Sonoran Desert and adjacent regions. The dotted line shows the full distribution of *J. cardiophylla* based on available georeferenced collections, and black dots are localities of populations sampled for the study.

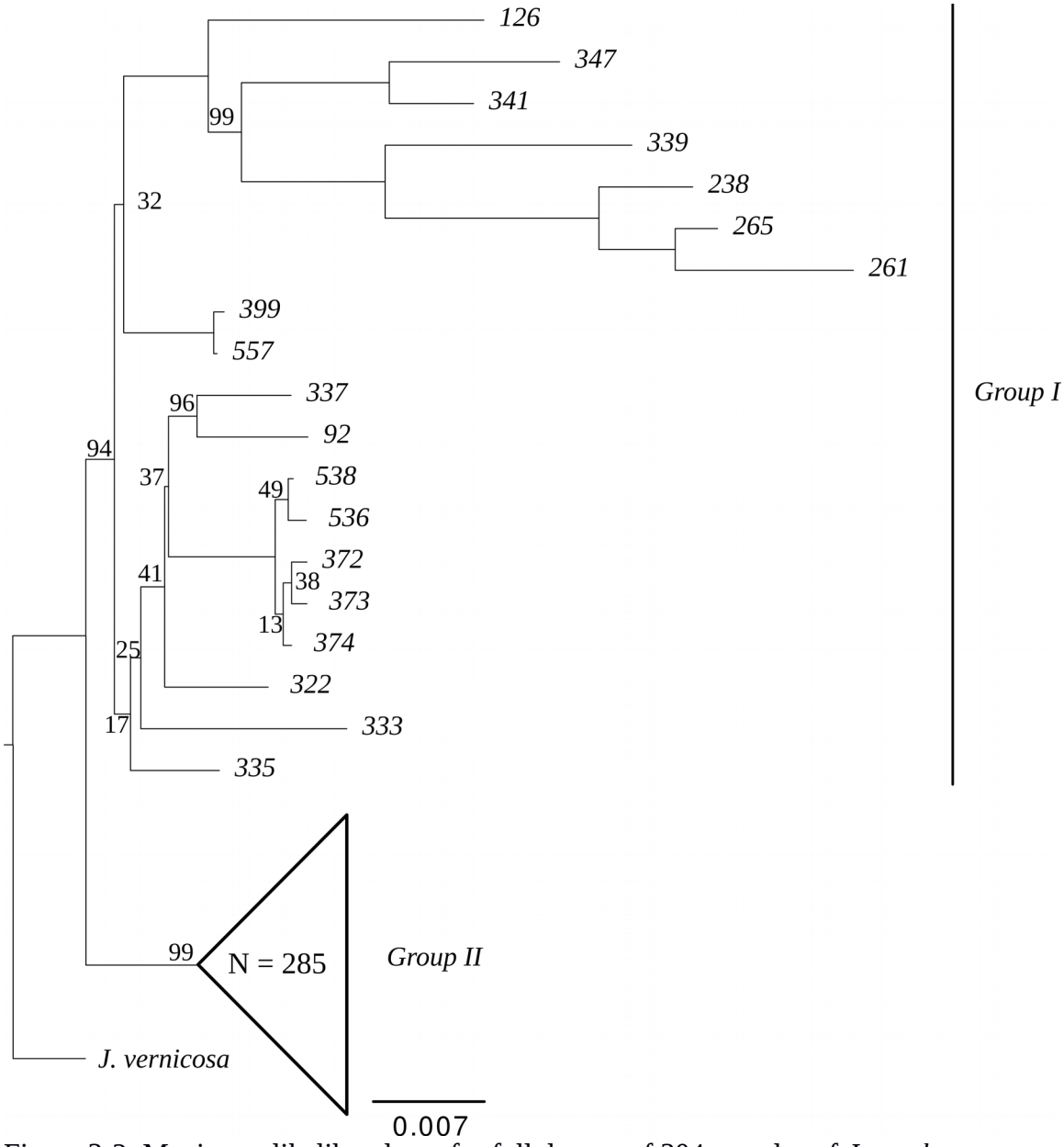


Figure 3-2: Maximum likelihood tree for full dataset of 304 samples of *Jatropha cardiophylla* and rooted with *J. vernicosa*. Values on branches are bootstrap support (BS = 100 not shown.) Numbers at the tips of branches are identifiers of sampled individuals

Dataset	N	Min	Loci	SNPs	π	Ho	D	MRCA
All samples*	305	275	72	1,625	0.0029	0.014	**-2.75**	N/A
Group I	19	17	33	593	0.0213	0.021	-1.58	4.49 (3.99-5.07)
Group II	285	250	202	1,986	0.0007	0.036	**-2.28**	0.33 (0.30-0.37)
Arizona	15 (*10*)	13 (*9*)	101 (*505*)	1,415 (*724*)	0.0190 (*0.0007*)	0.033 (*0.186*)	-1.49 (***-2.05***)	0.16 (0.14-0.18)
North Sonora	*152*	*137*	*396*	*1,915*	*0.0006*	*0.066*	***-2.25***	0.24 (0.21-0.27)
Central Sonora	73 (*62*)	66 (*56*)	20 (*41*)	319 (*228*)	0.0058 (*0.0008*)	0.022 (*0.068*)	**-2.48** (***-2.03***)	0.38 (0.26-0.50)
South Sonora	64 (*60*)	58 (*54*)	374 (*491*)	2,860 (*2,636*)	0.0024 (*0.0009*)	0.061 (*0.077*)	**-2.50 (*-2.23*)**	0.27 (0.24-0.29)

Table 3-1: RADseq datasets of *Jatropha cardiophylla* showing: N – the number of samples, min – the minimum number of individuals in which a loci must be present to be maintained, Loci – the number of loci recovered, SNPs – single nucleotide polymorphisms, π – mean pairwise sequence difference, Ho – observed heterozygosity, D – Tajima's D, and MRCA – estimated coalescence time in millions of yeas (95% highest probability distribution in parentheses) to most recent common ancestor using constant population tree prior. Italic values are for datasets assembled with group II individuals only. Bold values are significant at the $p < 0.05$ level. *All samples after the removal of 14 unsuitable samples and including *J. vernicosa*.

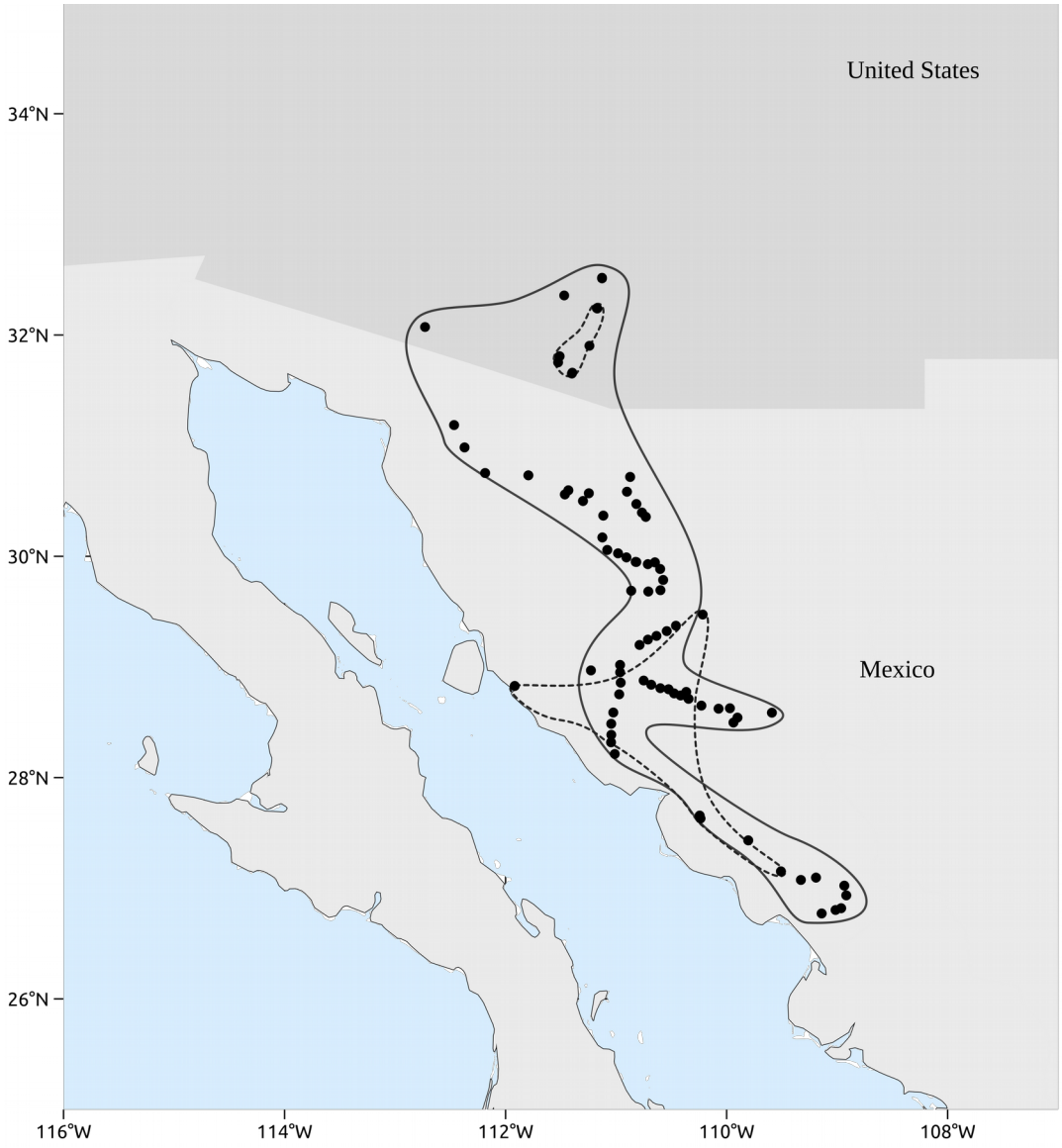


Figure 3-3: Geographic distribution of Group I (dashed lines) and Group II (solid line). Points show all collection localities.

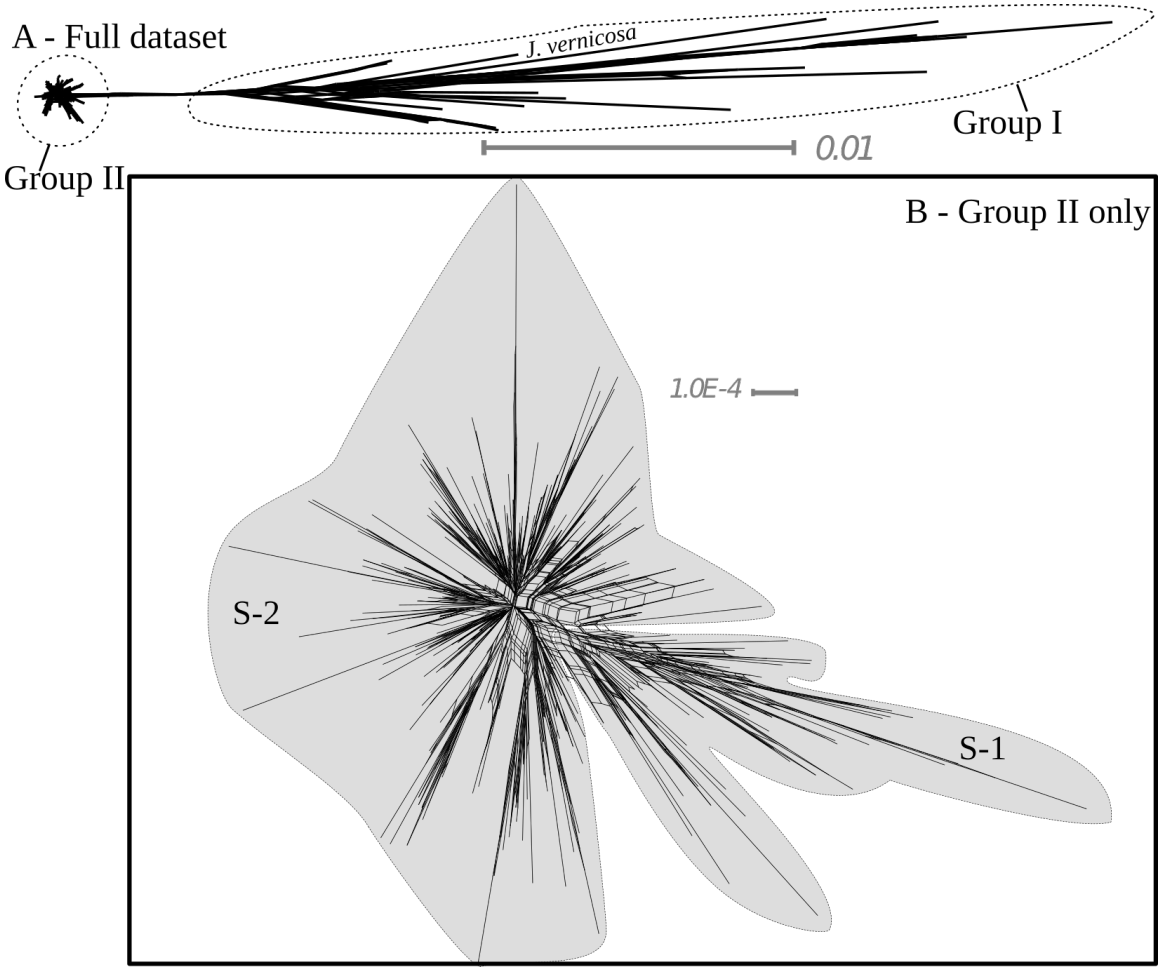


Figure 3-4: Phylogenetic networks from SplitsTree for: A) the full dataset and B) group II individuals only. Scale bar shows the uncorrected P-distance.

