Florida International University
FIU Digital Commons

FIU Electronic Theses and Dissertations | University Graduate School

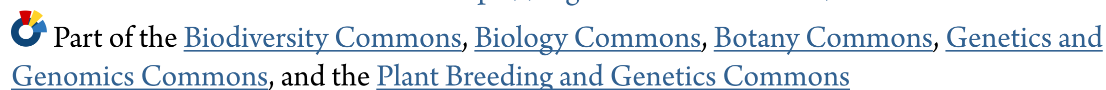
4-27-2018

The Evolution and Domestication Genetics of the Mango Genus, Mangifera (Anacardiaceae)

Emily Warschefsky
Florida International University, ewars001@fiu.edu

DOI: 10.25148/etd.FIDC006564

Follow this and additional works at: https://digitalcommons.fiu.edu/etd


Part of the Biodiversity Commons, Biology Commons, Botany Commons, Genetics and Genomics Commons, and the Plant Breeding and Genetics Commons

Recommended Citation
Warschefsky, Emily, "The Evolution and Domestication Genetics of the Mango Genus, Mangifera (Anacardiaceae)" (2018). *FIU Electronic Theses and Dissertations*. 3824.
https://digitalcommons.fiu.edu/etd/3824

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

EVOLUTION AND DOMESTICATION GENETICS OF THE MANGO GENUS, *MANGIFERA* (ANACARDIACEAE)

A dissertation submitted in partial fulfillment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOLOGY

by

Emily Warschefsky

2018

To: Dean Michael R. Heithaus
College of Arts, Sciences and Education

This dissertation, written by Emily Warschefsky, and entitled Evolution and Domestication Genetics of the Mango Genus, *Mangifera* (Anacardiaceae), having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Carl Lewis

Joel Trexler

Giri Narasimhan

Heather Bracken-Grissom

Maureen Donnelly, Co-Major Professor

Eric von Wettberg, Co-Major Professor

Date of Defense: April 27, 2018

The dissertation of Emily Warschefsky is approved.

Dean Michael R. Heithaus
College of Arts, Sciences and Education

Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2018

COPYRIGHT PAGE

The following chapters have been published in their entirety in peer-reviewed journals. According to the stated policy of each journal, they can be included here without obtaining formal copyright permission by using a proper citation.

CHAPTER II

Back to the Wilds: Tapping Evolutionary Adaptations for Resilient Crops Through Systematic Hybridization With Crop Wild Relatives. Warschefsky, E. R.V. Penmetsa, D.R. Cook, and E.J.B. von Wettberg. American Journal of Botany 101 (10): 1791–1800, 2014. doi:10.3732/ajb.1400116

CHAPTER III

Rootstocks: Diversity, Domestication, and Impacts on Shoot Phenotypes. Warschefsky, E.J., L.L. Klein , M.H. Frank, D.H. Chitwood, J.P. Londo, E.J.B. von Wettberg, and A.J. Miller. Trends in Plant Science 21(5): 418–437. doi 10.1016/j.tplants.2015.11.008.

ALL OTHER WORK

© Copyright 2018 by Emily Warschefsky

All rights reserved.

DEDICATION

To my parents, who instilled in me a love of questions and a yearning to answer them.

ACKNOWLEDGMENTS

First, thank you to my committee members (past and present) for sharing their expertise: Dr. Richard Campbell for his devotion to the mango and its wild relatives and for putting together a magnificent germplasm collection at Fairchild Tropical Botanic Garden (FTBG), Dr. Carl Lewis for the introduction to Indonesia and support in the lab at FTBG, Dr. Giri Narasimhan for his advice in bioinformatic endeavors, and Dr. Joel Trexler for his ecological and evolutionary insights. I want to extend a special thank you to Dr. Heather Bracken-Grissom for welcoming me into her laboratory space and community along with providing expertise and guidance on phylogenomics. Thank you to my co-advisor Dr. Maureen Donnelly for her spirited reassurance, support, and editing skills during the final stages of my dissertation. And finally, thank you to my co-advisor, Dr. Eric von Wettberg, for his boundless enthusiasm and for encouraging me to pursue my own research interests.

I am also grateful for the amazing community of labmates I have had the pleasure of working with over the years and who enriched my graduate school experience: Tonya Fotinos, Klara Scharnagl, Vanessa Sanchez, Damien Nesbeth, Seth Manthey, Eddie Marques, Jorge Pérez-Moreno, and Kelsey Reider. I also want to say a huge thank you to Laura Timm, my labmate and partner in navigating RADseq library preparation and data analysis. Thank you also to Danielle Coogan and Rachel Matthews who performed countless DNA extractions for my project.

I would like to acknowledge the following support that I have received from many grants, travel awards, and fellowships throughout my time as a PhD student: NSF Doctoral Dissertation Improvement Grant, USDA National Institute of Food and

Agriculture AFRI Fellowship, Fairchild Graduate Studies Dissertation Grant, Singapore Botanic Gardens Fellowship, Kelly Foundation Graduate Award for Tropical Botany, American Society of Plant Taxonomists Graduate Student Research Grant, Botanical Society of America Graduate Student Research Award, International Center for Tropical Botany Travel Award, Society for the Study of Evolution Graduate Student Travel Award, Botanical Society of America International Botanical Congress Travel Award, International Botanical Congress Travel Award, FIU Tropics Travel Award. I would also like to acknowledge funding for research assistantships and graduate fellowships provided by: Dr. Eric von Wettberg (Feed the Future Innovation Lab), USDA Project Grow, Fairchild Graduate Student Fellowship Program, FIU Dissertation Evidence Acquisition Fellowship, and FIU Dissertation Year Fellowship.

I am extremely grateful to the many institutions, colleagues, and collaborators that have helped me collect samples and hosted me during my travels – without them my research would not have been possible. Thank you to all of the staff who helped me at Fairchild Tropical Botanic Garden, Miami Dade Fruit and Spice Park, and the USDA's Chapman Field Station. Thank you to the many staff members at Bogor Botanic Garden, Purwodadi Botanic Garden, Singapore Botanic Garden, Pasoh Forest, and the Forestry Research Institute of Malaysia for their kindness and hospitality during my field trips. I would especially like to thank my Anacardiaceae collaborators, John D. Mitchell of the New York Botanic Garden, Susan Pell of the U.S. Botanic Garden, and Allison Miller of St. Louis University, who have so kindly provided their expertise throughout my graduate career.

Finally, thank you to my collaborators and coauthors on Chapters I and II for their contributions to these impactful reviews: Dan Chitwood, Doug Cook, Margaret Frank, Laura Klein, Jason Londo, Allison Miller, R. Varma Penmetsa, and Eric von Wettberg.

ABSTRACT OF THE DISSERTATION

EVOLUTION AND DOMESTICATION GENETICS OF THE MANGO GENUS,

MANGIFERA (ANACARDIACEAE)

by

Emily Warschefsky

Florida International University, 2018

Miami, Florida

Professor Eric von Wettberg, Co-Major Professor

Professor Maureen Donnelly, Co-Major Professor

Domesticated species are vital to global food security and have also been foundational to the formulation and advancement of evolutionary theory. My dissertation employs emerging molecular genomic tools to provide an evolutionary context for crop improvement. I begin by providing a contemporary perspective on two components of domestication biology that have long been used to improve crop production: wild relatives of crop species and grafted rootstocks. First, I propose a method to systematically introgress diversity from crop wild relatives into crop breeding programs. Then, I explore rootstocks, the lesser-known half of the perennial crop equation, documenting prevalence and diversity, cataloging rootstock traits under selection, and discussing recent advances in rootstock biology. Both crop wild relatives and rootstocks remain largely underutilized resources and hold great promise for agricultural innovation.

While humans have domesticated thousands of plant species, research has largely focused on annual crops, to the exclusion of perennials. To improve our understanding of

how tree species respond to domestication, I examine the evolution and domestication of one of the world's most important perennial tropical fruit crops, the mango, *Mangifera indica*, and its wild and semi-domesticated relatives. I generated a dataset suitable for studying *Mangifera* across evolutionary time using double digest restriction site associated DNA sequencing. I present a multilocus phylogeny that informs the classification of *Mangifera* and reveals, for the first time, the evolutionary relationships of wild, semi-domesticated, and domesticated species in the genus. Narrowing my focus to the intraspecific level, I examine how the introduction of *M. indica* into regions of the world impacted its genetic diversity. My results show *M. indica* maintained high levels of genetic diversity during its introduction into the Americas. However, the novel diversity I detect in Southeast Asian mango cultivars suggests that *M. indica* has a more complex domestication history than previously assumed. I also find evidence that *M. indica* hybridized with multiple congeners following its introduction into Southeast Asia, forming two hybrid lineages that may be maintained by clonal polyembryonic reproduction. Collectively, my research provides a comprehensive framework for understanding the evolution and domestication of a tropical tree crop of global economic importance.

TABLE OF CONTENTS

CHAPTER PAGE

PREFACE ... 1

I. INTRODUCTION ... 2
Literature Cited ... 11

II. BACK TO THE WILDS: TAPPING EVOLUTIONARY ADAPTATIONS FOR RESILIENT CROPS THROUGH SYSTEMATIC HYBRIDIZATION WITH CROP WILD RELATIVES ... 17
Abstract ... 18
Introduction ... 19
The Importance and Prevalence of Hybridization ... 22
Agriculture and Hybridization ... 27
Hybridization of Crops and their Wild Relatives to Confer Adaptive Traits ... 29
Obstacles to the Usage of Crop Wild Relatives ... 32
A Proposal for Future Work ... 35
Literature Cited ... 41
Figure Captions ... 52
Figures ... 53
Appendix Captions ... 54
Appendices ... 55
Appendices Literature Cited ... 63

III. ROOTSTOCKS: DIVERSITY, DOMESTICATION, AND IMPACTS ON SHOOT PHENOTYPES ... 76
Abstract ... 77
Getting to the Root of the Matter ... 77
Rootstock Diversity and Domestication ... 81
Recent Advances and Future Directions in Rootstock Biology ... 89
Concluding Remarks and Future Perspectives ... 97
Literature Cited ... 98
Tables ... 115
Figure Captions ... 122
Figures ... 123
Appendices Captions ... 124
Appendices ... 125

IV: A FRUITFUL PHYLOGENY: RAD-SEQ REVEALS THE EVOLUTIONARY RELATIONSHIPS BETWEEN WILD AND CULTIVATED SPECIES IN THE MANGO GENUS *(MANGIFERA* L., ANACARDIACEAE) ... 127
Abstract ... 128
Introduction ... 129

Materials and Methods....135
Results....139
Discussion....142
Literature Cited....150
Tables....161
Figure Captions....164
Figures....165
Appendices Captions....169
Appendices....170

V: POPULATION GENOMIC ANALYSIS OF CULTIVATED MANGO (*MANGIFERA INDICA* L.) SUGGESTS A COMPLEX HISTORY OF DOMESTICATION....212
Abstract....213
Introduction....214
Methods....220
Results....225
Discussion....229
Literature Cited....236
Tables....245
Figure Captions....249
Figures....250
Appendices Captions....256
Appendices....257

VI: INTERSPECIFIC HYBRIDIZATION IN *MANGIFERA*....265
Abstract....266
Introduction....267
Methods....270
Results....274
Discussion....278
Literature Cited....283
Tables....288
Figure Captions....299
Figures....300
Appendices Captions....308
Appendices....309

VII: CONCLUSIONS AND FUTURE DIRECTIONS....313
Literature Cited....326

VITA....330

LIST OF TABLES

TABLE PAGE

CHAPTER III

3.1 Grafted perennial crop species and examples of selected rootstock traits..........115

3.2 The twenty most-produced, grafted, woody perennial crop species, rootstock species used for their cultivation, and rootstock traits targeted during selection 118

CHAPTER IV

4.1 Sampling for phylogenetic analysis. For each proposed subgenus and section, the number of species included and total number of species included in the group is given in parentheses. The number of replicates included in each of the three subsets (phyloBIG, phyloMED, phyloSM) is provided..........161

4.2 Summary statistics for (A) phyloBIG, (B) phyloMED, and (C) phyloSM phylogenetic datasets (201, 98, and 41 individuals, respectively. The total number of loci recovered, percent of loci that are parsimony informative (PI), adjusted average variants per locus, adjusted average PI sites per locus, and percent of missing loci are given. Values are shaded in a red to green color scale, with red indicating less desirable characteristics (e.g., lower numbers of loci, higher percentages of missing data) and green indicating more desirable characteristics..........162

CHAPTER V

5.1 Samples included in datasets FULL, FULL_33, CULT, and CULT_106...........245

5.2 Measures of diversity for 106 mango cultivars from eight populations calculated from SNP loci (CULT_106). For each column, warmer colors reflect lower values. *Ho*: observed heterozygosity; *He*: heterozygosity within populations, aka ‘gene diversity’; *Fis*: inbreeding coefficient; Ar: allelic richness; *π:* nucleotide diversity; %Poly: percent polymorphic; Ap: private alleles.. 246

5.3 Pairwise differentiation between eight populations of 106 mango cultivars (CULT_106) calculated from 220 SNP loci. Values of F_{ST} are given above the diagonal, and p-values are given below the diagonal (light grey). 247

5.4 Hierarchical analysis of molecular variance among individuals (a) within populations nested in regions, (b) within populations nested in continents, (c) within regions nested in continents calculated from dataset CULT_106. 248

CHAPTER VI

6.1 Samples in the *M. odorata* dataset. Sample ID consists of the individual ddRAD Sample ID, the study collection number and putative identification (cultivar or species name). The collection location (FTBG = Fairchild Tropical Botanic Garden, FTG = Fairchild Tropical Garden Herbarium, SBG = Singapore Botanic Garden, KRP = Purwodadi Botanic Garden, KRB = Bogor Botanic Garden, USDA = USDA Subtropical Horticulture Research Station at Chapman Field, FSP = Miami Dade Fruit and Spice Park, GBTB = Gardens by the Bay, FRIM = Forestry Research Institute of Malaysia, PA = Pasoh Forest Arboretum, PF = Pasoh Forest Research Station) and accession number within the respective collection are provided. The ddRAD library, sublibrary, and individual sample ID are given, as well as the number of raw reads for each individual. 288

6.2 Samples in the *M. casturi* dataset. Sample ID consists of the individual ddRAD Sample ID, the study collection number and putative identification (cultivar or species name). The collection location (FTBG = Fairchild Tropical Botanic Garden, FTG = Fairchild Tropical Garden Herbarium, SBG = Singapore Botanic Garden, KRP = Purwodadi Botanic Garden, KRB = Bogor Botanic Garden, USDA = USDA Subtropical Horticulture Research Station at Chapman Field, FSP = Miami Dade Fruit and Spice Park, GBTB = Gardens by the Bay, FRIM = Forestry Research Institute of Malaysia, PA = Pasoh Forest Arboretum, PF = Pasoh Forest Research Station) and accession number within the respective collection are provided. The ddRAD library, sublibrary, and individual sample ID are given, as well as the number of raw reads for each individual. 293

6.3 Population genetic analysis of *M. odorata* and its parental taxa, *M. indica* and *M. foetida*. *Ho*: observed heterozygosity; *Hs*: heterozygosity within populations, aka 'gene diversity'; *Fis*: inbreeding coefficient; *π:* nucleotide diversity; rbarD: multilocus linkage disequilibrium; Ap%: percent private alleles. 297

6.4 Population genetic analysis of *M. casturi* and its putative parental taxa, *M. indica* and *M. quadrifida*. *Ho*: observed heterozygosity; *Hs*: heterozygosity within populations, aka 'gene diversity'; *Fis*: inbreeding coefficient; *π:* nucleotide diversity; rbarD: multilocus linkage disequilibrium; Ap%: percent private alleles. 298

LIST OF FIGURES

FIGURE PAGE

CHAPTER II

2.1 An outline of the 5-step approach we advocate for utilizing crop wild relatives in a systematic and thorough fashion in breeding programs.................................52

CHAPTER III

3.1 Primary targets of rootstock selection. Rootstocks used in perennial agriculture (A) have been selected from a pool of wild germplasm and bred for (B) their ability to graft to cultivated scions, (C) the root phenotype, and (D) their ability to impact the phenotype of the grafted scion. 122

CHAPTER IV

4.1 Taxonomy of *Mangifera* as proposed by Kostermans and Bompard (1993). Two subgenera and a total of six sections were delimited based on morphological characteristics. The number of species proposed to be in each group is shown in parentheses, with 11 described species of unknown placement. 164

4.2 Summary of (A) number of loci (B) Percent missing loci (C) Average Variants per locus, (D) Average Parsimony informative sites per locus, and (E) Percent of SNPs that are Parsimony informative across phyloBIG, phyloMED, and phyloSM datasets clustered at three different thresholds (85%, 90%, 95%) and for four levels of min_individuals (4, 20%, 50%, 90%). 164

4.3 Multidimensional scaling of weighted Robinson-Foulds Distance and Branch Score Distance between maximum likelihood topologies produced for A) phyloBIG B) phyloMED and C) phyloSM datasets. 164

4.4 Maximum likelihood phylogenetic hypothesis of *Mangifera* based on the phyloSM_m20_c85 dataset, with support from 1000 bootstrap replicates shown on branches. Highlighting indicates the five main clades recovered; Clade I (dark green) includes *M. indica, M. lalijiwa, M. laurina, M. casturi, M. zeylanica, M. pentandra, M. altissima, M. gedebe,* and multiple unidentified taxa; Clade II (light green) includes *M. odorata, M. foetida, M. pajang, M. macrocarpa* and one unidentified taxon; clade III (yellow) includes *M. quadrifida, M. gracilipes, M. magnifica, M. monandra, M.*

griffithii, M. rufocostata, and *M. subsessilifolia*; Clade IV (orange) contains *Bouea* species, and Clade V (red) includes two taxa previously included in *Mangifera*, *M. caesia* and *M. superba.* The names of species that are domesticated to some extent (including wild-collected, incipiently domesticated, cultivated, and highly domesticated) are shown in blue. 164

CHAPTER V

5.1 Map of the human-mediated migration of the mango. Colors represent the geographic populations of mango cultivars analyzed in this study and correspond to labels used throughout the results. The mango is thought to have originated and been domesticated in India, Nepal, Bangladesh, and Bhutan (red). It was first dispersed into Indochina (blue) and Malesia (green), then into East and West Africa (purple), South America (Brazil, orange), the Caribbean (pink), Mexico (yellow), and Florida (brown). 249

5.2 UPGMA dendrogram of mango cultivars from eight populations, closely related wild *Mangifera* species, and unidentified accessions (FULL_33). Individuals are color-coded according to their region of origin (map inset). Unidentified accessions and *Mangifera* species are labeled in grey...................249

5.3 Inferred population structure for the full dataset (FULL) as visualized with the software distruct with (a) two and (b) three ancestral populations. Each vertical bar represents a single individual that is assigned ancestry to one or more of the four ancestral populations as indicated by the colors. Groups are labeled by region (for *M. indica*) or as sp (species) or unk (unknown) (top) and population (for *M. indica)*, species name, or MSP (Malesian species) or FSP (Floridian species) (bottom). 249

5.4 Inferred population structure for mango cultivars (dataset CULT) as visualized with the software distruct with four ancestral populations. Each vertical bar represents a single individual that is assigned ancestry to one or more of the four populations (as indicated by the colors). Groups are labeled by region (top) and population (bottom). 249

5.5 Principal component analysis of mango cultivars (dataset CULT) from eight populations (map inset). Axes are labeled with the percent of variation explained by the corresponding principal component. Two individuals identified as outliers in UPGMA and STRUCTURE analyses were not plotted here...249

CHAPTER VI

6.1 Visualization of inferred population structure for the *M. odorata* dataset produced using the software distruct with K = 2 populations. Individuals

within the dataset are represented by a vertical bar and colors represent assignment to each population. Individuals are labeled by species name *(M. foetida, M. odorata, M. indica)* and, for *M. indica,* region of origin (India, Southeast Asia [SEA]). 299

6.2 Principal component analysis of 88 samples of *M. foetida, M. odorata,* and *M. indica* from the *M. odorata* dataset, with individual samples colored according to a) species only (red = *M. foetida,* orange = *M. odorata,* green = *M. indica*) or b) species and region (red = *M. foetida*, orange = *M. odorata,* green = Indian *M. indica,* yellow = Southeast Asian *M. indica*). 299

6.3 Plot of hybrid index (0 = *M. foetida,* 1 = *M. indica*) and interspecific heterozygosity for 17 samples of *M. odorata* calculated in the program INTROGRESS. 299

6.4 Neighbor network tree for *M. odorata* (orange), *M. foetida* (red), and *M. indica* (green) inferred using SPLITSTREE. 299

6.5 Visualization of inferred population structure for the *M. casturi* dataset produced using the software distruct with K = 2 populations. Individuals within the dataset are represented by a vertical bar and colors represent assignment to each population. Individuals are labeled by species name *(M. quadrifida, M. casturi, M. indica).* 299

6.6 Principal component analysis of 65 samples of *M. quadrifida, M. casturi,* and *M. indica* from the *M. casturi* dataset, with individual samples colored according to a) species only (brown = *M. quadrifida,* purple = *M. casturi,* green = *M. indica*). 299

6.7 Plot of hybrid index (0 = *M. quadrifida,* 1 = *M. indica*) and interspecific heterozygosity for 4 samples of *M. casturi* calculated in the program INTROGRESS. 299

6.8 Neighbor network tree for *M. casturi* (purple), *M. quadrifida* (brown), and *M. indica* (green) inferred using SPLITSTREE. 299

LIST OF APPENDICES

APPENDIX PAGE

CHAPTER II

2.1. Crops of putative/confirmed hybrid ancestry 54

CHAPTER III

3.1 Trends Box...124

3.2 Outstanding Questions Box 124

CHAPTER IV

4.1 Information for samples included in this study, including sample name, Species ID (which may differ from that of the sample name), and collection location (FTBG = Fairchild Tropical Botanic Garden, FTG = Fairchild Tropical Garden Herbarium, SBG = Singapore Botanic Garden, KRP = Purwodadi Botanic Garden, KRB = Bogor Botanic Garden, PF = Pasoh Forest, PA = Pasoh Arboretum, FS = Miami Dade Fruit and Spice Park, GBTB = Gardens by the Bay, FRIM = Forestry Research Institute of Malaysia, JFL/CL/SKP/DV = private individuals). Information on provenance is provided where available. Number of raw reads for each sample is given and inclusion in phyloBIG/MED/SM is given as presence/absence (1/0)...........169

4.2. Oligonucleotide sequences for eight sets of barcoded adapters used for ddRADseq. Adapters were ordered as single-stranded oligonucleotides, and the forward and reverse strands are given as P1.1 and P1.2, respectively. The unique barcode sequence is shown in lowercase within the full sequence and provided in a separate column for clarity. 169

4.3 Oligonucleotide sequences for twelve indexed PCR primers. One primer (P1) was universal and used for all samples, while the second primer (P2) included a unique index sequence. 169

4.4 Maximum likelihood phylogenies for 30 datasets included in this study (see figure for dataset name). Individual sample names are colored to correspond to clades outlined in Fig. 4. (dark green = clade I, light green = clade II, yellow = clade III, orange = clade IV, red = clade V, black = outgroup) with bootstrap support for 100 (4.01-4.09) or 1000 (4.10-4.30) replicates. 169

4.5 Quartet-based 50% majority rules phylogenies estimated with SVD Quartets for phyloSM_c85/90/95_m4 datasets. Individual sample names are colored to correspond to clades outlined in Fig. 4. (dark green = clade I, light green = clade II, yellow = clade III, orange = clade IV, red = clade V, black = outgroup) with bootstrap support from 100 replicates. 169

CHAPTER V

5.1 Information for the 158 samples included in this study. Sample ID consists of the individual ddRAD Sample ID, the study collection number and putative identification (cultivar or species name). The collection location (FTBG = Fairchild Tropical Botanic Garden, FTG = Fairchild Tropical Garden Herbarium, SBG = Singapore Botanic Garden, KRP = Purwodadi Botanic Garden, KRB = Bogor Botanic Garden, USDA = USDA Subtropical Horticulture Research Station at Chapman Field, FSP = Miami Dade Fruit and Spice Park, GBTB = Gardens by the Bay, FRIM = Forestry Research Institute of Malaysia, PA = Pasoh Forest Arboretum, PF = Pasoh Forest Research Station) and accession number within the respective collection are provided. The ddRAD library, sublibrary, and individual sample ID are given, as well as the number of raw reads for each individual. Each *M. indica* sample was classified within a Continent, Region, Population, and Subpopulation. Individuals of unknown provenance (Malesian species = MSP, Floridian species = FSP) or samples of *Mangifera* species were categorized accordingly. Nine individuals that were removed from the study as a result of low sequence coverage are not shown. 256

5.2 Plot of ΔK metric as described by Evanno et al. (2005) for the full dataset (FULL). The optimal number of ancestral populations is deemed to be that which has the highest value of ΔK. 256

5.3 Plot of ΔK metric as described by Evanno et al. (2005) for 108 mango cultivars (CULT). The optimal number of ancestral populations is deemed to be that which has the highest value of ΔK. 256

CHAPTER VI

6.1 Plot of ΔK metric as described by Evanno et al. (2005) for the *M. odorata* dataset. The optimal number of ancestral populations is deemed to be that which has the highest value of ΔK. 308

6.2 Plot of expected distribution of rbarD for unlinked loci using 999 permutations (grey bars) and actual rbarD for *M. odorata* samples. 308

6.3 Plot of ΔK metric as described by Evanno et al. (2005) for the *M. casturi* dataset. The optimal number of ancestral populations is deemed to be that which has the highest value of ΔK. 308

6.4 Plot of expected distribution of rbarD for unlinked loci using 999 permutations (grey bars) and actual rbarD for (A) *M. casturi* and (B) *M. quadrifida* samples. 308

PREFACE

The following chapters have been or will be submitted for publication and are formatted according to journal specifications:

CHAPTER II

Back to the Wilds: Tapping Evolutionary Adaptations for Resilient Crops Through Systematic Hybridization With Crop Wild Relatives. Warschefsky, E. R.V. Penmetsa, D.R. Cook, and E.J.B. von Wettberg. American Journal of Botany 101 (10): 1791–1800, 2014. doi:10.3732/ajb.1400116

CHAPTER III

Rootstocks: Diversity, Domestication, and Impacts on Shoot Phenotypes. Warschefsky, E.J., L.L. Klein , M.H. Frank, D.H. Chitwood, J.P. Londo, E.J.B. von Wettberg, and A.J. Miller. Trends in Plant Science 21(5): 418–437. doi 10.1016/j.tplants.2015.11.008.

CHAPTER IV

Systematic Biology

CHAPTER V

Molecular Ecology

CHAPTER VI

The American Journal of Botany

CHAPTER I

INTRODUCTION

Introduction

Over the past 12,000 years, humans have domesticated thousands of species from across the plant kingdom (Meyer et al. 2012; Meyer and Purugganan 2013; Gaut et al. 2015). The advent of domestication during the Neolithic Revolution laid the foundation for agricultural systems and human population growth (Smith and Zeder 2013; Gepts 2014a,b). Today, society as we know it is contingent on the reliability of food sources, the great majority of which come from domesticated plants and animals. Yet along with being integral components of modernity, domesticated species have served a principal role in our understanding of evolutionary biology. In his seminal work, *On the Origin of Species,* Darwin (1859) used domesticated species to illustrate the malleability of phenotypes under selection and successfully reasoned that human selection during domestication mirrors that which occurs in nature. Over the course of the subsequent 150 years, domesticated species have continued to provide tractable systems in which to study selection and other evolutionary processes including gene flow, genome evolution, adaptation, diversification, and convergent evolution (e.g., Arnold 2004; Kovach et al. 2007; Purugganan and Fuller 2009; Meyer and Purugganan 2013; Olsen and Wendel 2013; Washburn et al. 2016).

The Importance of Diversity

In the face of changing climatic conditions, agricultural production must keep pace with a rapidly expanding human population (The Hague Conference 2010; Beddington et al. 2012; Hatfield et al. 2014). To do so, modern breeding techniques must introduce novel traits into diverse crop systems that will enhance overall plant health and

production while simultaneously reducing the need for pesticides and fertilizers (Haussmann et al. 2004). The most sustainable way to achieve this goal is by taking advantage of pre–existing genetic variation found in genebanks and the wild relatives of crop species to produce disease resistant, locally adapted crop varieties (Haussmann et al. 2004; Warschefsky et al. 2014; Chapter II).

As early as the 1920s, scientists recognized that wild relatives of domesticated plants could serve as important genetic resources for crop breeding (Vavilov 1940; Tanksley and McCouch 1997; Meilleur and Hodgkin 2004). Compared to their domesticated counterparts, crop wild relatives often exhibit enhanced resistance to biotic stressors such as disease and pests, and environmental conditions like drought, salinity, and cold (Hajjar and Hodgkin 2007). Although advantageous traits from crop wild relatives have been used to improve major crops for more than 70 years (Meilleur and Hodgkin 2004), the genetic diversity of crop wild relatives remains underutilized (Ford-Lloyd et al. 2011).

Understanding the impacts of domestication on crop genetic diversity and characterizing the standing genetic variation within cultivated germplasm is critical to crop improvement efforts (e.g., Maxted and Guarino 2000; Iqbal et al. 2001; Burke et al. 2002; Mohammadi and Prasanna 2003; Esquinas-Alcázar 2005; Doebley et al. 2006; Ferreira 2006; Pickersgill 2007; Gross and Olsen 2010; Miller and Gross 2011; Kassa et al. 2012). The dogma of domestication dictates that crops undergo recurrent population bottlenecks as particular traits are selected for or against, causing an often-severe decrease in crop genetic diversity (Abbo et al. 2003; Doebley et al. 2006; Miller and Gross 2011; Bourguiba et al. 2012). Compounding this primary loss of diversity, many

crops are later dispersed into new regions through a series of founder events that further erode the genepool. Such drastic reductions in diversity diminish the ability of a crop to adapt to novel environments and fend off pests and diseases (e.g., Abbo et al. 2003; Esquinas-Alcázar 2005).

The Perennial Problem

While a great deal of research has been devoted to understanding the impacts of domestication on crop genetic diversity, much of the formative work focused on annual crops like cereals and grain legumes (e.g., Singh et al. 1991; Wang et al. 1999; Matsuoka et al. 2002; Li et al. 2006; Londo et al. 2006; Huang et al. 2012; Hufford et al. 2013; Saintenac et al. 2013). Recent research shows that many fundamental concepts of domestication are challenged by non–staple crop systems like perennials (Zohary 1992; Emswhiller 2006; Zeder 2006; Pickersgill 2007; Miller and Gross 2011; Meyer et al. 2012; McClure et al. 2014; Migicovsky and Myles 2017). For example, perennial crop species tend to be more robust to domestication-associated bottlenecks and maintain higher levels of genetic diversity throughout the domestication process than their annual counterparts (Miller and Gross 2011). In addition, hybridization occurs more frequently in perennial species than it does in annuals (Savolainen and Pyhäjärvi 2007; Miller and Gross 2011), suggesting that perennial crop diversity may be bolstered by genetic introgression from closely related wild species. Cases of wild-to-crop introgression have been observed in multiple perennial species including apple (Cornille et al. 2012, 2014) and date palm (Gros-Balthazard et al. 2017).

While perennial species may not experience a significant loss of genetic diversity because of population bottlenecks, they are nevertheless at risk of detrimentally low levels of diversity. Typically, tree crops are outcrossing and highly heterozygous, producing variable offspring that cannot be inbred to create true–breeding lines (Savolainen and Pyhäjärvi 2007; Miller and Gross 2011). Clonal propagation techniques, such as grafting, are therefore invaluable to perennial crop cultivation, allowing individual genotypes to be maintained and propagated *en masse* (Mudge et al. 2009; Warschefsky et al. 2016; Chapter III). Accordingly, in commercial settings perennial crops are often grown in monoculture; while there may be heterozygosity and genetic diversity within a given clone, population level variation is virtually nonexistent (Miller and Gross 2011). As landraces (locally adapted cultivars that have been improved through traditional agricultural practices) and so-called 'heirloom' varieties of perennial crops are replaced by elite cultivars that have been popularized for demanding commercial markets, perennial crops risk losing valuable genetic diversity (Miller and Gross 2011; McClure et al. 2014; Migicovsky and Myles 2017).