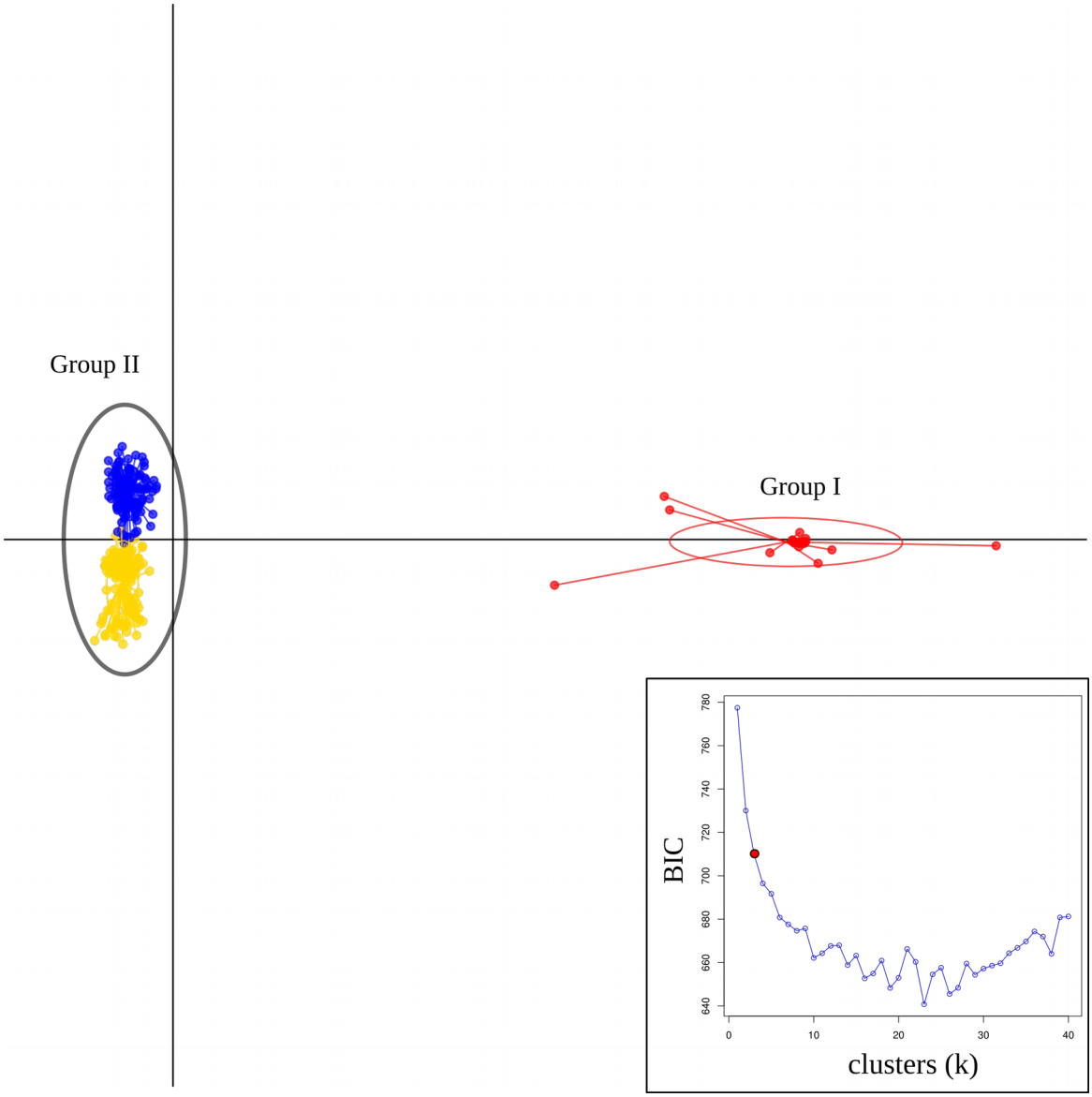


Figure 3-5: Scatter plot from discriminant analysis of principal components (DAPC) for the full dataset, excluding *Jatropha vernicosa* (k = 3). Inset: Results of k-means group selection

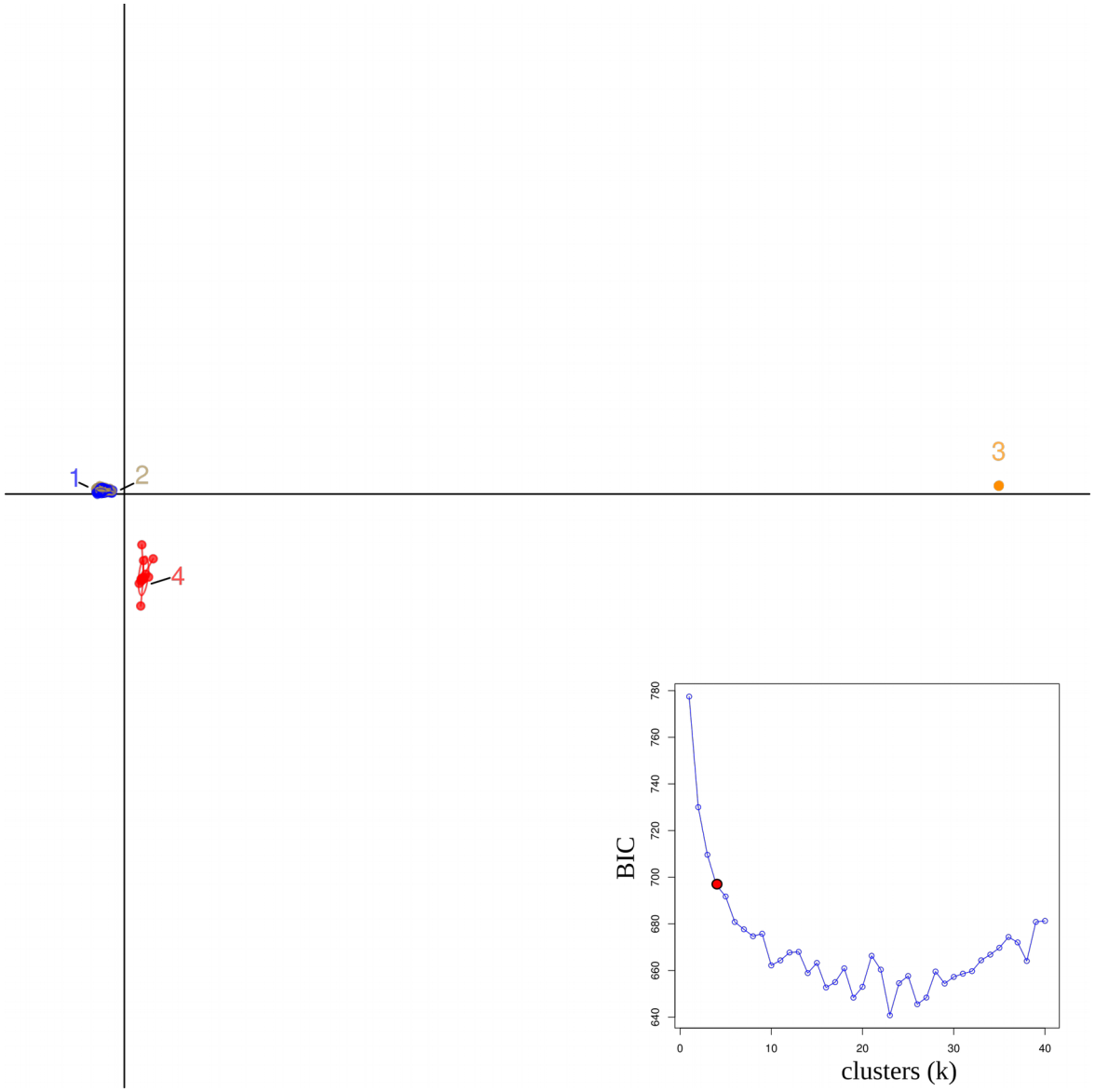


Figure 3-6: Scatter plot from discriminant analysis of principal components (DAPC) for the full dataset, excluding *Jatropha vernicosa* (k = 4). Inset: Results of k-means group selection. The red dot indicates the optimum number of groups.

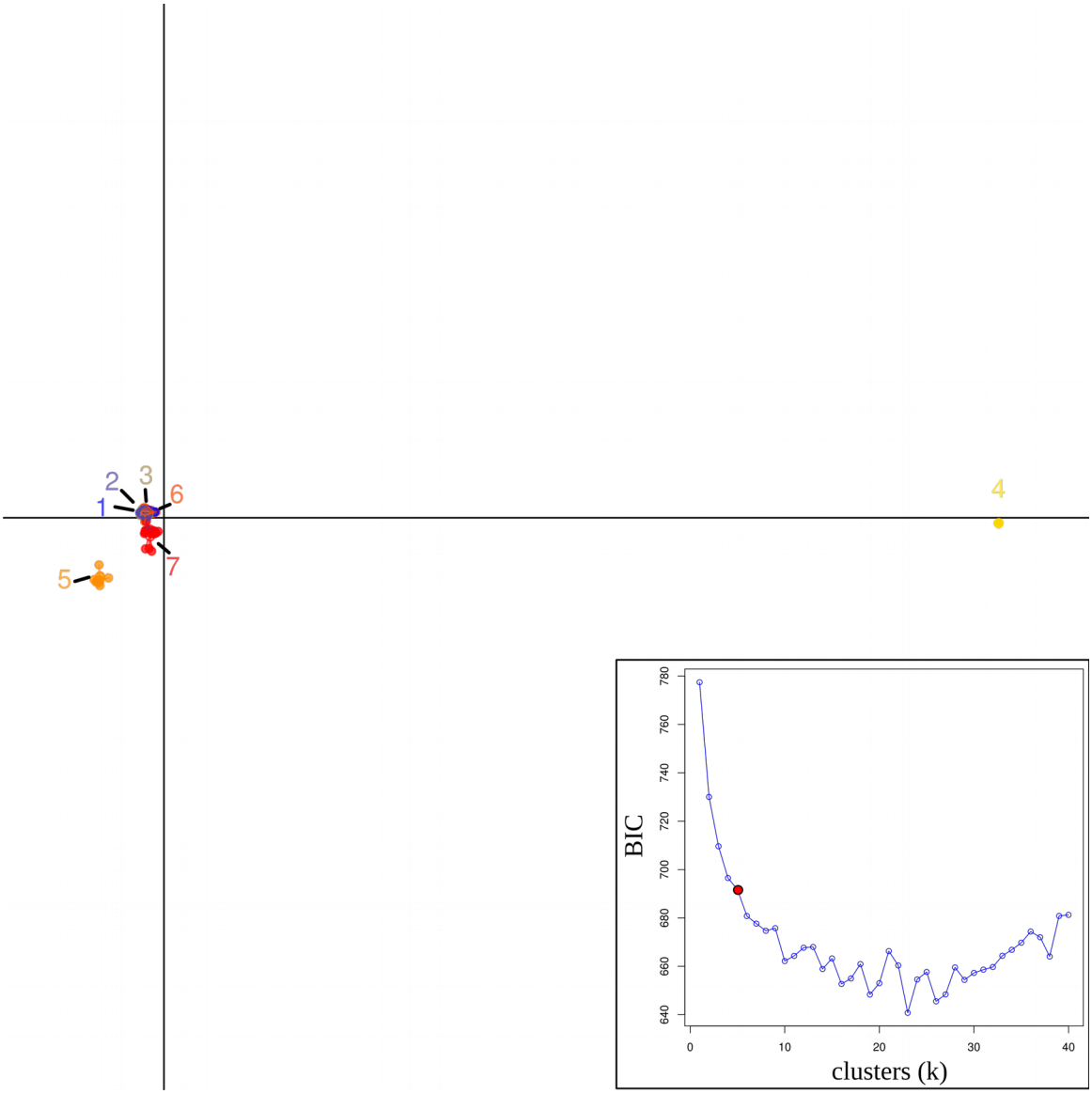


Figure 3-7: Scatter plot from discriminant analysis of principal components (DAPC) for the full dataset, excluding *Jatropha vernicosa* (k = 7). Inset: Results of k-means group selection. The red dot indicates the optimum number of groups.

	F_{ST}
A	
Group I x Group II	0.47

B	AZ	NS	CS
NS	0.03		
CS	0.04	0.01	
SS	0.08	0.09	0.06

C	D-1	D-2
D-2	0.10	
D-3	0.30	0.33

Table 3-2 Genetic distances (Nei's F_{ST}) between: A – Groups, B – geographic divisions using Group II individuals only, and C- clusters identified by DAPC for Group II only.

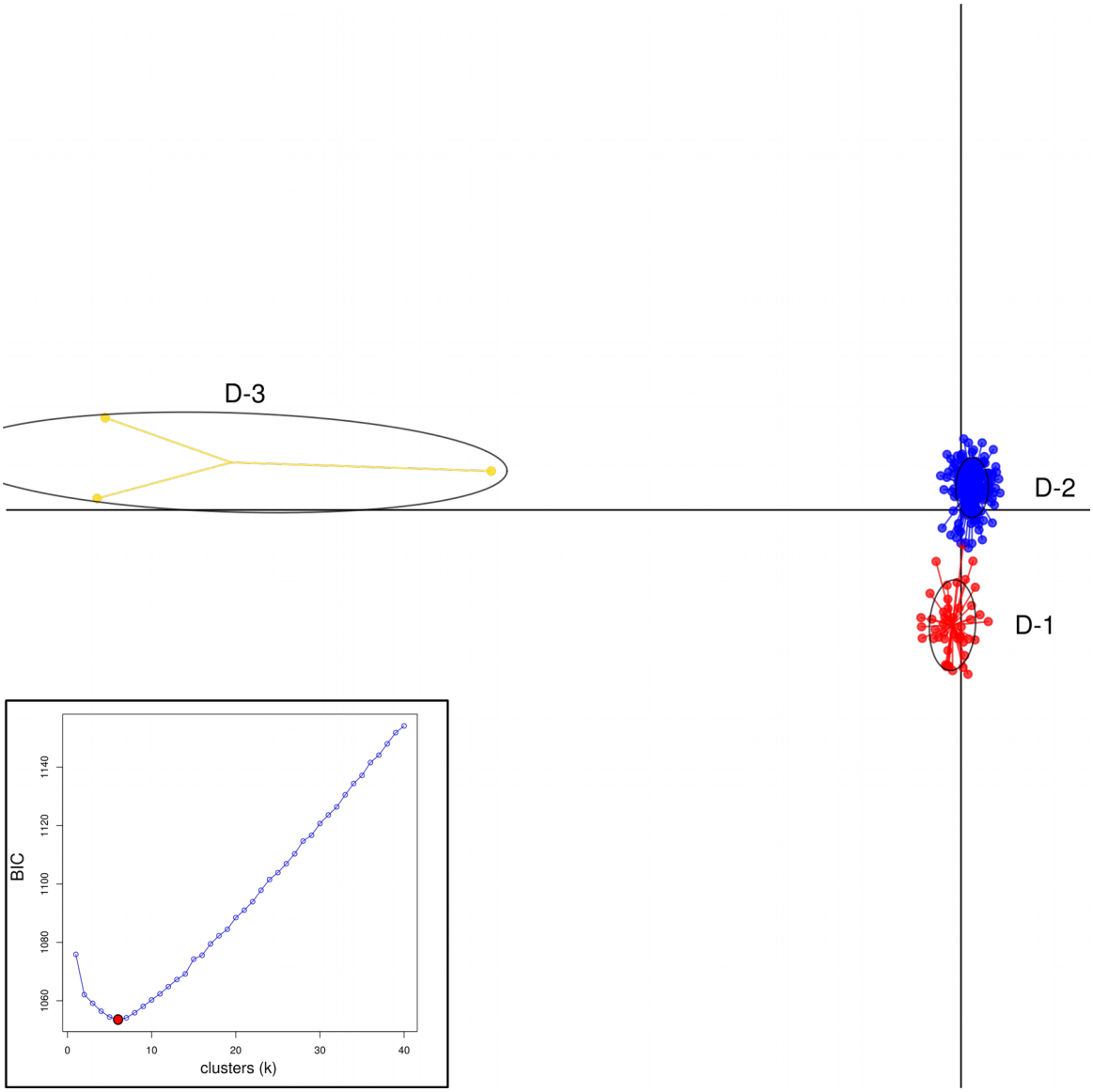
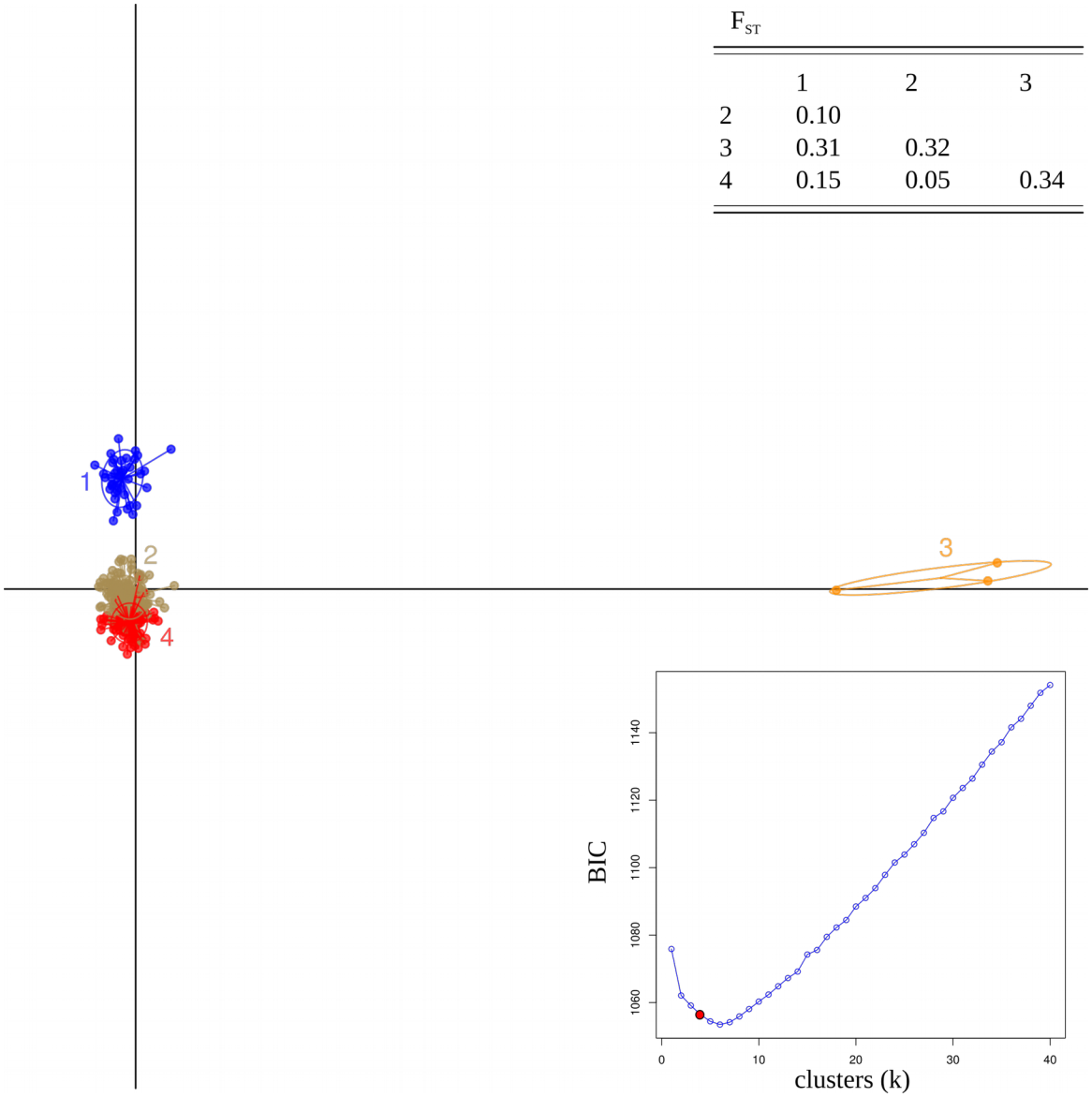


Figure 3-8: Scatter plot from discriminant analysis of principal components (DAPC) within Group II (k = 3). Inset: Results of k-means group selection. The red dot indicates the optimum number of groups.



F_{ST}

	1	2	3
2	0.10		
3	0.31	0.32	
4	0.15	0.05	0.34

Figure 3-9: Scatter plot from discriminant analysis of principal components (DAPC) within Group II (k = 4) with pairwise Nei's F_{ST} for identified clusters in upper right corner. Inset: Results of k-means group selection. The red dot indicates the optimum number of groups.

F_{ST}

	1	2	3	4
2	0.11			
3	0.12	0.06		
4	0.31	0.33	0.32	
5	0.14	0.04	0.04	0.34

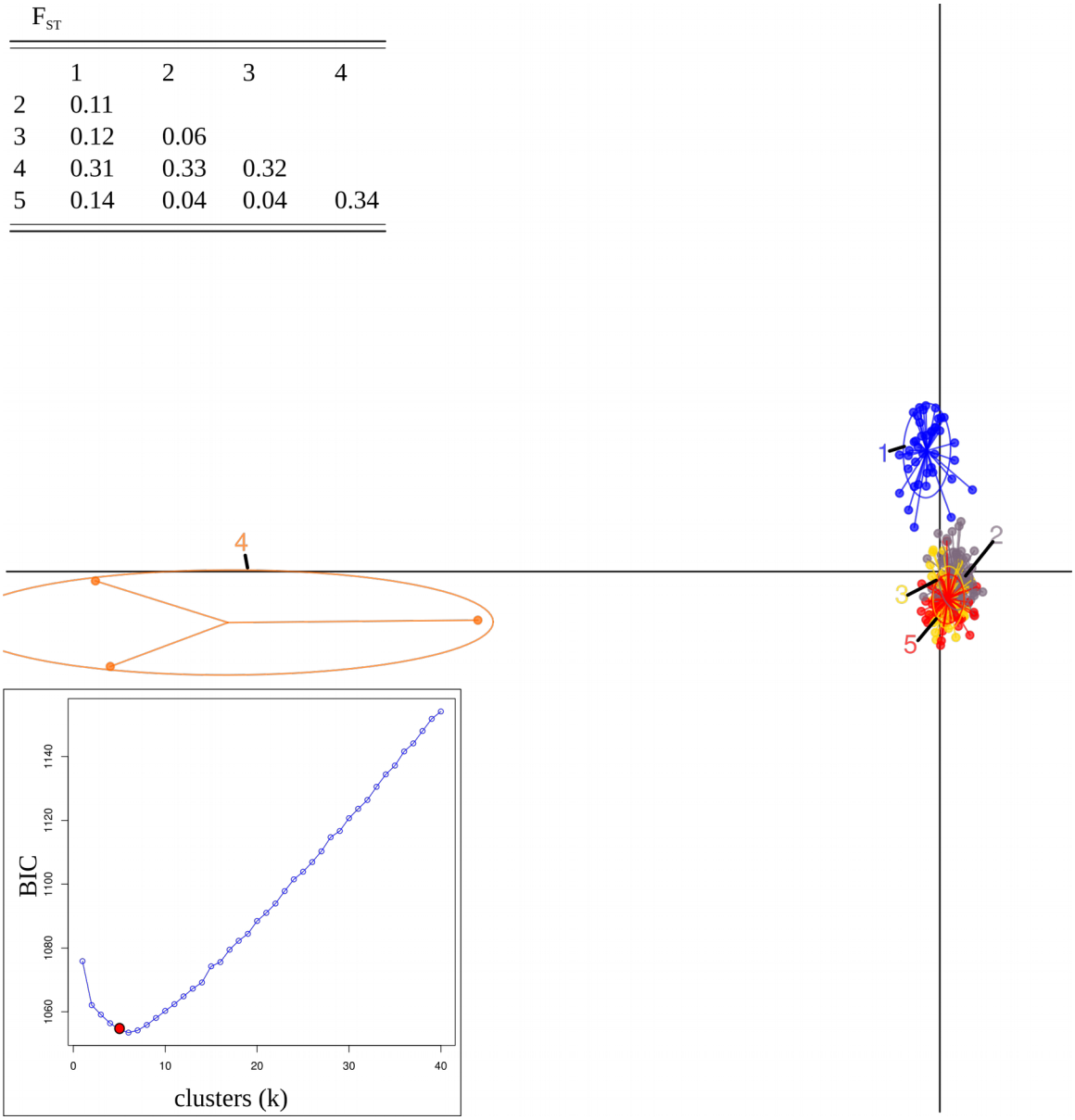
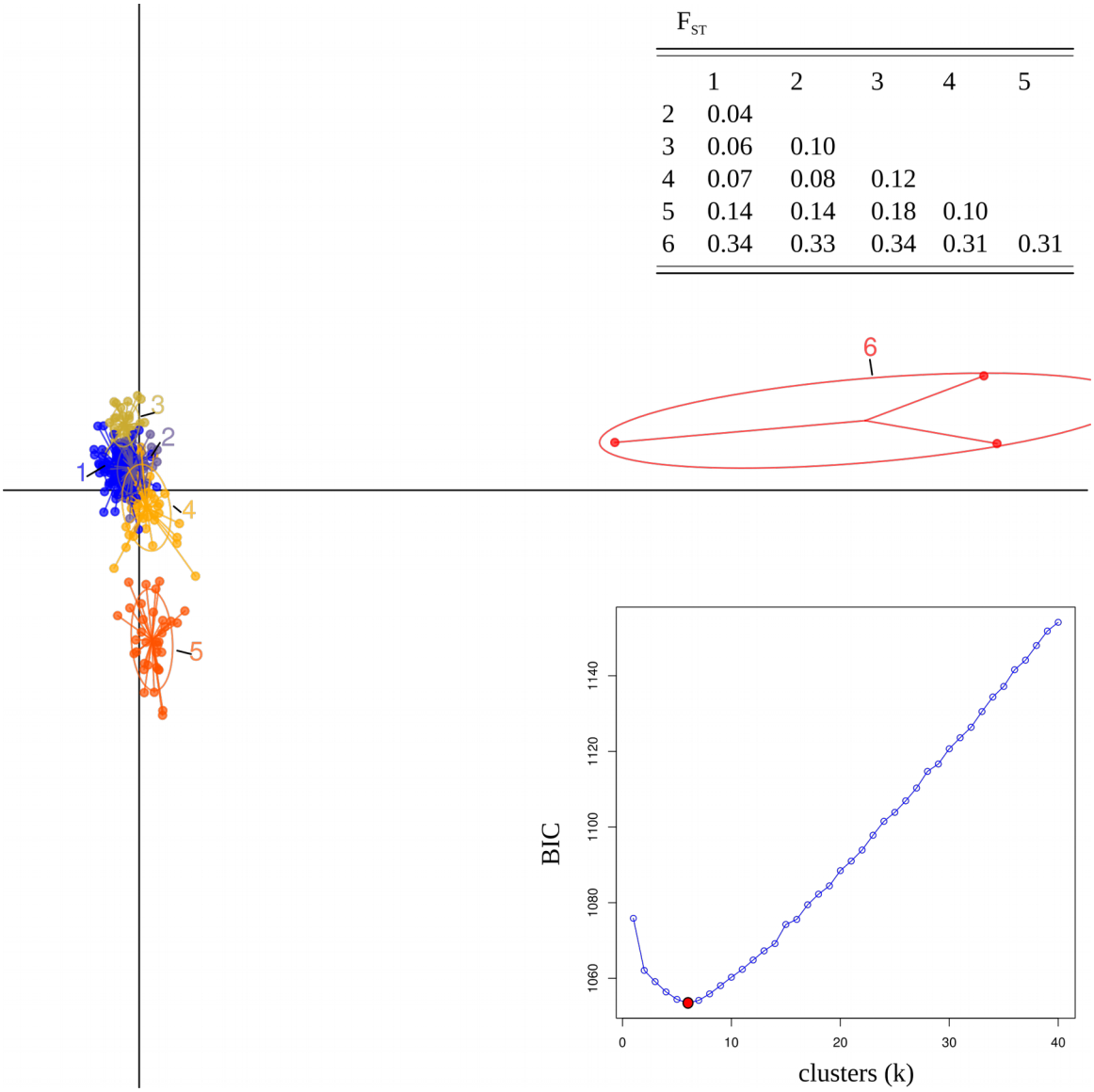


Figure 3-10: Scatter plot from discriminant analysis of principal components (DAPC) within Group II (k = 5) with pairwise Nei's F_{ST} for identified clusters in upper left corner. Inset: Results of k-means group selection. The red dot indicates the optimum number of groups.