***The Potential of* Mangifera**

Among the many domesticated species of the poison ivy family (Anacardiaceae), including pistachio *(Pistacia vera)*, cashew *(Anacardium occidentale)*, pepper tree *(Schinus* spp.*)*, and jocote *(Spondias purpurea)*, the mango *(Mangifera indica)* provides a novel system in which to study perennial domestication. Wild *M. indica* is thought to have originated in the foothills of the Himalayas in what is now northeastern India, Myanmar, Bangladesh, and Nepal (Mukherjee 1972; Kostermans and Bompard 1993).

However, it is unclear whether any wild populations remain in the region (IUCN 2012), and it remains possible that these populations are not truly wild, but instead are feral escapees from cultivation. Following its initial domestication in India more than 4,000 years ago, humans dispersed *M. indica* in two directions: west, through Africa and on to the Americas; and east, into Southeast Asia (the center of diversity of *Mangifera*) (Popenoe 1920; Mukherjee 1949; Kostermans and Bompard 1993). In addition to *M. indica*, some 26 species of *Mangifera* have edible fruits (Kostermans and Bompard 1993); these species range from incipiently domesticated to cultivated landraces (Clement 1999). Given the spectrum of domestication present in *Mangifera* and the mango's unique pattern of dispersal, the genus is an ideal system to examine patterns of domestication in perennial species and to test whether *M. indica* follows trends seen in other perennials: maintaining high levels of genetic diversity despite population bottlenecks and a propensity for interspecific hybridization.

Outline of Dissertation

I begin my dissertation with two literature reviews that focus on important issues in contemporary domestication biology. In Chapter II, I advocate for systematic efforts to introgress diversity from crop wild relatives into crop plants in order to incorporate useful adaptations and to increase the resilience and productivity of agriculture in the 21st century. Many of our most important food crops suffer from a lack of genetic diversity that hinders their ability to adapt to new environmental conditions and combat pests and diseases (e.g., Abbo et al. 2003; Esquinas-Alcázar 2005). While the potential for genetic gains from the use of crop wild relatives is well documented (e.g., Pimentel et al., 1997;

Tanksley and McCouch 1997; Maxted and Kell 2009), most breeding efforts have been trait-specific and have dealt with wild species in a limited and *ad hoc* manner. I begin by using primary literature to illustrate the capacity of homoploid hybridization and genetic introgression to generate genetic and phenotypic novelty (Rieseberg et al. 1999, 2007; Seehausen 2013). I then provide an overview of previous efforts to introduce genetic diversity from crop wild relatives into domesticated species (e.g., Acosta-Gallegos et al. 2007), and discuss the challenges associated with using crop wild relatives in breeding programs. Finally, I propose a multistep framework for using naturally occurring genetic variation in crop wild relatives to improve crop performance.

Grafting is an ancient agricultural practice that is widely used in perennial crops to join resilient root systems (rootstocks) to shoots (scions) that produce the harvested product (e.g., fleshy or dry fruits) (Mudge et al. 2009). While a growing body of literature is beginning to demonstrate the differences in the domestication process between annual and perennial crops, the work has almost exclusively focused on scions of woody perennial species. In Chapter III, I investigate rootstocks, the lesser-known half of the perennial crop equation. First, I provide an overview of natural grafting, which occurs in a number of species and may have inspired the advent of grafting in horticulture (Mudge et al. 2009). I then document the widespread use of grafting in perennial agriculture and survey the diversity of species used as rootstocks for the 20 most produced perennial crop species. Examining rootstock traits under selection, which include traits inherent to the root system as well as traits that modulate scion phenotypes. I close the chapter by exploring developing areas of rootstock research, including rootstock–scion molecular interactions and root microbiomes.

My dissertation continues with three data chapters that focus on domestication in the genus *Mangifera*. In Chapter IV, I provide a phylogenetic perspective of the diversity of *Mangifera,* an economically important genus of tropical fruit trees, and highlight its potential as a novel system in which to study domestication. Analyzing restriction site associated DNA sequencing (RADseq) data from 208 individuals representing approximately 36 taxa, I provide the first phylogenetic hypothesis for *Mangifera* and examine infrageneric classification and the relationships between cultivated and wild species. I also validate the use of RADseq as a tool for genus-level phylogenomic analysis in non–model systems and explore the impacts of bioinformatic parameters on tree topology and branch support.

In Chapter V, I survey the genetic diversity maintained in the mango genebank at Fairchild Tropical Botanic Garden (FTBG) and use these data to investigate how domestication and human-mediated migration has impacted the genetic diversity of *M. indica*. The U.S. is the leading importer of mango products (FAOSTAT 2013), and while U.S. production accounts for less than 1% of the worldwide total, many of the world's most important commercial cultivars were developed in South Florida (Knight and Schnell 1994; Evans 2008). Today, the mango is the most-produced tropical fruit in the world (FAO 2003) and innumerable cultivars of the so-called 'King of Fruits' are grown around the globe. The mango genebank at FTBG contains over 300 accessions from around the world including 26 accessions of wild collected *Mangifera* species. I analyze RADseq data from 108 mango cultivars representing eight geographic regions along with 50 accessions that were either unidentified or from closely related *Mangifera* species. Comparing different measures of genetic diversity in the eight regions, I assess the

severity of the genetic bottleneck associated with the mango's introduction into Africa and the Americas. I also analyze the geographic distribution of genetic diversity in *M. indica* to identify regions of the world with high and/or novel variation.

In Chapter VI, I examine the occurrence and consequences of interspecific hybridization between native Southeast Asian *Mangifera* species and *M. indica,* which was introduced to the region. Homoploid hybridization can have multiple different outcomes, including genomic introgression (Anderson 1949; Arnold 1992; Arnold et al. 2012) and, rarely, the formation of new evolutionary lineages that can be deemed hybrid species (Schumer et al. 2014). I use RADseq data to examine the genetic diversity of 17 individuals of a known hybrid, *M. odorata,* looking for evidence of introgression between its parental taxa, *M. indica* and *M. foetida.* I also test whether genetic data supports a hybrid origin of *M. casturi,* a species only known from cultivation (Kostermans and Bompard 1993) and classified as extinct in the wild (IUCN 2012). Following Chapter VI, I conclude my dissertation with a chapter outlining the major findings, contributions, and future directions of my research.

Intellectual Merit

The research I present here advances our efforts to make systematic use of crop wild relatives in breeding programs and provides insight into how woody perennial crops respond to domestication. The methodology applies advanced sequencing techniques to phylogeny, population genetics, and hybridization genetics in a non–model system of domestication. My work also provides unprecedented insights into the genetic composition and distribution of diversity in one of the world's most important tropical

fruit species, the mango (*M. indica*), which is a valuable food source and export in many developing countries. Moreover, this dissertation is designed to inform management, pre–breeding programs, and rootstock selection of the particular genebank resources from which I sampled. As a whole, my dissertation contributes to global efforts to produce diverse crop species that are able to cope with changing and erratic climatic conditions in order to meet the nutritive demands of a growing human population.

Literature Cited

Abbo, S., J. Berger, and N. C. Turner. 2003. Evolution of cultivated chickpea: four bottlenecks limit diversity and constrain adaptation. Funct. Plant Biol. 30:1081–1087.

Acosta-Gallegos, J. A., J. D. Kelly, and P. Gepts. 2007. Prebreeding in common bean and use of genetic diversity from wild germplasm. Crop Sci. 47:S-44-S-59. Crop Science Society of America, Madison, WI.

Anderson, E. 1949. Introgressive hybridization. John Wiley & Sons, Inc., New York, NY.

Arnold, M. 1992. Natural hybridization as an evolutionary process. Annu. Rev. Ecol. Syst. 23:237–261.

Arnold, M. L. 2004. Natural hybridization and the evolution of domesticated, pest and disease organisms. Mol. Ecol. 13:997–1007.

Arnold, M. L., E. S. Ballerini, and A. N. Brothers. 2012. Hybrid fitness, adaptation and evolutionary diversification: lessons learned from Louisiana Irises. Heredity (Edinb). 108:159–66.

Beddington, J., M. Asaduzzaman, M. Clark, A. Fernández, M. Guillou, M. Jahn, L. Erda, T. Mamo, N. Van Bo, C. A. Nobre, R. Scholes, R. Sharma, and J. Wakhungu. 2012. Achieving food security in the face of climate change: Final report from the Commission on Sustainable Agriculture and Climate Change. CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS). Copenhagen, Denmark. Available online at: www.ccafs.cgiar.org/commission.

Bourguiba, H., J.-M. Audergon, L. Krichen, N. Trifi-Farah, A. Mamouni, S. Trabelsi, C. D'Onofrio, B. M. Asma, S. Santoni, and B. Khadari. 2012. Loss of genetic diversity

as a signature of apricot domestication and diffusion into the Mediterranean Basin. BMC Plant Biol. 12:49. BMC Plant Biology.

Burke, J. M., S. Tang, S. J. Knapp, and L. H. Rieseberg. 2002. Genetic analysis of sunflower domestication. Genetics 161:1257–1267.

Clement, C. R. 1999. 1492 and the loss of Amazonian crop genetic resources. I. The relation between domestication and human population decline. Econ. Bot. 53:188–202.

Cornille, A., T. Giraud, M. J. M. Smulders, I. Roldán-Ruiz, and P. Gladieux. 2014. The domestication and evolutionary ecology of apples. Trends Genet. 30:57–65.

Cornille, A., P. Gladieux, M. J. M. Smulders, I. Roldán-Ruiz, F. Laurens, B. Le Cam, A. Nersesyan, J. Clavel, M. Olonova, L. Feugey, I. Gabrielyan, X.-G. Zhang, M. I. Tenaillon, and T. Giraud. 2012. New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. PLoS Genet. 8:e1002703.

Darwin, C. 1859. The origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. 1st ed. Random House, New York.

Doebley, J. F., B. S. Gaut, and B. D. Smith. 2006. The molecular genetics of crop domestication. Cell 127:1309–1321.

Emswhiller, E. 2006. Genetic Data and Plant Domestication. Pp. 99–122 *in* M. A. Zeder, E. Emshwiller, D. G. Bradley, and B. D. Smith, eds. Documenting domestication: New genetic and archaeological paradigms. University of California Press, Berkeley, CA.

Esquinas-Alcázar, J. 2005. Science and society: Protecting crop genetic diversity for food security: Political, ethical and technical challenges. Nat. Rev. Genet. 6:946–53.

Evans, E. A. 2008. Recent trends in world and U.S. mango production, trade, and consumption.

FAO. 2003. Tropical Fruits. *in* Medium-term prospects for agricultural commodities. Food and Agriculture Organization of the United Nations, Rome, Italy. p. 1–6.

FAOSTAT. 2013. Food and Agriculture Organization of the United Nations Statistics Division. Available from: http://www.fao.org/faostat/.

Ferreira, M. E. 2006. Molecular analysis of gene banks for sustainable conservation and increased use of crop genetic resources. Pp. 1–5 *in* J. Ruane and A. Sonnino, eds. The role of biotechnology in exploring and protecting agricultural genetic resources. Food and Agriculture Organization of the United Nations. Rome, Italy.

Ford-Lloyd, B. V., M. Schmidt, S. J. Armstrong, O. Barazani, J. Engels, R. Hadas, K. Hammer, S. P. Kell, D. Kang, K. Khoshbakht, Y. Li, C. Long, B.-R. Lu, K. Ma, V. T. Nguyen, L. Qiu, S. Ge, W. Wei, Z. Zhang, and N. Maxted. 2011. Crop wild relatives—Undervalued, underutilized and under threat? Bioscience 61:559–565.

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

Gepts, P. 2014a. Domestication of plants. Pp. 474–486 *in* N. Van Alfen, ed. Encyclopedia of agriculture and food systems. Elsevier, San Diego, CA.

Gepts, P. 2014b. The contribution of genetic and genomic approaches to plant domestication studies. Curr. Opin. Plant Biol. 18:51–9. Elsevier Ltd.

Gros-Balthazard, M., M. Galimberti, A. Kousathanas, C. Newton, S. Ivorra, L. Paradis, Y. Vigouroux, R. Carter, M. Tengberg, V. Battesti, S. Santoni, L. Falquet, J. C. Pintaud, J. F. Terral, and D. Wegmann. 2017. The discovery of wild date palms in Oman reveals a complex domestication history involving centers in the Middle East and Africa. Curr. Biol. 27:2211–2218.e8.

Gross, B. L., and K. M. Olsen. 2010. Genetic perspectives on crop domestication. Trends Plant Sci. 15:529–37.

Hajjar, R., and T. Hodgkin. 2007. The use of wild relatives in crop improvement: a survey of developments over the last 20 years. Euphytica 156:1–13.

Hatfield, J., G. Takle, R. Grotjahn, P. Holden, R. C. Izaurralde, T. Mader, E. Marshall, and D. Liverman. 2014. Ch. 6: Agriculture. *in* J. M. Melillo, T.C. Richmond, and G.W. Yohe, eds. Climate change impacts in the United States: The third national climate assessment.150–174.

Haussmann, B. I. G., H. K. Parzies, T. Presterl, Z. Susic, and T. Miedaner. 2004. Plant genetic resources in crop improvement. Plant Genet. Resour. Charact. Util. 2:3–21.

Huang, P., J. Molina, J. M. Flowers, S. Rubinstein, S. A. Jackson, M. D. Purugganan, and B. A. Schaal. 2012. Phylogeography of Asian wild rice, *Oryza rufipogon*: a genome-wide view. Mol. Ecol. 21:4593–604.

Hufford, M. B., P. Lubinksy, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand, and J. Ross-Ibarra. 2013. The genomic signature of crop-wild introgression in maize. PLoS Genet. 9:e1003477.

Iqbal, M. J., O. U. K. Reddy, K. M. El-Zik, and A. E. Pepper. 2001. A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. TAG Theor. Appl. Genet. 103:547–554.

IUCN. 2012. World Conservation Monitoring Centre 1998. *IUCN Red List of Threatened Species. Version 2012.2.* Available at: www.iucnredlist.org.

Kassa, M. T., R. V. Penmetsa, N. Carrasquilla-Garcia, B. K. Sarma, S. Datta, H. D. Upadhyaya, R. K. Varshney, E. J. B. von Wettberg, and D. R. Cook. 2012. Genetic patterns of domestication in pigeonpea (*Cajanus cajan* (L.) Millsp.) and wild *Cajanus* relatives. PLoS One 7:e39563.

Knight, R. J., and R. J. Schnell. 1994. Mango introduction in Florida and the "Haden" cultivar's significance to the modern industry. Econ. Bot. 48:139–145.

Kostermans, A. J. G. H., and J. M. Bompard. 1993. The mangoes: Their botany, nomenclature, horticulture, and utilization. Academic Press, San Diego, CA.

Kovach, M. J., M. T. Sweeney, and S. R. McCouch. 2007. New insights into the history of rice domestication. Trends Genet. 23:578–587.

Li, C., A. Zhou, and T. Sang. 2006. Rice domestication by reducing shattering. Science 311:1936–9.

Londo, J. P., Y. Chiang, K. Hung, T. Chiang, and B. A. Schaal. 2006. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. Proc. Natl. Acad. Sci. 103:9578–9583.

Matsuoka, Y., Y. Vigouroux, M. M. Goodman, J. Sanchez G, E. Buckler, and J. Doebley. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. Proc. Natl. Acad. Sci. U.S.A. 99:6080–4.

Maxted, N., and L. Guarino. 2000. Chapter 3: Planning Plant Genetic Conservation. Pp. 37–78 *in* R. D. Smith, J. B. Dickie, S. H. Linington, H. W. Pritchard, and R. J. Probert, eds. Seed conservation: Turning science into practise. Kew Publishing, London.

Maxted, N., and S. Kell. 2009. Establishment of a global network for the in situ conservation of crop wild relatives: Status and needs. 266 pp. FAO Commission on Genetic Resources for Food and Agriculture, Rome, Italy.

McClure, K. A., J. Sawler, K. M. Gardner, D. Money, and S. Myles. 2014. Genomics: A potential panacea for the perennial problem. Am. J. Bot. 101:1–11.

Meilleur, B. A., and T. Hodgkin. 2004. In situ conservation of crop wild relatives: Status and trends. Biodivers. Conserv. 13:663–684.

Meyer, R. S., A. E. DuVal, and H. R. Jensen. 2012. Patterns and processes in crop domestication: An historical review and quantitative analysis of 203 global food crops. New Phytol. 196:29–48.

Meyer, R. S., and M. D. Purugganan. 2013. Evolution of crop species: genetics of domestication and diversification. Nat. Rev. Genet. 14:840–52.

Migicovsky, Z., and S. Myles. 2017. Exploiting wild relatives for genomics-assisted breeding of perennial crops. Front. Plant Sci. 8:1–16.

Miller, A. J. and B. L. Gross. 2011. From forest to field: Perennial fruit crop domestication. Am. J. Bot. 98:1389–414.

Mohammadi, S. A. A., and B. M. Prasanna. 2003. Analysis of genetic diversity in crop plants—Salient statistical tools. Crop Sci. 43:1235–1248.

Mudge, K., J. Janick, S. Scofield, and E. E. Goldschmidt. 2009. A history of grafting. Pp. 437–494 *in* J. Janick, ed. Horticultural Reviews vol. 35. John Wiley & Sons, Inc.

Mukherjee, S. K. 1972. Origin of mango *(Mangifera indica)*. Econ. Bot. 26:260–264.

Mukherjee, S. K. 1949. The mango and its wild relatives. Sci. Cult. 26:5–9.

Olsen, K. M., and J. F. Wendel. 2013. Crop plants as models for understanding plant adaptation and diversification. Front. Plant Sci. 4:290.

Pickersgill, B. 2007. Domestication of plants in the Americas: insights from Mendelian and molecular genetics. Ann. Bot. 100:925–40.

Pimentel, D., C. Wilson, C. McCullum, R. Huang, P. Dwen, J. Flack, Q. Tran, T. Saltman, and B. Cliff. 1997. Economic and environmental benefits of biodiversity. Bioscience 47:747–757.

Popenoe, W. 1920. The mango. Pp. 79–145 *in* Manual of tropical and subtropical fruits. The Macmillian Company, New York.

Purugganan, M. D., and D. Q. Fuller. 2009. The nature of selection during plant domestication. Nature 457:843–8.

Rieseberg, L. H., M. A. Archer, and R. K. Wayne. 1999. Transgressive segregation, adaptation and speciation. Heredity 83:363–372.

Rieseberg, L. H., S.-C. Kim, R. A. Randell, K. D. Whitney, B. L. Gross, C. Lexer, and K. Clay. 2007. Hybridization and the colonization of novel habitats by annual sunflowers. Genetica 129:149–65.

Saintenac, C., D. Jiang, S. Wang, and E. Akhunov. 2013. Sequence-based mapping of the polyploid wheat genome. G3 Genes, Genomes, Genet. 3:1105–14.

Savolainen, O., and T. Pyhäjärvi. 2007. Genomic diversity in forest trees. Curr. Opin. Plant Biol. 10:162–7.

Schumer, M., G. G. Rosenthal, and P. Andolfatto. 2014. How common is homoploid hybrid speciation? Evolution 1553–1560.

Seehausen, O. 2013. Conditions when hybridization might predispose populations for adaptive radiation. J. Evol. Biol. 26:279–281.

Singh, S. P., P. Gepts, and D. G. Debouck. 1991. Races of common bean. Econ. Bot. 45:379–396.

Smith, B. D., and M. A. Zeder. 2013. The onset of the Anthropocene. Anthropocene 4:8–13. Elsevier B.V.

Tanksley, S. D., and S. R. McCouch. 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. Science 277:1063–1066.

The Hague Conference. 2010. The Hague Conference on Agriculture, Food Security and Climate Change. Rome, Italy.

Vavilov. 1940. The new systematics of cultivated plants. Pp. 549–566 *in* J. Huxley, ed. The New Systematics. Clarendon, Oxford.

Wang, R., A. Stec, J. Hey, L. Lukens, and J. Doebley. 1999. The limits of selection during maize domestication. Nature 398:236–239.

Warschefsky, E. J., L. L. Klein, M. H. Frank, D. H. Chitwood, J. P. Londo, E. J. B. von Wettberg, and A. J. Miller. 2016. Rootstocks: Diversity, domestication, and impacts on shoot phenotypes. Trends Plant Sci. 21:418–437. Elsevier Ltd.

Warschefsky, E., R. V. Penmetsa, D. R. Cook, and E. J. B. von Wettberg. 2014. Back to the wilds: Tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. Am. J. Bot. 101:1791–800.

Washburn, J. D., K. A. Bird, G. C. Conant, and J. C. Pires. 2016. Convergent evolution and the origin of complex phenotypes in the age of systems biology. Int. J. Plant Sci. 177:305–318.

Zeder, M. a. 2006. Central questions in the domestication of plants and animals. Evol. Anthropol. Issues, News, Rev. 15:105–117.

Zohary, D. 1992. Is the European plum, *Prunus domestica* L., a *P. cerasifera* L. allopolyploid? Euphytica 60:75–77.

CHAPTER II

BACK TO THE WILDS: TAPPING EVOLUTIONARY ADAPTATIONS FOR RESILIENT CROPS THROUGH SYSTEMATIC HYBRIDIZATION WITH CROP WILD RELATIVES

Emily Warschefsky, R. Varma Penmetsa, Douglas R. Cook, and Eric J. B. von Wettberg

This chapter was published in:

American Journal of Botany 101 (10): 1791–1800, 2014. doi:10.3732/ajb.1400116

Abstract

The genetic diversity of our crop plants has been substantially reduced during the process of domestication and breeding. This reduction in diversity necessarily constrains our ability to expand a crop's range of cultivation into environments that are more extreme than those in which it was domesticated, including into "sustainable" agricultural systems with reduced inputs of pesticides, water, and fertilizers. Conversely, the wild progenitors of crop plants typically possess high levels of genetic diversity, which underlie an expanded (relative to domesticates) range of adaptive traits that may be of agricultural relevance, including resistance to pests and pathogens, tolerance to abiotic extremes, and reduced dependence on inputs. Despite their clear potential for crop improvement, wild relatives have rarely been used systematically for crop improvement, and in no cases, have full sets of wild diversity been introgressed into a crop. Instead, most breeding efforts have focused on specific traits and dealt with wild species in a limited and typically ad hoc manner. Although expedient, this approach misses the opportunity to test a large suite of traits and deploy the full potential of crop wild relatives in breeding for the looming challenges of the 21st century. Here we review examples of hybridization in several species, both intentionally produced and naturally occurring, to illustrate the gains that are possible. We start with naturally occurring hybrids, and then examine a range of examples of hybridization in agricultural settings.

Introduction

All domesticated species, both plants and animals, are impacted in unintended, often negative ways during domestication and breeding (e.g., Ladizinsky, 1985; Spillane and Gepts, 2001; Hyten et al., 2006; Taberlet et al., 2008; Gross and Olsen, 2010; Meyer et al., 2012; Olsen and Wendel, 2013). In particular, many crops lack genetic diversity and possess properties that reduce fitness in the natural environment. This problem derives both from demographic processes (e.g., genetic drift, population bottlenecks) and from changes in the nature of selection during breeding and cultivation that elevate the frequency of alleles with unique value in the agricultural environment and that permit the persistence of deleterious alleles (e.g., through selection tradeoffs and selection relaxation) (Olsen and Wendel, 2013). The combination of the loss of adaptive alleles through drift and fixation of deleterious alleles through altered selection necessarily constrains our ability to expand the cultivation of domesticated species into environments beyond those in which domestication occurred, e.g., into more extreme climates, into marginal soils, into degraded agricultural landscapes, or into "sustainable" systems with reduced agricultural inputs. As part of this special issue, "Speaking of Food," we argue that there is a need for systematic efforts to introgress broad subsets of wild relative diversity into our crop plants to incorporate the range of useful adaptations for disease resistance, abiotic stress tolerance, and other agronomic challenges that are required in order to increase the resiliency and productivity of agriculture in the 21st century. Here we review the ecological and evolutionary literature on the effects of hybridization to show the capacity of hybridization to generate phenotypic novelty, then detail examples of hybridization of crop wild relatives with domesticated plants.

Wild species have an important role to play in meeting the challenges for 21st century agriculture, which must become increasingly efficient to meet humankind's demand for a more plentiful and nutritious food supply (e.g., Tanksley and McCouch, 1997; Pimentel et al., 1997; Haussmann et al., 2004; Maxted and Kell, 2009; Tester and Langridge, 2010; Ford-Lloyd et al., 2011; McCouch et al., 2013). Such challenges are particularly acute in the developing world, where extreme climatic conditions, marginal soils, and reduced inputs limit productivity, create increased risk, and diminish livelihoods through reduced income and malnutrition. Yet the impact of a properly implemented and well-used resource of wild germplasm would extend beyond the developing world. Many of the crop phenotypes important to cultivation in the developing world (e.g., tolerance to heat and drought, reduced dependence on inputs [e.g., nitrogen, phosphate, pesticides, water], and increased seed nutrient density) are also key to meeting the global demand for crops that incorporate traits for climate-resilience, increased sustainability, and increased nutritional value.

The potential for genetic gains from use of crop wild relatives is well documented (e.g., Pimentel et al., 1997; Tanksley and McCouch, 1997; Maxted and Kell, 2009). Nevertheless, crop wild relatives have been used sparingly and typically in an ad hoc manner in many crop breeding programs (Hajjar and Hodgkin, 2007; Maxted and Kell, 2009; Brumlop et al., 2013). Impediments to the systematic use of wild material in crop improvement programs include the often poor agronomic performance of crop–wild hybrids and their immediate backcrosses, and the labor intensive process of constructing large-scale, representative populations that are suitable for phenotypic assessment. For perennial crop species, which can take many years to reach reproductive age, such

repeated backcrossing is prohibitively time consuming. Moreover, for many crops, the use of wild germplasm is further constrained by the limited state of international germplasm collections. Compounding the problem, many crop wild relatives are at risk of extinction from habitat loss, habitat fragmentation, changing land use and management practices, climate change, and introgression from agricultural relatives (Ford-Lloyd et al., 2011).

The fact that crop wild relatives are under-used in crop improvement programs presents an opportunity. One can restructure germplasm resources, essentially de novo, guided by appropriate ecological and population genetic theory; when properly implemented, such collections would represent a diversity of source habitats and encompass the breadth of segregating genetic variation and adaptations characteristic of the target species. Genomics, phenotyping, and computational approaches can subsequently be used to infer natural adaptations in situ, for example, based on knowledge of population structure, allele frequency, and recombination history, combined with knowledge about selective constraints in individual populations. Such analyses can motivate targeted phenotyping activities and ultimately nominate candidate genes for adaptive traits, leading to increased understanding of the autecology of crop wild relatives. In parallel to the analysis of gene function in situ, purpose-built populations that are hybrids between crops and their wild relatives provide powerful tools for trait dissection, and as such they become the vehicle by which the genetic (genomic) basis of valuable agronomic traits can be understood. Examples of such populations include nested association mapping (NAM) (e.g., Yu et al., 2008; McMullen et al., 2009) and multiparent advanced generation intercross (MAGIC) (e.g., Cavanagh et al., 2008)

panels, to which the logic of association genetics can be applied (e.g., Huang and Han, 2013; Korte and Farlow, 2013), as well as advanced backcross introgression lines (Tanksley and McCouch, 1997) that capture genome intervals and their adaptive traits from wild relatives within the essential crop genome. The value of combining ecology, population genetics, genomics, and phenotyping is well documented in model species, such as *Arabidopsis*, *Drosophila*, mice and maize (e.g., Yu et al., 2008; Ayroles et al., 2009; McMullen et al., 2009; Atwell et al., 2010, Tian et al., 2011; Flint and Eskin, 2012; MacKay et al., 2012; Korte and Farlow, 2013), but has not been used widely in support of crop species and their wild relatives (e.g., Huang and Han, 2013). To make the most of this approach, however, we must understand more about the complex effects of hybridization. To that end, we review examples of hybridization in several species, both intentionally produced and naturally occurring, to illustrate the gains that are possible.

The Importance and Prevalence of Hybridization

One of the most reviewed and most debated sources of variation in sexually reproducing organisms is hybridization, or reproduction among members of genetically distinct groups (i.e., populations within species, or distinct but closely related species) (Ellstrand and Schierenbeck, 2000; Mallet, 2005; Soltis and Soltis, 2009; Abbott et al., 2013; Schumer et al., 2014). Hybridization of crops and their wild relatives has long been an important source of variation in breeding, despite its ad hoc application. We argue there is a need for systematic efforts to introgress a broad range of wild relative diversity into our crop plants, with the goal of creating a genetic toolbox from which natural adaptations for traits such as disease resistance, tolerance to climatic extremes (especially

temperature and moisture), and productivity in otherwise marginal soils can be identified and deployed. First, we summarize the extensive literature that illustrates the potential of hybridization and introgression to generate phenotypic novelty, including both plant and animal examples. We then document that most examples of intentional introgression from wild relatives into cultivated species have focused on a narrow range of traits and a limited range of the variation present in crop wild relatives. Finally, we argue that a growing understanding of both the genetic architecture of domestication and the genomic consequences of hybridization makes it feasible to systematically introgress substantial amounts of the diversity present in wild relatives into cultivated genetic backgrounds. From systematic introgressions, it is feasible to quickly recover both wild relative stress tolerance and cultivated agronomic traits of interest through advance generation backcrosses (Tanksley and McCouch, 1997) and nested association mapping or multiparent advance generation intercross populations (e.g., Cavanagh et al., 2008; Yu et al., 2008; McMullen et al., 2009).

Hybridization occurs between individuals with varying levels of genetic differentiation and via multiple mechanisms, and it is therefore not surprising that such interbreeding events can have drastically different consequences (e.g., Barton and Hewitt, 1985; Mallet, 2005; Soltis and Soltis, 2009; Abbott et al., 2013). From reinforcement of isolating mechanisms, to low levels of genetic introgression between slightly divergent populations, to the formation of distinct hybrid species, hybridization is thought of as both a creative and a restrictive force in evolution (e.g., Anderson and Stebbins, 1954; Barton and Hewitt, 1985; Mallet, 2007; Genner and Turner, 2012; Abbott et al., 2013;

Schumer et al., 2014). It is the potential for the production of novelty that makes hybridization such an intriguing—and potentially useful—phenomenon.

In some cases, hybridization can lead to saltational evolution (Mallet, 2007). In plants, this process often occurs via polyploidization, wherein an individual is produced that has the complete genomes of both parental species (Soltis and Soltis, 2009). While few hybrid animal species arise in this fashion, in plants this mode of saltational evolution is a common mechanism of hybrid speciation (Wood et al., 2009; Otto and Whitton, 2000). Thus, historical polyploid events are known to have played an important role in the evolution of angiosperms (Cui et al., 2006), with subsequent diploidization to the modern genomes. Similar but more recent polyploid events underlie the evolution of several modern crop species (see below and Appendix 2.1). In other instances, hybridization does not result in genome duplication, but leads to repeated rounds of natural backcrossing and selection, resulting in the introduction of genome segments that contain novel adaptive traits. Segmental introgressions have been important, for example, in the adaptation of highland maize varieties based on gene flow from highland-adapted wild species (Hufford et al., 2013).

Regardless of whether hybridization results in a new lineage, gene flow between divergent lineages, or fusion of lineages, it can generate multilocus genotypes that are not present in either parent, leading to offspring with particular traits that exceed those of either parental population. In fact, this effect of transgressive segregation is the rule rather than the exception (Rieseberg et al., 1999). Furthermore, the very nature of hybridization may predispose hybrid lineages to have novel traits through the restructuring of genetic interactions and by altering predispositions for reproductive

isolation (Seehausen, 2013). Although hybrids tend to occur in small numbers and may often be maladapted, strong selection and genetic drift can lead to colonization of new niche space. Given these facts, it is no surprise that hybrid vigor has been cited as an impetus for the evolution of invasiveness (e.g., Ellstrand and Schierenbeck, 2000) and for adaptive radiations, during which phenotypic novelty emerges at a rapid rate (e.g., Seehausen, 2004; East African Great Lake cichlid fish, Joyce et al., 2011; Genner and Turner, 2012; Keller et al., 2013; Hawaiian silverswords, Baldwin, 1997; Barrier et al., 1999), and it is clear why hybridization is such a powerful tool for the improvement of crops.

Examples of hybridization are widespread in plants and have become increasingly common in animals (e.g., Mallet, 2005, 2007; Abbott et al., 2013). Although the rate of hybridization among related species tends to be low, the number of species that hybridize is relatively high (Mallet, 2005). An estimated 10% of all animal species and 25% of all plant species undergo hybridization (Mallet, 2005), and genome-wide scans of an increasingly large number of organisms reveal that their genomes are subject to introgression (e.g., Baack and Rieseberg, 2007; Arnold and Martin, 2010; Green et al., 2010; Abbott et al., 2013).

Hybridization is known to be common in some groups of animals—for example, 75% of the ducks of the British Isles (Gillham and Gillham, 1996; Mallet, 2005) and over 25% of all tit species (Paridae) can hybridize (Harrap and Quinn, 1996). In recent years, the prevalence of introgressive hybridization in animals has become even more apparent, with examples arising from across the animal kingdom, including both recent and historical hybridization in the primate family (e.g., Pastorini et al., 2009; Green et al.,

2010; Zinner et al., 2011). Interestingly, an estimated 20% of the Neandertal genome survived in modern humans through hybridization and introgression, with 1–3% admixture on an individual basis (Vernot and Akey, 2014). In the past decade, there have emerged several examples of homoploid hybrid speciation in animals, where hybridization between two species has led to a third with a distinct morphology or niche (Mallet, 2007). These hybridization events can lead to important novel phenotypes, such as the emergence of the specialized feeding forms of tephritid fly *Rhagoletes mendax ×zephyria* on invasive honeysuckle plants (*Lonicera* spp.) arising from hybridization between parental species specialized on blueberry (*R. mendax*) and snowberry (*R. zephyria*) (Schwarz et al., 2005). As is the case in plants, introgression between domesticated animals and their wild relatives is known to occur (e.g., pigs and wild boars; Goedbloed et al., 2013), and these wild relatives are considered valuable genetic resources for livestock improvement efforts (Taberlet et al., 2008). While this paper aims to explore the capacity for hybridization to generate phenotypic novelty in crop plants, examples from the animal kingdom demonstrate the widespread prevalence of hybridization and suggest that livestock may also benefit from similar genomically based breeding programs.

Naturally occurring hybridization in plants has been known since the time of Linnaeus (e.g., Gustafsson, 1979) and featured prominently in *On the Origin of Species* (chapter 8) and perhaps in Darwin's conception of species (e.g., Kottler, 1978). Ever since, hybridization has been an often discussed, reviewed, and debated topic in plant evolution (e.g., Stebbins, 1950; Abbott, 1992; Arnold, 1992; Rieseberg, 1995, 1997; Rhymer and Simberloff, 1996; Levin et al., 1996; Ellstrand and Schierenbeck, 2000;

Barton, 2001; Seehausen, 2004; Soltis and Soltis, 2009; Arnold and Martin, 2010; Abbott et al., 2013; Schumer et al., 2014). Cases of natural hybridization in plants, both homoploid and polyploid, give us insight into the range of phenotypic and ecological effects hybridization can have, and can act as models for harnessing the power of hybridization in agriculture.

A striking and well-known example of homoploid hybrid speciation is that of *Helianthus* sunflowers of the United States. Rieseberg (1991) identified three natural homoploid hybrid species, *H. anomalus, H. deserticola,* and *H. paradoxicus*, as the offspring of *H. annuus* and *H. petiolaris*. These hybrid *Helianthus* are exemplars of transgressive segregation, highlighting the expanded potential of hybrid species, in this case through colonization of extreme habitats where neither parental species can survive (sand dune, desert floor, and salt flats, respectively) (Rieseberg et al., 2007). Additionally, these species show that repeated hybridization events between the same parental species can have vastly different outcomes. As Arnold et al. (2012) discussed, work investigating another homoploid hybrid complex, the Louisiana iris has demonstrated the variability of hybrid fitness, which is dependent on both genotype and environment. These lessons from natural hybrids bear particular relevance to crop improvement efforts that aim to produce crops adapted to changing climatic conditions.

Agriculture and Hybridization

In agriculture, hybridization between crops and wild relatives has long been a major research focus. One goal is to introgress adaptive traits from wild relatives into cultivated forms as part of breeding programs, which we detail below. A second focus

has been on crops of hybrid origin. A number of crops have formed by way of hybridization, including many of polyploid origin (e.g., Nagaharu, 1935; Udall and Wendel, 2006; Vaughan et al., 2007; Appendix 2.1). Meyer et al. (2012) cite 37 of 203 crops as having a change in ploidy as part of their domestication, or about 15%, similar to estimates of speciation events across angiosperms involving shifts in ploidy (Wood et al., 2009). Some notable examples of polyploid crops include wheat (Peng et al., 2011), bananas (Heslop-Harrison and Schwarzacher, 2007), strawberries (Folta and Davis, 2006), and vegetable and oilseed brassicas (Prakash et al., 2011). Some more recently developed crops involve hybridization events that have occurred far beyond the region of domestication and rather recently, such as the formation of grapefruit in Barbados in the 18th century as a homoploid hybrid of the sweet orange (*Citrus sinensis*) and shaddock (*C. maxima*) (Kumamoto et al., 1987). A third area of intense recent interest has been in the escape of transgenes from genetically modified crops into weedy wild relatives. This field has been growing and has been reviewed several times (Pilson and Prendeville, 2004; Armstrong et al., 2005; Ellstrand et al., 2013). Although we do not aim to thoroughly review this topic, we mention it due to its importance and the way it complements intentional introgression from wild into cultivated backgrounds. For example, studies by Mercer and colleagues of crop introgression into wild (weedy) populations found shifts in growth rate and flowering time that likely impact the ability of transgenes to persist in populations (Mercer et al., 2006a, b, 2007). Vacher et al. (2011) found similar results from introgression of genetically modified crops into populations of weedy *Brassica* crop relatives. Snow and colleagues showed that the transgenes introgressed into weedy relatives can persist in weedy populations (e.g., Snow et al.,

2010). Importantly, our ability to detect crop–wild hybridization at both a fine scale and broad scope is improving as the costs of sequencing decline. For example, Hufford et al. (2013) found widespread genomic signatures of crop and wild alleles moving quite frequently between cultivated maize and wild teosinte in southern Mexico. Such gene flow is likely more common than previously appreciated in crops that are grown in proximity to wild relatives, even those that primarily self-pollinate.

Hybridization of Crops and their Wild Relatives to Confer Adaptive Traits

Early examples of targeted introgression can be traced to the work of Vavilov (1926, 1951). Since that time, crop wild relatives have been used to confer adaptive traits in a variety of crops, with the most widespread use occurring in a limited number of annual crops such as wheat, rice, barley, cassava, potato, and tomato. Maxted and Kell (2009) report 291 articles that identify and attempt to transfer useful traits from 185 wild relative taxa to 29 crop species. More than 50% of these traits are for disease and pest resistance, with traits for abiotic stress tolerance accounting for an additional 10–15%. Yield improvement also accounts for perhaps another 20%, although this can be hard to differentiate from other categories. Another review by Brumlop et al. (2013) of 104 molecular assisted breeding papers published from 1995–2012 found that approximately 74% of these studies were focused on introgression of traits that confer disease resistance, with the rest focused on traits involved in abiotic stress tolerance, improved yield, and growth habit. Although generations of breeders have performed crop–wild crosses across a large number of taxa and involving thousands of individual crosses, these crosses are still limited in comparison to the range of variation present in wild relatives of all of our

cultivated plant species. Records of pedigrees for many crops show that most breeding programs have had a narrowing of the crop genetic basis over the past few decades as breeders tend to reuse a few favored parents to establish new elite varieties (e.g., Kumar et al., 2003). Efforts to improve plant health and production through the use of introgression from crop wild relatives have substantial economic value, which was estimated at nearly $115 billion globally over 15 yr ago (Pimentel et al., 1997); we can only assume that this value has risen since that time. Given the many challenges posed by climate change, the scale of usage of crop wild relatives must be increased dramatically to keep up with changing conditions.

A noteworthy example of targeted introgression of a wild relative comes from common bean, *Phaseolus vulgaris* (reviewed by Acosta-Gallegos et al., 2007). Breeders have successfully introgressed genes conferring resistance to insects (e.g., bruchid beetle seed predators, *Apion* pod weevils) and pathogens (e.g., *Fusarium*), as well as higher nitrogen, iron, and calcium seed content from existing collections of wild *Phaseolus* (reviewed in Acosta-Gallegos et al., 2007). These efforts have contributed to both higher yields and improved nutritional quality and have also lessened the environmental impact of crop production by facilitating reduced pesticide, herbicide, and fertilizer use. While these gains are significant, the collection of wild relatives in *Phaseolus* was not built systematically or with an intention of building a resource that reflects the complete range of habitats in which wild *Phaseolus* thrives. Without such a systematic search, it may be difficult to find the full range of alleles for particularly valuable traits, like acrelin insecticidal proteins in *Phaseolus*, which occur at low frequencies in natural populations of crop wild relatives (Acosta-Gallegos et al., 1998). Furthermore, Acosta-Gallegos et al.

(2007) argued persuasively that the wild diversity can be better used if converted or incorporated (sensu Simmonds, 1993) into the common bean breeding pool. This goal is most easily accomplished with marker assisted breeding or genomic selection (e.g., Nakaya and Isobe, 2012) and an understanding of the genetic basis of domestication traits (in *Phaseolus*, these traits include pod shattering, growth habit, and photoperiod insensitivity) that must be recovered in crop–wild crosses to create a new cultigen that is suitable for future agricultural conditions.

The example of common beans extends to other crops critical to food security in the developing world, such as the 19 crop species for which the Consultative Group on International Agricultural Research (CGIAR) coordinates breeding efforts. Hajjar and Hodgkin (2007) reviewed the use of crop wild relatives by CGIAR in 16 of their mandate crops. Through both an examination of literature and interviews with CGIAR breeders and germplasm managers, the authors extensively surveyed breeder usage in the CGIAR system and uncovered patterns not obvious from published literature alone. As in the literature based reviews by Maxted and Kell (2009) and Brumlop et al. (2013), they found that over 80% of usage has been for disease and pest resistance. However, in 13 of the 16 mandate crops some traits besides resistance have been successfully transferred from crop wild relatives, representing a rise in the usage of wild relatives in breeding since an earlier review (Prescott-Allen and Prescott-Allen, 1986). This trend toward greater usage of wild relatives is consistent with the broader breeding community (Dulloo et al., 2013). However, these reviews illustrate that the majority of examples of crop–wild hybridization in breeding have been ad hoc in their usage of wild germplasm. None of these efforts have screened existing wild relatives for more than a few traits, and none

have used crop wild relative collections that were systematically built to represent the range of adaptations found in natural populations.

Obstacles to the Usage of Crop Wild Relatives

In addition to highlighting the potential benefits of crop wild relatives, the described studies also discussed obstacles that limit the use of crop wild relatives in breeding programs, including their poor agronomic performance (Haussmann et al., 2004). Poor performance can take many forms. For example, crop wild relatives often lack important domestication traits, such as shattering pods or shifted germination timing (e.g., Acosta-Gallegos et al., 2007), or broader environmental adaptations (e.g., Haussmann et al., 2004). In some crops, such as chickpea, phenological differences make the temperate wild relative unsuited to subtropical or tropical conditions (e.g., Abbo et al., 2003; Berger et al., 2006), and the same issues are at play for tropical crops grown in temperate regions, such as maize and common bean (reviewed by Jung and Müller [2009], Buckler et al. [2009], and Acosta-Gallegos et al. [2007]). It can be difficult to remove such undesirable traits from crop–wild hybrid lines. For attempts to introgress a targeted trait with a fairly simple genetic basis, such as a resistance gene, backcrossing can be time-consuming and difficult. Even after three generations, regions of a wild chromosome spanning many centimorgans may remain around an average selected gene (Stam and Zeven, 1981; Welz and Geiger, 2000; Haussmann et al., 2004). Linked regions that negatively influence agronomic performance, pleiotropy, and other complications make the task harder (e.g., Xu et al., 2006). Loci associated with domestication are similar to barrier loci (sensu Abbott et al., 2013), reducing gene flow between

populations; these loci are central to the genetics of speciation and may reduce the fitness of hybrids. Although molecular-assisted breeding and increasingly genomic selection can be of great assistance if the candidate gene is known, these techniques remain time consuming (e.g., Young, 1999; Varshney et al., 2005; Xu and Crouch, 2008; Kumar et al., 2011).

Another obstacle in the use of wild relatives is their poor representation in international germplasm collections. Maxted and Kell (2009) estimated that only 2–6% of international germplasm collections are of crop wild relatives, with landraces and varieties making up the vast majority of accessions for most crops. Although collections for a few crops and their wild relatives are large, wild relatives of many crops have been poorly collected or have been almost ignored, and some such as faba bean even lack well-identified wild relatives (e.g., Kaur et al., 2014). Two striking examples with which we are familiar are grain legumes of considerable importance in the semiarid tropics and many temperate areas: chickpea and peanut. Berger et al. (2003) estimated that for the immediate wild ancestor of chickpea, *Cicer reticulatum*, the existing international collections of more than 150 named accessions stem from only 18 independent accessions; the large number of accessions counted in these collections appears to derive from proliferation of these original 18 accessions as distinct lineages. This practice grossly inflates the adequacy of the collection, because most accessions are redundant. For peanut, an allotetraploid with an A and a B genome, there is only a single individual available of the B genome parent, *Arachis ipaensis* in the USDA and ICRISAT collections, despite over 40 collecting trips organized by USDA and other collectors (Holbrook and Stalker, 2003). Moreover, peanut appears to derive from a single

hybridization event, creating an unusually strong genetic bottleneck in the crop. To combat this genetic deficiency, synthetic allotetraploid hybrids have been created from related A genome accessions and the sole B genome representative (Fonceka et al., 2012). Many more *Arachis* species are poorly collected and at high risk of extinction (e.g., Jarvis et al., 2003). Yet even for well-collected crops like *Phaseolus*, collections of wild relatives are likely not geographically exhaustive; gap analyses still indicate regions and taxa that are underrepresented (e.g., Ramirez-Villegas et al., 2010). Assessments of the adequacy of current wild collections have commonly been based on the number of wild accessions in germplasm repositories. However, this measure often overestimates diversity in the collections because initially collected samples are generally assigned additional accession identifiers during distribution and evaluation. Reliance on numerical coverage in collections has also shifted the focus of future collection efforts to taxa with lower numeric or limited geographic representation while overlooking the inadequacies of current redundant collections. Furthermore, nearly all older collections of wild relatives have incomplete passport information and most have all of the seeds from a particular geographic location bulked into a single bag, making it difficult to impossible to determine patterns of within and among population variation in crop wild relatives (Greene and Hart, 1999). In addition to the inadequacies of many ex situ germplasm collections, many crop wild relatives occur in geopolitically unstable areas where collection has long been complicated, and where in situ conservation is at best challenging.

An additional obstacle to the use of wild relatives is the unpredictability of both a wild individual's phenotype under agronomic conditions and the phenotype of crop–wild

hybrids. Phenotypes of wild individuals are often assessed in agricultural settings, a largely uninformative practice when the overall wild phenotype is specifically adapted for fitness in wild but not cultivated settings. For instance, when plant phenology (e.g., flowering time and/or vernalization) differ substantially between wild and cultivated material then phenotypic comparisons are problematic and it may be first necessary to "correct" the timing of development to cultivated set points before initiating phenotypic assessment. Further complicating the issue, genotype–environment interactions can make the phenotype expressed under agricultural conditions different from what it would be under natural conditions. Predicting the phenotypes of crop–wild hybrids also remains complicated. In very few cases, even for model organisms such as *Drosophila* or *Arabidopsis*, do we understand the genotype–phenotype map well enough to fully predict phenotypes of crosses or advanced introgressions. However, an important first step toward building this capacity for crop–wild hybrids is understanding the major loci that have been under strong artificial selection during domestication. For an increasing number of crops, major domestication loci have been identified (e.g., reviews by Doebley et al. [2006], Gross and Olsen [2010], Meyer et al. [2012]). In advanced backcross lines, breeders can recover crop alleles of the major domestication loci, speeding the recovery of the essential crop phenotype and retaining adaptive variation from the wild relatives.

A Proposal for Future Work

We propose a multistep framework for utilizing naturally occurring variation in wild relatives of crops (Fig. 2.1). It is increasingly possible to digitize genotype–environment interactions in wild progenitor populations and from there predict the effect

of wild alleles in cultivated backgrounds. We propose five steps to better use crop wild relatives.

(1) Build comprehensive collections of wild relatives**—**These must span as much of the spatial and ecological range of wild relatives as possible to maximize the extent of adaptive variation likely to be captured by the collection. The sites of collection should be fully characterized to understand the major axes of environmental variation present, including climate, soil, and co-occurring species. This task is similar to efforts to close gaps in germplasm collections (e.g., Dempewolf et al., 2014), except that it places the emphasis not on under-sampled regions but on maximizing the range of adaptive variation present in the collection across multiple ecological axes.

(2) Sequence wild relative genomes**—**It is increasingly feasible to generate full genome sequences for crop wild relatives. Recently, genomes have been published for many minor crops, such as chickpea (Varshney et al., 2013) and pigeonpea (Varshney et al., 2011), and efforts by the African Orphan Crops Consortium are underway to sequence 100 traditional African food crops (http://www.mars.com/global/african-orphan-crops.aspx). Ideally, hundreds of accessions from the target wild species should be sequenced as a prelude to constructing functional subsets of crop wild relative diversity. Toward this end, low cost genotyping (e.g., genotyping by sequencing, Romay et al., 2013) allows cost effective recovery of genetic data, which can yield population genetic parameters that guide the prioritization of genotypes for full resequencing. When combined with sampling strategies that emphasize population-level coverage, selection of subsets of sampled accessions enriched for adaptive alleles can be achieved by focusing on high frequency alleles within individual sampled populations.

(3) Create and phenotype sets of purpose-driven hybrid populations**—**Nested association mapping (Yu et al., 2008; McMullen et al., 2009) and advance backcross introgression populations (Tanksley and McCouch, 1997) are synergistic for trait discovery and breeding. In these crosses one can remove barriers (e.g., phenology, growth habit, pod shattering) that otherwise impede the use of wild germplasm in breeding and dissect the genetic basis of adaptive traits. These populations must be carefully phenotyped for a range of high-priority traits related to crop production (e.g., ability to tolerate changing climatic conditions, resistance to emerging disease threats) using standardized phenotyping procedures in replicated trials. Phenotyping by international partnerships under a range of conditions would ensure both high-power of trait–genomic associations and their relevance to disparate crop production environments. Because these prebreeding populations lend themselves to direct incorporation into breeding programs, they can be maintained by participatory breeding networks (Murphy et al., 2005; Ceccarelli, 2006).

(4) Develop a predictive network of genotype–phenotype associations**—**A genotype–phenotype map for crops and their wild relatives will identify genes and genome regions from wild species that improve yield and resilience in the crop. The association between genomic and environmental variation in natural populations (from step 1), combined with trait–genotype associations established through phenotyping of wild–cultivated introgression lines (from step 3) will enable the identification of agronomically valuable alleles with great precision, and initiate their deployment in crop improvement programs.

(5) Deploy identified phenotypes into crop breeding pipelines**—**The ultimate goal of such activities is to select high value genome intervals from wild species for improvement of elite crop genotypes. Advanced backcross introgression (ABI) lines, preconstructed as a library of partially overlapping introgressed segments in otherwise elite cultivated genomes, can provide a ready-to-go breeding resource once these wild genome intervals are identified. Creating such ABI germplasm resources in advance of (or in parallel to) trait discovery will speed the delivery of wild traits into elite backgrounds, while the immortal nature of such a resource ensures that subsequent discovery of new traits can benefit from a preexisting ABI pipeline for trait delivery. Ideally, such ABI libraries would be created with multiple elite genotypes that together encompass traits for the primary agroclimatic zones of the crop under consideration.

The problem of perennials and wild relatives with narrow distributions**—**In many ways, perennial crops challenge our ideas about the evolutionary processes involved in domestication (Miller and Gross, 2011), and they also present unique challenges to crop improvement efforts. In particular, the extended juvenile stage of such individuals makes repeated backcrossing an extremely time-consuming endeavor. Perennial species are also more likely to be obligately outcrossing than their annual counterparts (Barrett et al., 1996; Petit and Hampe, 2006), preventing the production of inbred, homozygous lines required of some proposed techniques (e.g., nested association mapping). Nevertheless, steps 1 and 2 of this plan outline entirely feasible and extremely important goals for perennial species. Furthermore, some progress toward identifying genomic regions containing adaptive loci could be achieved in perennial species by examining the correlation between environmental variables and allele frequencies across

a geographic gradient (Coop et al., 2010; Friesen and von Wettberg, 2010; Pyhäjärvi et al., 2013). These combined efforts could have a twofold benefit to perennial crops, many of which are cultivated by means of grafting, providing both breeding material as well as novel rootstock material.

Crop wild relatives with narrow distributions may pose a different challenge for utilization in breeding: the lack of adaptation to a variety of environments. Local adaptation is common in plants, but not ubiquitous (e.g., Linhart and Grant, 1996; Hereford, 2009). In plants with more limited distributions, this pattern tends to be weaker (Hereford, 2009). Yet in the case of some crops, like maize and chickpea (e.g., Moeller et al., 2007; Abbo et al., 2003), wild relatives with limited ranges still demonstrate variation among populations consistent with local adaptation (Pyhäjärvi et al., 2013; von Wettberg et al., unpublished data). Even if the wild relative range is extremely limited and local adaptation is minimal, hybridization can likely yield some expansion of the range of genetic variation. Ultimately, crops like maize and chickpea perform well across huge regions (both are grown on six continents) not because the crop (or a wild relative) has a superior breadth of habitat adaptations, but rather because the agricultural habitat is highly contrived, relatively uniform, and managed through intensive inputs. When these contrivances are not or cannot be met, yields in agricultural systems suffer. This is where the true value of systematic hybridization of crop wild relatives can provide the most value, even with a limited increase in adaptive breadth.

***The continual need for conservation*—**No review of the use of crop wild relatives should ignore the fact that these species are nearly universally threatened (e.g., Ford-Lloyd et al., 2011; Maxted et al., 2012). Many are rare due to habitat loss,

fragmentation, and degradation. Crop wild relatives are also threatened by climate change (e.g., Jarvis et al., 2008; Ford-Lloyd et al., 2011) and agricultural intensification and development. For example, at least one of 20 wild chickpea populations that we discovered was lost just 6 mo later (von Wettberg et al., personal observation). Although wild relatives of crops are specifically targeted by conservation efforts (Meilleur and Hodgkin, 2004; Hunter and Heywood, 2011) in many regions, there is little in situ conservation of these resources (e.g., Maxted et al., 1997), in part, because national parks or other conservation areas were established with aims independent and unrelated to the preservation of crop genetic resources. Crop wild relatives also face genetic risks, such as introgression from cultivated forms, or in the case of medicinally or pharmaceutically useful species, direct overharvesting (e.g., Nantel et al., 1996; Law and Salick, 2005). Few of these issues are easily tractable. However, prioritizing the most threatened crop wild relatives is essential (e.g., Vincent et al., 2013). Furthermore, collaborations with local researchers and organizations can build local consensus about the numerous benefits of protecting crop wild relatives both in situ and ex situ, and perhaps be more effective than efforts from the outside alone. Despite inherent difficulties, international efforts to systematically collect crop wild relatives represent the first step toward building more climate resilient crops that can meet the demands of agriculture in the 21st century.

Literature Cited

Abbo, S., J. Berger, and N. C. Turner. 2003. Viewpoint: Evolution of cultivated chickpea: Four bottlenecks limit diversity and constrain adaptation. *Functional Plant Biology* 30: 1081–1087.

Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, et al. 2013. Hybridization and speciation. *Journal of Evolutionary Biology* 26: 229–246.

Abbott, R. J. 1992. Plant invasions, interspecific hybridization and the evolution of new plant taxa. *Trends in Ecology & Evolution* 7: 401–405.

Acosta-Gallegos, J. A., J. D. Kelly, and P. Gepts. 2007. Prebreeding in common bean and use of genetic diversity from wild germplasm. *Crop Science* 47: S44–S59.

Acosta-Gallegos, J. A., C. Quintero, J. Vargas, O. Toro, J. Tohme, and C. Cardona. 1998. A new variant of arcelin in wild common bean, *Phaseolus vulgaris* L. from southern Mexico. *Genetic Resources and Crop Evolution* 45: 235–242.

Anderson, E., and G. L. Stebbins. 1954. Hybridization as an evolutionary stimulus. *Evolution* 8: 378–388.

Armstrong, T. T., R. G. Fitzjohn, L. E. Newstrom, A. D. Wilton, and W. G. Lee. 2005. Transgene escape: What potential for crop–wild hybridization? *Molecular Ecology* 14: 2111–2132.

Arnold, M. L. 1992. Natural hybridization as an evolutionary process. *Annual Review of Ecology and Systematics* 23: 237–261.

Arnold, M. L., E. S. Ballerini, and A. N. Brothers. 2012. Hybrid fitness, adaptation and evolutionary diversification: Lessons learned from Louisiana Irises. *Heredity* 108: 159–166.

Arnold, M. L., and N. H. Martin. 2010. Hybrid fitness across time and habitats. *Trends in Ecology & Evolution* 25: 530–536.

Atwell, S., Y. S. Huang, B. J. Vilhjálmsson, G. Willems, M. Horton, Y. Li, D. Meng, at al. 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465: 627–631.

Ayroles, J. F., M. A. Carbone, E. A. Stone, K. W. Jordan, R. F. Lyman, M. M. Magwire, S. M. Rollman, et al. 2009. Systems genetics of complex traits in *Drosophila melanogaster. Nature Genetics* 41: 299–307.

Baack, E. J., and L. H. Rieseberg. 2007. A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics & Development* 17: 513–518.

Baldwin, B. G. 1997. Adaptive radiation of the Hawaiian silversword alliance: Congruence and conflict of phylogenetic evidence molecular and non-molecular investigations. *In* T. J. Givnish and K. J. Sytsma [eds.], Molecular evolution and adaptive radiation, 103–128. Cambridge University Press, Cambridge, UK.

Barrier, M., B. G. Baldwin, R. H. Robichaux, and M. D. Purugganan. 1999. Interspecific hybrid ancestry of a plant adaptive radiation: Allopolyploidy of the Hawaiian silversword alliance (Asteraceae) inferred from floral homeotic gene duplications. *Molecular Biology and Evolution* 16: 1105–1113.

Barrett, S. C. H., L. D. Harder, and A. C. Worley. 1996. The comparative biology of pollination and mating in flowering plants. *Philosophical Transactions of the Royal Society of London. B* 351: 1271–1280.

Barton, N. H. 2001. The role of hybridization in evolution. *Molecular Ecology* 10: 551–568.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. *Annual Review of Ecology and Systematics* 16: 113–148.

Berger, J., S. Abbo, and N. C. Turner. 2003. Ecogeography of annual wild *Cicer* species: The poor state of the world collection. *Crop Science* 43: 1076–1090.

Berger, J. D., M. Ali, P. S. Basu, B. D. Chaudhary, S. K. Chaturvedi, P. S. Deshmukh, P. S. Dharmaraj, et al. 2006. Genotype by environment studies demonstrate the critical role of phenology in adaptation of chickpea (*Cicer arietinum* L.) to high and low yielding environments of India. *Field Crops Research* 98: 230–244.

Brumlop, S., W. Reichenbecher, B. Tappeser, and M. R. Finckh. 2013. What is the SMARTest way to breed plants and increase agrobiodiversity? *Euphytica* 194: 53–66.

Buckler, E. S., J. B. Holland, P. J. Bradbury, C. B. Acharya, P. J. Brown, C. Browne, E. Ersoz, et al. 2009. The genetic architecture of maize flowering time. *Science* 325: 714–718.

Cavanagh, C., M. Morell, I. Mackay, and W. Powel. 2008. From mutations to MAGIC: Resources for gene discovery, validation and delivery in crop plants. *Current Opinion in Plant Biology* 11: 215–221.

Ceccarelli, S. 2006. Decentralized-participatory plant breeding: Lessons from the south—Perspectives in the north. *In* D. Desclaux and M. Hédont [eds.], Eco-PB Proceedings on Participatory Plant Breeding: Relevance for Organic Agriculture?, La Besse,

France, 2006, 8–15. European Consortium for Organic Plant Breeding [http://www.eco-pb.org], Driebergen, The Netherlands.

Coop, G., D. Witonsky, A. Di Rienzo, and J. K. Pritchard. 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185: 1411–1423.

Cui, L., P. K. Wall, J. H. Leebens-Mack, B. G. Lindsay, D. E. Soltis, J. J. Doyle, P. S. Soltis, et al. 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Research* 16: 738–749.

Dempewolf, H., R. J. Eastwood, L. Guarino, C. K. Khoury, J. V. Müller, and J. Toll. 2014. Adapting agriculture to climate change: A global initiative to collect, conserve, and use crop wild relatives. *Agroecology and Sustainable Food Systems* 38: 369–377.

Doebley, J. F., B. S. Gaut, AND B. D. Smith. 2006. The molecular genetics of crop domestication. *Cell* 127: 1309–1321.

Dulloo, M. E., I. Thormann, E. Fiorino, S. De Felice, V. R. Rao, and L. Snook. 2013. Trends in research using plant genetic resources from germplasm collections: From 1996 to 2006. *Crop Science* 53: 1217–1227.

Ellstrand, N. C., P. Meirmans, J. Rong, D. Bartsch, A. Ghosh, T. J. De Jong, P. Haccou, et al. 2013. Introgression of crop alleles into wild or weedy populations. *Annual Review of Ecology, Evolution and Systematics* 44: 325–345.

Ellstrand, N. C., and K. A. Schierenbeck. 2000. Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences, USA* 97: 7043–7050.

Flint, J., and E. Eskin. 2012. Genome-wide association studies in mice. *Nature Reviews. Genetics* 13: 807–817.

Folta, K. M., and T. M. Davis. 2006. Strawberry genes and genomics. *Critical Reviews in Plant Sciences* 25: 399–415.

Fonceka, D., H. A. Tossim, R. Rivallan, H. Vignes, E. Lacut, F. De Bellis, I. Faye, et al. 2012. Construction of chromosome segment substitution lines in peanut (*Arachis hypogaea* L.) using a wild synthetic and QTL mapping for plant morphology. *PLoS ONE* 7: e48642.

Ford-Lloyd, B. V., M. Schmidt, S. J. Armstrong, O. Barazani, J. Engels, R. Hadas, K. Hammer, et al. 2011. Crop wild relatives— Undervalued, underutilized and under threat? *Bioscience* 61: 559–565.