F_{ST}

	1	2	3	4	5
2	0.04				
3	0.06	0.10			
4	0.07	0.08	0.12		
5	0.14	0.14	0.18	0.10	
6	0.34	0.33	0.34	0.31	0.31

Figure 3-11: Scatter plot from discriminant analysis of principal components (DAPC) within Group II (k = 6) with pairwise Nei's F_{ST} for identified clusters in upper left corner. Inset: Results of k-means group selection. The red dot indicates the optimum number of groups.

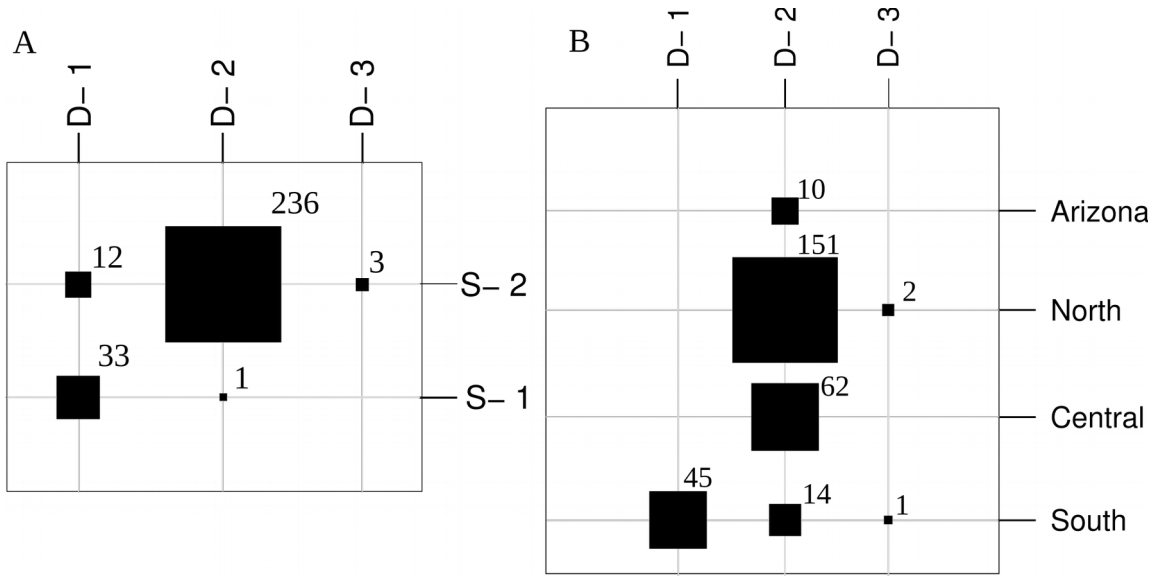


Figure 3-12: A - Contingency table illustrating A - the number of shared individuals between clusters from Group II identified by discriminant analysis of principal components with k=3 (columns marked ‘D’) and SplitsTree network analysis (rows marked ‘S’), and B - the geographic distribution of clusters identified by DAPC.

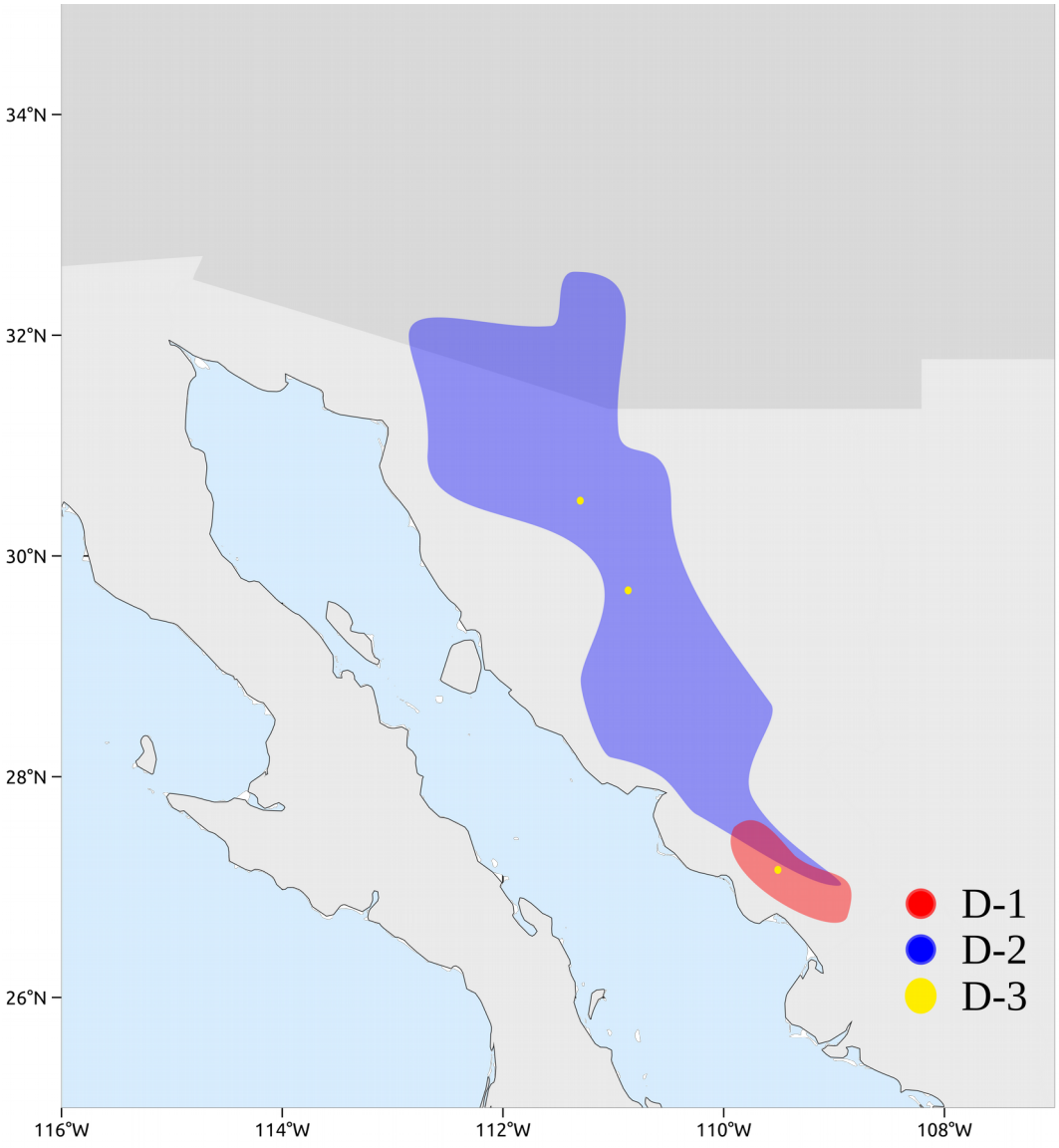


Figure 3-13: Geographic distribution of clusters from Group II identified by discriminant analysis of principal components (k=3).

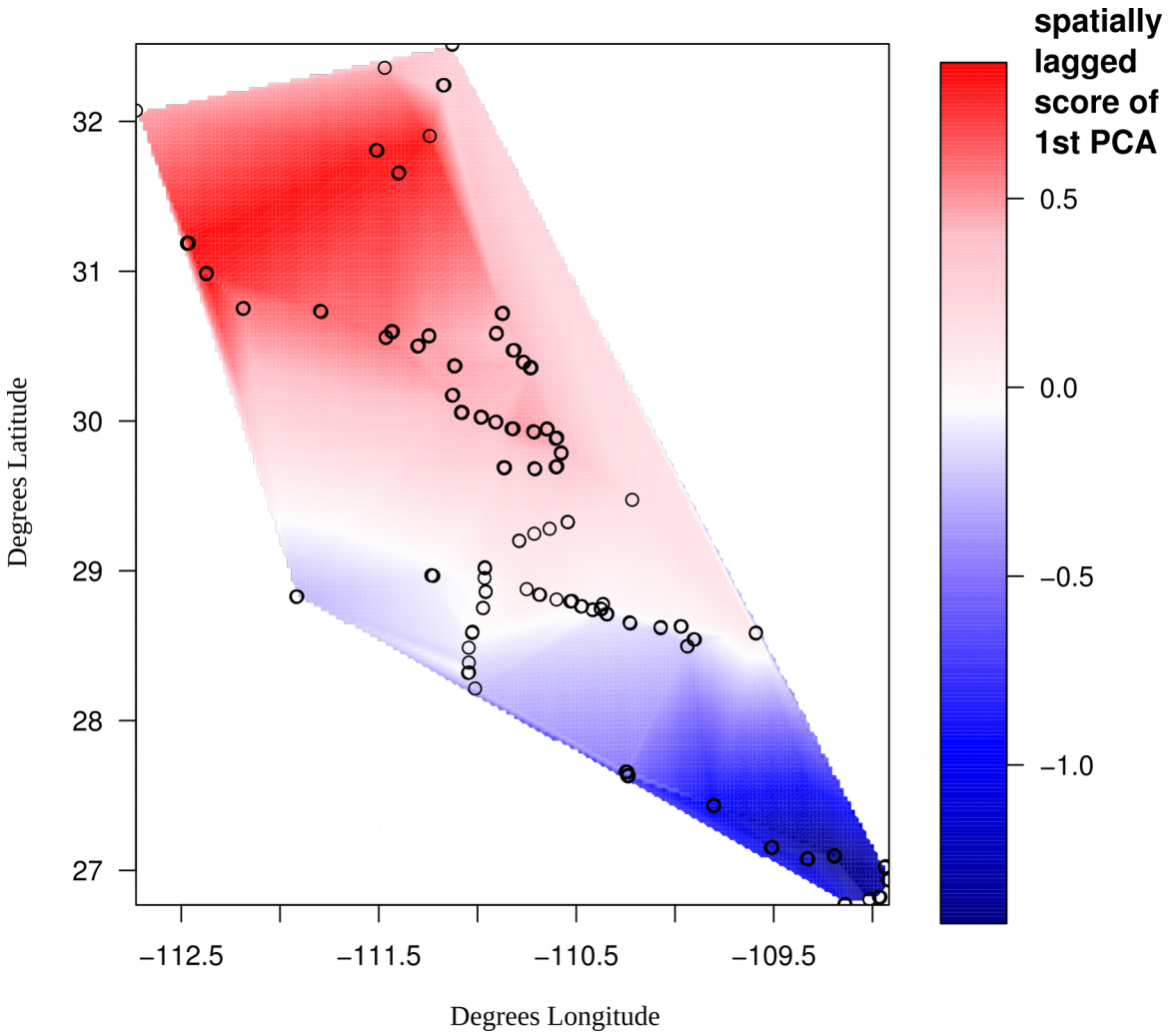


Figure 3-14: Interpolated mapping of the lagged scores of first principal component from spatial PCA for full dataset. Positive and negative spatial autocorrelation of genetic variation are shown in red and blue respectively. Circles show collection localities.

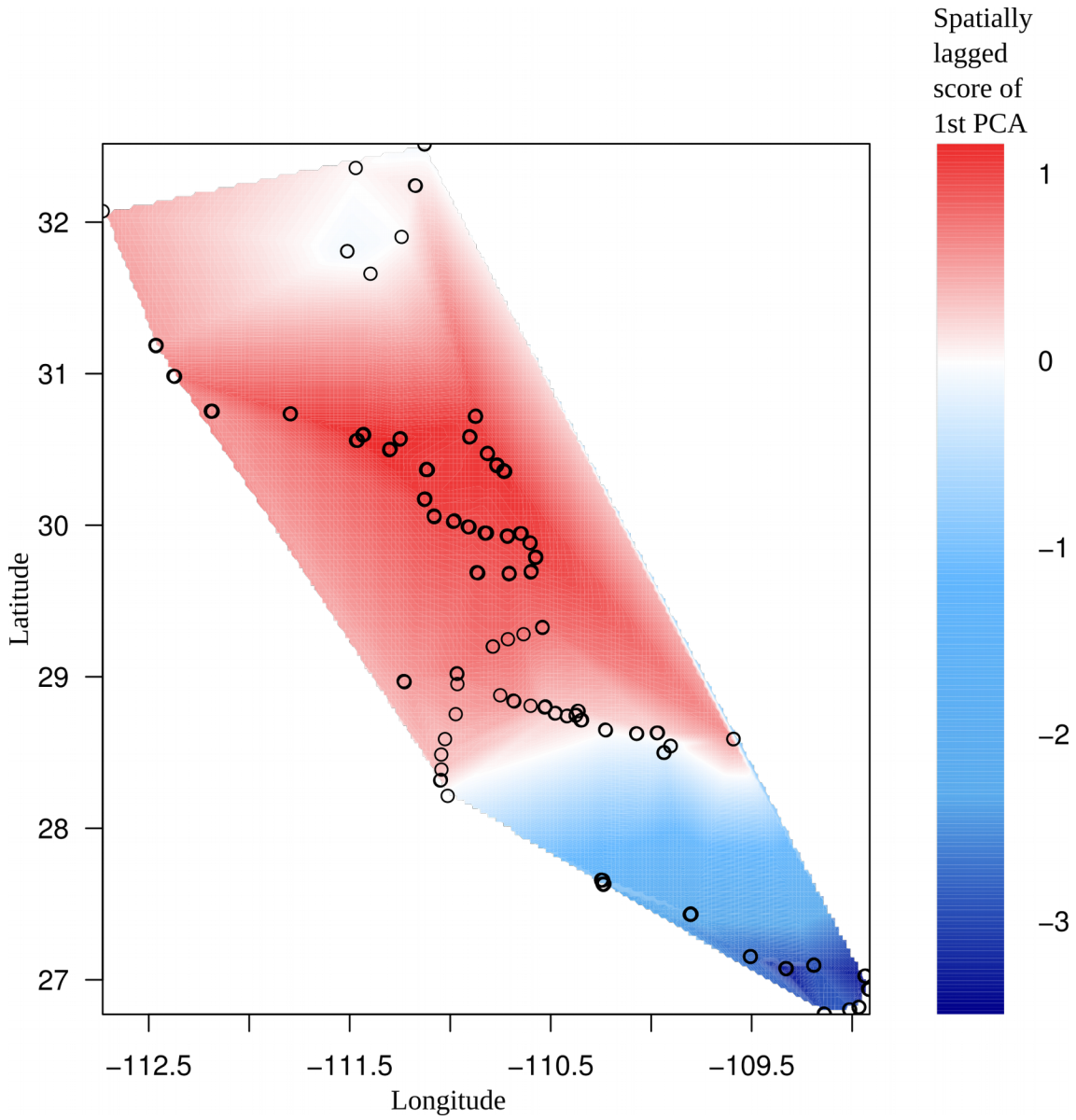


Figure 3-15: Interpolated mapping of the lagged scores of first principal component from spatial PCA for Group II dataset. Positive and negative spatial autocorrelation of genetic variation are shown in red and blue respectively. Circles show collection localities.