Friesen, M. L., and E. J. von Wettberg. 2010. Adapting genomics to study the ecology and evolution of agricultural systems. *Current Opinion in Plant Biology* 13: 119–125.

Genner, M. J., and G. F. Turner. 2012. Ancient hybridization and phenotypic novelty within Lake Malawi's cichlid fish radiation. *Molecular Biology and Evolution* 29: 195–206.

Gillham, E., and B. Gillham. 1996. Hybrid ducks: A contribution towards an inventory. Published by the authors [location unknown].

Goedbloed, D. J., H. J. Megens, P. van Hooft, J. M. Herrero-Medrano, W. Lutz, P. Alexandri, and R. P. Crooijmans. 2013. Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations. *Molecular Ecology* 22: 856–866.

Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328: 710–722.

Greene, S. L., and T. S. Hart. 1999. Implementing a geographic analysis in germplasm conservation. *In* S. L. Greene and L. Guarino [eds.], Linking genetic resources to geography: Strategies for conserving crop biodiversity, 25–38. Crop Science Society of America, Special Publication 27, Madison, Wisconsin, USA.

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

Gustafsson, Å. 1979. Linnaeus' peloria: The history of a monster. *Theoretical and Applied Genetics* 54: 241–248.

Hajjar, R., and T. Hodgkin. 2007. The use of wild relatives in crop improvement: A survey of developments over the last 20 years. *Euphytica* 156: 1–13.

Harrap, S., and D. Quinn. 1996. Tits, nuthatches and treecreepers. Christopher Helm Publishers, London, UK.

Haussmann, B. I. G., H. K. Parzies, T. Presterl, Z. Susic, and T. Miedaner. 2004. Plant genetic resources in crop improvement. *Plant Genetic Resources* 2: 3–21.

Hereford, J. 2009. A quantitative survey of local adaptation and fitness trade-offs. *American Naturalist* 173: 579–588.

Heslop-Harrison, J. S., and T. Schwarzacher. 2007. Domestication, genomics and the future for banana. *Annals of Botany* 100: 1073–1084.

Holbrook, C. C., and H. T. Stalker. 2003. Peanut breeding and genetic resources. *Plant Breeding Reviews* 22: 297–356.

Hufford, M. B., P. Lubinksy, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand, and J. Ross-Ibarra. 2013. The genomic signature of crop– wild introgression in maize. *PLOS Genetics* 9: e1003477.

Huang, X., and B. Han. 2013. Natural variations and genome-wide association studies in crop plants. *Annual Review of Plant Biology* 65: 531–551.

Hunter, D., and V. Heywood [eds.]. 2011. Crop wild relatives: A manual of *in situ* conservation. Earthscan, London, UK.

Hyten, D. L., Q. Song, Y. Zhu, I. Y. Choi, R. L. Nelson, J. M. Costa, J. E. Specht, et al. 2006. Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Sciences, USA* 103: 16666–16671.

Jarvis, A., M. E. Ferguson, D. E. Williams, L. Guarino, P. G. Jones, H. T. Stalker, J. F. M. Valls, et al. 2003. Biogeography of wild *Arachis:* Assessing conservation status and setting future priorities. *Crop Science* 43: 1100–1108.

Jarvis, A., A. Lane, and R. J. Hijmans. 2008. The effect of climate change on crop wild relatives. *Agriculture, Ecosystems & Environment* 126: 13–23.

Joyce, D. A., D. H. Lunt, M. J. Genner, G. F. Turner, R. Bills, and O. Seehausen. 2011. Repeated colonization and hybridization in Lake Malawi cichlids. *Current Biology* 21: 526–527.

Jung, C., and A. E. Müller. 2009. Flowering time control and applications in plant breeding. *Trends in Plant Science* 14: 563–573.

Kaur, S., N. O. Cogan, J. W. Forster, and J. G. Paull. 2014. Assessment of genetic diversity in faba bean based on single nucleotide polymorphism. *Diversity* 6: 88–101.

Keller, I., C. E. Wagner, L. Greuter, S. Mwaiko, O. M. Selz, A. Sivasundar, S. Wittwer, and O. Seehausen. 2013. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology* 22: 2848–2863.

Korte, A., and A. Farlow. 2013. The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9: 29.

Kottler, M. J. 1978. Charles Darwin's biological species concept and theory of geographic speciation: The transmutation notebooks. *Annals of Science* 35: 275–297.

Kumamoto, J., R. W. Scora, H. W. Lawton, and W. A. Clerx. 1987. Mystery of the forbidden fruit: Historical epilogue on the origin of the grapefruit, *Citrus paradisi* (Rutaceae). *Economic Botany* 41: 97–107.

Kumar, J., A. K. Choudhary, R. K. Solanki, and A. Pratap. 2011. Towards marker-assisted selection in pulses: A review. *Plant Breeding* 130: 297–313.

Kumar, S., S. Gupta, S. Chandra, and B. B. Singh. 2003. How wide is the genetic base of pulse crops? *In* M. Ali, B. B. Singh, S. Kumar, and V. Dhar [eds.], Proceedings of the National Symposium on Crop Diversification and Natural Resources Management: Pulses in New Perspective, 2003, 211-221. Indian Institute Pulses Research, Kanpur, India.

Ladizinsky, G. 1985. Founder effect in crop–plant evolution. *Economic Botany* 39: 191–199.

Law, W., and J. Salick. 2005. Human-induced dwarfing of Himalayan snow lotus, *Saussurea laniceps* (Asteraceae). *Proceedings of the National Academy of Sciences, USA* 102: 10218–10220.

Levin, D. A., J. Francisco Ortega, and R. K. Jansen. 1996. Hybridization and the extinction of rare plant species. *Conservation Biology* 10: 10–16.

Linhart, Y. B., and M. C. Grant. 1996. Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology and Systematics* 27: 237–277.

Mackay, T. F., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, D. Zhu, S. Casillas, et al. 2012. The *Drosophila melanogaster* genetic reference panel. *Nature* 482: 173–178.

Mallet, J. 2005. Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* 20: 229–237.

Mallet, J. 2007. Hybrid speciation. *Nature* 446: 279–283. Maxted, N., B. V. Ford-Lloyd, and J. G. Hawkes, [eds.]. 1997. Plant genetic conservation: The *in situ* approach. Chapman & Hall, London, UK.

Maxted, N., S. Kell, B. Ford-Lloyd, E. Dulloo, and A. Toledo. 2012. Toward the systematic conservation of global crop wild relative diversity. *Crop Science* 52: 774–785.

Maxted, N., and S. P. Kell. 2009. Establishment of a global network for the in situ conservation of crop wild relatives: Status and needs. FAO Commission on Genetic Resources for Food and Agriculture, Rome, Italy.

McCouch, S., G. J. Baute, J. Bradeen, P. Bramel, P. K. Bretting, E. Buckler, J. M. Burke, et al. 2013. Agriculture: Feeding the future. *Nature* 499: 23–24.

McMullen, M. D., S. Kresovich, H. S. Villeda, P. Bradbury, H. Li, Q. Sun, S. Flint-Garcia, et al. 2009. Genetic properties of the maize nested association mapping population. *Science* 325: 737–740.

Meilleur, B. A., and T. Hodgkin. 2004. *In situ* conservation of crop wild relatives: status and trends. *Biodiversity and Conservation* 13: 663–684.

Mercer, K. L., D. A. Andow, D. L. Wyse, and R. G. Shaw. 2007. Stress and domestication traits increase the relative fitness of crop–wild hybrids in sunflower. *Ecology Letters* 10: 383–393.

Mercer, K. L., R. G. Shaw, and D. L. Wyse. 2006a. Increased germination of diverse crop-wild hybrid sunflower seeds. *Ecological Applications* 16: 845–854.

Mercer, K. L., D. L. Wyse, and R. G. Shaw. 2006b. Effects of competition on the fitness of wild and crop–wild hybrid sunflower from a diversity of wild populations and crop lines. *Evolution* 60: 2044–2055.

Meyer, R. S., A. E. Duval, and H. R. Jensen. 2012. Patterns and processes in crop domestication: An historical review and quantitative analysis of 203 global food crops. *New Phytologist* 196: 29–48.

Miller, A. J., and B. L. Gross. 2011. From forest to field: Perennial fruit crop domestication. *American Journal of Botany* 98: 1389–1414.

Moeller, D. A., M. I. Tenaillon, and P. Tiffin. 2007. Population structure and its effects on patterns of nucleotide polymorphism in teosinte (*Zea mays* ssp. *parviglumis*). *Genetics* 176: 1799–1809.

Murphy, K., D. Lammer, S. Lyon, B. Carter, and S. S. Jones. 2005. Breeding for organic and low-input farming systems: An evolutionary– participatory breeding method for inbred cereal grains. *Renewable Agriculture and Food Systems* 20: 48–55.

Nagaharu, U. 1935. Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japanese Journal of Botany* 7: 389–452.

Nakaya, A., and S. N. Isobe. 2012. Will genomic selection be a practical method for plant breeding? *Annals of Botany* 110: 1303–1316.

Nantel, P., D. Gagnon, and A. Nault. 1996. Population viability analysis of American ginseng and wild leek harvested in stochastic environments. *Conservation Biology* 10: 608–621.

Olsen, K. M., and J. F. Wendel. 2013. A bountiful harvest: Genomic insights into crop domestication phenotypes. *Annual Review of Plant Biology* 64: 47–70.

Otto, S. P., and J. Whitton. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* 34: 401–437.

Pastorini, J., A. Zaramody, D. J. Curtis, C. M. Nievergelt, and N. I. Mundy. 2009. Genetic analysis of hybridization and introgression between wild mongoose and brown lemurs. *BMC Evolutionary Biology* 9: 32.

Peng, J. H., D. Sun, and E. Nevo. 2011. Domestication evolution, genetics and genomics in wheat. *Molecular Breeding* 28: 281–301.

Petit, R., and A. Hampe. 2006. Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution and Systematics* 37: 187–214.

Pilson, D., and H. R. Prendeville. 2004. Ecological effects of transgenic crops and the escape of transgenes into wild populations. *Annual Review of Ecology, Evolution and Systematics* 35: 149–174.

Pimentel, D., C. Wilson, C. McCullum, R. Huang, P. Dwen, J. Flack, Q. Tran, et al. 1997. Economic and environmental benefits of biodiversity. *Bioscience* 47: 747–757.

Prakash, S., X.-M. Wu, and S. R. Bhat. 2011. History, evolution, and domestication of *Brassica* crops. *In* J. Janick [ed.], Plant breeding reviews, vol. 35, 19–84. John Wiley, Hoboken, New Jersey, USA.

Prescott-Allen, C., and R. Prescott-Allen. 1986. The first resource: Wild species in the North American economy. Yale University Press, New Haven, Connecticut, USA.

Pyhäjärvi, T., M. B. Hufford, S. Mezmouk, and J. Ross-Ibarra. 2013. Complex patterns of local adaptation in teosinte. *Genome Biology and Evolution* 5: 1594–1609.

Ramírez-Villegas, J., C. Khoury, A. Jarvis, D. G. Debouck, and L. Guarino. 2010. A gap analysis methodology for collecting crop genepools: A case study with *Phaseolus* beans. *PLOS ONE* 5: e13497.

Rhymer, J. M., and D. Simberloff. 1996. Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics* 27: 83–109.

Rieseberg, L. H. 1991. Homoploid reticulate evolution in Helianthus (Asteraceae): Evidence from ribosomal genes. *American Journal of Botany* 78: 1218–1237.

Rieseberg, L. H. 1995. The role of hybridization in evolution: Old wine in new skins. *American Journal of Botany* 82: 944–953.

Rieseberg, L. H. 1997. Hybrid origins of plant species. *Annual Review of Ecology and Systematics* 28: 359–389.

Rieseberg, L. H., M. A. Archer, and R. K. Wayne. 1999. Transgressive segregation, adaptation and speciation. *Heredity* 83: 363–372.

Rieseberg, L. H., S. C. Kim, R. A. Randell, K. D. Whitney, B. L. Gross, C. Lexer, and K. Clay. 2007. Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica* 129: 149–165.

Romay, M. C., M. J. Millard, J. C. Blaubitz, J. A. Peiffer, K. L. Swarts, T. M. Casstevens, R. J. Elxhire, et al. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biology* 14: R55.

Schumer, M., G. G. Rosenthal, and P. Andolfatto. 2014. How common is homoploid hybrid speciation? *Evolution* 68: 1553–1560.

Schwarz, D., B. M. Matta, N. L. Shakir-Botteri, and B. A. McPheron. 2005. Host shift to an invasive plant triggers rapid animal hybrid speciation. *Nature* 436: 546–549.

Seehausen, O. 2004. Hybridization and adaptive radiation. *Trends in Ecology & Evolution* 19: 198–207.

Seehausen, O. 2013. Conditions when hybridization might predispose populations for adaptive radiation. *Journal of Evolutionary Biology* 26: 279–281.

Simmonds, N. W. 1993. Introgression and incorporation: Strategies for the use of crop genetic resources. *Biological Reviews of the Cambridge Philosophical Society* 68: 539–562.

Snow, A. A., T. M. Culley, L. G. Campbell, P. M. Sweeney, S. G. Hegde, and N. C. Ellstrand. 2010. Long-term persistence of crop alleles in weedy populations of wild radish (*Raphanus raphanistrum*). *New Phytologist* 186: 537–548.

Soltis, P. S., and D. E. Soltis. 2009. The role of hybridization in plant speciation. *Annual Review of Plant Biology* 60: 561–588.

Spillane, C., and P. Gepts. 2001. Evolutionary and genetic perspectives on the dynamics of crop genepools. *In* H. D. Cooper, C. Spillane, and T. Hodgkin [eds.], Broadening the genetic base of crop production, 25–70. CABI, Wallingford, UK.

Stam, P., and A. C. Zeven. 1981. The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. *Euphytica* 30: 227–238.

Stebbins, C. L. JR. 1950. Variation and evolution in plants. Columbia University Press, New York, New York, USA.

Taberlet, P., A. Valentini, H. R. Rezaei, S. Naderi, F. Pompanon, R. Negrini, and P. Ajmone-Marsan. 2008. Are cattle, sheep, and goats endangered species? *Molecular Ecology* 17: 275–284.

Tanksley, S. D., and S. R. McCouch. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277: 1063–1066.

Tester, M., and P. Langridge. 2010. Breeding technologies to increase crop production in a changing world. *Science* 327: 818–822.

Tian, F., P. J. Bradbury, P. J. Brown, H. Hung, Q. Sun, S. Flint-Garcia, T. R. Rocheford, et al. 2011. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature Genetics* 43: 159–162.

Udall, J. A., and J. F. Wendel. 2006. Polyploidy and crop improvement. *Crop Science* 46: S3–S14.

Vacher, C., T. M. Kossler, M. E. Hochberg, and A. E. Weis. 2011. Impact of interspecific hybridization between crops and weedy relatives on the evolution of flowering time in weedy phenotypes. *PLOS ONE* 6: e14649.

Varshney, R. K., W. Chen, Y. Li, A. K. Bharti, R. K. Saxena, J. A. Schlueter, M. T. A. Donoghue, et al. 2011. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource poor farmers. *Nature Biotechnology* 30: 83–89.

Varshney, R. K., A. Graner, and M. E. Sorrells. 2005. Genomics assisted breeding for crop improvement. *Trends in Plant Science* 10: 621–630.

Varshney, R. K., C. Song, R. K. Saxena, S. Azam, S. Yu, A. G. Sharpe, S. Cannon, et al. 2013. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nature Biotechnology*.

Vaughan, D. A., E. Balazs, and J. S. Heslop-Harrison. 2007. From crop domestication to super-domestication. *Annals of Botany* 100: 893–901.

Vavilov, N. I. 1926. Origin and geography of cultivated plants, translated by D. Löve. Cambridge University Press, Cambridge, UK.

Vavilov, N. I. 1951. The origin, variation, immunity and breeding of cultivated plants: Selected writings of N. I. Vavilov, translated by K. Starr Chester. Chronica Botanica, vol. 13. Ronald Press, New York, New York, USA.

Vernot, B., and J. M. Akey. 2014. Resurrecting surviving Neandertal lineages from modern human genomes. *Science* 343: 1017–1021.

Vincent, H., J. Wiersema, S. Kell, H. Fielder, S. Dobbie, N. P. Castañeda Álvarez, L. Guarino, et al. 2013. A prioritized crop wild relative inventory to help underpin global food security. *Biological Conservation* 167: 265–275.

Welz, H. G., And H. H. Geiger. 2000. Genes for resistance to northern corn leaf blight in diverse maize populations. *Plant Breeding* 119: 1–14.

Wood, T. E., N. Takebayashi, M. S. Barker, I. Mayrose, P. B. Greenspoon, and L. H. Rieseberg. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences, USA* 106: 13875–13879.

Xu, K., X. Xu, T. Fukao, P. Canlas, R. Maghirang-Rodriguez, S. Heuer, A. M. Ismail, et al. 2006. *Sub1A* is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* 442: 705–708.

Xu, Y., and J. H. Crouch. 2008. Marker-assisted selection in plant breeding: From publications to practice. *Crop Science* 48: 391–407.

Young, N. D. 1999. A cautiously optimistic vision for marker-assisted breeding. *Molecular Breeding* 5: 505–510.

Yu, J., J. B. Holland, M. D. McMullen, and E. S. Buckler. 2008. Genetic design and statistical power of nested association mapping in maize. *Genetics* 178: 539–551.

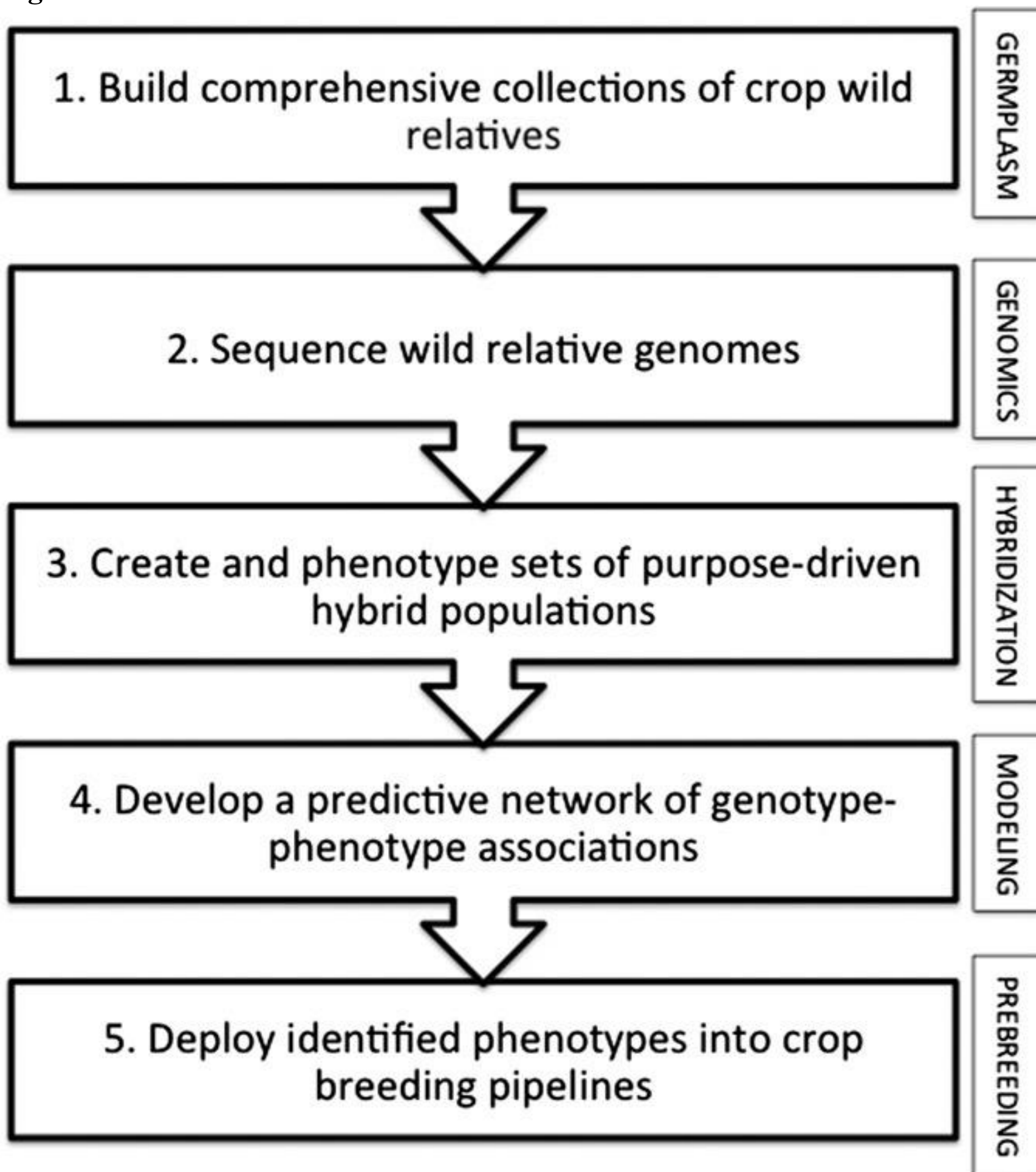
Zinner, D., M. L. Arnold, and C. Roos. 2011. The strange blood: Natural hybridization in primates. *Evolutionary Anthropology* 20: 96–103.

Figure Captions

Figure 2.1. An outline of the 5-step approach we advocate for utilizing crop wild relatives in a systematic and thorough fashion in breeding programs.

Figures

Figure 2.1.



Appendices Captions

Appendix 2.1. Crops of putative/confirmed hybrid ancestry.

Appendices

Appendix 2.1.

Species	Common Name	Family	Mode of Hybridization[1]	Confirmed or Putative[2]	Parental species	Ploidy and Chromosome Count	References
Abelmoschus esculentus (L.) Moench	Okra	Malvaceae	ALLO	P	Uncertain	Polyploid (tetraploid) usually 2n=4x=130 Variable ploidy	Joshi and Hardas, 1956; Schafleitner et al., 2013
Actinidia deliciosa (A. Chev.) C.F.Liang & A.R.Ferguson	Kiwifruit	Actinidiaceae	ALLO	P	*Actinidia chinensis* Planch. and Unknown	Polyploid (hexaploid) 2n=6x=174	Atkinson et al., 1997
Agave fourcroydes Lem.	Henequen	Asparagaceae	ALLO	C	Uncertain	Polyploid (usu. pentaploid, triploid) 2n=5x(3x)=150(90) Variable ploidy, polyploid event not recent	Robert et al., 2008; Hughes et al., 2007
Agave sisalana Perrine	Sisal	Asparagaceae	ALLO	C	Uncertain	Polyploid (usu. pentaploid, hexaploid) 2n=5x(6x)=150(180) Variable ploidy, polyploid event not recent	Robert et al., 2008
Allium ampeloprasum L.	Great headed garlic	Amaryllidaceae	INTERHY	P	*Allium ampeloprasum* L.	Homoploid	Guenaoui et al., 2013
Allium cepa L.	Common onion	Amaryllidaceae	INTERHY	P	Uncertain: *Allium vavilovii* Popov & Vved., *A. galanthum* Kar. & Kir. or *A. fistulosum* L.	Homoploid	Gurushidze et al., 2007
Allium cornutum Clementi	Triploid onion	Amaryllidaceae	ALLO	C	*Allium cepa* L., *A. roylei* Stearn, unknown	Triparental Polyploid (triploid) 2n=3x=24	Fredotovic et al., 2014
Ananas comosus (L.) Merr is	Pineapple	Bromeliaceae	INTERIN	P	*Ananas ananassoides* (Baker) L.B. Smith	Homoploid	Duval et al., 2003
Annona x *atemoya*	Atemoya	Annonaceae	INTERHY	C	*Annona cherimola* Mill. and *A. squamosa* L.	?	Perfectti et al., 2004; Jalikop, 2010
Arachis hypogaea L.	Peanut	Fabaceae	ALLO	C	*Arachis duranensis* Krapov. & W.C. Greg. and *A. ipaënsis* Krapov. & W.C. Greg.	Polypoid (tetraploid) 2n=4x=40	Kochert et al., 1996; Bertioli et al., 2011

Armoracia rusticana P.Gaertn ., B.Mey. & Scherb.	Horseradish	Brassicaceae	INTERHY	P	Uncertain	?	Courter and Rhodes, 1969
Artocarpus altilis (Parkinson ex F.A.Zorn) Fosberg	Breadfruit	Moraceae	INTERIN	P	*Artocarpus mariannensis* Trécul	?	Zerega et al., 2005; Jones et al., 2013
Avena sativa L.	Oat	Poaceae	ALLO	C	Uncertain	Polyploid (hexaploid) 2n=6x=42	Linares et al., 1998; Oliver et al., 2013
Brassica carinata A.Braun	Ethiopian mustard	Brassicaceae	ALLO	C	*Brassica oleracea* L. and *B. nigra* (L.) K.Koch	Polyploid (tetraploid) 2n=4x=19	Arias et al., 2014
Brassica juncea (L.) Czern.	Indian mustard	Brassicaceae	ALLO	C	*Brassica nigra* (L.) K.Koch and *B. rapa* L.	Polyploid (tetraploid) 2n=4x=18	Arias et al., 2014
Brassica napus L.	Rapeseed, Rutabega	Brassicaceae	ALLO	C	*Brassica rapa* L. and *B. oleracea* L.	Polyploid (tetraploid) 2n=4x=19	Arias et al., 2014
Cajanus cajan (L.) Millsp.	Pigeon Pea	Fabaceae	INTERIN, INTRAIN	P	Wild *Cajanus cajan* and other species	Homoploid	Kassa et al., 2012
Cannabis sativa L.	Hemp	Cannabaceae	INTRAIN	P	*Cannabis sativa* L. 'Indica' and 'Sativa' types	Homoploid	de Meijer and van Soest, 1992
Carica pentagona Heilborn	Babaco	Caricaceae	INTERHY	C	Uncertain (*Carica stipulata* V.M.Badillo, *Vasconcellea pubescens* A.DC., *Vasconcellea weberbaueri* (Harms) V.M. Badillo)	?	Van Droogenbroeck et al., 2002; Van Droogenbroeck et al., 2006
Carya illinoinensis (Wangenh.) K.Koch	Pecan	Juglandaceae	INTERHY	P	Uncertain	?	Grauke et al., 2011
Castanea dentata (Marshall) Borkh	Chestnut	Fagaceae	INTERIN	C	*Castanea pumila* (L.) Mill.	Homoploid Also ongoing efforts to introgress blight resistance from *Castanea mollissima* Blume (see Jacobs et al., 2013)	Li and Dane, 2013
Castanea sativa Mill.	Chestnut	Fagaceae	INTERIN	C	*Castanea sativa* Eurosiberian and Mediterranean populations	Homoploid	Villani et al., 1999; Mattioni et al., 2013
Chenopodium quinoa Willd.	Quinoa	Chenopodiaceae	ALLO	P	Uncertain	Polyploid (tetraploid)	Heiser, 1974; Ward, 2000; Maughan et al., 2004
Cicer arietinum L.	Chickpea (pea-shaped)	Fabaceae	INTRAHY	P	Cicer arietinum L. Desi and Kabuli Germplasm	?	Upadhyaya et al., 2008; Keneni et al., 2011
Cichorium intybus L.	Radicchio	Asteraceae	INTERIN	C	Wild *Cichorium intybus* L.	Homoploid	Kiaer et al., 2009

Citrus aurantiifolia (Christm.) Swingle	Key lime	Rutaceae	INTERHY	C	*Citrus medica* L. and *C.* subg. *Papeda*	?	Ollitrault and Navarro, 2012; Penjor et al., 2014; Nicolosi et al., 2000; Moore, 2001
Citrus aurantium L.	Sour oranges	Rutaceae	INTERHY	C	*Citrus maxima* (Burm.) and *C. reticulata* Blanco	?	Wu et al., 2014; Moore, 2001
Citrus clementina hort.	Clementine	Rutaceae	INTERHY	C	*Citrus sinensis* (L.) Osbeck and *C. reticulata* Blanco	?	Wu et al., 2014
Citrus limon (L.) Osbeck	Lemon, lime	Rutaceae	INTERHY	C	*Citrus medica* L., *C. aurantiifolia* (Christm.) Swingle, and uncertain	?	Nicolosi et al., 2000; Moore, 2001
Citrus paradisi Macfad.	Grapefruit	Rutaceae	INTERHY	C	*Citrus sinensis* (L.) Osbeck and *C. maxima* (Burm.)	?	Wu et al., 2014; Moore, 2001
Citrus reticulata Blanco	Mandarin	Rutaceae	INTERIN	C	*Citrus maxima* (Burm.)	?	Wu et al., 2014
Citrus sinensis (L.) Osbeck	Sweet orange (blood, common)	Rutaceae	INTERHY	C	Uncertain	?	Wu et al., 2014; Moore, 2001
Cocos nucifera L.	Coconut	Arecaceae	INTRAIN	C	*Cocos nucifera* L. Indo-Atlantic and Pacific lineages	Homoploid	Gunn et al., 2011
Coffea arabica L.	Coffee	Rubiaceae	ALLO	C	*Coffea eugenioides* S.Moore and *C. canephora* Pierre ex A.Froehner	Polyploid (tetraploid) 2n=4x=44	Lashermes et al., 1999
Corylus avellana L.	Hazelnut	Betulaceae	INTRAIN	C	Wild *Corylus avellana* L. in Southern Europe	Homoploid	Campa et al., 2011; Boccacci et al., 2013
Cucurbita pepo L.	Winter Squash, Pumpkin	Cucurbitaceae	INTRAIN	P	*Cucurbita pepo* var. *texana* (Scheele) D.S.Decker	Homploid	Kirkpatrick and Wilson, 1988
Daucus carota subsp. *sativus* (Hoffm.) Arcang.	Carrot	Apiaceae	INTRAIN	C	*Daucus carota* L. subsp. *carota*	Homoploid	Iorizzo et al., 2013; Simon, 2000
Dioscorea L. spp.	Yam	Dioscoreaceae	INTERHY, INTROG	P	Uncertain	Variable Multiple species of putative hybrid (perhaps allopolyploid) origin including *Dioscorea cayennensis* subsp. *rotundata* (Poir.) J.Miège. and *D. cayennensis* Lam.	Terauchi et al., 1992; Dansi et al., 1999; Bhattacharjee et al., 2011; Mignouna et al., 2002
Diospyros kaki L.f.	Persimmon	Ebenaceae	ALLO	P	Uncertain	Polyploid (hexaploid)	Yonemori et al., 2008