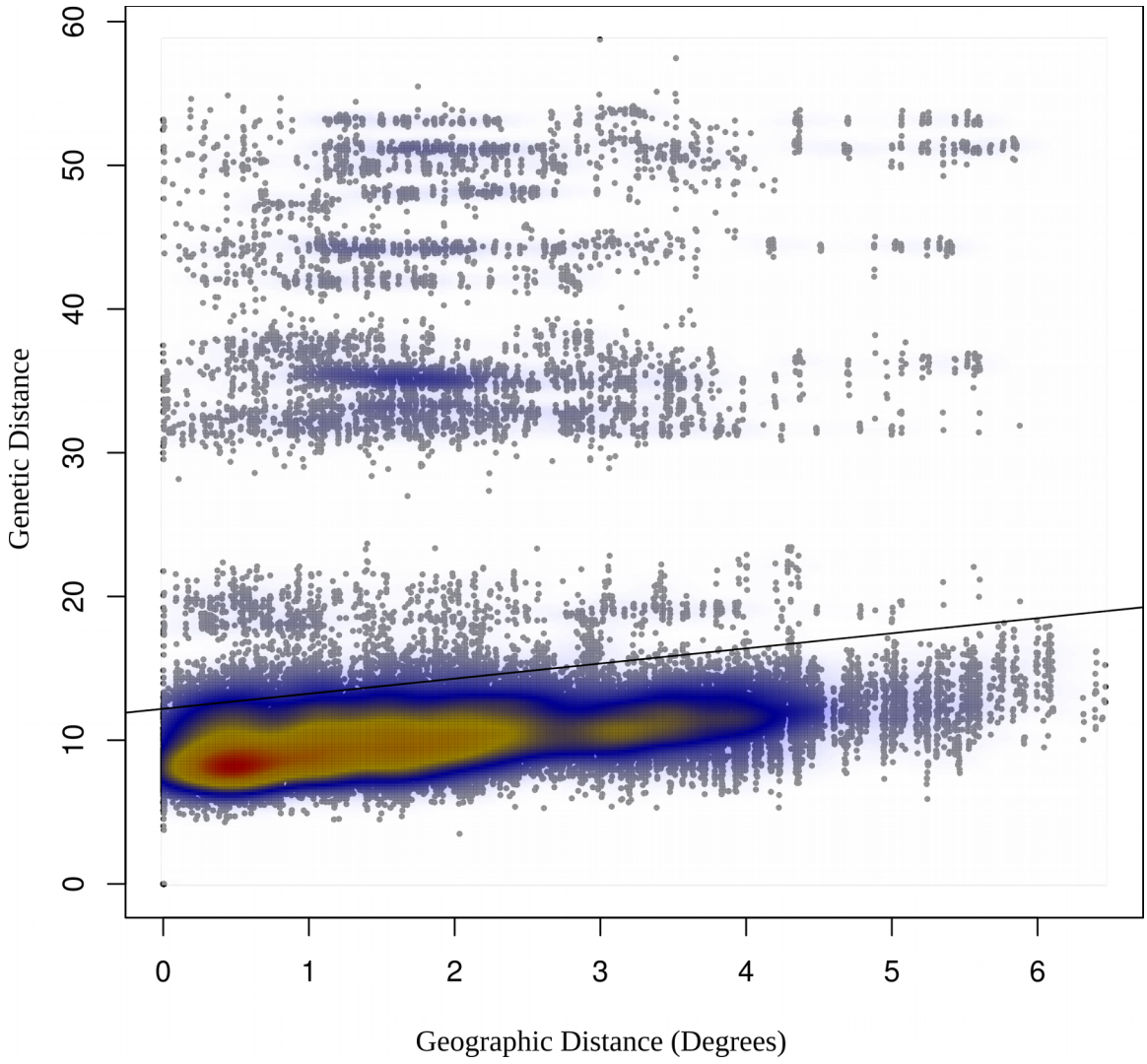


Figure 3-16: Kernel density plot showing the relationship between geographic and genetic distance for *Jatropha cardiophylla* based on the full data excluding *J. vernicosa*. Warmer colors indicate a higher density of points.

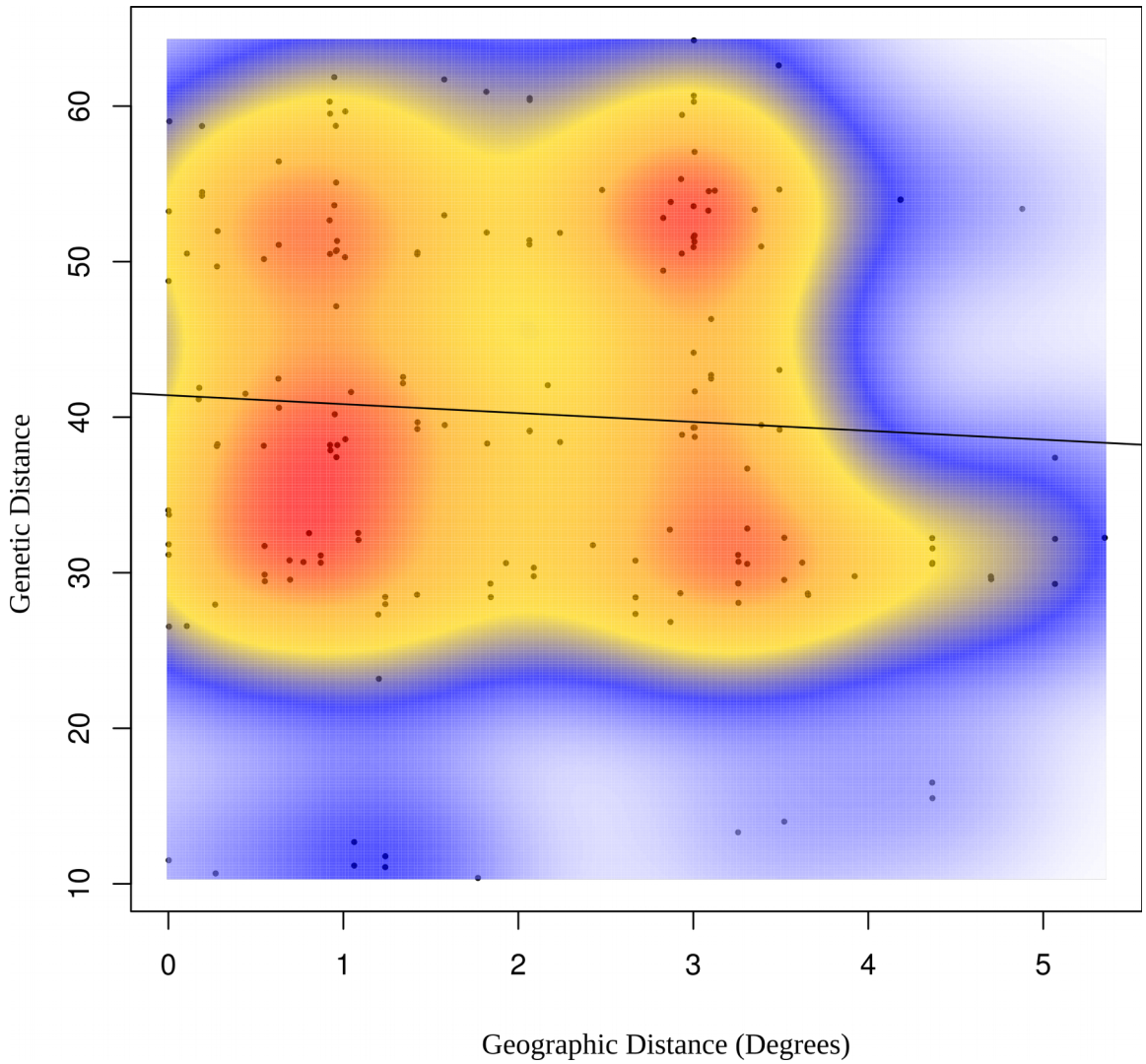


Figure 3-17: Kernel density plot showing the relationship between geographic and genetic distance for the 19 individuals of *Jatropha cardiophylla* identified as Group I in clustering analyses. Warmer colors indicate a higher density of points.

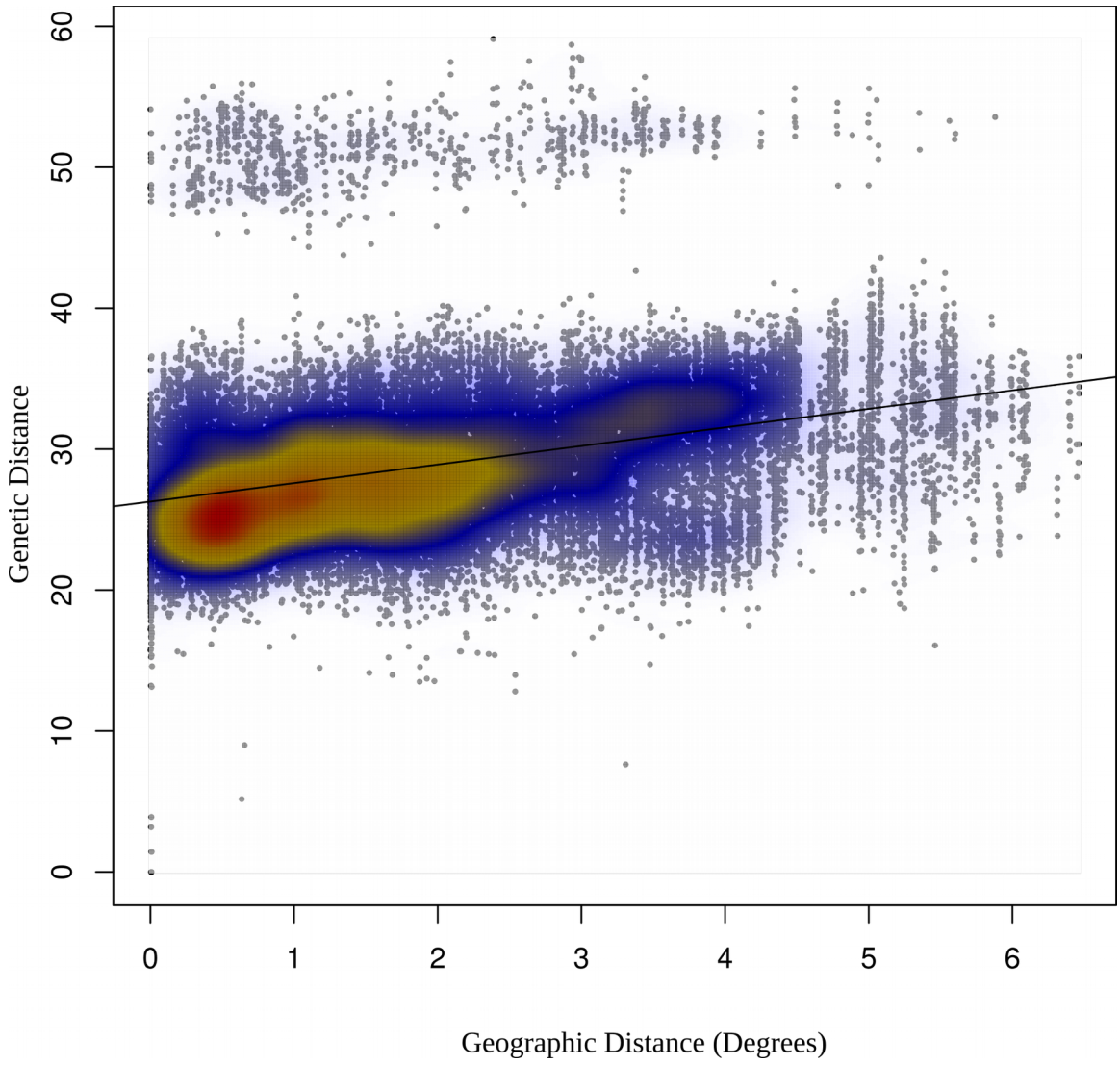


Figure 3-18: Kernel density plot showing the relationship between geographic and genetic distance for the 285 individuals of *Jatropha cardiophylla* identified as Group II in clustering analyses. Warmer colors indicate a higher density of points.

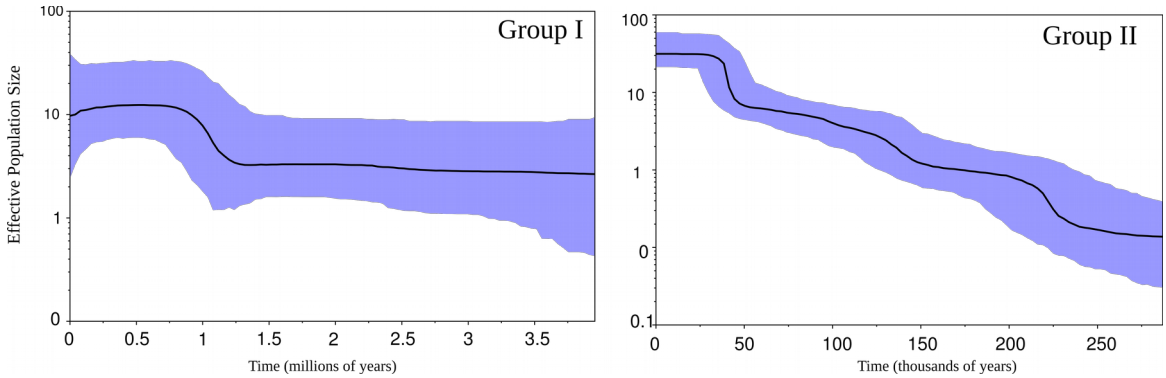


Figure 3-19: Bayesian skyline plots showing changes in effective population size (N_e) for Groups I and II. Note the log scale for the y-axis. Black line is the median estimate of N_e, and the blue envelope is the 95% highest posterior density.

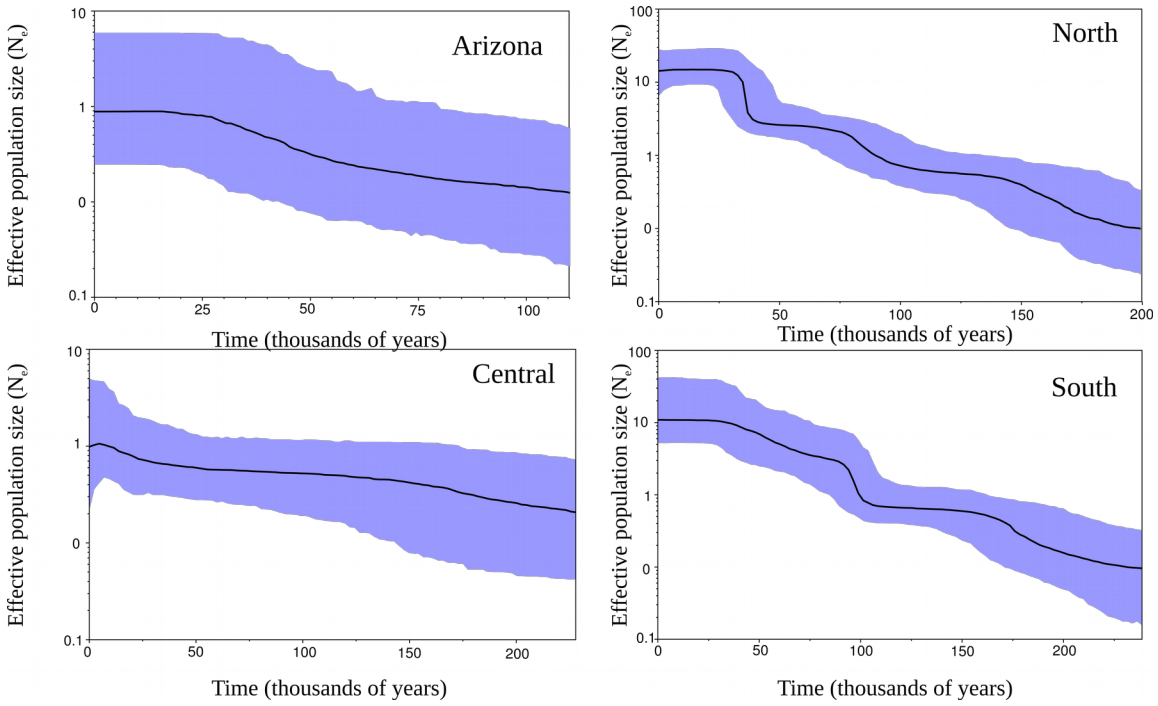


Figure 3-20: Bayesian skyline plots showing changes in effective population size (N_e) for geographic areas using Group II individuals only. Note the log scale for the y-axis. Black line is the median estimate of N_e, and the blue envelope is the 95% highest posterior density.

Environmental variable	WorldClim Code	% Contribution	Permutation importance
Mean Temperature - Coldest Qt.	BIO11	31.6	32.1
Annual Precipitation	BIO12	28.1	23
Mean Temperature - Wettest Qt.	BIO13	13	2.5
Precipitation Seasonality	BIO15	11.7	17.2
Precipitation - Driest Qt.	BIO17	7	7.9
Temperature - Annual Range	BIO7	4.5	8
Mean Temperature - Warmest Qt.	BIO8	1.8	1.4
Mean Temperature - Diurnal Range	BIO2	1	2.9
Isothermality	BIO3	0.8	4.6
Precipitation - Coldest Qt.	BIO19	0.4	0.3

Table 3-3: Environmental variables used to construct the ecological niche model of *Jatropha cardiophylla*, with the percent contribution and permuted contribution (via jackknifing) of each variable to the model. Qt = Quarter.

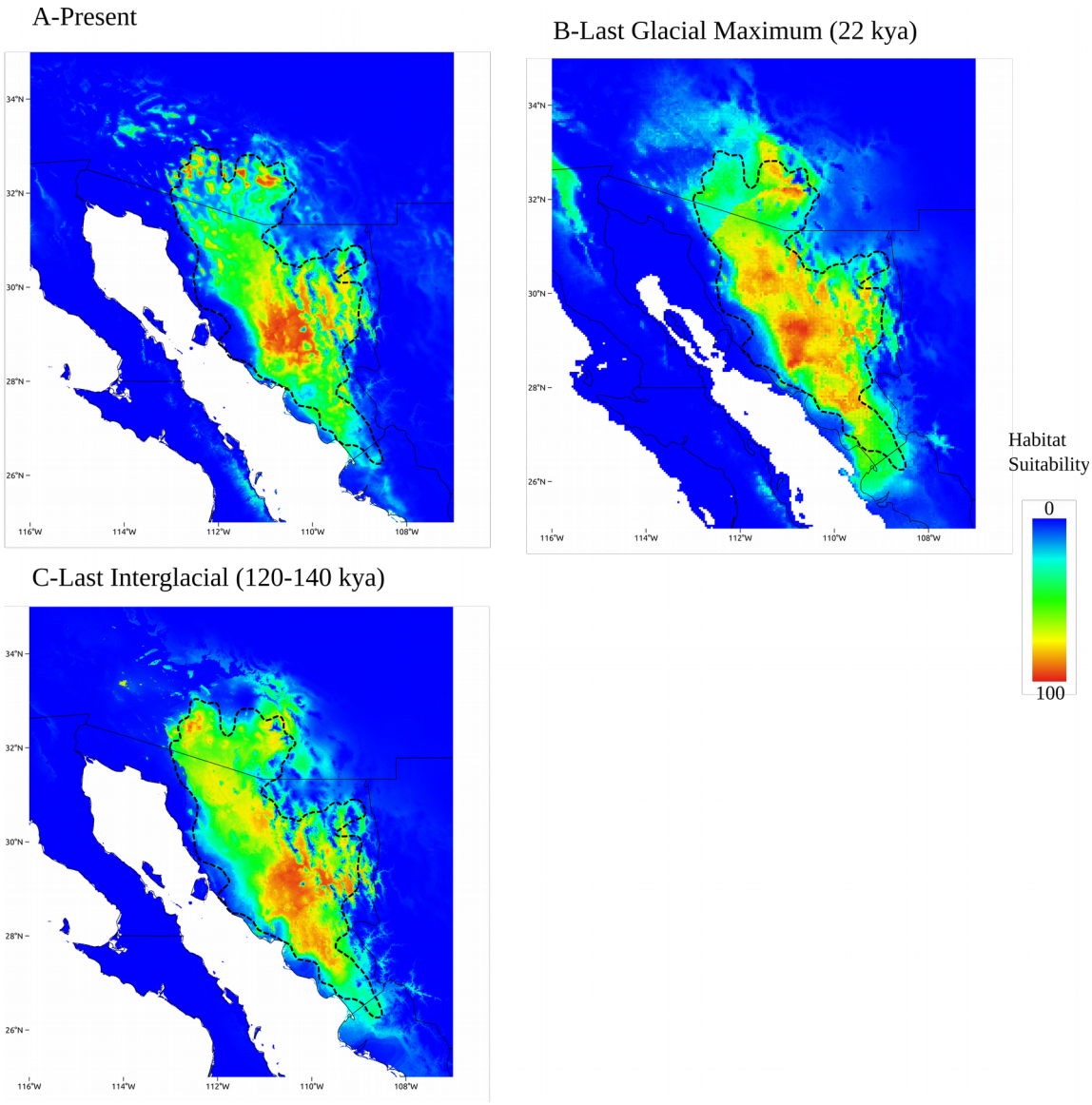


Figure 3-21: Maps showing the distribution of suitable habitat of *Jatropha cardiophylla* at: A- present date, B – Last Glacial Maximum, and C – Last Interglacial. The black dotted line shows the current distribution of *J. cardiophylla*. Habitat suitability is the estimated probability of occurrence.

Literature Cited

Aird, D., Chen, W.S., Ross, M., Connolly, K., Meldrim, J., Russ, C., Fisher, S., Jaffe, D., Nusbaum, C., Gnirke, A., 2011. Analyzing and minimizing bias in Illumina sequencing libraries. Genome Biol. 11, P3.

Ali, J.R., 2012. Colonizing the Caribbean: Is the GAARlandia land-bridge hypothesis gaining a foothold? J. Biogeogr. 39, 431–433. doi:10.1111/j.1365-2699.2011.02674.x

Anderson, R.S., Van Devender, T.R., 1995. Vegetation history and paleoclimates of the coastal lowlands of Sonora, Mexico -pollen records from packrat middens. J. Arid Environ. 30, 295–306.

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., Hohenlohe, P.A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet 17 , 81–92.

Archer, F.I., Adams, P.E., Schneiders, B.B., 2017. strataG: An R package for manipulating, summarizing and analysing population genetic data. Mol. Ecol. Resour. 17, 5–11. doi:10.1111/1755-0998.12559

Avise, J.C. 2000. Phylogeography: the history and formation of species. Harvard Univ. Press, Cambridge, MA.

Axelrod, D., 1979. Age and origin of Sonoran Desert vegetation. Occas. Pap. Calif. Acad. Sci.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3, 1–7.

Bacon, C.D., Silvestro, D., Jaramillo, C., Smith, B.T., Chakrabarty, P., Antonelli, A., 2015. Biological evidence supports an early and complex emergence of the Isthmus of Panama. Proc. Natl. Acad. Sci. 112, E3631–E3631. doi:10.1073/pnas.1511204112

Barido-Sottani, J., Bošková, V., Plessis, L. Du, Kühnert, D., Magnus, C., Mitov, V., Müller, N.F., Pečerska, J., Rasmussen, D.A., Zhang, C., Drummond, A.J., Heath, T.A., Pybus, O.G., Vaughan, T.G., Stadler, T., 2018. Taming the BEAST - A community teaching material resource for BEAST 2. Syst. Biol. 67, 170–174. doi:10.1093/sysbio/syx060

Bartish, I. V., Antonelli, A., Richardson, J.E., Swenson, U., 2011. Vicariance or long-distance dispersal: historical biogeography of the pantropical subfamily Chrysophylloideae (Sapotaceae). J. Biogeogr. 38, 177–190. doi:10.1111/j.1365-2699.2010.02389.x

Beck, J.B., Semple, J.C., 2015. Next-generation sampling: pairing genomics with herbarium specimens provides species-level signal in *Solidago* (Asteraceae). Appl. Plant Sci. 3, 1500014.

Bell, K.C., Hafner, D.J., Leitner, P., Matocq, M.D., 2010. Phylogeography of the ground squirrel subgenus *Xerospermophilus* and assembly of the Mojave Desert biota. J. Biogeogr. 37, 363–378. doi:10.1111/j.1365-2699.2009.02202.x

Betancourt, J.L, Van Devender, T.R., 1990. Packrat middens: the last 40,000 years of biotic change. University of Arizona Press, Tucson.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. a, Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537. doi:10.1371/journal.pcbi.1003537

Bowers, J.E., Turner, R.M., Burgess, T.L., 2004. Temporal and spatial patterns in emergence and early survival of perennial plants in the Sonoran Desert. Plant Ecol. 172, 107–119. doi:10.1023/B:VEGE.0000026026.34760.1b

Brown, D.E., Brennan, T.C., Unmack, P.J., 2007. A digitized biotic community map for plotting and comparing North American plant and animal distributions. Canotia 3, 1–12.

Bryson Jr, R.W., Riddle, B.R., 2012. Tracing the origins of widespread highland species: a case of Neogene diversification across the Mexican sierras in an endemic lizard. Biol. J. Linn. Soc. 105, 382–394. doi:10.1111/j.1095-8312.2011.01798.x

Bryson, R.W., García-Vázquez, U.O., Riddle, B.R., 2012. Diversification in the Mexican horned lizard *Phrynosoma orbiculare* across a dynamic landscape. Mol. Phylogenet. Evol. 62, 87–96. doi:10.1016/j.ympev.2011.09.007

Burnham, K.P., Anderson, D.R., 2002. Model selection and multimodel inference: a practical information-theoretic approach (2nd ed.). Springer, New York, USA. doi:10.1016/j.ecolmodel.2003.11.004

Bushnell, B. 2016. BBMap short read aligner. URL *https://sourceforge.net/projects/bbmap/*

Castoe, T.A., Spencer, C.L., Parkinson, C.L., 2007. Phylogeographic structure and historical demography of the western diamondback rattlesnake (*Crotalus atrox*): A perspective on North American desert biogeography. Mol. Phylogenet. Evol. 42, 193–212. doi:10.1016/j.ympev.2006.07.002

Catchen, J.M., Hohenlohe, P.A.., Bassham, S., Amores, A., Cresko, W.A.., 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22, 3124–3140.

Carreno, A.L., Helenes, J., 2002. Geology and ages of the islands, in: a new island biogeography of the Sea of Cortez. pp. 14–40.

Cervantes, A., Fuentes, S., Gutiérrez, J., Magallón, S., Borsch, T., 2016. Successive arrivals since the Miocene shaped the diversity of the Caribbean Acalyphoideae (Euphorbiaceae). J. Biogeogr. 43, 1773–1785. doi:10.1111/jbi.12790

Chakrabarty, P., 2006. Systematics and historical biogeography of Greater Antillean Cichlidae. Mol. Phylogenet. Evol. 39, 619–27. doi:10.1016/j.ympev.2006.01.014

Chanderbali, A.S., Werff, H. Van Der, Renner, S.S., 2001. Phylogeny and historical biogeography of Lauraceae: evidence from the chloroplast and nuclear genomes. Annal 88, 104–134.

Chessel, D., Dufour, A.B., Thioulouse, J., 2004. The ade4 package: One-table methods. R News 4, 5–10.

Chifman, J., Kubatko, L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics (Oxford, England), 30: 3317-3324.