Ficus carica L.	Fig	Moraceae	INTERIN	P	Uncertain	?	Aradhya et al., 2010
Fragaria ananassa (Duchesne ex Weston) Duchesne ex Rozier	Strawberries	Rosaceae	INTERHY	C	*Fragaria virginiana* Mill. (octoploid), *F. chiloensis* (L.) Mill. (octoploid)	Homoploid relative to parentals (octoploid) 2n=8x=56 Uncertain which species formed the octoploid progenitors	Evans, 1977; Hirakawa et al., 2014
Garcinia mangostana L.	Mangosteen	Clusiaceae	ALLO	P	*Garcinia celebica* L. and *G. malaccensis* Hook.f. ex T.Anderson	Polyploid (tetraploid) Recent work shows this may not be of hybrid origin (Nazre, 2014)	Richards, 1990
Gossypium hirsutum L.	Upland Cotton	Malvaceae	?	C	Uncertain, referred to as 'A' and 'D'	Polyploid (formed <1MYA) 2n =4x=52 Polyploidization likely led to agronomically significant traits (Applequist et al., 2001)	Wendel and Cronn 2003
Hibiscus sabdariffa L.	Roselle	Malvaceae	ALLO	P	Uncertain	Polyploid (tetraploid) 2n=4x=72	Menzel and Wilson, 1966; Satya et al., 2013
Hordeum vulgare L.	Barley	Poaceae	INTROG	C	*Hordeum spontaneum* K.Koch	Homoploid	Badr et al., 2000; Dai et al., 2012
Humulus lupulus L.	Hops	Cannabaceae	INTRAIN	C	*Humulus lupulus* L. North American and European Germplasm	Homoploid	Reeves and Richards, 2011; Stajner et al., 2008; Seefelder et al., 2000
Ipomoea batatas (L.) Lam.	Sweet Potato	Convolvulaceae	INTRAIN; INTERIN?	P	*Ipomoea batatas* (L.) Lam. Central American and South American Germplasm	Homoploid relative to parentals	Roullier et al., 2013
Juglans regia L.	Walnut	Juglandaceae	INTERHY	C	*Juglans sigillata* Dode	Homoploid	Gunn et al., 2010
Lactuca sativa L.	Lettuce	Asteraceae	INTRAHY	P	*Lactuca serriola* L. and other *L.* spp.	Homoploid	de Vries, 1997
Lagenaria siceraria (Molina) Standl.	Bottle Gourd	Cucurbitaceae	INTRAIN	C	*Lagenaria siceraria* (Molina) Standl. African/American and Asian Germplasm	Homoploid	Clarke et al., 2006
Lens culinaris Medik. ssp. *culinaris*	Lentil	Fabaceae	INTRAIN	P	Wild lentil, *Lens culinaris* subsp. *orientalis* (Boiss.) Ponert	Homoploid	Erskine et al., 2011

Macadamia integrifolia Maiden & Betche	Macadamia	Proteaceae	INTERHY, INTERIN	C	*Macadamia tetraphylla* L.A.S.Johnson, and other *M.* spp.	Homoploid	Hardner et al., 2009; Steiger et al., 2003; Aradhya et al., 1998
Malus domestica Borkh.	Apple	Rosaceae	INTERHY	C	*Malus sieversii* (Ledeb.) M.Roem., *M. sylvestris* (L.) Mill., and possibly others	Homoploid	Cornille et al., 2012
Mentha piperita L.	Peppermint	Lamiaceae	ALLO	C	*Mentha aquatica* L. and *M. spicata* L.	Polyploid (12-ploid) 2n=12x=66 or 72	Harley and Brighton, 1977; Gobert et al., 2002
Musa paradisiaca L.	Banana	Musaceae	ALLO	C	*Musa acuminata* Colla, *M. balbisiana* Colla	Polyploid (usually triploid) 2n=3x=33	Simmonds and Shepherd, 1955; Heslop-Harrison and Schwarzacher, 2007; De Langhe et al., 2010
Nicotiana tabacum L.	Tobacco	Solanaceae	ALLO	C	Uncertain (*Nicotiana sylvestris* Speg. & S. Comes and *N. tomentosiformis* Goodsp.)	Polyploid (tetraploid) 2n=4x=48	Kenton et al., 1993; Murad et al., 2002
Olea europaea L.	Olive	Oleaceae	INTRAIN	P	Wild *Olea europaea* L., Eastern and Western Germplasm	Homoploid	Kaniewski et al., 2012; Besnard et al., 2013; Breton et al., 2006; Rubio de Casas et al., 2006; Besnard et al., 2007; Besnard et al., 2000
Opuntia L. spp.	Opuntia	Cactaceae	INTERHY, ALLO	P	Including *Opuntia ficus-indica* (L.) Mill.	Polyploid, homoploid	Hughes et al., 2007; Griffith, 2004
Oryza sativa L.	Rice	Poaceae	INTRAIN, INTERIN	P	*Oryza sativa* L. 'Japonica' and 'Indica' Germplasm, *Oryza rufipogon* Griff.	Homoploid	Caicedo et al., 2007; Gao and Innan, 2008
Oxalis tuberosa Molina	Oca	Oxalidaceae	ALLO	P	Uncertain	Polyploid (octaploid) 2n=8x=64	Emswhiller and Doyle, 2002; Emshwiller 2002; Emswhiller et al., 2009
Pennisetum glaucum (L.) R.Br.	Pearl Millet	Poaceae	INTRAIN	C	Wild *Pennisetum glaucum* (L.) R.Br.	Homoploid	Oumar et al., 2008
Persea americana Mill.	Avocado (Hass and other cultivars)	Lauraceae	INTRAIN	C	*Persea americana* Mill. 'Guatamalensis', 'Drymifolia', and' Americana'	Homoploid	Chen et al., 2008; Davis et al., 1998; Ashworth and Clegg, 2003; Douhan et al., 2011

Phoenix dactylifera L.	Date palm	Arecaceae	INTERHY	P	Uncertain	Homoploid	El Hadrami et al., 2011; Bennaceur et al., 1991
Piper methysticum G.Forst.	Kava	Piperaceae	ALLO	P	*Piper wichmannii* C. DC. and *P. gibbiflorum* C.DC.	Polyploid (decaploid) 2n=10x=130	Singh, 2004; Lebot et al., 1991
Pistacia vera L.	Pistachio	Anacardiaceae	INTERIN	P	*Pistacia atlantic*a Desf., *P. chinensis* subsp. *integerrima* (J. L. Stewart ex Brandis) Rech. f.	Homoploid	Kafkas et al., 2001
Pisum abyssinicum A.Braun	Pea	Fabaceae	INTERHY	C	Uncertain (*Pisum fulvum* Sibth. & Sm. and other *P. spp.*)	Homoploid	Vershinin et al., 2003
Pisum sativum L.	Pea	Fabaceae	INTERHY	C	Uncertain (*Pisum sativum* subsp. *elatius* (M.Bieb.) Asch. & Graebn. and other *P. spp.*)	Homoploid	Vershinin et al., 2003
Plinia cauliflora (Mart.) Kausel	Jaboticaba	Myrtaceae	Intraspecific hybridization, INTERHY	P	*Plinia* 'Jaboticaba' and 'Cauliflora' Germplasm; *P. peruviana* (Poir.) Govaerts	Homoploid	Balerdi et al., 2006
Prunus cerasus L.	Cherry	Rosaceae	ALLO	C	*Prunus avium* (L.) L. and *P. fruticosa* Pall.	Polyploid (tetraploid) 2n=4x=32	Tavaud et al., 2004; Olden and Nybom, 1968
Prunus domestica L.	Plum	Rosaceae	ALLO	C	Uncertain (P. cerasifera Ehrh. and P. spinosa L.) Japanese Plum is also of hybrid origin (see Hartmann and Neumuller 2009). Also hybridizes with other cultivated *Prunus* spp.	Polyploid (hexaploid) 2n=6x=48	Zohary, 1992; Hartmann and Neumuller, 2009
Prunus dulcis (Mill.) D.A.Webb	Almond	Rosaceae	INTERIN	C	*Prunus orientalis* (Mill.) Koehne and other *P.* spp.	Homoploid	Delplancke et al., 2012; Delplancke et al., 2013
Pyrus L. species	Pear	Rosaceae	INTERHY	C	Many species, Also introgression with semidomesticated populations (see Iketani et al. 2009)	Homoploid	Silva et al., 2014
Raphanus raphanistrum subsp. *sativus* (L.) Domin	Radish	Brassicaceae	INTRAIN	C	*Raphanus raphanistrum* L. subsp. *raphanistrum*	Homoploid	Ridley et al., 2008

Rheum L. cultivated species	Rhubarb	Polygonaceae	INTERHY	P	Unclear,	Homoploid relative to parentals (tetraploid) Hybrids include: *Rheum rhaponticum* L., *R. rhabarbarum* L., *R. palmatum* L.	Foust and Marshall, 1991; Kuhl and Deboer, 2008
Rubus L. spp.	Red raspberry, Blackberry, Tayberry, Boysenberry, etc.	Rosaceae	ALLO, INTERHY	C	Many	Polyploid	Alice et al., 2014; Alice and Campbell, 1999
Saccharum spp.	Sugarcane	Poaceae	ALLO	C	*Saccharum officinarum* L. and *S. spontaneum* L.	Polyploid Variable, 2n=10-13x=100-130	Grivet et al., 1995; D'Hont et al., 1996
Secale cereale L.	Rye	Poaceae	INTERHY	C	Uncertain (*Secale montanum* Guss., *S. vavilovii* Grossh.)	Homoploid	Bartos et al., 2008; Korzun et al., 2001; Hillman, 1978; Tang et al., 2011; Salamini et al., 2002
Sechium edule (Jacq.) Sw.	Chayote	Cucurbitaceae	INTERIN	P	*Sechium compositum* (Donn. Sm.) C. Jeffrey	Homoploid	Newstrom, 1991
Setaria italica (L.) P.Beauv.	Foxtail millet	Poaceae	INTERIN	C	*Setaria viridis* (L.) P.Beauv.	Homoploid	Till-Bottraud et al., 1992
Solanum L. spp. Section *Petota*	Potato	Solanaceae	INTERHY, ALLO, INTERIN	C	Including *Solanum tuberosum* L., *S. ajanhuiri* Juz. & Bukasov, *S. curtilobum* Juz. & Bukasov, *S. juzepczukii* Bukasov	Homoploid and Polyploid	Rodriguez et al., 2010
Solanum lycopersicum L.	Tomato	Solanaceae	INTRAIN, INTERIN	C	*Solanum lycopersicum* var. *cerasiforme* (Dunal) D.M. Spooner, G.J. Anderson & R.K. Jansen and *S. pimpinellifolium* L.	Homoploid	Blanca et al., 2012; Causse et al., 2013; Rick, 1950
Solanum melongena L.	Eggplant	Solanaceae	INTERHY, INTERIN, INTRAIN	C	*Solanum undatum* Lam. and others; wild *S. melongena* L. (=*S. insanum* L.)	Homoploid Hybrid origin is not confirmed, but introgression is well documented	Knapp et al., 2013; Meyer et al., 2012
Solanum muricatum Aiton	Pepino dulce	Solanaceae	INTERHY, INTERIN	Likely	*Solanum* species in Series *Caripensia*	Homoploid, Polyphyletic origin and extensive, ongoing introgression with wild species	Blanca et al., 2007

Spondias purpurea L.	Jocote	Anacardiaceae	INTERIN	P	*Spondias mombin* L.	Homoploid	Miller and Schaal, 2005
Theobroma cacao L.	Cacao (Trinitario-type)	Malvaceae	INTRAHY	C	*Theobroma cacao* L. 'Forastero' and 'Criollo' Germplasm	Homoploid	Yang et al., 2013
Triticum aestivum L.	Bread Wheat, Spelt	Poaceae	ALLO	C	*Triticum turgidum* L. (tetraploid) with *Aegilops tauschii* Coss.	Polyploid (hexaploid) 2n=6x=42	Matsuoka, 2011; Dvorak, 2012
Triticum turgidum L.	Emmer Wheat, Durum Wheat	Poaceae	ALLO	C	*Triticum urartu* Thumanjan ex Gandilyan and *Aegilops speltoides* Tausch	Polyploid (tetraploid) 2n=4x=28	Dvorak et al., 2012; Matsuoka, 2011; Yamane and Kawahara, 2005
Vaccinium corymbosum L.	Highbush Blueberry	Ericaceae	INTERHY, INTERIN	P	*Vaccinium tenellum* Aiton, *V. darrowii* Camp, (*V. virgatum* Aiton, *V. angustifolium* Aiton)	Uncertain Possible hybrid origin during the Plesitocene	Vander Kloet, 1980; Bruederle et al., 1994; Lyrene et al., 2003; Boches et al., 2006
Vanilla tahitensis J.W. Moore	Tahitian vanilla	Orchidaceae	ALLO	C	*Vanilla planifolia* Jacks. ex Andrews and *V. odorata* C.Presl	Polyploid Variable, 2n=2x(4x)=32(64)	Lubinsky et al., 2008
Vitis rotundifolia Michx.	Grape	Vitaceae	INTERIN	C	Many*Vitis* spp.	Homoploid	Reisch et al., 2012; This et al., 2006
Zea mays L.	Maize	Poaeceae	INTRAIN	C	Wild *Zea mays* L. (teosinte, =subsp. *parviglumis* Iltis & Doebley)	Homoploid	Van Heerwaarden et al., 2011; Hufford et al., 2013

[1] ALLO = Allopolyploid, INTERHY = Interspecific hybrid, INTERIN = Interspecific introgression, INTRAHY = Intraspecific hybrid, INTRAIN = Intraspecific introgression
[2] C = Confirmed, P = Putative

Appendices Literature Cited

Appendix 2.1 Literature Cited

Alice, L.A., and C.S. Campbell. 1999. Phylogeny of *Rubus* (Rosaceae) based on nuclear ribosomal DNA internal transcribed spacer sequences. *American Journal of Botany* 86: 81–97.

Alice, L.A., T. Eriksson, B. Eriksen, and C.S. Campbell. 2014. Hybridization and gene flow between distantly related species of *Rubus* (Rosaceae): Evidence from nuclear ribosomal dna internal transcribed spacer region sequences. *Systematic Botany* 26: 769–778.

Applequist, W.L., R. Cronn, and J.F. Wendel. 2001. Comparative development of fiber in wild and cultivated cotton. *Evolution & Development* 3: 3–17.

Aradhya, M.K., E. Stover, D. Velasco, and A. Koehmstedt. 2010. Genetic structure and differentiation in cultivated fig (*Ficus carica* L.). *Genetica* 138: 681–94.

Aradhya, M.K., L.K. Yee, F.T. Zee, and R.M. Manshardt. 1998. Genetic variability in Macadamia. *Genetic Resources and Crop Evolution* 45: 19–32.

Arias, T., M.A. Beilstein, M. Tang, M.R. McKain, and J.C. Pires. 2014. Diversification times among *Brassica* (Brassicaceae) crops suggest hybrid formation after 20 million years of divergence. *American Journal of Botany* 101: 86–91.

Ashworth, V.E.T.M., and M.T. Clegg. 2003. Microsatellite markers in avocado (*Persea americana* Mill.): Genealogical relationships among cultivated avocado genotypes. *Journal of Heredity* 94: 407–415.

Atkinson, R.G., G. Cipriani, D.J. Whittaker, and R.C. Gardner. 1997. Systematics and Evolution The allopolyploid origin of kiwifruit, *Actinidia deliciosa* (Actinidiaceae). *Plant Systematics and Evolution* 205: 111–124.

Badr, A., K. Müller, R. Schäfer-Pregl, H. El Rabey, S. Effgen, H.H. Ibrahim, C. Pozzi, et al. 2000. On the origin and domestication history of barley *(Hordeum vulgare)*. *Molecular Biology and Evolution* 17: 499–510.

Balerdi, C.F., R. Rafie, and J. Crane. 2006. Jaboticaba (*Myrciaria cauliflora*, Berg.) a delicious fruit with an excellent market potential. *Proceedings of the Florida State Horticultural Society* 119: 66–68.

Bartos, J., E. Paux, R. Kofler, M. Havránková, D. Kopecký, P. Suchánková, J. Safár, et al. 2008. A first survey of the rye *(Secale cereal*e) genome composition through BAC end sequencing of the short arm of chromosome 1R. *BMC Plant Biology* 8: 95.

Bennaceur, M., C. Lanaud, M.H. Chevallier, and N. Bounaga. 1991. Genetic diversity of the date palm (*Phoenix dactylifera* L.) from Algeria revealed by enzyme markers. *Plant Breeding* 107: 56–69.

Bertioli, D.J., G. Seijo, F.O. Freitas, J.F.M. Valls, S.C.M. Leal-Bertioli, and M.C. Moretzsohn. 2011. An overview of peanut and its wild relatives. *Plant Genetic Resources* 9: 134–149.

Besnard, G., and A. Bervillé. 2000. Multiple origins for Mediterranean olive (*Olea europaea* L. ssp. *europaea*) based upon mitochondrial DNA polymorphisms. *Molecular Biology and Genetics* 323: 173–181.

Besnard, G., B. Khadari, M. Navascués, A. El Bakkali, N. Arrigo, V.B. De Caraffa, S. Santoni, et al. 2013. The complex history of the olive tree: from Late Quaternary diversification of Mediterranean lineages to primary domestication in the northern Levant. *Proceedings of the Royal Society B: Biological Sciences* 280: 20122833.

Besnard, G., R. Rubio de Casas, and P. Vargas. 2007. Plastid and nuclear DNA polymorphism reveals historical processes of isolation and reticulation in the olive tree complex (*Olea europaea*). *Journal of Biogeography* 34: 736–752.

Bhattacharjee, R., M. Gedil, A. Sartie, E. Otoo, D. Dumet, H. Kikuno, P.L. Kumar, and R. Asiedu. 2011. *Dioscorea*. *In* C. Kole [ed.], Wild crop relatives: Genomic and breeding resources, 71–96. Springer-Verlag, Berlin, Heidelberg.

Blanca, J., J. Cañizares, L. Cordero, L. Pascual, M.J. Diez, and F. Nuez. 2012. Variation revealed by SNP genotyping and morphology provides insight into the origin of the tomato. *PloS One* 7: e48198.

Blanca, J.M., J. Prohens, G.J. Anderson, E. Zuriaga, J. Canizares, and F. Nuez. 2007. AFLP and DNA sequence variation in an andean domesticate, Pepino (*Solanum muricatum*, Solanaceae): Implications for evolution and domestication. *American Journal of Botany* 94: 1219–1229.

Boccacci, P., M. Aramini, N. Valentini, L. Bacchetta, M. Rovira, P. Drogoudi, a. P. Silva, et al. 2013. Molecular and morphological diversity of on-farm hazelnut (*Corylus avellana* L.) landraces from southern Europe and their role in the origin and diffusion of cultivated germplasm. *Tree Genetics & Genomes* 9: 1465–1480.

Boches, P., N. V Bassil, and L. Rowland. 2006. Genetic diversity in the highbush blueberry evaluated with microsatellite markers. *Journal of the American Horticultural Society* 131: 674–686.

Breton, C., M. Tersac, and A. Bervillé. 2006. Genetic diversity and gene flow between the wild olive (oleaster, *Olea europaea* L.) and the olive: several Plio-Pleistocene

refuge zones in the Mediterranean basin suggested by simple sequence repeats analysis. *Journal of Biogeography* 33: 1916–1928.

Bruederle, L.P. and N. Vorsa. 1994. Genetic Differentiation of Diploid Blueberry, *Vaccinium* sect. Cyanococcus (Ericaceae). *Systematic Botany* 19: 337–349.

Caicedo, A.L., S.H. Williamson, R.D. Hernandez, A. Boyko, A. Fledel-Alon, T.L. York, N.R. Polato, et al. 2007. Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genetics* 3: 1745–1756.

Campa, A., N. Trabanco, E. Pérez-Vega, M. Rovira, and J.J. Ferreira. 2011. Genetic relationship between cultivated and wild hazelnuts (*Corylus avellana* L.) collected in northern Spain. *Plant Breeding* 130: 360–366.

Causse, M., N. Desplat, L. Pascual, M.-C. Le Paslier, C. Sauvage, G. Bauchet, A. Bérard, et al. 2013. Whole genome resequencing in tomato reveals variation associated with introgression and breeding events. *BMC Genomics* 14: 791.

Chen, H., P.L. Morrell, V.E.T.M. Ashworth, M. de la Cruz, and M.T. Clegg. 2008. Tracing the geographic origins of major avocado cultivars. *The Journal of Heredity* 100: 56–65.

Clarke, A.C., M.K. Burtenshaw, P.A. Mclenachan, D.L. Erickson, and D. Penny. 2006. Proceedings of the SMBE Tri-National Young Investigators' Workshop 2005: Reconstructing the Origins and Dispersal of the Polynesian Bottle Gourd *(Lagenaria siceraria)*. *Molecular Biology and Evolution* 23: 893–900.

Cornille, A., P. Gladieux, M.J.M. Smulders, I. Roldán-Ruiz, F. Laurens, B. Le Cam, A. Nersesyan, et al. 2012. New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genetics* 8: e1002703.

Courter, J.W., and A.M. Rhodes. 1969. Historical Notes on Horseradish. *Economic Botany* 186: 156–164.

D'Hont, A., L. Grivet, P. Feldmann, S. Rao, N. Berding, and J.C. Glaszmann. 1996. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Molecular & General Genetics* 250: 405–13.

Dai, F., E. Nevo, D. Wu, J. Comadran, M. Zhou, L. Qiu, Z. Chen, et al. 2012. Tibet is one of the centers of domestication of cultivated barley. *Proceedings of the National Academy of Sciences of the United States of America* 109: 16969-16973.

Dansi, A., H.D. Mignouna, J. Zoundjih, A. Sangare, R. Asiedu, and F.M. Quin. 1999. Morphological diversity, cultivar groudps and possible descent in the cultivated yams

(Dioscorea cayenensis/D. rotundata) complex in Benin Republic. *Genetic Resources and Crop Evolution* 46: 371–388.

Davis, J., D. Henderson, M. Kobayashi, M.T. Clegg, T. Michael, and P.C.K. Allen. 1998. Genealogical Relationships Among Cultivated Avocado as Revealed Through RFLP Analyses. *The Journal of Heredity* 89: 319–323.

Delplancke, M., N. Alvarez, L. Benoit, a Espíndola, H. I Joly, S. Neuenschwander, and N. Arrigo. 2013. Evolutionary history of almond tree domestication in the Mediterranean basin. *Molecular Ecology* 22: 1092–1104.

Delplancke, M., N. Alvarez, A. Espíndola, H. Joly, L. Benoit, E. Brouck, and N. Arrigo. 2012. Gene flow among wild and domesticated almond species: insights from chloroplast and nuclear markers. *Evolutionary Applications* 5: 317–329.

Douhan, G.W., E. Fuller, B. McKee, and E. Pond. 2011. Genetic diversity analysis of avocado *(Persea americana* Miller) rootstocks selected under greenhouse conditions for tolerance to phytophthora root rot caused by *Phytophthora cinnamomi*. *Euphytica* 182: 209–217.

Van Droogenbroeck, B., P. Breyne, P. Goetghebeur, E. Romeijn-Peeters, T. Kyndt, and G. Gheysen. 2002. AFLP analysis of genetic relationships among papaya and its wild relatives (Caricaceae) from Ecuador. *Theoretical and Applied Genetics* 105: 289–297.

Van Droogenbroeck, B., T. Kyndt, E. Romeijn-Peeters, W. Van Thuyne, P. Goetghebeur, J.P. Romero-motochi, and G. Gheysen. 2006. Evidence of natural hybridization and introgression between *Vasconcellea* species (Caricaceae) from southern Ecuador revealed by chloroplast, mitochondrial and nuclear DNA markers. *Annals of Botany* 97: 793–805.

Duval, M.F., G.S.C. Buso, F.R. Ferreira, J.L. Noyer, G.C. Eeckenbrugge, P. Hamon, and M.E. Ferreira. 2003. Relationships in *Ananas* and other related genera using chloroplast DNA restriction site variation. *Genome* 46: 990–1004.

Dvorak, J., K.R. Deal, M.-C. Luo, F.M. You, K. von Borstel, and H. Dehghani. 2012. The origin of spelt and free-threshing hexaploid wheat. *The Journal of Heredity* 103: 426–441.

Emshwiller, E. 2002. Ploidy Levels among species in the "*Oxalis tuberosa* Alliance" as Inferred by Flow Cytometry. *Annals of Botany* 89: 741–753.

Emshwiller, E., T. Theim, A. Grau, V. Nina, and F. Terrazas. 2009. Origins of domestication and polyploidy in oca (*Oxalis tuberosa*; Oxalidaceae). 3. AFLP data of oca and four wild, tuber-bearing taxa. *American Journal of Botany* 96: 1839–48.

Emshwiller, E., and J.J. Doyle. 2002. Origins of Domestication and Polyploidy in Oca (*Oxalis tuberosa*: Oxalidaceae). 2. Chloroplast-expressed glutamine synthetase data. *American Journal of Botany* 89: 1042–1056.

Erskine, W., A. Sarker, and M. Ashraf. 2011. Reconstructing an ancient bottleneck of the movement of the lentil (*Lens culinaris* ssp. *culinaris*) into South Asia. *Genetic Resources and Crop Evolution* 58: 373–381.

Evans, W.D. 1977. The use of synthetic octoploids in strawberry breeding. *Euphytica* 26: 497–503.

Foust, C.M., and D.E. Marshall. 1991. Culinary Rhubarb Production in North America: History and Recent Statistics. *HortScience* 16: 1360–1363.

Fredotovic, Z., I. Samanic, H. Weiss-Schneeweiss, J. Kamenjarin, T.-S. Jang, and J. Puizina. 2014. Triparental origin of triploid onion, *Allium* × *cornutum* (Clementi ex Visiani, 1842), as evidenced by molecular, phylogenetic and cytogenetic analyses. *BMC plant biology* 14: 24.

Gao, L.-Z., and H. Innan. 2008. Nonindependent domestication of the two rice subspecies, *Oryza sativa* ssp. *indica* and ssp. *japonica*, demonstrated by multilocus microsatellites. *Genetics* 179: 965–976.

Gobert, V., S. Moja, M. Colson, and P. Taberlet. 2002. Hybridization in the section *Mentha* (Lamiaceae) inferred from AFLP markers. *American Journal of Botany* 89: 2017–2023.

Grauke, L.J., M.A. Mendoza-Herrera, A.J. Miller, and B.W. Wood. 2011. Geographic patterns of genetic variation in native pecans. *Tree Genetics & Genomes* 7: 917–932.

Griffith, M.P. 2004. The origins of an important cactus crop, *Opuntia ficus-indica* (Cactaceae): New molecular evidence. *American Journal of Botany* 91: 1915–1921.

Grivet, L., A.D. Hont, D. Roques, P. Feldmann, C. Lanaud, and J.C. Glaszmann. 1995. RFLP Mapping i n Cultivated Sugarcane. *Genetics* 142: 987–1000.

Guenaoui, C., S. Mang, G. Figliuolo, and M. Neffati. 2013. Diversity in *Allium ampeloprasum*: from small and wild to large and cultivated. *Genetic Resources and Crop Evolution* 60: 97–114.

Gunn, B.F., M. Aradhya, J.M. Salick, A.J. Miller, Y. Yongping, L. Lin, and H. Xian. 2010. Genetic variation in walnuts (*Juglans regia* and *J. sigillata*; Juglandaceae): Species distinctions, human impacts, and the conservation of agrobiodiversity in Yunnan, China. *American Journal of Botany* 97: 660–71.

Gunn, B.F., L. Baudouin, and K.M. Olsen. 2011. Independent origins of cultivated coconut (*Cocos nucifera* L.) in the old world tropics. *PloS One* 6: e21143.

Gurushidze, M., S. Mashayekhi, F.R. Blattner, N. Friesen, and R.M. Fritsch. 2007. Phylogenetic relationships of wild and cultivated species of *Allium* section *Cepa* inferred by nuclear rDNA ITS sequence analysis. *Plant Systematics and Evolution* 269: 259–269.

El Hadrami, A., F. Daayf, and I. El Hadrami. 2011. Date Palm Genetics and Breeding. S. M. Jain, J. M. Al-Khayri, and D. V. Johnson [eds.], Date Palm Biotechnology. Springer, Dordrecht, Netherlands.

Hardner, C.M., C. Peace, A.J. Lowe, J. Neal, P. Pisanu, M. Powell, A. Schmidt, et al. 2009. Genetic Resources and Domestication of Macadamia. *In* J. Janick [ed.], Horticultural Reviews, Volume 35, 1–126. John Wiley & Sons, Inc.

Harley, R.M., and C.A. Brighton. 1977. Chromosome numbers in the genus *Mentha* L . *Botanical Journal of the Linnean Society* 74: 71–96.

Hartmann, W., and M. Neumuller. 2009. Plum Breeding. *In* S. M. Jain, and P. M. Priyadarshan [eds.], Breeding Plantation Tree Crops: Temperate Species, 161–231. Springer New York, New York, NY.

Van Heerwaarden, J., J. Doebley, W.H. Briggs, J.C. Glaubitz, M.M. Goodman, J. de Jesus Sanchez Gonzalez, and J. Ross-Ibarra. 2011. Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proceedings of the National Academy of Sciences of the United States of America* 108: 1088–1092.

Heiser, B. 1974. On the origin of the cultivated chenopods (Chenopodium). *Genetics* 78: 503–505.

Heslop-Harrison, J.S., and T. Schwarzacher. 2007. Domestication, genomics and the future for banana. *Annals of botany* 100: 1073–84.

Hillman, G. 1978. On the Origins of Domestic Rye – *Secale cereale*: The Finds from Aceramic Can Hasan III in Turkey. *Anatolian Studies* 28: 157-174.

Hirakawa, H., K. Shirasawa, S. Kosugi, K. Tashiro, S. Nakayama, M. Yamada, M. Kohara, et al. 2014. Dissection of the octoploid strawberry genome by deep sequencing of the genomes of *Fragaria* species. *DNA Research* 21: 169–181.

Hufford, M.B., P. Lubinksy, T. Pyhajarvi, M.T. Devengenzo, N.C. Ellstrand, and J. Ross-Ibarra. 2013. The genomic signature of crop-wild introgression in maize. *PLoS Genetics* 9: e1003477.

Hughes, C.E., R. Govindarajulu, A. Robertson, D.L. Filer, S. a Harris, and C.D. Bailey. 2007. Serendipitous backyard hybridization and the origin of crops. *Proceedings of the National Academy of Sciences of the United States of America* 104: 14389–14394.

Iketani, H., T. Yamamoto, H. Katayama, C. Uematsu, N. Mase, and Y. Sato. 2010. Introgression between native and prehistorically naturalized (archaeophytic) wild pear (*Pyrus* spp.) populations in Northern Tohoku, Northeast Japan. *Conservation Genetics* 11: 115–126.

Iorizzo, M., D. a Senalik, S.L. Ellison, D. Grzebelus, P.F. Cavagnaro, C. Allender, J. Brunet, et al. 2013. Genetic structure and domestication of carrot (*Daucus carota* subsp. *sativus*) (Apiaceae). *American Journal of Botany* 100: 930–938.

Jacobs, D.F., H.J. Dalgleish, and C.D. Nelson. 2013. A conceptual framework for restoration of threatened plants: the effective model of American chestnut *(Castanea dentata)*. *New Phytologist* 197: 378–393.

Jalikop, S.H. 2010. Tree and fruit traits of progenies from the cross between *(Annona cherimola* Mill. × *A. squamosa* L.) × *A. reticulata* L. and approaches for the introgression of valuable genes from *A. reticulata* L. *Euphytica* 173: 161–171.

Jones, A.M.P., S.J. Murch, J. Wiseman, and D. Ragone. 2013. Morphological diversity in breadfruit (*Artocarpus*, Moraceae): insights into domestication, conservation, and cultivar identification. *Genetic Resources and Crop Evolution* 60: 175–192.

Joshi, A.B., and M.W. Hardas. 1956. Alloploid Nature of Okra, *Abelmoschus esculentus* (L.) Moench. *Nature* 178: 1190.

Kafkas, S., and R. Perl-Treves. 2001. Morphological and molecular phylogeny of *Pistacia* species in Turkey. *Theoretical and Applied Genetics* 102: 908–915.

Kaniewski, D., E. Van Campo, T. Boiy, J.-F. Terral, B. Khadari, and G. Besnard. 2012. Primary domestication and early uses of the emblematic olive tree: palaeobotanical, historical and molecular evidence from the Middle East. *Biological Reviews of the Cambridge Philosophical Society* 87: 885–99.

Kassa, M.T., R.V. Penmetsa, N. Carrasquilla-Garcia, B.K. Sarma, S. Datta, H.D. Upadhyaya, R.K. Varshney, et al. 2012. Genetic patterns of domestication in pigeonpea *(Cajanus cajan* (L.) Millsp.) and wild *Cajanus* relatives. *PloS one* 7: e39563.