Clayton, J.W., Soltis, P.S., Soltis, D.E., 2009. Recent long-distance dispersal overshadows ancient biogeographical patterns in a pantropical angiosperm family (Simaroubaceae, Sapindales). Syst. Biol. 58, 395–410. doi:10.1093/sysbio/syp041

Coates, A.G., Collins, L.S., Aubury, M.P., Berggren, W.A., 2004. The geology of the Darien, Panama, and the late Miocene-Pliocene collision of the Panama arc with northwestern South America. Bull. Geol. Soc. Am. 116, 1327–1344. doi:10.1130/B25275.1

Commission for Environmental Cooperation, 1997 Ecological regions of North America: toward a common perspective. Revised 2006. Commission for Environmental Cooperation, Montreal, Quebec, Canada.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. doi:10.1093/bioinformatics/btr330

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772.

Davis, C.C., Bell, C.D., Mathews, S., Donoghue, M.J., 2002. Laurasian migration explains Gondwanan Disjunctions: evidence from Malpighiaceae. Proc. Natl. Acad. Sci. 99, 6833–6837. doi:10.1073/pnas.

Dehgan, B., 1984. Phylogenetic significance of interspecific hybridization in *Jatropha* (Euphorbiaceae). Syst. Bot. 9, 467–478.

Dehgan, B., 2012. *Jatropha* (Euphorbiaceae), Flora Neot. The New York Botanical Garden Press.

Dehgan, B., Schutzman, B., 1994. Contributions toward a monograph of neotropical *Jatropha*: phenetic and phylogenetic analyses. Ann. Missouri Bot. Gard. 81, 349–367.

Dehgan, B., Webster, G.L., 1979. Morphology and infrageneric relationships of the genus *Jatropha*. Univeristy Calif. Publ. Bot. 74, 1–73.

De La Torre, A.R., Li, Z., Van De Peer, Y., Ingvarsson, P.K., 2017. Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. Mol. Biol. Evol. 34, 1363–1377. doi:10.1093/molbev/msx069

DeNova, J.A., Medina, R., Montero, J.C., Weeks, A., Rosell, J. a., Olson, M.E., Eguiarte, L.E., Magallón, S., 2012. Insights into the historical construction of species-rich Mesoamerican seasonally dry tropical forests: the diversification of *Bursera* (Burseraceae, Sapindales). New Phytol. 193, 276–287. doi:10.1111/j.1469-8137.2011.03909.x

Dilcher, L., Manchester, S., 1988. Investigations of Angiosperms from the Eocene of North America: a fruit belonging to the Euphorbiaceae. Tert. Res. 9, 45–58.

Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., Wagner, H.H., 2018. adespatial: multivariate multiscale spatial analysis.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88. doi:10.1371/journal.pbio.0040088

Dupin, J., Matzke, N.J., Särkinen, T., Knapp, S., Olmstead, R.G., Bohs, L., Smith, S.D., 2017. Bayesian estimation of the global biogeographical history of the Solanaceae. J. Biogeogr. 44, 887–899. doi:10.1111/jbi.12898

Durand, E.Y., Patterson, N., Reich, D., Slatkin, M., 2011. Testing for ancient admixture between closely related populations. Mol. Biol. Evol. 28, 2239–2252.

Eaton, D.A.R., 2014. PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30, 1844–1849.

Eaton, D.A.R., Hipp, A.L., González-Rodríguez, A., Cavender-Bares, J., 2015. Historical introgression among the American live oaks and the comparative nature of tests for introgression. Evolution (N. Y). 69, 2587–2601.

Eaton, D.A.R., Overcast, I., 2016. iPyrad: interactive assembly and analysis of RADseq data sets. [WWW Document]. URL http://ipyrad.readthedocs.io/

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

Eaton, D.A.R., Spriggs, E.L., Park, B., Donoghue, M.J., 2017. Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. Syst. Biol. 66, 399–412.

Escudero, M., Eaton, D.A.R., Hahn, M., Hipp, A.L., 2014. Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: A case study in *Carex* (Cyperaceae). Mol. Phylogenet. Evol. 79, 359–367.

Excoffier, L., Heckel, G., 2006. Computer programs for population genetics data analysis: A survival guide. Nat. Rev. Genet. 7, 745–758. doi:10.1038/nrg1904

Fairless, D., 2007. The little shrub that could—maybe. Nature, 449, 652-655.

Fehlberg, S.D., Fehlberg, K.M., 2017. Spatial genetic structure in brittlebush (*Encelia farinosa*, Asteraceae) in the southwestern deserts of North America: a comparison of

nuclear and chloroplast DNA sequences. Plant Syst. Evol. 303, 1367–1382. doi:10.1007/s00606-017-1463-2

Ferrari, L., Orozco-Esquivel, T., Manea, V., Manea, M., 2012. The dynamic history of the Trans-Mexican Volcanic Belt and the Mexico subduction zone. Tectonophysics 522–523, 122–149. doi:10.1016/j.tecto.2011.09.018

Fu, Y.X., 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147, 915–925.

Gándara, E., Sosa, V., 2014. Spatio-temporal evolution of *Leucophyllum pringlei* and allies (Scrophulariaceae): a group endemic to North American xeric regions. Mol. Phylogenet. Evol. 76, 93–101. doi:10.1016/j.ympev.2014.02.027

García-Palomo, A., Macías, J.L., Tolson, G., Valdez, G., Mora, J.C., 2002. Volcanic stratigraphy and geological evolution of the Apan region, east-central sector of the Trans-Mexican Volcanic Belt. Geofis. Int. 41, 133–150.

González-Trujillo, R., Correa-Ramírez, M.M., Ruiz-Sanchez, E., Salinas, E.M., Jiménez, M.L., León, F.J.G.-D., 2016. Pleistocene refugia and their effects on the phylogeography and genetic structure of the wolf spider *Pardosa sierra* (Araneae: Lycosidae) on the Baja California Peninsula. J. Arachnol. 44, 367–379. doi:10.1636/R15-84.1

Goudet, J., Jombart, T., 2015. hierfstat: Estimation and Tests of Hierarchical F-Statistics.

Govaerts, R.,Frodin, D.G., Radcliffe-Smith, A., 2000. World checklist and bibliography of Euphobiaceae (and Pandaceae). The Royal Botanic Gardens, Kew.

Grafen, A., 1989. The Phylogenetic Regression. Philos. Trans. R. Soc. B Biol. Sci. 326, 119–157.

Graham, C.F., Glenn, T.C., McArthur, A.G., Boreham, D.R., Kieran, T., Lance, S., Manzon, R.G., Martino, J., Pierson, T., Rogers, S.M., Wilson, J.Y., Somers, C.M., 2015. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). Mol. Ecol. Resour. 1-12.

Graham, M.R., Hendrixson, B.E., Hamilton, C. , Bond, J.E., 2015. Miocene extensional tectonics explain ancient patterns of diversification among turret-building tarantulas (*Aphonopelma mojave* group) in the Mojave and Sonoran deserts. J. Biogeogr. n/a-n/a. doi:10.1111/jbi.12494

Gugger, P.F., Gonzalez-Rodriguez, A., Rodriguez-Correa, H., Sugita, S., Cavender-Bares, J., 2011. Southward Pleistocene migration of Douglas-fir into Mexico: Phylogeography, ecological niche modeling, and conservation of “rear edge” populations. New Phytol. 189, 1185–1199. doi:10.1111/j.1469-8137.2010.03559.x

Gutiérrez-García, T.A., Vázquez-Domínguez, E., 2013. Consensus between genes and stones in the biogeographic and evolutionary history of Central America. Quat. Res. (United States) 79, 311–324. doi:10.1016/j.yqres.2012.12.007

Hafner, D.J., Riddle, B.R., 2011. Boundaries and barriers of North American warm deserts: an evolutionary perspective. Palaeogeogr. palaeobiogeography Biodivers. Sp. time 73–112.

Harvey, M.G., Judy, C.D., Seeholzer, G.F., Maley, J.M., Graves, G.R., Brumfield, R.T., 2015. Similarity thresholds used in DNA sequence assembly from short reads can reduce the comparability of population histories across species. PeerJ 3, e895.

Hearn, D.J., 2006. *Adenia* (Passifloraceae) and its adaptive radiation: Phylogeny and growth form diversification. Syst. Bot. 31, 805–821.

Hedges, S.B., 2006. Paleogeography of the Antilles and origin of West Indian terrestrial vertebrates. Ann. Missouri Bot. Gard. 93, 231–244. doi:10.3417/0026-6493(2006)93[231:POTAAO]2.0.CO;2

Heled, J., Drummond, A.J., 2015. Calibrated birth-death phylogenetic time-tree priors for Bayesian inference. Syst. Biol. 64, 369–383. doi:10.1093/sysbio/syu089

Herrera, S., Shank, T.M., 2016. RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. Mol. Phylogenet. Evol. 100, 70–79.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978. doi:10.1002/joc.1276

Hipp, A.L., Eaton, D.A.R., Cavender-Bares, J., Fitzek, E., Nipper, R., Manos, P.S., 2014. A framework phylogeny of the American oak clade based on sequenced RAD data. PLoS One 9.

Huang, H., Knowles, L.L., 2014. Unforeseen Consequences of Excluding Missing Data from Next-Generation Sequences: Simulation Study of RAD Sequences. Syst. Biol. 0, 1–9.

Holsinger, K.E., Weir, B.S., 2009. Genetics in geographically structured populations: Defining, estimating and interpreting F_{ST}. Nat. Rev. Genet. doi:10.1038/nrg2611

Huelsenbeck, J.P., Nielsen, R., Bollback, J.P., 2003. Stochastic mapping of morphological characters. Syst. Biol. 52, 131–158. doi:10.1080/10635150309342

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267. doi:10.1093/molbev/msj030

Iturralde-Vinent, M.A., 2006. Meso-Cenozoic Caribbean Paleogeography: implications for the historical biogeography of the region. Int. Geol. Rev. 48, 791–827. doi:10.2747/0020-6814.48.9.791

Iturralde-Vinent, M.A., MacPhee, R.D.E., 1999. Paleogeography of the Caribbean region: implications for Cenozoic biogeography. Bull. Am. Museum Nat. Hist. 238, 1–95. doi:10.2747/0020-6814.48.9.791

Jaeger, J.R., Riddle, B.R., Bradford, D.F., 2005. Cryptic Neogene vicariance and Quaternary dispersal of the red-spotted toad (*Bufo punctatus*): insights on the evolution of North American warm desert biotas. Mol. Ecol. 14, 3033–3048. doi:10.1111/j.1365-294X.2005.02645.x

Janzen, D.H., 1988. Tropical dry forest, in: Wilson, E.O.& F.M.P. (Ed.), Biodiversity. National Academy Press. Washington D.C., pp. 130–137.

Jaramillo, C., Montes, C., Cardona, A., Silvestro, D., Antonelli, A., Bacon, C.D., 2017. Comment (1) on “Formation of the Isthmus of Panama” by O ’ Dea et al. Sci. Adv. 3, e1602321. doi:10.1126/sciadv.1602321

Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405. doi:10.1093/bioinformatics/btn129

Jombart, T., Ahmed, I., 2011. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. Bioinformatics 27, 3070–3071. doi:10.1093/bioinformatics/btr521

Jombart, T., Devillard, S., Dufour, A., Pontier, D., 2008. Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity (Edinb). 101, 92–103. doi:10.1038/hdy.2008.34

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066. doi:10.1093/nar/gkf436

Kishino, H., Miyata, Z.T., Hasegawa, M., 1990. Maximum Likelihood Inference of Protein Phylogeny and the Origin of Chloroplasts. J. Mol. Evol. 31, 151–160.

Knaus, B.J., Grunwald, N.J., 2016. vcfR: an R package to manipulate and visualize VCF format data. bioRxiv 041277. doi:10.1101/041277

Landis, M.J., Matzke, N.J., Moore, B.R., Huelsenbeck, J.P., 2013. Bayesian analysis of biogeography when the number of areas is large. Syst. Biol. 62, 789–804. doi:10.1093/sysbio/syt040

Laport, R.G., Minckley, R.L., Ramsey, J., 2012. Phylogeny and cytogeography of the North American Creosote Bush (*Larrea tridentata*>, Zygophyllaceae). Syst. Bot. 37, 153–164. doi:10.1600/036364412X616738

Lavin, M., 2006. Floristic and geographical stability of discontinuous seasonally dry tropical forests explains patterns of plant phylogeny and endemism, in: Pennington, R.T., Ratter, J.A. (Eds.), Neotropical savannahs and seasonally dry forests: plant diversity, biogeography and conservation. CRC Press, pp. 425–440.

Lavin, M., Schrire, B.D., Lewis, G., Pennington, R.T., Delgado-Salinas, A., Thulin, M., Hughes, C.E., Matos, A.B., Wojciechowski, M.F., 2004. Metacommunity process rather than continental tectonic history better explains geographically structured phylogenies in legumes. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 359, 1509–1522. doi:10.1098/rstb.2004.1536

Leal, I.R., Wirth, R., Tabarelli, M., 2007. Seed dispersal by ants in the semi-arid caatinga of north-east Brazil. Ann. Bot. 99, 885–894. doi:10.1093/aob/mcm017

Linder, H.P., 2013. Phylogeography. J. Biogeogr. 44, 243–244. doi:10.1016/B978-0-12-374984-0.01161-X

Loera, I., Ickert-Bond, S.M., Sosa, V., 2017. Pleistocene refugia in the Chihuahuan Desert: The phylogeographic and demographic history of the gymnosperm *Ephedra compacta*. J. Biogeogr. 2706–2716. doi:10.1111/jbi.13064

Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536.

Maddison, W.P., Knowles, L.L., 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55, 21–30.