Keneni, G., E. Bekele, M. Imtiaz, K. Dagne, E. Getu, and F. Assefa. 2011. Genetic Diversity and Population Structure of Ethiopian Chickpea *(Cicer arietinum* L.) Germplasm Accessions from Different Geographical Origins as Revealed by Microsatellite Markers. *Plant Molecular Biology Reporter* 30: 654–665.

Kenton, A., A.S. Parokonny, Y.Y. Gleba, and M.D. Bennett. 1993. Characterization of the *Nicotiana tabacum* L. genome by molecular cytogenetics. *Molecular & General Genetics* 240: 159–169.

Kiaer, L.P., F. Felber, A. Flavell, R. Guadagnuolo, D. Guiatti, T.P. Hauser, a. M. Olivieri, et al. 2009. Spontaneous gene flow and population structure in wild and cultivated chicory, *Cichorium intybus* L. *Genetic Resources and Crop Evolution* 56: 405–419.

Kirkpatrick, K.J., and H.D. Wilson. 1988. Interspecific gene flow in Cucurbita: *C. texana* vs. *C. pepo*: *American journal of Botany* 75: 519–527.

Vander Kloet, S.P. 1980. The taxonomy of the highbush blueberry, *Vaccinium corymbosum*. *Canadian Journal of Botany* 58: 1187–1201.

Knapp, S., M.S. Vorontsova, and J. Prohens. 2013. Wild relatives of the eggplant *(Solanum melongena* L.: Solanaceae): new understanding of species names in a complex group. *PloS one* 8: e57039.

Kochert, G., H.T. Stalker, M. Gimenes, L. Galgaro, C. Romero Lopes, and K. Moore. 1996. RFLP and Cytogenetic Evidence on the Origin and Evolution of Allotetraploid Domesticated Peanut, *Arachis hypogaea* (Leguminosae). *American Journal of Botany* 83: 1282–1291.

Korzun, V., S. Malyshev, and V. Voylokov. 2001. A genetic map of rye *(Secale cereale* L.*)* combining RFLP, isozyme, protein, microsatellite and gene loci. *Theoretical and Applied Genetics* 102: 709–717.

Kuhl, J.C., and V.L. Deboer. 2008. Genetic Diversity of Rhubarb Cultivars. *Journal of the American Horticultural Society* 133: 587–592.

De Langhe, E., E. Hribová, S. Carpentier, J. Dolezel, and R. Swennen. 2010. Did backcrossing contribute to the origin of hybrid edible bananas? *Annals of Botany* 106: 849–857.

Lashermes, P., M.C. Combes, J. Robert, P. Trouslot, A. D'Hont, F. Anthony, and A. Charrier. 1999. Molecular characterisation and origin of the *Coffea arabica* L. genome. *Molecular & General Genetics* 261: 259–266.

Lebot, V., M.K. Aradhya, and R.M. Manshardt. 1991. Geographic Survey of Genetic Variation in Kava *(Piper methysticum* Forst. f. and *P. wichmannii* C. DC.*)*. *Pacific Science* 45: 169–185.

Li, X., and F. Dane. 2013. Comparative chloroplast and nuclear DNA analysis of *Castanea* species in the southern region of the USA. *Tree Genetics & Genomes* 9: 107–116.

Linares, C., E. Ferrer, and A, Fominaya. 1998. Discrimination of the closely related A and D genomes of the hexaploid oat *Avena sativa* L. *Proceedings of the National Academy of Sciences of the United States of America* 95: 12450–12455.

Lubinsky, P., K.M. Cameron, M.C. Molina, M. Wong, S. Lepers-Andrzejewski, A. Gómez-Pompa, and S.-C. Kim. 2008. Neotropical roots of a Polynesian spice: the hybrid origin of Tahitian vanilla, *Vanilla tahitensis* (Orchidaceae). *American Journal of Botany* 95: 1040–1047.

Lyrene, P.M., N. Vorsa, and J.R. Ballington. 2003. Polyploidy and sexual polyploidization in the genus *Vaccinium*. *Euphytica* 133: 27–36.

Matsuoka, Y. 2011. Evolution of polyploid *Triticum* wheats under cultivation: the role of domestication, natural hybridization and allopolyploid speciation in their diversification. *Plant & Cell Physiology* 52: 750–764.

Mattioni, C., M.A. Martin, P. Pollegioni, M. Cherubini, and F. Villani. 2013. Microsatellite markers reveal a strong geographical structure in European populations of *Castanea sativa* (Fagaceae): evidence for multiple glacial refugia. *American Journal of Botany* 100: 951–961.

Maughan, P.J., a Bonifacio, E.N. Jellen, M.R. Stevens, C.E. Coleman, M. Ricks, S.L. Mason, et al. 2004. A genetic linkage map of quinoa *(Chenopodium quinoa)* based on AFLP, RAPD, and SSR markers. *Theoretical and Applied Genetics* 109: 1188–1195.

De Meijer, E.P.M., and L.J.M. van Soest. 1992. The CPRO *Cannabis* germplasm collection. *Euphytica* 62: 201–211.

Menzel, M.Y., and F.D. Wilson. 1966. Hybrids and Genome Relations of *Hibiscus sabdariffa, H. meeusei, H. radiatus* and *H. acetosella*. *American Journal of Botany* 53: 270–275.

Meyer, R.S., K.G. Karol, D.P. Little, M.H. Nee, and A. Litt. 2012. Phylogeographic relationships among Asian eggplants and new perspectives on eggplant domestication. *Molecular Phylogenetics and Evolution* 63: 685–701.

Mignouna, D., a. Mank, N. Ellis, N. Van Den Bosch, R. Asiedu, C. Ng, and J. Peleman. 2002. A genetic linkage map of Guinea yam *(Dioscorea rotundata* Poir.*)* based on AFLP markers. *Theoretical and Applied Genetics* 105: 716–725.

Miller, A., and B. Schaal. 2005. Domestication of a Mesoamerican cultivated fruit tree, *Spondias purpurea*. *Proceedings of the National Academy of Sciences of the United States of America* 102: 12801–6.

Moore, G. A. 2001. Oranges and lemons: clues to the taxonomy of *Citrus* from molecular markers. *Trends in Genetics* 17: 536–540.

Murad, L., K.Y. Lim, V. Christopodulou, R. Matyasek, C.P. Lichtenstein, A. Kovarik, and A.R. Leitch. 2002. The origin of tobacco's T genome is traced to a particular lineage within *Nicotiana tomentosiformis* (Solanaceae). *American Journal of Botany* 89: 921–928.

Nazre, M. 2014. New evidence on the origin of mangosteen *(Garcinia mangostana* L.*)* based on morphology and ITS sequence. *Genetic Resources and Crop Evolution* February.

Newstrom, L.E. 1991. Evidence for the origin of chayote, *Sechium edule* (Curcurbitaceae). *Economic Botany* 3: 410–428.

Nicolosi, E., Z.N. Deng, A. Gentile, S. La Malfa, G. Continella, and E. Tribulato. 2000. *Citrus* phylogeny and genetic origin of important species as investigated by molecular markers. *Theoretical and Applied Genetics* 100: 1155–1166.

Olden, E.J., and N. Nybom. 1968. On the origin of *Prunus cerasus* L. *Hereditas* 59: 327–345.

Oliver, R.E., N. a Tinker, G.R. Lazo, S. Chao, E.N. Jellen, M.L. Carson, H.W. Rines, et al. 2013. SNP discovery and chromosome anchoring provide the first physically-anchored hexaploid oat map and reveal synteny with model species. *PloS One* 8: e58068.

Ollitrault, P., and L. Navarro. 2012. Citrus. *In* M. L. Badenes, and D. H. Byrne [eds.], Fruit Breeding: Handbook of Plant Breeding, 623–662. Springer US, Boston, MA.

Oumar, I., C. Mariac, J.-L. Pham, and Y. Vigouroux. 2008. Phylogeny and origin of pearl millet *(Pennisetum glaucum* [L.] R. Br*)* as revealed by microsatellite loci. *Theoretical and Applied Genetics* 117: 489–497.

Penjor, T., T. Mimura, R. Matsumoto, M. Yamamoto, and Y. Nagano. 2014. Characterization of limes *(Citrus aurantifolia)* grown in Bhutan and Indonesia using high-throughput sequencing. *Scientific Reports* 4: 4853.

Perfectti, F., and L. Pascual. 2004. Geographic variation for isozymes in cherimoya *(Annona cherimola* Mill.*)*. *Genetic Resources and Crop Evolution* 51: 837–843.

Reeves, P.A., and C.M. Richards. 2011. Species delimitation under the general lineage concept: an empirical example using wild North American hops (Cannabaceae: *Humulus lupulus*). *Systematic Biology* 60: 45–59.

Reisch, B.I., C.L. Owens, and P.S. Cousins. 2012. Grape. *In* M. L. Badenes, and D. H. Byrne [eds.], Fruit Breeding, Handbook of Plant Breeding, 225–262. Springer US, Boston, MA.

Richards, A.J. 1990. Studies in *Garcinia*, dioecious tropical forest trees: the origin of the mangosteen *(G. mangostana* L.*)*. *Botanical Journal of the Linnean Society* 103: 301–308.

Rick, C.M. 1950. The Role of Natural Hybridization in the Derivation of Cultivated Tomatoes of Western South America.

Ridley, C.E., S.-C. Kim, and N.C. Ellstrand. 2008. Bidirectional history of hybridization in California wild radish, *Raphanus sativus* (Brassicaceae), as revealed by chloroplast DNA. *American Journal of Botany* 95: 1437–1442.

Robert, M.L., K.Y. Lim, L. Hanson, F. Sanchez-teyer, M.D. Bennett, A.R. Leitch, and I.J. Leitch. 2008. Wild and agronomically important Agave species (Asparagaceae) show proportional increases in chromosome number, genome size, and genetic markers with increasing ploidy. *Botanical Journal of the Linnean Society* 158: 215–222.

Rodríguez, F., M. Ghislain, A.M. Clausen, S.H. Jansky, and D.M. Spooner. 2010. Hybrid origins of cultivated potatoes. *Theoretical and Applied Genetics* 121: 1187–1198.

Roullier, C., A. Duputié, P. Wennekes, L. Benoit, V.M. Fernández Bringas, G. Rossel, D. Tay, et al. 2013. Disentangling the origins of cultivated sweet potato *(Ipomoea batatas* (L.) Lam.*)*. *PloS one* 8: e62707.

Rubio de Casas, R., G. Besnard, P. Schönswetter, L. Balaguer, and P. Vargas. 2006. Extensive gene flow blurs phylogeographic but not phylogenetic signal in *Olea europaea* L. *Theoretical and Applied Genetics* 113: 575–583.

Salamini, F., H. Ozkan, A. Brandolini, R. Schäfer-Pregl, and W. Martin. 2002. Genetics and geography of wild cereal domestication in the near east. *Nature Reviews Genetics* 3: 429–441.

Satya, P., M. Karan, C.S. Kar, a. K. Mahapatra, and B.S. Mahapatra. 2013. Assessment of molecular diversity and evolutionary relationship of kenaf *(Hibiscus cannabinus* L.*)*, roselle *(H. sabdariffa* L.*)* and their wild relatives. *Plant Systematics and Evolution* 299: 619–629.

Schafleitner, R., S. Kumar, C.-Y. Lin, S.G. Hegde, and A. Ebert. 2013. The okra *(Abelmoschus esculentus)* transcriptome as a source for gene sequence information and molecular markers for diversity analysis. *Gene* 517: 27–36.

Seefelder, S., H. Ehrmaier, G. Schweizer, and E. Seigner. 2000. Genetic diversity and phylogenetic relationships among accessions of hop *Humulus lupulus* as determined by amplifed fragment length polymorphism fingerprinting compared with pedigree data. *Plant Breeding* 119: 257–263.

Silva, G.J., T.M. Souza, R.L. Barbieri, and A. Costa de Oliveira. 2014. Origin, Domestication, and Dispersing of Pear *(Pyrus* spp.). *Advances in Agriculture* 2014: 1–8.

Simmonds, N.W., and K. Shepherd. 1955. Taxonomy and Origins of Cultivated Bananas. *Journal of the Linnean Society of London, Botany* 55: 301–312.

Simon, P.W. 2000. Domestication, Historical Development, and Modern Breeding of Carrot. *In:* J. Janick [ed.], Plant Breeding Reviews 19, 157–190. John Wiley & Sons.

Singh, Y.N. 2004. Kava: From Ethnology to Pharmacology. CRC Press.

Stajner, N., Z. Satovic, A. Cerenak, and B. Javornik. 2008. Genetic structure and differentiation in hop *(Humulus lupulus* L.*)* as inferred from microsatellites. *Euphytica* 161: 301–311.

Steiger, D.L., P.H. Moore, F. Zee, Z. Liu, and R. Ming. 2003. Genetic relationships of macadamia cultivars and species revealed by AFLP markers. *Euphytica* 1892: 269–277.

Tang, Z.X., K. Ross, Z.L. Ren, Z.J. Yang, H.Y. Zhang, T. Chikmawati, and J.P. Gustafson. 2011. *Secale*. *In* C. Kole [ed.], Wild crop relatives: Genomic and breeding resources, cereals, 367–396. Springer-Verlag, Berlin, Heidelberg.

Tavaud, M., a Zanetto, J.L. David, F. Laigret, and E. Dirlewanger. 2004. Genetic relationships between diploid and allotetraploid cherry species *(Prunus avium, Prunus* x *gondouinii* and *Prunus cerasus)*. *Heredity* 93: 631–8.

Terauchi, R., V. a Chikaleke, G. Thottappilly, and S.K. Hahn. 1992. Origin and phylogeny of Guinea yams as revealed by RFLP analysis of chloroplast DNA and nuclear ribosomal DNA. *Theoretical and Applied Genetics* 83: 743–51.

This, P., T. Lacombe, and M.R. Thomas. 2006. Historical origins and genetic diversity of wine grapes. *Trends in Genetics* 22: 511–9.

Till-Bottraud, I., X. Reboud, P. Brabant, M. Lefranc, B. Rherissi, F. Vedel, and H. Darmency. 1992. Outcrossing and hybridization in wild and cultivated foxtail millets: consequences for the release of transgenic crops. *Theoretical and Applied Genetics* 83: 940–6.

Upadhyaya, H.D., S.L. Dwivedi, M. Baum, R.K. Varshney, S.M. Udupa, C.L.L. Gowda, D. Hoisington, and S. Singh. 2008. Genetic structure, diversity, and allelic richness in composite collection and reference set in chickpea *(Cicer arietinum* L.*)*. *BMC Plant Biology* 8: 106.

Vershinin, A. V, T.R. Allnutt, M.R. Knox, M.J. Ambrose, and T.H.N. Ellis. 2003. Transposable elements reveal the impact of introgression, rather than transposition, in *Pisum* diversity, evolution, and domestication. *Molecular Biology and Evolution* 20: 2067–75.

Villani, F., A. Sansotta, M. Cherubini, D. Cesaroni, and V. Sbordoni. 1999. Genetic structure of natural populations of *Castanea sativa* in Turkey : evidence of a hybrid zone. *Journal of Evolutionary Biology* 12: 233–244.

De Vries, I.M. 1997. Origin and domestication of *Lactuca sativa* L . *Genetic Resources and Crop Evolution* 44: 165–174.

Ward, S.M. 2000. Allotetraploid segregation for single-gene morphological characters in quinoa *(Chenopodium quinoa* Willd.*)*. *Euphytica* 116: 11–16.

Wendel, J.F., and R. Cronn. 2003. Crytpic trysts, genomic mergers, and plant speciation. 133–142.

Wu, G.A., S. Prochnik, J. Jenkins, J. Salse, U. Hellsten, F. Murat, X. Perrier, et al. 2014. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature Biotechnology*.

Yamane, K., and T. Kawahara. 2005. Intra- and interspecific relationships among diploid *Triticum-Aegilops* species (Poaceae) based on base-pair substitutions, indels, andmicrosatellites in chloroplast noncoding sequences. *American Journal of Botany* 92: 1887–1898.

Yang, J.Y., M. Scascitelli, L. a. Motilal, S. Sveinsson, J.M.M. Engels, N.C. Kane, H. Dempewolf, et al. 2013. Complex origin of Trinitario-type *Theobroma cacao* (Malvaceae) from Trinidad and Tobago revealed using plastid genomics. *Tree Genetics & Genomes* 9: 829–840.

Yonemori, K., C. Honsho, S. Kanzaki, H. Ino, A. Ikegami, A. Kitajima, A. Sugiura, and D.E. Parfitt. 2008. Sequence analyses of the ITS regions and the matK gene for determining phylogenetic relationships of Diospyros kaki (persimmon) with other wild Diospyros (Ebenaceae) species. *Tree Genetics & Genomes* 4: 149–158.

Zerega, N.J.C., D. Ragone, and T.J. Motley. 2005. Systematics and Species Limits of Breadfruit *(Artocarpus*, Moraceae). *Systematic Biology* 30: 603–615.

Zohary, D. 1992. Is the European plum, *Prunus domestica* L., a *P. cerasifera* EHRH. X *P. spinosa* L. allo-polyploid? *Euphytica* 60: 75-77.

CHAPTER III

ROOTSTOCKS: DIVERSITY, DOMESTICATION, AND IMPACTS ON SHOOT PHENOTYPES

Emily J. Warschefsky, Laura L. Klein , Margaret H. Frank, Daniel H. Chitwood, Jason P. Londo, Eric J.B. von Wettberg, and Allison J. Miller

This chapter was published as a feature review in:

Trends in Plant Science 21(5): 418-437, 2016. doi 10.1016/j.tplants.2015.11.008.

Abstract

Grafting is an ancient agricultural practice that joins the root system (rootstock) of one plant to the shoot (scion) of another. It is most commonly employed in woody perennial crops to indirectly manipulate scion phenotype. While recent research has focused on scions, here we investigate rootstocks, the lesser-known half of the perennial crop equation. We review natural grafting, grafting in agriculture, rootstock diversity and domestication, and developing areas of rootstock research, including molecular interactions and rootstock microbiomes. With growing interest in perennial crops as valuable components of sustainable agriculture, rootstocks provide one mechanism by which to improve and expand woody perennial cultivation in a range of environmental conditions.

Getting to the Root of the Matter

Roots anchor plants in the ground, acquire water and nutrients from the soil, serve as storage organs, and are the primary zone of contact with soil organisms. Root systems vary substantially in architecture and function, both within and between species, and they are a crucial component in coordinating plant responses to a range of abiotic and biotic stressors, including pathogens, water and nutrient shortages, and potentially toxic compounds such as salt or heavy metals (e.g., [1–4]). In perennial crops and some annuals, grafting is used to join resilient root systems (rootstocks) to shoots (scions) that produce the harvested product (e.g., fleshy or dry fruits).

The vast majority of woody perennial plant cultivation involves clonal propagation [5–7], a technique that facilitated the domestication of the earliest woody

crops including olive, grape, and fig [8]. In these and many other species, grafting is an important part of the propagation process. Grafting typically joins two plant organs (root system and shoot) from different individuals that form vascular connections and survive in a unique symbiotic relationship as a genetic chimera [8]. The development of grafting around 1800 BCE facilitated a 'second wave' of woody perennial domestication and resulted in the wide-scale cultivation of new woody crops, including many Rosaceae (apple, pear, plum, and cherry), and the improvement of previously ungrafted, clonally propagated perennials [8,9].

In long-lived woody plants, grafting is a common means to clonally propagate desirable scions, thus side-stepping challenges traditionally associated with breeding of woody perennials, including prolonged juvenile phases and primarily outcrossing reproductive systems [5]. It is becoming increasingly apparent that the use of genetically distinct individuals as rootstocks serves to improve perennial crops, with different rootstocks conferring unique traits in both belowground and aboveground components of the plant [8]. In addition to reducing the time to fruit set, grafting can result in trees of shorter stature, both traits favored by early farmers. In modern agriculture, grafting has greatly increased the efficiency of perennial crop breeding by allowing root and shoot traits to be selected independently rather than requiring both sets of traits to be present in a single genetic individual. Here we explore our current state of knowledge of rootstocks, the lesser-known half of the perennial crop equation. We review natural grafting, grafting in perennial agriculture, rootstock diversity and domestication, and recent advances and future directions of rootstock research, including molecular interactions and rootstock microbiomes.

Natural Grafting

Grafting occurs in natural populations of some species, a phenomenon that may have inspired the development of grafting in agriculture and horticulture [8,10]. In nature, grafting can occur between stems or roots of the same individual or the same species or even between congeners or plants of different families [11–14]. Species that naturally graft tend to spread vegetatively and grow in dense stands, often in dry, somewhat harsh environments or those with loose soils that promote shallow, far-reaching root systems (e.g., *Acer saccharinum, Betula lutea, Pinus* spp., *Populus tremuloides, Pseudotsuga* spp., *Thuja* spp., *Tilia americana, Tsuga heterophylla, Ulmus americana,* and several tropical species, [13,15–18]).

Seminal research in natural grafting documents the transfer of water, nutrients, compounds, dyes, silvicides, pathogens, and even genetic material between individuals through grafts [19–21]. More recently, several studies have demonstrated the transfer of fungal and bacterial pathogens through natural graft junctions, including Dutch elm disease, oak wilt, laurel wilt, tomato wilt, and citrus variegated chlorosis [20,22–27]. Chemicals used in management or treatment (e.g., ammonium sulfamate, glyphosate, propiconazole) can also move through graft junctions [19,24,25]. Furthermore, experimental work with tobacco grafts shows the transfer of partial or whole nuclear and plastid genomes short distances across graft junctions [28–30] (see below). Similarly, naturally occurring plastid and nuclear genome transfers have been documented between the tropical tree species *Amborella trichopoda* and its epiphytes, with interspecific cellular contact occurring at wound sites [31]. Horizontal gene transfer is also present between parasitic plants and their hosts [32], but the connection of haustoria to host plant

vasculature represents a very different mechanism compared with graft junctions. These findings have exciting implications for agriculture and suggest a novel mechanism for asexual speciation under certain circumstances [15,30].

Grafting in Woody Perennial Agriculture

The phenomenon of natural grafting was coopted for use in cultivation and today is an essential part of agriculture, horticulture, and silviculture. Grafting typically employs two individuals, one or both of which are clonally propagated, depending on the desired outcome: sexually produced (seed grown) rootstock and clonal scions are often used in traditional agricultural settings but also for some industrial-scale crops (e.g., *Coffea, Juglans*); clonal rootstocks and sexually produced scions are typically used during the cultivar breeding and selection process; and, when uniformity is desired, both rootstock and scion are clonally produced. In more advanced grafting practices, a third individual (interstock) is sometimes used to join a rootstock and scion that may otherwise be incompatible [8].

A review of the available literature indicates that more than 70 woody perennial crop species propagated for their edible fruits are grown on rootstocks (Table 3.1), in addition to those species used for fodder, fiber, oil, and timber. Rootstocks are widely used for economically important perennial fruit and nut species: 20 of the 25 most-produced fruit and nut crops [33] are grafted in certain circumstances (Table 3.2); the remaining five crops are monocots, for which grafting is not a viable method of propagation. The value of rootstocks has become evident even for annuals and several

recent reviews have explored various aspects of grafting and rootstock–scion interactions (e.g., [8,34–42]).

Research into the domestication of woody perennial crops has lagged behind that of annuals due to the logistical difficulties of working with large, long-lived species that require immense amounts of time, space, and money to cultivate and maintain [5–7]. Grafting adds an additional layer of complexity to this work, as the performance of multiple scion–phenotype combinations must be evaluated over many years and, ideally, in many different environments. Until recently, rootstock research has focused on important horticultural goals such as improving scion phenotype [37,43] and identifying pest- and pathogen-resistant rootstocks [44,45]. However, advances in molecular techniques have made it possible to achieve a more intricate understanding of the processes involved in grafting and the role that rootstocks play in perennial crop domestication.

Rootstock Diversity And Domestication

Although grafting has been an important part of growing woody perennial crops for at least 2000 years, surprisingly little is known about the plant species that are used as rootstocks. It is clear that rootstocks for different crops are at different stages of domestication. In the early stages of rootstock use, rootstocks are used primarily as a means of clonal propagation of the scion and are chosen based on their availability, with little selective pressure on specific traits. Instead, growers deal with undesirable traits using intensive and costly techniques such as pruning, fertilization, and pesticide application. As rootstock domestication advances, traits including productivity and

disease resistance are often targeted. Rootstocks have a long history and an important role in agriculture, but many questions about rootstock diversity and domestication remain. How many genotypes and species are used as rootstocks for a given crop and how closely related are they to the scion species? What are the geographic origins, current distributions, and frequency of use of these genotypes? Are there morphological or genetic signatures of domestication in taxa used for rootstocks?

Rootstock Diversity

Despite a growing body of literature documenting the diversity and phylogeography of cultivated plants and their evolution under domestication, including woody perennials used as scions [6,46], rootstock species are rarely considered. Looking at some of the most economically important grafted plants in terms of tonnage (Table 3.2), some trends in rootstock diversity and domestication become apparent: (i) rootstock species are often closely related to but genetically distinct from the scion species they support; (ii) for a single crop, multiple species and their hybrid derivatives are often used to generate rootstocks, although (iii) ultimately relatively few rootstock genotypes are employed for a given crop in most contemporary agricultural systems; (iv) rootstock selection is a function of both the scion genotype with which it is grafted and the environment in which the grafted plant will be grown [47]; and (v) rootstocks are selected not only for traits inherent in the root system but also for traits imparted to the scion (Fig. 3.1, Key Figure).

For example, the introduction of the North American aphid *Phylloxera* into Europe in the mid- 1800s devastated the grape *(Vitis vinifera)* industry on the European

continent [48]. Grafting *V. vinifera* scions onto *Phylloxera*-resistant rootstocks allowed *V. vinifera* to grow in the presence of *Phylloxera* and today grafting is commonplace in grape, with native North American grapevine species functioning as indispensable resources for the development of abiotic and biotic stress-resistant rootstocks [49,50]. *Vitis riparia* and *Vitis rupestris* were initially selected for *Phylloxera* resistance and for their capacity to self-root. Subsequent integration of *Vitis berlandieri* into rootstock development programs expanded the potential range of vineyards because of its tolerance of chalky soils like those found in the Champagne and Cognac regions [49]. Today, grapevine scions are primarily grafted on these three species and their hybrid derivatives, although other species are also used [51–53] (Table 3.2).

Given that graft compatibility can occur across broad phylogenetic distances, crop wild relatives are of great significance to grafted perennial crops. This underscores the importance of maintaining significant living collections of perennial crop wild relatives that represent a range of variation in morphology, phenology, and ecology [54], not only for scions but also for rootstock development. Much like in crop breeding, rootstock breeding and selection efforts often target wild and semidomesticated species, feral individuals, and landraces that are thought to be disease and stress resistant and that are adapted to local environments.

Evolution of Rootstock Species Under Domestication

Rootstock species are considered to be undergoing domestication because they are part of a mutualistic relationship between humans and plants that enhances the fitness of both the domesticator and the domesticate [55]. Although rootstock breeding clearly

targets specific traits (Fig. 3.1), to our knowledge no formal description of a 'domestication syndrome' exists for rootstock species. Further, few comparative morphological or genetic studies of rootstocks and their wild ancestors exist to infer signatures of rootstock domestication. Many rootstocks are derived from perennial, outcrossing wild species (Table 3.2) and exhibit some of the hallmarks of woody plant domestication, including high levels of heterozygosity and extensive clonal propagation [5–7]. However, rootstock species are unlike most other domesticated perennials in that, while some traits under selection are directly expressed in the rootstock itself (e.g., pathogen resistance), others are expressed in the grafted scion (e.g., dwarfing) (see below).

In general, grafting affects three major processes in a plant: uptake and transport of water and nutrients, hormone production and transport, and the large-scale movement of proteins, mRNAs, and small RNAs (sRNAs). These processes have implications for both belowground and aboveground functioning, but the interconnectedness of the variables at work in rootstock–scion interactions (rootstock genotype, scion genotype, environment) obscures individual contributions to phenotypic variation. Certainly, the genotypes of both rootstock and scion play an important role in these interactions, and different combinations of stock and scion are known to vary in their phenotypic effects (e.g., [56–60]). Additional factors impacting rootstock–scion interactions include the age of the grafted individuals, the grafting technique employed, seasonality, time since grafting [15,40], genotype x environment and genotype x genotype x environment interactions, root morphology and architecture [42], the degree of rootstock–scion compatibility, and root–microbe interactions. Below, we consider the primary traits

targeted during rootstock selection (Fig. 3.1) and assess our current understanding of their genetic underpinnings.

Graft Union Formation and Graft Compatibility

The development and integration of highly efficient root systems in crops through grafting is possible only if the rootstock and scion are graft compatible. Consequently, the primary selection factor for any rootstock is its ability to form the tissue that serves as the junction between the rootstock and scion: the graft union [15,34]. The healing of the graft union can take anywhere from days in the case of herbaceous plants to more than 1 year in the case of some woody perennials [40] and in some cases graft incompatibility may not become apparent for several years [61,62]. The quality of vascular connections formed in the graft union varies between rootstock–scion combinations and can impact water transport from root to shoot for long periods of time or permanently [15,34,36,40]. Work in Arabidopsis suggests that intertissue communication, cell interdigitation, and auxin responses are all important for the success of graft unions [63]. Despite being the single most important factor required for grafting, the mechanisms of graft compatibility and incompatibility are still not well understood at the physiological or molecular level [15].

Root Structure and Function

Assuming a successful graft union, rootstocks are selected in part for traits inherent to the root system itself, primarily resistance to soil-borne pests and pathogens (e.g., [35]), and tolerance to abiotic stressors such as salinity, drought, and flooding [40]

(Tables 3.1 and 3.2). How do processes occurring in the root impact stress response in other parts of the plant? In many ways our understanding of root structure and function has lagged behind that of easier-to-observe aboveground organs (e.g., [64,65]), but several recent reviews report that root anatomy can mediate responses to a range of abiotic and biotic stressors [3,4,66–70]. For example, shifts in key anatomic traits – from root cortical aerenchyma to xylem diameter and conductance to variation in root hairs to endo- and exodermal lignification and suberization that reduce root water loss – all have the potential to increase late-season water availability [69]. While many crops exhibit substantial variability in root anatomy, selection of improved root systems – those that are deep rooting, less metabolically active, or more water conserving – comes with costs such as less root mass to forage for soil phosphorus at shallow depths, reduced efficiency at exploiting ephemeral nutrient patches, and exposure to more challenging abiotic conditions (e.g., low temperatures, salinity, compaction, aluminum or manganese toxicity) at deeper soil levels [71]. Nonetheless, for many crops in many soils, a deep-rooting, water-conserving root phenotype is likely to have several advantages (e.g., [68]).

Rootstock Modulation of Scion Phenotypes

In addition to selecting for phenotypes expressed in the roots, rootstocks are also selected based on their effects on the scion, including precocity (early bearing), production, disease resistance, and fruit quality (Tables 3.1 and 3.2; Fig. 3.1). One of the most sought-after phenotypes – rootstock-induced reduction in scion vigor, or ‘dwarfing’ – causes a decrease in tree volume, height, canopy diameter, and circumference [40], reducing the need for pruning in commercial orchards. Scion vigor is known to be

affected by numerous factors including root hydraulic pressure, water uptake efficiency, hormone production, nutrient uptake, stomatal conductance, and intercellular CO_2 levels [40,56,72] and even within a single species, *Malus pumila*, there exists evidence supporting different mechanisms underlying dwarfing [39]. It is likely that in the case of dwarfing, as with many other rootstock-induced traits, multiple independent molecular pathways can result in similar scion phenotypes. Much is known about many other scion traits in general; however, the impact of rootstocks on these scion phenotypes remains unclear. For example, in apple tree scions the genetic underpinnings of tree architecture [73,74], hydraulic efficiency [75], and biennial bearing [76] have been documented. Work to date has demonstrated that rootstock genotype plays a role in shaping variation in these traits in the scion [77–79]; however, the relative roles of rootstock and scion as well as the mechanism underlying rootstock influence remain insufficiently understood. Expanding these studies to include questions such as the effect of rootstock diversity or molecular signaling during scion modulation presents exciting areas of future research.