Marshall, L.G., Webb, D.S., Sepkoski, J.J., Raup, D.M., 1982. Mammalian evolution and the Great American Interchange. Science (80-.). 215, 1532–1357. doi:10.1126/science.215.4538.1351

Mastretta-Yanes, A., Moreno-Letelier, A., Piñero, D., Jorgensen, T.H., Emerson, B.C., 2015. Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt. J. Biogeogr. 42, 1586–1600. doi:10.1111/jbi.12546

Matos-Maraví, P., Núñez Águila, R., Peña, C., Miller, J.Y., Sourakov, A., Wahlberg, N., 2014. Causes of endemic radiation in the Caribbean: evidence from the historical biogeography and diversification of the butterfly genus Calisto (Nymphalidae: Satyrinae: Satyrini). BMC Evol. Biol. 14, 199. doi:10.1186/s12862-014-0199-7

Matzke, N.J., 2013. BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. R Packag. version 0.2 1, http://CRAN.R-project.org/package=BioGeoBEARS.

Matzke, N.J., 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Syst. Biol. 63, 951–970. doi:10.1093/sysbio/syu056

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T., 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. Mol. Phylogenet. Evol. 66, 526–538.

McLoughlin, S., 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. Aust. J. Bot. 49, 271–300.

Metcalfe, S.E., 2006. Late Quaternary Environments of the northern Deserts and central Transvolcanic Belt of Mexico. Ann. Missouri Bot. Gard. 93, 258–273.

Michalak, I., Zhang, L.B., Renner, S.S., 2010. Trans-Atlantic, trans-Pacific and trans-Indian Ocean dispersal in the small Gondwanan Laurales family Hernandiaceae. J. Biogeogr. 37, 1214–1226. doi:10.1111/j.1365-2699.2010.02306.x

Miller, M. A., Pfeiffer, W., Schwartz, T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*, 1–8.

Montes, C., Cardona, A., Jaramillo, C., Pardo, A., Silva, J.C., Valencia, V., Ayala, C., Pérez-Angel, L.C., Rodriguez-Parra, L.A., Ramirez, V., Niño, H., 2015. Middle

Miocene closure of the Central American Seaway. Science (80-.). 348, 226–229. doi:10.1126/science.aaa2815

Mooney, H.A., Bullock, S.H., Medina, E., 1995. Introduction, in: Bullock, S.H., Mooney, H.A., Medina, E. (Eds.), Seasonally dry tropical forests. Cambridge University Press, Cambridge.

Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 37, 17–23.

Morley, R.J., 2003. Interplate dispersal paths for megathermal angiosperms. Perspect. Plant Ecol. Evol. Syst. 6, 5–20. doi:10.1078/1433-8319-00039

Myers, E.A., Bryson, R.W., Hansen, R.W., Aardema, M.L., Lazcano, D., Burbrink, F.T., 2018. Exploring Chihuahuan Desert diversification in the gray-banded kingsnake, *Lampropeltis alterna* (Serpentes: Colubridae). Mol. Phylogenet. Evol. 131, 211–218. doi:10.1016/j.ympev.2018.10.031

Myers, E.A., Hickerson, M.J., Burbrink, F.T., 2017. Asynchronous diversification of snakes in the North American warm deserts. J. Biogeogr. 44, 461–474. doi:10.1111/jbi.12873

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G. A., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403, 853–8. doi:10.1038/35002501

Nei, M., 1987. Molecular evolutionary genetics. Columbia University Press.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.

Nie, Z.-L., Sun, H., Manchester, S.R., Meng, Y., Luke, Q., Wen, J., 2012. Evolution of the intercontinental disjunctions in six continents in the *Ampelopsis* clade of the grape family (Vitaceae). BMC Evol. Biol. 12, 17. doi:10.1186/1471-2148-12-17

Nieto-Blázquez, M.E., Antonelli, A., Roncal, J., 2017. Historical Biogeography of endemic seed plant genera in the Caribbean: Did GAARlandia play a role? Ecol. Evol. 7, 10158–10174. doi:10.1002/ece3.3521

O'Dea, A., Lessios, H.A., Coates, A.G., Eytan, R.I., Restrepo-Moreno, S.A., Cione, A.L., Collins, L.S., De Queiroz, A., Farris, D.W., Norris, R.D., Stallard, R.F., Woodburne, M.O., Aguilera, O., Aubry, M.P., Berggren, W.A., Budd, A.F., Cozzuol, M.A.,

Coppard, S.E., Duque-Caro, H., Finnegan, S., Gasparini, G.M., Grossman, E.L., Johnson, K.G., Keigwin, L.D., Knowlton, N., Leigh, E.G., Leonard-Pingel, J.S., Marko, P.B., Pyenson, N.D., Rachello-Dolmen, P.G., Soibelzon, E., Soibelzon, L., Todd, J.A., Vermeij, G.J., Jackson, J.B.C., 2016. Formation of the Isthmus of Panama. Sci. Adv. 2, 1–11. doi:10.1126/sciadv.1600883

Otto-Bliesner, B.L., Marshall, S.J., Overpeck, J.T., Miller, G.H., Hu, A., Members, C.L.I.P., 2006. Simulating Arctic climate warmth and ice sheet sensitivity for the Last Interglacial. Science (80-.). 1751–1754. doi:10.1007/s13595-014-0446-5

Paradis, E., 2010. pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics 26, 419–20. doi:10.1093/bioinformatics/btp696

Paradis, E., Claude, J., Strimmer, K., 2004. ape: Analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290. doi:10.1093/bioinformatics/btg412

Pennington, R.T., Lavin, M., Oliveira-Filho, A.T., 2009. Woody plant diversity, evolution, and ecology in the tropics: perspectives from seasonally dry tropical forests. Annu. Rev. Ecol. Evol. Syst. 40, 437–457. doi:10.1146/annurev.ecolsys.110308.120327

Pennington, R.T., Prado, D.E., Pendry, C. a., 2000. Neotropical seasonally dry forests and Quaternary vegetation changes. J. Biogeogr. 27, 261–273. doi:10.1046/j.1365-2699.2000.00397.x

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7, e37135. doi:10.1371/journal.pone.0037135

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Modell. 190, 231–259. doi:10.1016/j.ecolmodel.2005.03.026

Provost, K., Mauck, W., York, N., Smith, B.T., 2018. Genomic divergence in allopatric Northern Cardinals of the North American warm deserts is associated with behavioral differentiation. bioRxiv 1–27. doi:10.1101/347492

QGIS Development Team, 2014. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org. Qgisorg. doi:http://www.qgis.org/

Radcliffe-Smith, A., 1997. Notes on Madagascan Euphorbiaceae V: *Jatropha*. Kew Bull. 52, 177–181.

Radcliffe-Smith, A., & Esser, H.J., 2001. Genera *Euphorbiacearum*. Royal Botanic Gardens, Kew, UK

Rambaut A, Suchard MA, X.D.& D.A., 2014. Tracer v1.6 [WWW Document]. URL http://beast.bio.ed.ac.uk/Tracer

Raven, P., Axelrod, D., 1974. Angiosperm biogeography and past continental movements. Ann. Missouri Bot. Gard. 61, 539–673.

Rebernig, C. a., Schneeweiss, G.M., Bardy, K.E., Schönswetter, P., Villaseñor, J.L., Obermayer, R., Stuessy, T.F., Weiss-Schneeweiss, H., 2010. Multiple Pleistocene refugia and Holocene range expansion of an abundant southwestern American desert plant species (*Melampodium leucanthum*, Asteraceae). Mol. Ecol. 19, 3421–3443. doi:10.1111/j.1365-294X.2010.04754.x

Ree, R.H., Smith, S. a, 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57, 4–14. doi:10.1080/10635150701883881

Renaud, G., Stenzel, U., Maricic, T., Wiebe, V., & Kelso, J. 2015. DeML: Robust demultiplexing of Illumina sequences using a likelihood-based approach. Bioinformatics, 31, 770–772.

Revell, L.J., 2012. phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3, 217–223. doi:10.1111/j.2041-210X.2011.00169.x

Richards, C.L., Carstens, B.C., Knowles, L., 2007. Distribution modeling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. J. Biogeogr. 34, 1833–1845. doi:10.1111/j.1365-2699.2007.01814.x

Ronquist, F., 1997. Dispersal-Vicariance Analysis: a new approach to the quantification of historical biogeography. Syst. Biol. 46, 195. doi:10.2307/2413643

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. a, Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–42.

Rosen, D.E., 1975. A vicariance model of Caribbean biogeography. Syst. Biol. 24, 431–464. doi:10.1093/sysbio/24.4.431

Rosenberg, N.A., Nordborg, M., 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nat. Rev. Genet. 3, 380–390. doi:10.1038/nrg795

Rubin, B.E.R., Ree, R.H., Moreau, C.S., 2012. Inferring phylogenies from RAD sequence data. PLoS One 7.

Scheinvar, E., Gámez, N., Castellanos-Morales, G., Aguirre-Planter, E., Eguiarte, L.E., 2016. Neogene and Pleistocene history of *Agave lechuguilla* in the Chihuahuan Desert. J. Biogeogr. 1–13. doi:10.1111/jbi.12851

Schönswetter, P., Stehlik, I., Holderegger, R., Tribsch, A., 2005. Molecular evidence for glacial refugia of mountain plants in the European Alps. Mol. Ecol. 14, 3547–3555. doi:10.1111/j.1365-294X.2005.02683.x

Schrire, B.D., Lavin, M., Barker, N.P., Forest, F., 2009. Phylogeny of the tribe Indigofereae (Leguminosae-Papilionoideae): geographically structured more in succulent-rich and temperate settings than in grass-rich environments. Am. J. Bot. 96, 816–852. doi:10.3732/ajb.0800185

Schweyen, H., Rozenberg, A., Leese, F., 2014. Detection and removal of PCR duplicates in population genomic ddRAD studies by addition of a degenerate base region (DBR) in sequencing adapters 146–160.

SEINet Portal Network [WWW Document], 2018. URL http//:swbiodiversity.org/seinet/index.php

Sepulchre, P., Arsouze, T., Donnadieu, Y., Dutay, J.-C., Jaramillo, C., Le Bras, J., Martin, E., Montes, C., Waite, A.J., 2013. Consequences of shoaling of the Central American Seaway determined from modeling Nd isotopes. Paleoceanography 29, 176–189. doi:10.1002/2013pa002501

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Shreve, F.B., 1922. Conditions indirectly affecting vertical distribution on desert mountains. Ecology 3, 269-274

Shreve, F., & Wiggins, I.L., 1964. Vegetation and flora of the Sonoran Desert (Vol. 1). Stanford University Press.

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

Standley, P.C., 1923. Trees and Shrubs of Mexico. Contrib. from United States Natl. Herb. 23, 637. doi:10.1038/107603a0

Sun, M., Sun, X., Zhao, Y., Wang, O., Li, Z., Hu, Z., Mei, P., 1989. Cenezoic paleobiota of the continental shelf of East China Sea. Ed. Division of Comprehensive Studies on Ocean Geology of the Ministry of Geology and Minerals of P. R. China and Institute of Geology of the Chinese Academy of Geological Sciences. Geological Publishing House, Beijing

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 595, 585–595.

Tiffney, B.H., 2000. Geographic and climatic influences on the Cretaceous and Tertiary history of Euramerican floristic similarity. Acta Univ. Carolinae 44, 5–16.

Tiffney, B.H., 1985. Perspectives on the origin o f the floristic similarity between eastern asia and eastern north america. J. Arnold Arbor. 66, 73–94.

Tokuoka, T., 2007. Molecular phylogenetic analysis of Euphorbiaceae *sensu stricto* based on plastid and nuclear DNA sequences and ovule and seed character evolution. J. Plant Res. 511–522.

Trejo, I., Dirzo, R., 2002. Floristic diversity of Mexican seasonally dry tropical forests. Biodivers. Conserv. 2063–2084.

Van Devender, T.R., 1973. Late Pleistocene plants and animals of the Sonoran Desert: a survey of ancient packrat middens in southwestern Arizona. Ph.D. Dissertation, University of Arizona, Tucson, Arizona.

Van Devender, T.R., Burgess, T.L., Piper, J.C., Turner, R.M., 1994. Paleoclimatic implications of Holocene plant remains from the Sierra Bacha, Sonora, Mexico. Quat. Res. doi:10.1006/qres.1994.1011

van Ee, B.W., Berry, P.E., Riina, R., Gutiérrez Amaro, J.E., 2008. Molecular phylogenetics and biogeography of the Caribbean-Centered *Croton* subgenus

Moacroton (Euphorbiaceae s.s.), The Botanical Review. doi:10.1007/s12229-008-9003-y

Vásquez-Cruz, M., Sosa, V., 2016. New insights on the origin of the woody flora of the Chihuahuan Desert: The case of *Lindleya*. Am. J. Bot. 103, 1694–1707. doi:10.3732/ajb.1600080

Vargas, O.M., Ortiz, E.M., Simpson, B.B., 2017. Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). New Phytol. 214, 1736–1750.

Waltari, E., Hijmans, R.J., Peterson, T., Nyari, A.S., Perkins, S.L., Guralnick, R.P., 2007. Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. PLoS One 2, e563. doi:10.1371/journal.pone.0000563

Webb, C.O., Ackerly, D.D., Kembel, S.W., 2008. Phylocom: software for the analysis of phylogenetic community structure and trait evolution. Bioinformatics 24, 2098–2100. doi:10.1093/bioinformatics/btn358

Weeks, A., Daly, D.C., Simpson, B.B., 2005. The phylogenetic history and biogeography of the frankincense and myrrh family (Burseraceae) based on nuclear and chloroplast sequence data. Mol. Phylogenet. Evol. 35, 85–101. doi:10.1016/j.ympev.2004.12.021

Wessinger, C.A., Freeman, C.C., Mort, M.E., Rausher, M.D., Hileman, L.C., 2016. Multiplexed shotgun genotyping resolves species relationships within the North American genus *Penstemon*. Am. J. Bot. 103, 1–11.

Wurdack, K.J., Hoffman, P., Chase, M.W., 2005. Molecular phylogenetic analysis of uniovulate Euphorbiaceae (Euphorbiaceae sesu stricto) using plastic RBCL and TRNL-F DNA Sequences. Am. J. Bot. 92, 1397–1420.

Xi, Z., Ruhfel, B.R., Schaefer, H., Amorim, A.M., Sugumaran, M., Wurdack, K.J., Endress, P.K., Matthews, M.L., Stevens, P.F., Mathews, S., Davis, C.C., 2012. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. Proc. Natl. Acad. Sci. U. S. A. 109, 17519–24. doi:10.1073/pnas.1205818109