In addition to altering tree architecture, rootstocks are widely used to confer resistance to pests and pathogens that affect the scion, including physiological disorders (reviewed in [40]). For instance, anthracnose resistance of avocado scions has been shown to be induced by rootstocks [80]. In this case, resistance is linked to increased diene concentrations, which may be due to improved scion nutrition. Improved nutrient and water uptake and transport by the rootstock to the scion is also thought to play a role in resistance to physiological disorders such as physiological pitting and stem-end browning in kiwi and stem-end rind breakdown in citrus [81,82].

Genetic Underpinnings of Rootstock Traits Under Selection

Disentangling genetic and environmental components of desirable rootstock traits would allow marker-assisted selection to facilitate rootstock breeding [42]. Several studies have progressed toward this goal by generated segregating F1 mapping populations of rootstocks to which a common scion is grafted to identify the genetic basis of economically important traits of rootstocks that are expressed in the scion. Unlike annual crops, which are commonly inbred and genetically homozygous, in woody perennials genetic mapping often occurs in the F1 generation, for which the parents are typically highly heterozygous. In apple, studies exploring rootstock genetic contributions to dwarfing phenotypes led to the identification of the Dwarfing 1 (Dw1) and Dwarfing 2 (Dw2) loci [83,84]. Another study documented the genetic basis of absorption and translocation of nutrients by apple rootstocks and demonstrated significant rootstock effects on the transport of Ca, Cu, K, Mg, Mn, Na, P, S, Zn, and Mo [85]. In a similar experiment in grafted tomato, three to eight loci controlling salt tolerance from the rootstock genome were linked to increased yield in the scion [86–88]. Additional studies in grapevine identified loci in rootstocks that influence tolerance to lime-induced iron deficiency [89] and scion transpiration, leaf area, and water-use efficiency [90]. These pioneering studies and others provide convincing evidence of a genetic basis underlying rootstock modulation of scion phenotypes. Future work, including fine mapping, is needed to achieve a detailed understanding of the rootstock-genetic architectures of agriculturally important traits exhibited in the root and/or in the grafted scion. In addition, multisite, multiyear studies will facilitate deeper understanding of genotype x environment interactions.

Scion Modulation of Rootstocks

While most agricultural grafting involves using rootstocks to influence phenotypes expressed in the scion, it is worth noting that grafting can also use scions to affect root system phenotypes, an approach that may be useful in some root/tuber crops. For example, grafting was developed in Indonesia in the late 20th century for cassava, using an inedible wild relative as the scion to improve yield [8]. In this 'Mukibat' system an arboreal *Manihot* species, *M. glaziovii*, is grafted onto the cultivated cassava *Manihot esculenta* [91]. This pairing increased the total yield of tubers by approximately 100% as well as tuber size [92]. Yield quality in sweet potato and potato have also benefited significantly through the application of grafted scions [93–95]. In tuber crops, effects on the rootstock are of particular significance for crop yield; in fruit crops scion effects on the rootstock have received less attention. However, scion effects on rootstocks are likely to be ubiquitous and large, as the flow of sugar, hormones, and nucleic acids into the root system has substantial effects on root growth, carbohydrate storage, and phenology [96,97] (as documented in Molecular Interactions below). In general, examining scion effects on rootstocks remains an important but woefully understudied component of rootstock–scion interactions.

Recent Advances and Future Directions in Rootstock Biology

Rootstocks can confer enhanced tolerance to abiotic and biotic stressors, providing a valuable mechanism to improve and expand perennial crop cultivation and global food production in the face of changing climatic conditions [42,37]. While these grafting-induced benefits are well understood from a physiological perspective, we have

yet to build an integrated understanding of the molecular mechanisms that coordinate rootstock–scion communication and ultimately lead to enhanced crop traits. Many important aspects of rootstock biology are just beginning to take shape, including long-distance molecular signaling and the capacity of rootstocks to modulate interaction between plant and soil microbiomes.

Molecular Interactions

Grafting has enabled mass cultivation and improvement of woody perennial crops, but the generation of genetic chimeras through grafting also provides an important tool for understanding fundamental questions in plant biology. Multiple recent works have begun to shed light on one such longstanding question: whether grafting induces heritable changes in the scion [98]. Revolutionary work from the Bock laboratory examining genomic interactions between sexually incompatible *Nicotiana* species showed that entire chloroplast and nuclear genomes can be bidirectionally transferred across the graft union, resulting in asexual hybrids between the rootstock and scion genotypes [29,30,99]. This phenomenon appears to be localized to tissues near the graft union and is therefore not heritable except in the rare event that an adventitious bud forms from one of these cells in the graft junction. Nevertheless, these studies demonstrate that, through as-yet-unknown mechanisms, large pieces of DNA or entire plastid genomes can traverse the graft junction, suggesting that it may be possible for these macromolecules to travel further into the scion under certain grafting conditions [100]. In addition to the movement of DNA itself, interspecific grafting within the Solanaceae has been shown to cause heritable changes in DNA methylation patterns in

the scion [101]. This research also found changes in methylation of rootstock material, indicating that this is a reciprocal process across the graft junction. Collectively, this work suggests that future studies should focus on the graft transmissability of heritable material and its impacts on plant form, physiology, and evolution.

Additionally, grafted plants offer unique arenas to investigate other pathways of long-distance communication between cells. While long-distance signaling may indirectly involve hormones, metabolites [102–104], or water and nutrient availability, other molecules – proteins, transcripts, and sRNAs – provide a direct link to underlying genetic mechanisms [105,106]. The extent to which these direct versus indirect long-distance signals coordinate grafting-induced improvements in the reciprocal half of the plant remains unclear. However, recent research lends substantial support for the direct involvement of mobile, mature sRNAs, which act as signals between the root and shoot targeting a wide range of transcripts and eliciting far-ranging graft-transmissible effects, from phosphate starvation response [107] to tuberization [97,108] to pest and pathogen resistance [109]. When mutant Arabidopsis rootstocks defective in sRNA biogenesis were grafted to wild-type scions, mature 22- and 24-nucleotide sRNAs accumulated in the roots, indicating that these sRNAs had been produced in the shoot and subsequently traversed the graft junction [110]. This experiment unequivocally demonstrated that mature sRNAs, and not simply sRNA precursors, are capable of long-distance transport, and has helped to answer longstanding questions regarding systemic, whole plant phenomena such as acquired virus resistance [111].

Beyond sRNAs, a growing body of work indicates that portions of the transcriptome itself are graft transmissible and the functional movement of individual

transcripts (e.g., those inducing tuberization in a photoperiod-sensitive manner [96] or mediating morphological changes in traits such as leaf complexity [112,113]) has been demonstrated. Furthermore, recent work combining interspecific grafting with high-throughput RNA sequencing has revealed that a large fraction of the transcribed genome undergoes long-distance transport [114]. While the exact quantification of the mobile transcriptome varied from just over 2000 non-cell-autonomous transcripts between related Arabidopsis ecotypes [114] to almost half of the annotated gene space between Arabidopsis and the parasitic plant *Cuscuta* [115], these experiments clearly demonstrate that plant transcriptomes are spatially promiscuous.

The paradigm shift from a model of cell-autonomous to massively mobile transcript localization in plant molecular signaling raises new questions about mRNA transport and non-cell-autonomous mRNA function. Is there a 'zip code' that marks transcripts for long-distance movement and directs their end localization and, if so, how conserved across genetically distinct rootstocks and scions is the mechanism? Do mobile transcripts function to influence growth and development in their new location? Elegant work has identified RNA motifs that are required for the long-distance transport of GIBBERELLIC ACID-INSENSITIVE (GAI) transcripts [116]; whether these motifs can be universally extended to explain mass transcript trafficking remains to be seen. Additional research has complemented transcriptomic profiling of graft-transmissible mRNAs with proteomics to demonstrate that many of the transported RNAs are indeed translated at their new location, suggesting that these mobile transcripts are capable of functioning after long-distance transport [114].

Growing support for long-distance, graft-transmissible molecular signaling in conjunction with rapid advances in genotyping and phenotyping technologies that allow us to hone in on the genetic mechanisms underlying enhanced abiotic and biotic stress tolerance has sparked interest in a new agronomic application of this ancient technique [117]. Transgrafting – the physical joining of a genetically engineered rootstock with a wild-type scion (or vice versa) – enables targeted crop protection without genetic alteration of the product [38]. This practice has been explored in both annuals such as watermelon *(Citrullus lanatus)* [118] and perennials such as apple *(Malus domestica)* [119], grape *(V. vinifera)* [120,121], citrus *(Citrus* spp.) [122], and cherry *(Prunus* spp.) [123] and shows promise for combating abiotic stressors such as salt and drought [122] as well as detrimental diseases affecting both scion and rootstock [118,120,121,123]. The efficacy of transgrafting is illustrated by a case study in grape, where the crippling decomposition of *V. vinifera* vasculature by *Xylella fastidiosa* (the causative agent of Pierce's disease) is apparently completely halted by the genetic fortification of rootstock cell walls [124]. Importantly, while modification of the *V. vinifera* rootstock was sufficient to confer protection on the scion, PCR assays demonstrate that this resistance was achieved without the movement of stable genetic material [38]. The widespread adoption of transgrafting may allow targeted crop protection without the direct modification of crop products.

While the vast majority of studies looking for graft-transmissible molecular signals have been performed in annual model systems, perennials provide a more agriculturally relevant basis for this line of research due to the extensive use of grafting in commercial vineyards and orchards. These long-lived ‘fields’ represent a valuable

resource for exploring perennial-specific questions. For example, how does the composition and quantity of the mobile transcriptome change seasonally or from year to year? How does the environment, under real-world conditions, modulate the plasticity of long-distance communication in plants? Finally, beyond the movement of molecules, grafted perennials can also be used to study the graft transmission of both bacteria and endophytes.

Microbiomes of Rootstocks

Just as the study of rootstock genetics and domestication is in the early stages of understanding, so too are the effects of rootstocks on the plant microbiome. The emerging field of microbiome research provides evidence that rootstock–scion interactions are almost certainly influenced in part by the beneficial root microbiome, which includes fungal endophytes and plant growth-promoting (PGP) bacteria found within and around the root system. These microbes can influence uptake of micronutrients, generate hormones, create a root zone environment that is hostile to pests and pathogens, and impact plant phenotypes including disease [125], nitrogen, phosphorus and iron limitation, and resistance to heat, drought, and salt [126]. Additionally, members of the root microbiome can enter the plant and be transported via xylem to aerial tissues where they can act as biocontrol agents and impact stress response [127].

Research investigating the interactions of root microbiomes with shoot performance, specifically in grafted plants, is scarce. However, evidence from own-rooted grapevine has demonstrated that the bacterium *Burkholderia phytofirmans* strain

PsJN can colonize the root system and, on transmission to the shoot, protects the vine from pathogenic Botrytis and Pseudomonas while also modulating carbohydrate metabolism and increasing freezing tolerance [128–131]. PGP bacteria have also been implicated in altering plant photosynthesis rate, transpiration, stomatal conductance, and internal leaf CO_2 [127].

Multiple biotic and abiotic factors impact the diversity and composition of the root microbiome. Some studies have suggested that abiotic factors play a large role in determining microbiome community structure, which is particularly relevant to cultivated species where the abiotic environment is often manipulated to enhance plant growth. For example, microbiome community structure associated with arid grassland ecosystems is driven not by the complexity of plant functional groups but rather by water availability [132]. Similarly, in studies examining Cannabis varieties, soil differences are implicated as a major contributor to microbiome community composition as a whole [133]. Compared with non-farmed desert land, farmed deserts show dramatic changes in microbiomes due to irrigation and concomitant loss of extremophile species [134–136]. These results indicate that abiotic factors, including those manipulated in agricultural settings, are often major drivers of microbiome communities in plants. However, evidence for plant genotype playing a role in microbiome community composition indicates that abiotic conditions are not the sole factor determining the microbiome. Experiments evaluating different grapevine-associated soil bacteria show that some *(Pseudomonas, Acinetobacter, Sphingobacterium, Enterobacter,* and *Delftia* sp.) have the ability to protect plants against simulated drought and produce biomass despite low water availability [127]. The magnitude of these benefits is dependent on stress treatment and

the sensitivity of individual grapevine genotypes to drought. This research indicates that drought-sensitive rootstocks could be supplemented and modified by inoculating the soil with the beneficial bacterial strains identified in the microbiome of drought-tolerant rootstocks, increasing performance under drought conditions [127]. Similar evidence of genotype-specific fungal communities in the soil has been observed in crop fields of potato and wheat [137] and in greenhouse studies with Cannabis [133].

Ongoing studies demonstrate the ability of many plant species to actively select for the bacterial composition of the rhizosphere [138]. Roots modulate their microbiomes by exuding a complex mix of amino acids, organic acids, and sugars, a cocktail that functions both as a defense against pathogens and as a recruitment tool to foster the growth of beneficial microbes [139]. Concentrations and compositions of root exudates differ between plant species [140,141] and between varieties within species [142] and are also known to change in response to abiotic conditions. This variation provides selective niches that determine the species composition of plant microbiomes, perhaps similar to the 'arms-race' mechanism of plant pathogen–host evolution [143]. Although most of the research examining what root-based effects determine microbiome composition and structure have been conducted in annual crop systems, it is likely that perennial species similarly generate selective environments for beneficial microbes.

Current work reveals that the root microbiome can enhance plant productivity under stressful conditions, prevent infection from pathogenic bacteria, modify nitrogen availability and carbon storage, and have many other major biological impacts. However, it also raises new questions about rootstock microbiomes. Which combination of microbiome species and rootstock genotypes optimally alters the phenotype of the grafted

plant? How permanent is the microbiome? Can inoculation be used to further leverage rootstock performance? How much of the root microbiome is transferred to the shoot microbiome? To what extent is recruitment of specific communities possible?

Concluding Remarks and Future Perspectives

This synthesis of our present knowledge of grafting and rootstock biology comes as there is growing interest in sustainably enhancing crop productivity to address challenges posed by global population growth and climate change [2,32,44]. For woody perennial species, which have long generation times and are often self-incompatible, traditional breeding practices employed in annual crops are usually infeasible. Rootstocks provide agriculturists with a mechanism by which to improve perennial crops and increase their productivity under harsh environmental conditions while simultaneously limiting agricultural inputs (irrigation, fertilizer, pesticides). In addition, this work provides a reference for comparison of grafting in annual crops, such as tomato and melons, for which the process is widely used to combat abiotic and biotic stresses as well as to boost scion vigor.

We advocate additional research in the molecular, evolutionary, and domestication processes of rootstock species using newly emerging technologies and analyses including high-throughput genomics and phenomics (see Appendix 3.2, Outstanding Questions) [4,144]. The resulting data will address pertinent questions for rootstock biology, including rootstock diversity, the evolution of clonal, perennial crops under artificial selection, mechanisms underlying rootstock–scion interactions and graft compatibility, and the impact of root systems on economically important traits in the

scion. Of particular interest is the development and maintenance of diverse living germplasm collections for woody perennials used as scions and rootstocks, as well as the construction of crosses and grafting experiments needed to facilitate additional work examining the genetic basis of traits in grafted crops. This complex task requires the identification of genes contributing to phenotypic variation in both the rootstock and the scion, genes that may be carried by the rootstock, the scion, or both. Genome-wide association mapping [145] and sequence-first population genomic approaches [146] offer promising avenues of exploration in perennials, which are often long-lived and highly heterozygous. Comprehensive germplasm collections, coupled with dynamic technological and analytical advances, have the potential to yield significant advances in grafted crops, which represent a key component of sustainable agriculture.

Literature Cited

1. Lynch, J. (1995) Root architecture and plant productivity. Plant Physiol. 109, 7–13

2. Lynch, J.P. (2007) Turner review no. 14: roots of the second green revolution. Aust. J. Bot. 55, 493–512

3. Hodge, A. et al. (2009) Plant root growth, architecture and function. Plant Soil 321, 153–187

4. Kuijken, R.C.P. et al. (2015) Root phenotyping: from component trait in the lab to breeding. J. Exp. Bot. Published online June 12, 2015. http://dx.doi.org/10.1093/jxb/erv239

5. Zohary, D. and Spiegel-Roy, P. (1975) Beginnings of fruit growing in the old world. Science 187, 319–327

6. McKey, D. et al. (2010) The evolutionary ecology of clonally propagated domesticated plants. New Phytol. 186, 318–332

7. Miller, A.J. and Gross, B.L. (2011) From forest to field: perennial fruit crop domestication. Am. J. Bot. 98, 1389–1414

8. Mudge, K. et al. (2009) A history of grafting. Hortic. Rev. 35, 437–494

9. Meyer, R.S. et al. (2012) Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. New Phytol. 196, 29–48

10. Juniper, B.E. and Mabberley, D.J. (2006) Apples and grafting. In *The Story of the Apple*, pp. 91–114, Timber Press

11. La Rue, C.D. (1934) Root grafting in trees. Am. J. Bot. 21, 121–126

12. Graham, B.F. and Bormann, F.H. (1966) Natural root grafts. Bot. Rev. 32, 255–292

13. Eis, S. (1972) Root grafts and their silvicultural implications. Can. J. For. Res. 2, 111–120

14. Jayawickrama, K.J.S. et al. (1991) Rootstock effects in grafted conifers: a review. New For. 5, 157–173

15. Goldschmidt, E.E. (2014) Plant grafting: new mechanisms, evolutionary implications. Front. Plant Sci. 5, 1–9

16. Beddie, A.D. (1941) Natural grafts in New Zealand trees. Trans. Proc. R. Soc. N. Z. 71, 197–203

17. Jelínková, H. et al. (2009) Molecular and dendrochronological analysis of natural root grafting in *Populus tremuloides* (Salicaceae). Am. J. Bot. 96, 1500–1505

18. Jones, F.A. et al. (2011) The roots of diversity: below ground species richness and rooting distributions in a tropical forest revealed by DNA barcodes and inverse modeling. PLoS ONE 6, e24506

19. Bormann, F.H. and Graham, B.F. (1959) The occurrence of natural root grafting in eastern white pine, *Pinus strobus* L., and its ecological implications. Ecology 40, 677–691

20. Epstein, A.H. (1978) Root graft transmission of tree pathogens. Annu. Rev. Phytopathol. 16, 181–192

21. Tattar, T.A. (2009) Injection, infusion, and systemic movement in trees. In *Proceedings of the National Oak Wilt Symposium,* June 4–7, 2007, Austin, TX, pp. 169–173, Texas Forest Service

22. Bertrand, B. et al. (2000) Genetic study of *Coffea canephora* coffee tree resistance to *Meloidogyne incognita* nematodes in Guatemala and *Meloidogyne sp.* nematodes in El Salvador for selection of rootstock varieties in Central America. Euphytica 113, 79–86

23. He, C.X. et al. (2000) Distribution of *Xylella fastidiosa* in citrus rootstocks and transmission of citrus variegated chlorosis between sweet orange plants through natural root grafts. Plant Dis. 84, 622–626

24. Ciccotti, A.M. et al. (2007) Transmission of "*Candidatus Phytoplasma mali*" by root bridges under natural and experimental conditions. Bull. Insectol. 60, 387–388

25. Blaedow, R.A. and Juzwik, J. (2010) Spatial and temporal distribution of *Cereatocystis fagacearum* in roots and root grafts of oak wilt affected red oaks. Arboric Urban For. 36, 28–34

26. Rivard, C.L. et al. (2010) Grafting tomato with interspecific rootstock to manage diseases caused by *Sclerotium rolfsii* and southern root-knot nematode. Plant Dis. 94, 1015–1021

27. Ploetz, R.C. et al. (2012) Responses of avocado to laurel wilt, caused by *Raffaelea lauricola*. Plant Pathol. 61, 801–808

28. Bock, R. (2010) The give-and-take of DNA: horizontal gene transfer in plants. Trends Plant Sci. 15, 11–22

29. Stegemann, S. et al. (2012) Horizontal transfer of chloroplast genomes between plant species. Proc. Natl. Acad. Sci. U.S.A. 109, 2434–2438

30. Fuentes, I. et al. (2014) Horizontal genome transfer as an asexual path to the formation of new species. Nature 511, 232–235

31. Rice, D.W. et al. (2013) Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. Science 342, 1468–1473

32. Davis, C.C. and Xi, Z. (2015) Horizontal gene transfer in parasitic plants. Curr. Opin. Plant Biol. 26, 14–19

33. FAOSTAT (2013) Production, Food and Agriculture Organization of the United Nations Statistics Division

34. Aloni, B. et al. (2010) Hormonal signaling in rootstock–scion interactions. Sci. Hortic. (Amsterdam) 127, 119–126

35. Louws, F.J. et al. (2010) Grafting fruiting vegetables to manage soilborne pathogens, foliar pathogens, arthropods and weeds. Sci. Hortic. (Amsterdam) 127, 127–146

36. Martínez-Ballesta, M.C. et al. (2010) Physiological aspects of rootstock–scion interactions. Sci. Hortic. (Amsterdam) 127, 112–118

37. Schwarz, D. et al. (2010) Grafting as a tool to improve tolerance of vegetables to abiotic stresses: thermal stress, water stress and organic pollutants. Sci. Hortic. (Amsterdam) 127, 162–171

38. Haroldsen, V.M. et al. (2012) Mobility of transgenic nucleic acids and proteins within grafted rootstocks for agricultural improvement. Front. Plant Sci. 3, 39

39. Gregory, P.J. et al. (2013) Contributions of roots and rootstocks to sustainable, intensified crop production. J. Exp. Bot. 64, 1209–1222

40. Koepke, T. and Dhingra, A. (2013) Rootstock scion somatogenetic interactions in perennial composite plants. Plant Cell Rep. 32, 1321–1337

41. Tramontini, S. et al. (2013) Rootstock control of scion response to water stress in grapevine. Environ. Exp. Bot. 93, 20–26

42. Albacete, A. et al. (2015) Unravelling rootstock scion interactions to improve food security. J. Exp. Bot. 66, 2211–2226

43. Rouphael, Y. et al. (2010) Impact of grafting on product quality of fruit vegetables. Sci. Hortic. (Amsterdam) 127, 172–179

44. Seemüller, E. et al. (2008) Apple proliferation resistance of *Malus sieboldii*-based rootstocks in comparison to rootstocks derived from other *Malus* species. Eur. J. Plant Pathol. 121, 109–119

45. Grauke, L.J. and Starr, J.L. (2014) Phenotypic screening of pecan seedling rootstocks in search of nematode resistance. Trees 28, 1333–1341

46. Vavilov, N.I. (1992) *Origin and Geography of Cultivated Plants* (Love, D., transl.), Cambridge University Press

47. Cummins, J.N. and Aldwinckle, H.S. (1983) Breeding apple rootstocks. Plant Breed. Rev. 1, 294–394

48. Pouget, R. (1990) Histoire de la Lutte Contre le Phylloxera de la Vigne en France, Institut National de la Recherche Agronomique (in French)

49. Galet, P. (1979) A Practical Ampelography. Translated and Adapted by L. T. Morton, Cornell University Press

50. Pongrácz, D.P. (1983) Rootstocks for Grape-Vines, David Philip

51. Vasanthaiah, H.K.N. et al. (2011) Muscadiniana. In *Wild Crop Relatives: Genomic and Breeding Resources. Temperate Fruits* (Cole, K., ed.), pp. 65–77, Springer-Verlag

52. Burger, P. et al. (2009) Grape breeding. In *Breeding Plantation Tree Crops: Tropical Species* (Jain, S.M. and Priyadarshan, P.M., eds), pp. 161–189, Springer

53. Soneji, J.R. and Nageswara Rao, M. (2011) *Vitis*. In *Wild Crop Relatives: Genomic and Breeding Resources. Temperate Fruits* (Cole, K., ed.), pp. 223–239, Springer-Verlag

54. Maxted, N. and Kell, S.P. (2009) *Establishment of a Global Network for the in situ Conservation of Crop Wild Relatives: Status and Needs,* FAO Commission on Genetic Resources for Food and Agriculture

55. Zeder, M.A. (2015) Core questions in domestication research. Proc. Natl. Acad. Sci. U.S.A. 112, 3191–3198

56. Tworkoski, T. and Miller, S. (2007) Rootstock effect on growth of apple scions with different growth habits. Sci. Hortic. (Amsterdam) 111, 335–343

57. Albrecht, U. et al. (2012) Influence of rootstock variety on Huanglongbing disease development in field-grown sweet orange (*Citrus sinensis* [L.] Osbeck) trees. Sci. Hortic. (Amsterdam) 138, 210–220

58. Lanauskas, J. et al. (2012) Rootstock effect on the performance of sweet cherry cv. Lapins. Hort. Sci. 39, 55–60

59. Nadernejad, N. et al. (2013) Effect of different rootstocks on PAL activity and phenolic compounds in flowers, leaves, hulls and kernels of three pistachio (*Pistacia vera* L.) cultivars. Trees 27, 1681–1689

60. Amiri, M.E. et al. (2014) Influence of rootstock on mineral uptake and scion growth of “Golden Delicious” and “Royal Gala” apples. J. Plant Nutr. 37, 16–29

61. Andrews, P.K. and Serrano Marquez, C. (1993) Graft incompatibility. Hort. Rev. 15, 183–232

62. Pereira, I. et al. (2014) Growth characteristics and phenylalanine ammonia-lyase activity in peach grafted on different *Prunus spp*. Biol. Plant 58, 114–120

63. Melnyk, C.W. et al. (2015) A developmental framework for graft formation and vascular reconnection in *Arabidopsis thaliana*. Curr. Biol. 25, 1306–1318

64. Eissenstat, D.M. and Yanai, R.D. (1997) The ecology of root lifespan. In *Advances in Ecological Research.* (27th edn), Academic Press

65. Eshel, A. and Beeckman, T., eds (2013) Plant Roots: The Hidden Half (4th edn), CRC Press

66. Vance, C.P. et al. (2003) Phosphorus acquisition and use: critical adaptations by plants for securing a nonrenewable resource. New Phytol. 157, 423–447

67. Lynch, J.P. (2011) Root phenes for enhanced soil exploration and phosphorus acquisition: tools for future crops. Plant Physiol. 156, 1041–1049

68. Lynch, J.P. (2013) Steep, cheap and deep: an ideotype to optimize water and N acquisition by maize root systems. Ann. Bot. 112, 347–357

69. Lynch, J.P. et al. (2014) Root anatomical phenes associated with water acquisition from drying soil: targets for crop improvement. J. Exp. Bot. 65, 6155–6166

70. Lynch, J.P. and Wojciechowski, T. (2015) Opportunities and challenges in the subsoil: pathways to deeper rooted crops. J. Exp. Bot. Published online January 11, 2015. http://dx.doi.org/ 10.1093/jxb/eru508

71. Bengough, A.G. et al. (2011) Root elongation, water stress, and mechanical impedance: a review of limiting stresses and beneficial root tip traits. J. Exp. Bot. 62, 59–68

72. Tworkoski, T. and Fazio, G. (2015) Effects of size-controlling apple rootstocks on growth, abscisic acid, and hydraulic conductivity of scion of different vigor. Int. J. Fruit Sci. 15, 369–381

73. Segura, V. et al. (2008) Dissection apple tree architecture into genetic, ontogenetic and environmental effects: mixed linear modeling of repeated spatial and temporal measures. New Phytol. 178, 302–314

74. Segura, V. et al. (2009) Dissecting apple tree architecture into genetic, ontogenetic and environmental effects: QTL mapping. Tree Genet. Genomes 5, 165–179

75. Lauri, P-E. et al. (2011) Genetic determinism of anatomical and hydraulic traits within an apple progeny. Plant Cell Environ. 34, 1276–1290

76. Guitton, B. (2012) Genetic control of biennial bearing in an apple. J. Exp. Bot. 63, 131–149

77. Costes, E. et al. (2006) Analyzing fruit tree architecture: implications for tree management and fruit production. Hortic. Rev. 32, 1–61

78. Seleznyova, A.N. et al. (2003) Application of architectural analysis and AMAPmod methodology to study dwarfing phenomenon: the branch structure of ‘Royal Gala’ apple grafted on dwarfing and non-dwarfing rootstock/interstock combinations. Ann. Bot. 91, 665–672

79. Maguylo, K. and Lauri, P.E. (2004) Growth and fruiting characteristics of eight apple genotypes assessed as unpruned trees on ‘M. 9’ rootstock and as own-rooted

trees in southern France. In *VIII International Symposium on Canopy, Rootstocks and Environmental Physiology in Orchard Systems*, pp. 93–99, International Society for Horticultural Science

80. Willingham, S.L. et al. (2001) Rootstock influences postharvest anthracnose development in 'Hass' avocado. Aust. J. Agric. Res. 52, 1017–1022

81. Thorp, T.G. et al. (2007) Effect of inter-specific rootstocks on inorganic nutrient concentrations and fruit quality of 'Hort16A' kiwifruit (*Actinidia chinensis* Planch. var. *chinensis*). J. Hortic. Sci. Biotechnol. 82, 829–838

82. Ritenour, M.A. et al. (2004) Effect of rootstock on stem-end rind breakdown and decay of fresh citrus. Horttechnology 14, 315–319

83. Pilcher, R.L.R. et al. (2008) Genetic markers linked to the dwarfing trait of apple rootstock "Malling 9". J. Am. Soc. Hortic. Sci. 133, 100–106

84. Fazio, G. et al. (2014) Dw2, a new dwarfing locus in apple rootstocks and its relationship to induction of early bearing in apple scions. J. Am. Hortic. Soc. 139, 87–98

85. Fazio, G. et al. (2013) Phenotypic diversity and QTL mapping of absorption and translocation of nutrients by apple rootstocks. Asp. Appl. Biol. 119, 37–50

86. Estañ, M.T. et al. (2009) Identification of fruit yield loci controlling the salt tolerance conferred by *Solanum* rootstocks. Theor. Appl. Genet. 118, 305–312

87. Asins, M.J. et al. (2010) Genetic analysis of physiological components of salt tolerance conferred by *Solanum* rootstocks. What is the rootstock doing for the scion? Theor. Appl. Genet. 121, 105–115

88. Asins, M.J. et al. (2015) Genetic dissection of tomato rootstock effects on scion traits under moderate salinity. Theor. Appl. Genet. 128, 667–679

89. Bert, P.F. et al. (2013) Mapping genetic loci for tolerance to lime-induced iron deficiency chlorosis in grapevine rootstocks (*Vitis sp.*). Theor. Appl. Genet. 126, 451–473

90. Marguerit, E. et al. (2012) Rootstock control of scion transpiration and its acclimation to water deficit are controlled by different genes. New Phytol. 194, 416–429

91. Bruijn, G.D. and Dharmaputra, T.S. (1974) The Mukibat system, a high-yielding method of cassava production in Indonesia. Neth. J. Agric. Sci. 22, 89–100

92. Ahit, O.P. et al. (1981) Growth and development of cassava under the traditional and the Mukibat system of planting. Ann. Trop. Res. 3, 187–198

93. Lardizabal, R.D. and Thompson, P.G. (1990) Growth regulators combined with grafting increase flower number and seed production in sweet potato. Hortic. Sci. 25, 79–81

94. Abelenda, J.A. et al. (2011) From the model to the crop: genes controlling tuber formation in potato. Curr. Opin. Biotechnol. 22, 287–292

95. Song, G.Q. et al. (2015) Grafting of genetically engineered plants. J. Am. Soc. Hortic. Sci. 140, 203–213

96. Banerjee, A.K. et al. (2006) Dynamics of a mobile RNA of potato involved in a long-distance signaling pathway. Plant Cell 18, 3443–3457

97. Martin, A. et al. (2009) Graft-transmissible induction of potato tuberization by the microRNA miR172. Development 136, 2873–2881

98. Liu, Y. (2006) Historical and modern genetics of plant graft hybridization. Adv. Genet. 56, 101–129

99. Stegemann, S. and Bock, R. (2009) Exchange of genetic material between cells in plant tissue grafts. Science 324, 649–651

100. Liu, Y.S. et al. (2010) New insights into plant graft hybridization. Heredity 104, 1

101. Wu, R. et al. (2013) Inter-species grafting caused extensive and heritable alterations of DNA methylation in Solanaceae plants. PLoS ONE 8, e61995

102. Beveridge, C.A. et al. (1997) The shoot controls zeatin riboside export from pea roots. Evidence from the branching mutant rms4. Plant J. 11, 339–345

103. Van Norman, J.M. and Frederick, R.L. (2004) BYPASS1 negatively regulates a root-derived signal that controls plant architecture. Curr. Biol. 14, 1739–1746

104. . Domagalska, M.A. and Leyser, O. (2011) Signal integration in the control of shoot branching. Nat. Rev. Mol. Cell Biol. 12, 211–221

105. Jensen, P.J. et al. (2010) Rootstock-regulated gene expression patterns in apple tree scions. Tree Genet. Genomes 6, 57–72

106. Melnyk, C.W. and Meyerowitz, E.M. (2015) Plant grafting. Curr. Biol. 25, 183–188

107. Pant, B.D. et al. (2008) MicroRNA399 is a long-distance signal for the regulation of plant phosphate homeostasis. Plant J. 53, 731–738

108. Bhogale, S. et al. (2014) MicroRNA156: a potential graft-transmissible microRNA that modulates plant architecture and tuberization in *Solanum tuberosum* ssp. *andigena*. Plant Physiol. 164, 1011–1027

109. Jensen, P.J. et al. (2012) Rootstock-regulated gene expression patterns associated with fire blight resistance in apple. BMC Genomics 13, 9

110. Molnar, A. et al. (2010) Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells. Science 328, 872–875

111. Hamilton, A.J. and Baulcombe, D.C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286, 950–952

112. Kim, M. et al. (2001) Developmental changes due to long-distance movement of a homeobox fusion transcript in tomato. Science 293, 287–289

113. Kudo, H. and Harada, T. (2007) A graft-transmissible RNA from tomato rootstock changes leaf morphology of potato scion. Hortscience 42, 225–226

114. Thieme, C.J. et al. (2015) Endogenous Arabidopsis messenger RNAs transported to distant tissues. Nat. Plants 1, 15025

115. Kim, G. et al. (2014) Genomic-scale exchange of mRNA between a parasitic plant and its hosts. Science 345, 808–811

116. Huang, N.C. and Yu, T.S. (2009) The sequences of Arabidopsis GA-INSENSITIVE RNA constitute the motifs that are necessary and sufficient for RNA long-distance trafficking. Plant J. 59, 921– 929

117. Roy, S.J. et al. (2011) Genetic analysis of abiotic stress tolerance in crops. Curr. Opin. Plant Biol. 14, 232–239

118. Park, S.M. et al. (2005) Transgenic watermelon rootstock resistant to CGMMV (cucumber green mottle mosaic virus) infection. Plant Cell Rep. 24, 350–356

119. Smolka, A. et al. (2010) Effects of transgenic rootstocks on growth and development of non-transgenic scion cultivars in apple. Transgenic Res. 19, 933–948

120. Dutt, M. et al. (2007) Transgenic rootstock protein transmission in grapevines. Acta Hortic. 738, 749–754

121. Hemmer, C. et al. (2009) *Transgenic Rootstocks Expressing GFLV Coat Protein Gene in a Three Years Field Trial; Resistance Assessment, Impact on GFLV Diversity and Exchanges Between Rootstock and Scion,* Institut National de la Recherche Agronomique

122. Gong, X.Q. and Liu, J.H. (2013) Genetic transformation and genes for resistance to abiotic and biotic stresses in *Citrus* and its related genera. Plant Cell Tissue Organ Cult. 113, 137–147

123. Song, G.Q. et al. (2013) Engineering cherry rootstocks with resistance to *Prunus* necrotic ring spot virus through RNAi-mediated silencing. Plant Biotechnol. J. 11, 702–708

124. Aguero, C.B. et al. (2005) Evaluation of tolerance to Pierce's disease and Botrytis in transgenic plants of *Vitis vinifera* L. expressing the pear PGIP gene. Mol. Plant Pathol. 6, 43–51

125. St Laurent, A. et al. (2010) Rootstock genotype succession influences apple replant disease and root-zone microbial community composition in an orchard soil. Plant Soil 337, 259–272

126. Goh, C.H. et al. (2013) The impact of beneficial plant-associated microbes on plant phenotypic plasticity. J. Chem. Ecol. 39, 826– 839

127. Rolli, E. et al. (2015) Improved plant resistance to drought is promoted by the root-associated microbiome as a water stress-dependent trait. Environ. Microbiol. 17, 316–331

128. Barka, E.A. et al. (2002) Inhibitory effect of endophyte bacteria on Botrytis cinerea and its influence to promote the grapevine growth. Biol. Control 24, 135–142

129. Barka, E.A. et al. (2006) Enhancement of chilling resistance of inoculated grapevine plantlets with a plant growth-promoting rhizobacterium, *Burkholderia phytofirmans* strain PsJN. Appl. Environ. Microbiol. 72, 7246–7252

130. Fernandez, O. et al. (2012) Burkholderia phytofirmans PsJN acclimates grapevine to cold by modulating carbohydrate metabolism. Mol. Plant Microbe Interact. 25, 496–504

131. Theocharis, A. et al. (2012) *Burkholderia phytofirmans* PsJN primes *Vitis vinifera* L. and confers a better tolerance to low nonfreezing temperatures. Mol. Plant Microbe Interact. 25, 241–249

132. Zhang, X. et al. (2014) Water content differences have stronger effects than plant functional groups on soil bacteria in a steppe ecosystem. PLoS ONE 9, e115798

133. Winston, M.E. et al. (2014) Understanding cultivar-specificity and soil determinants of the Cannabis microbiome. PLoS ONE 9, e99641

134. Köberl, M. et al. (2011) Desert farming benefits from microbial potential in arid soils and promotes diversity and plant health. PLoS ONE 6, e24452

135. Timmusk, S. et al. (2011) Bacterial distribution in the rhizosphere of wild barley under contrasting microclimates. PLoS ONE 6, e17968

136. Marasco, R. et al. (2012) A drought resistance-promoting microbiome is selected by root system under desert farming. PLoS ONE 7, e48479

137. Viebahn, M. et al. (2005) Assessment of differences in ascomycete communities in the rhizosphere of field-grown wheat and potato. FEMS Microbiol. Ecol. 53, 245–253

138. Doornbos, R.F. et al. (2012) Impact of root exudates and plant defense signaling on bacterial communities in the rhizosphere. A review. Agron. Sustain. Dev. 32, 227–243

139. Moe, L.A. (2013) Amino acids in the rhizosphere: from plants to microbes. Am. J. Bot. 100, 1692–1705

140. Näsholm, T. et al. (2000) Uptake of organic nitrogen in the field by four agriculturally important plant species. Ecology 81, 1155– 1161

141. Uren, N.C. (2000) Types, amount, and possible functions of compounds released into the rhizosphere by soil-grown plants. In *The Rhizosphere: Biochemistry and Organic Substances at the Soil–Plant Interface* (Pinton, R. et al., eds), pp. 19–40, CRC Press

142. Reeve, J.R. et al. (2008) Soil-based cycling and differential uptake of amino acids by three species of strawberry *(Fragaria spp.)* plants. Soil Biol. Biochem. 40, 2547–2552

143. Anderson, J.P. et al. (2010) Plants versus pathogens: an evolutionary arms race. Funct. Plant Biol. 37, 499–512

144. McClure, K.A. et al. (2014) Genomics: a potential panacea for the perennial problem. Am. J. Bot. 101, 1–11

145. Myles, S. (2013) Improving fruit and wine: what does genomics have to offer? Trends Genet. 29, 190–196

146. Friesen, M.L. and von Wettberg, E.J. (2010) Adapting genomics to study the evolution and ecology of agricultural systems. Curr. Opin. Plant Biol. 13, 119–125

147. Clearwater, M.J. et al. (2004) Hydraulic conductance and rootstock effects in grafted vines of kiwifruit. J. Exp. Bot. 55, 1371–1382

148. Morton, J. (1987) Fruits of Warm Climates, J.F. Morton

149. Maranz, S. et al. (2008) Potential to harness superior nutritional qualities of exotic baobabs if local adaptation can be conferred through grafting. Agrofor. Syst. 72, 231–239

150. Ferreira-Silva, S.L. et al. (2008) Changes in physiological indicators associated with salt tolerance in two contrasting cashew rootstocks. Braz. J. Plant Physiol. 20, 51–59

151. Fu, X-Y. (2012) Effects of flooding on grafted *Annona* plants of different scion/rootstock combinations. Agric. Sci. 3, 249–256

152. George, A.P. and Nissen, R.J. (1987) Propagation of *Annona* species: a review. Sci. Agric. 33, 75–85

153. Zhou,Y.etal. (2014) Dwarfing of breadfruit *(Artocarpus altilis)* trees: opportunities and challenges. Am. J. Exp. Agric. 4, 1743–1763

154. Islam, M.M. et al. (2003) Effect of age of rootstock and time of grafting on the success of epicotyl grafting in jackfruit (*Artocarpus heterophyllus* L.). Asian J. Plant Sci. 2, 1047–1051

155. Bratsch, A. et al. (2003) Early growth characteristics of seven grafted varieties and non-grafted seedling pawpaw. Horttechnology 13, 423–427

156. Ploetz, R.C. (2004) Influence of temperature on *Pythium splendens* – induced root disease on carambola, *Averrhoa carambola*. Mycopathologia 157, 225–231

157. Mondal, T.K. (2009) Tea breeding. In *Breeding Plantation Tree Crops: Tropical Species* (Jain, S.M. and Priyadarshan, P.M., eds), pp. 545–587, Springer

158. Chan, Y-K. (2009) Breeding papaya (*Carica papaya* L.). In *Breeding Plantation Tree Crops: Tropical Species* (Jain, S.M. and Priyadarshan, P.M., eds), pp. 121–159, Springer

159. Grauke, L.J. and Thompson, T.E. (2003) Rootstock development in temperate nut crops. Acta Hortic. 622, 553–556

160. Oraguzie, N.C. et al. (1998) Examination of graft failure in New Zealand chestnut (Castanea spp.) selections. Sci. Hortic. (Amsterdam) 76, 89–103

161. Pereira-Lorenzo, S. et al. (2012) Chestnut. In *Fruit Breeding, Handbook of Plant Breeding* (Badenes, M.L. and Byrne, D.H., eds), pp. 623–662, Springer

162. Correia, P.J. et al. (2010) Tolerance of young (*Ceratonia siliqua* L.) carob rootstock to NaCl. Agric. Water Manag. 97, 910–916

163. Gmitter, F.G. et al. (2009) *Citrus* breeding. In *Breeding Plantation Tree Crops: Temperate Species* (Jain, S.M. and Priyadarshan, P. M., eds), pp. 105–134, Springer

164. Ollitrault, P. and Navarro, L. (2012) *Citrus*. In *Fruit Breeding, Handbook of Plant Breeding* (Badenes, M.L. and Byrne, D.H., eds), pp. 623–662, Springer

165. Nageswara Rao, M. et al. (2011) *Citrus*. In *Wild Crop Relatives: Genomic and Breeding Resources. Tropical and Subtropical Fruits* (Cole, K., ed.), pp. 43–59, Springer-Verlag

166. Bertrand, B. and Etienne, H. (2001) Growth, production, and bean quality of Coffea arabica as affected by interspecific grafting: consequences for rootstock breeding. Hortscience 36, 269–273

167. Molnar, T.J. (2011) *Corylus*. In *Wild Crop Relatives: Genomic and Breeding Resources. Forest Trees* (Kole, C., ed.), pp. 15–48, Springer-Verlag

168. Choo, W.K. (2000) Longan Production in Asia, Food and Agriculture Organization of the United Nations Regional Office for Asia and the Pacific

169. Koshita, Y. et al. (2006) The early growth and photosynthetic rate of Japanese persimmons (*Diospyros kaki* L.) grafted onto different interstocks. Sci. Hortic. (Amsterdam) 109, 138–141

170. Lim, T.K. (2012) *Diospyros digyna*. In: Edible Medicinal and NonMedicinal Plants: Fruits, pp. 425–427, Springer

171. Drenth, A. and Guest, D.I., eds (2004) Diversity and Management of Phytophthora in Southeast Asia, ACIAR Monograph No 114, ACIAR

172. López-Gómez, E. et al. (2007) Effect of rootstocks grafting and boron on the antioxidant systems and salinity tolerance of loquat plants (*Eriobotrya japonica* Lindl.). Environ. Exp. Bot. 60, 151– 158

173. Yakushiji, H. et al. (2012) Interspecific hybridization of fig (*Ficus carica* L.) and *Ficus erecta* Thunb., a source of *Ceratocystis* canker resistance. Euphytica 183, 39–47

174. Vahdati, K. et al. (2009) Screening for drought-tolerant genotypes of Persian walnuts (*Juglans regia* L.) during seed germination. Hortscience 44, 1815–1819

175. McGranahan, G. and Leslie, C. (2009) Breeding walnuts (*Juglans regia*). In *Breeding Plantation Tree Crops: Temperate Species* (Jain, S.M. and Priyadarshan, P.M., eds), pp. 249–273, Springer

176. Woeste, K. and Michler, C. (2011) *Juglans*. In *Wild Crop Relatives: Genomic and Breeding Resources. Forest Trees* (Cole, K., ed.), pp. 77–88, Springer-Verlag

177. Menzel, C. (2002) *The Lychee Crop in Asia and the Pacific,* Food and Agriculture Organization of the United Nations Regional Office for Asia and the Pacific

178. Hardner, C.M. et al. (2009) Genetic resources and domestication of *Macadamia*. Hortic. Rev. 35, 1–126

179. Mohandas, S. (2012) Arbuscular mycorrhizal fungi benefit mango (Mangifera indica L.) plant growth in the field. Sci. Hortic. (Amsterdam) 143, 43–48

180. Wasielewski, J. and Campbell, R.J. (2000) The potential of inverted seed grafts for dwarfing and precocity in members of the Sapotaceae grown in south Florida. Proc. Fla State Hortic. Soc. 113, 12–13

181. Bostan, S.Z. (2002) Interrelationships among pomological traits and selection of medlar. J. Am. Pomol. Soc. 56, 215–218

182. Vijayan, K. et al. (2011) Germplasm conservation in mulberry *(Morus spp.)*. Sci. Hortic. (Amsterdam) 128, 371–379

183. Mitras, K. et al. (2011) Performance of Three Rambutan Varieties (*Nephelim lappaceum* L.) on Various Nursery Media. ICRAF Working Paper Series No. 136, World Agroforestry Centre

184. Trifilò, P. et al. (2007) Rootstock effects on xylem conduit dimensions and vulnerability to cavitation of *Olea europaea* L. Trees 21, 549–556

185. Fabbri, A. et al. (2009) Olive breeding. In *Breeding Plantation Tree Crops: Tropical Species* (Jain, S.M. and Priyadarshan, P.M., eds), pp. 423–465, Springer

186. Rugini, E. et al. (2011) *Olea*. In *Wild Crop Relatives: Genomic and Breeding Resources. Temperate Fruits* (Cole, K., ed.), pp. 79– 117, Springer-Verlag

187. Estrada-Luna, A.A. et al. (2002) In vitro micrografting and the histology of graft union formation of selected species of prickly pear cactus (Opuntia spp.). Sci. Hortic. (Amsterdam) 92, 317– 327

188. Douhan, G.W. et al. (2011) Genetic diversity analysis of avocado (*Persea americana* Miller) rootstocks selected under greenhouse conditions for tolerance to *Phytophthora* root rot caused by *Phytophthora cinnamomi*. Euphytica 182, 209–217

189. Lahav, E. and Lavi, U. (2009) Avocado genetics and breeding. In *Breeding Plantation Tree Crops: Tropical Species* (Jain, S.M. and Priyadarshan, P.M., eds), pp. 247–285

190. Del Carmen Gijón, M. et al. (2010) Rootstock influences the response of pistachio (*Pistacia vera* L. cv. Kerman) to water stress and rehydration. Sci. Hortic. (Amsterdam) 125, 666–671

191. Hormaza, J.I. and Wünsch, A. (2011) *Pistacia*. In *Wild Crop Relatives: Genomic and Breeding Resources. Temperate Fruits* (Cole, K., ed.), pp. 119–128, Springer-Verlag

192. Balerdi, C.F. et al. (2006) Jaboticaba (Myrciaria cauliflora, Berg.) a delicious fruit with an excellent market potential. Proc. Fla State Hortic. Soc. 119, 66–68

193. Hernández, F. et al. (2010) Performance of *Prunus* rootstocks for apricot in Mediterranean conditions. Sci. Hortic. (Amsterdam) 124, 354–359

194. Ham, H. (2009) Apricot breeding. In *Breeding Plantation Tree Crops: Temperate Species* (Jain, S.M. and Priyadarshan, P.M., eds), pp. 83–103, Springer

195. Rato, A.E. et al. (2008) Soil and rootstock influence on fruit quality of plums (*Prunus domestica* L.). Sci. Hortic. (Amsterdam) 118, 218–222

196. Hartmann, W. and Neumüller, M. (2009) Plum breeding. In *Breeding Plantation Tree Crops: Temperate Species* (Jain, S. M. and Priyadarshan, P.M., eds), pp. 161–231, Springer

197. Yadollahi, A. et al. (2011) The response of different almond genotypes to moderate and severe water stress in order to screen for drought tolerance. Sci. Hortic. (Amsterdam) 129, 403–413

198. Gradziel, T.M. (2009) Almond *(Prunus dulcis)* breeding. In *Breeding Plantation Tree Crops: Temperate Species* (Jain, S.M. and Priyadarshan, P.M., eds), pp. 1–31, Springer

199. Tombesi, S. et al. (2011) Phenotyping vigour control capacity of new peach rootstocks by xylem vessel analysis. Sci. Hortic. (Amsterdam) 127, 353–357

200. Pommer, C. and Murakami, K. (2009) Breeding guava (*Psidium guajava* L.). In *Breeding Plantation Tree Crops: Tropical Species* (Jain, S.M. and Priyadarshan, P.M., eds), pp. 83–120, Springer

201. Nimisha, S. et al. (2013) Molecular breeding to improve guava (*Psidium guajava* L.): current status and future prospective. Sci. Hortic. (Amsterdam) 164, 578–588

202. Vazifeshenas, M. et al. (2009) Effects of different scion–rootstock combinations on vigor, tree size, yield and fruit quality of three Iranian cultivars of pomegranate. Fruits 64, 343–349

203. Stern, R.A. et al. (2013) Lavi 1 – a new *Pyrus betulifolia* rootstock for “Coscia” pear *(Pyrus communis)* in the hot climate of Israel. Sci. Hortic. (Amsterdam) 161, 293–299

204. Bell, R.L. and Itai, A. (2011) *Pyrus*. In *Wild Crop Relatives: Genomic and Breeding Resources. Temperate Fruits* (Kole, C., ed.), pp. 147–177, Springer-Verlag

205. Fischer, M. (2009) Pear breeding. *In Breeding Plantation Tree Crops: Temperate Species* (Jain, S.M. and Priyadarshan, P.M., eds), pp. 135–160, Springer

206. Mng'omba, S.A. et al. (2012) Scion and stock diameter size effect on growth and fruit production of *Sclerocarya birrea* (marula) trees. J. Hortic. For. 4, 153–160

207. El-Siddig, K. et al. (2006) Fruits for the Future 1. Revised Edition. Tamarind: *Tamarindus indica* L, International Centre for Underutilised Crops

208. Yin, J.P.T. (2004) Rootstock effects on cocoa in Sabah, Malaysia. Exp. Agric. 40, 445–452

209. Zhang, D. et al. (2011) *Theobroma*. In *Wild Crop Relatives: Genomic and Breeding Resources* (Cole, K., ed.), pp. 277– 296, Springer-Verlag

210. Monteiro, W.R. et al. (2009) Genetic improvement in cocoa. In *Breeding Plantation Tree Crops: Tropical Species* (Jain, S.M. and Priyadarshan, P.M., eds), pp. 589–626, Springer

211. Xu, C. et al. (2014) Technique of grafting with wufanshu (*Vaccinium bracteatum* Thunb.) and the effects on blueberry plant growth and development, fruit yield and quality. Sci. Hortic. (Amsterdam) 176, 290–296

212. Soneji, J.R. and Nageswara Rao, M. (2011) *Vitis*. *In Wild Crop Relatives: Genomic and Breeding Resources. Temperate Fruits* (Cole, K., ed.), pp. 223–239, Springer-Verlag

213. Reisch, B.I. et al. (2012) Grape. In *Fruit Breeding, Handbook of Plant Breeding* (Badenes, M.L. and Byrne, D.H., eds), pp. 225– 262, Springer

214. Pereira-Lorenzo, S. et al. (2009) Breeding apple (*Malus domestica* Borkh). In *Breeding Plantation Tree Crops: Temperate Species* (Jain, S.M. and Priyadarshan, P.M., eds), pp. 33–81, Springer

215. Ignatov, A. and Bodishevskaya, A. (2011) *Malus*. In *Wild Crop Relatives: Genomic and Breeding Resources. Temperate Fruits* (Cole, K., ed.), pp. 45–64, Springer-Verlag

216. Ferree, D.C. (1998) Performance of eight apomictic selections as apple rootstocks. Hortscience 33, 641–643

217. Carisse, O. and Khanizadeh, S. (2006) Relative resistance of newly released apple rootstocks to *Phytophthora cactorum*. Can. J. Plant Sci. 6, 199–204

218. Khanizadeh, S. et al. (2007) The effects of alar (acid succinamic 2,2-dimethylhydrazine) on the growth of 11 advanced apple rootstock selections. J. Food Agric. Environ. 5, 228–230

219. Atkinson, C.J. et al. (1999) Drought tolerance of apple rootstocks: production and partitioning of dry matter. Plant Soil 206, 223–235

220. Snoussi, H. et al. (2012) Assessment of the genetic diversity of the Tunisian Citrus rootstock germplasm. BMC Genet. 13, 16

221. Bally, I.S.E. et al. (2009) Mango breeding. In *Breeding Plantation Tree Crops: Tropical Species* (Jain, S.M. and Priyadarshan, P.M., eds), pp. 51–82, Springer

222. Dinesh, M.R. et al. (2011) *Mangifera*. In *Wild Crop Relatives: Genomic and Breeding Resources, Tropical and Subtropical Fruits* (Kole, C., ed.), pp. 61–74, Springer

223. Campbell, R.J. and Ledesma, N. (2013) Update on New *Mangifera* Species in Florida, USA. In *IX International Mango Symposium,* pp. 317–320, International Society for Horticultural Science

224. Elkins, R. (2012) Needs assessment for future US pear rootstock research directions based on the current state of pear production and rootstock research. J. Am. Pomol. Soc. 66, 153–163

225. Hancock, J.F. et al. (2008) Peaches. In *Temperate Fruit Crop Breeding* (Hancock, J.F., ed.), pp. 265–298, Springer

226. Byrne, D.H. et al. (2012) Peach. *In Fruit Breeding, Handbook of Plant Breeding* (Baldenes, M.L. and Byrne, D.H., eds), pp. 623– 662, Springer

227. Novaes, P. et al. (2011) Grafting for improving net photosynthesis of *Coffea arabica* in field in southeast of Brazil. Exp. Agric. 47, 53–68

228. Cohen, Y. (1991) Decline of persimmon (*Diospyros kaki* L.) trees on *Diospyros virginiana* rootstocks. Sci. Hortic. (Amsterdam) 48, 61–70

229. De Paiva, J.R. et al. (2009) Cashew (*Anacardium occidentale* L.) breeding: a global perspective. In *Breeding Plantation Tree Crops: Tropical Species* (Jain, S.M. and Priyadarshan, P.M., eds), pp. 287–324, Springer

230. Zhebentyayeva, T. et al. (2012) Apricot. In *Fruit Breeding, Handbook of Plant Breeding* (Badenes, M.L. and Byrne, D.H., eds), pp. 415–458, Springer

231. Ledbetter, C.A. (2008) Apricots. In *Temperate Fruit Crop Breeding* (Hancock, J.F., ed.), pp. 39–82, Springer Science

Tables

Table 3.1. Grafted perennial crop species and examples of selected rootstock traits

Scientific Name	Common Name	Family	Targeted trait	Refs.
Actinidia deliciosa (A.Chev.) C.F.Liang & A.R.Ferguson	Kiwifruit	Actinidiaceae	scion growth, water uptake	[150,151]
Adansonia digitata L.	Baobab	Malvaceae	nutrient content	[152]
Anacardium occidentale L.	Cashew	Anacardiaceae	salt tolerance	[153]
Annona spp.	Custard Apple	Annonaceae	flood tolerance	[151,154, 155]
Antidesma bunius (L.) Spreng.	Bignay	Phyllanthaceae	none known	[151]
Artocarpus altilis (Parkinson ex F.A.Zorn) Fosberg	Breadfruit	Moraceae	dwarfing	[156]
Artocarpus heterophyllus Lam.	Jackfruit	Moraceae	none known	[157]
Asimina triloba (L.) Dunal	Pawpaw	Annonaceae	precocity	[158]
Averrhoa carambola L.	Starfruit/ Carambola	Oxalidaceae	none known	[151,159]
Camellia spp.	Tea	Theaceae	none known	[160]
Carica papaya L.	Papaya	Caricaceae	none known	[151,161]
Carissa spp.	Carissa, Karanda	Apocynaceae	none known	[151]
Carya illinoinensis (Wangenh.) K.Koch	Pecan	Juglandaceae	nematode resistance	[46,162]
Casimiroa edulis La Llave	White Sapote	Rutaceae	none known	[151]
Castanea spp.	Chestnut	Fagaceae	graft compatibility	[162-164]
Ceratonia siliqua L.	Carob	Fabaceae	salt tolerance	[151,165]
Chrysophyllum cainito L.	Star Apple	Sapotaceae	none known	[151]
Citrus spp.	Citrus	Rutaceae	disease resistance	[58,166-168]
Coffea arabica L.	Coffee	Rubiaceae	fruit quality, production, scion growth	[169]
Corylus avellana L.	Hazelnut	Betulaceae	none known	[162,170]
Dimocarpus longan Lour.	Longan	Sapindaceae	none known	[151,171]
Diospyros kaki L.f.	Persimmon	Ebenaceae	dwarfing	[151,172]
Diospyros nigra (J.F.Gmel.) Perrier.	Black Sapote	Ebenaceae	none known	[173]
Durio spp.	Durian	Moraceae	none known	[151,174]
Eriobotrya japonica (Thunb.) Lindl.	Loquat	Rosaceae	boron and salt tolerance	[151,175]
Ficus carica L.	Fig	Moraceae	disease resistance	[176]
Fortunella spp.	Kumquat	Rutaceae	none known	[151]
Juglans regia L.	Walnut	Juglandaceae	drought tolerance	[162,177-179]
Litchi chinensis Sonn.	Lychee	Sapindaceae	none known	[151,180]
Macadamia spp.	Macadamia	Proteaceae	none known	[181]

Malpighia emarginata DC.	Barbados Cherry	Malpighiaceae	none known	[151]
Malus domestica Borkh.	Apple	Rosaceae	mineral uptake, scion growth	[61]
Mangifera indica L.	Mango	Anacardiaceae	root microbe interactions	[182]
Manilkara zapota (L.) P.Royen	Sapodilla	Sapotaceae	dwarfing, precocity	[151,183]
Melicoccus bijugatus Jacq.	Mamoncillo	Sapotaceae	none known	[151]
Mespilus germanica L.	Medlar	Rosaceae	fruit yield, quality	[184]
Morus alba L.	Mulberry	Moraceae	none known	[185]
Nephelium lappaceum L.	Rambutan	Sapindaceae	rootstock growth	[151,186]
Nephelium mutabile Blume	Pulasan	Sapindaceae	none known	[151]
Olea europaea L.	Olive	Oleaceae	drought tolerance, dwarfing	[187-189]
Opuntia ficus-indica (L.) Mill.	Opuntia	Cactaceae	graft compatibility	[190]
Passiflora edulis Sims	Passionfruit	Passifloraceae	none known	[151]
Persea americana Mill.	Avocado	Lauraceae	disease resistance	[161,191,192]
Pistacia vera L.	Pistachio	Anacardiaceae	drought tolerance	[46,193,194]
Plinia cauliflora (Mart.) Kausel	Jaboticaba	Myrtaceae	none known	[195]
Pouteria spp.	Canistel, Mamey Sapote	Sapotaceae	dwarfing, precocity	[151,183]
Prunus armeniaca	Apricot	Rosaceae	fruit yield, and quality	[196,197]
Prunus domestica L.	Plum	Rosaceae	fruit quality	[198,199]
Prunus dulcis (Mill.) D.A.Webb	Almond	Rosaceae	drought tolerance	[46,200, 201]
Prunus persica (L.) Batsch	Peach	Rosaceae	dwarfing	[151,202]
Prunus avium (L.) L., *P. cerasus* L.	Cherry	Rosaceae	fruit size, quality, and yield, scion vigor	[59]
Psidium guajava L.	Guava	Myrtaceae	none known	[151,203,204]
Punica granatum L.	Pomegranate	Lyrthraceae	fruit quality, yield, scion vigor	[205]
Pyrus communis L.	Pear	Rosaceae	fruit yield and production, heat tolerance	[206-208]
Quararibea cordata (Bonpl.) Vischer	Chupa-chupa	Bombacaceae	none known	[151]
Sandoricum koetjape (Burm.f.) Merr.	Santol	Meliaceae	none known	[151]
Sclerocarya birrea (A.Rich.) Hochst.	Marula	Anacardiaceae	fruit production, rootstock growth	[209]
Spondias dulcis Parkinson	Ambarella	Anacardiaceae	none known	[151]
Tamarindus indica L.	Tamarind	Fabaceae	none known	[151,210]
Theobroma cacao L.	Cocoa	Malvaceae	yield	[211-213]

Vaccinium spp.	Blueberries	Ericaceae	precocity, scion vigor	[214]
Vitis vinifera L.	Grape	Vitaceae	drought tolerance	[41,215-218]
Ziziphus spp.	Jujube	Rhamnaceae	none known	[151]

Table 3.2. The twenty most-produced, grafted, woody[1] perennial crop species, rootstock species used for their cultivation, and rootstock traits targeted during selection

Common name Primary species used as scion (Family)	Rootstock species	Method of rootstock propagation[1]	Primary targets of rootstock selection	Estimated global production (tonnes per year)[2]	Refs.
Apple *Malus domestica* Borkh.(Rosaceae)	*M. baccata, M. domestica, M. doumeri, M. halliana, M. hupehensis, M. sargenti, M. sieboldii, M. sieversii, M. sikkimensis, M. sylvestris, M. transitoria, M. toringoides, M. yunnanensis*	clonal	scion architecture and morphology, size control/dwarfing fruit quality, disease/pest resistance; abiotic tolerance: drought cold, soil conditions	80,822,521	[219–244]
Grape *Vitis vinifera* L. (Vitaceae)	*V. aestivalis, V. berlandieri, V. californica, V. labrusca, V. rotundifolia, V. rupestris, V. vinifera, V. vulpina*	clonal	scion vigor, disease/pest resistance; abiotic tolerance: drought, salt, acidic soils, iron chlorosis	77,181,122	[215-217]
Orange *Citrus x aurantium* L. *C. sinensis* (L.) Osbeck (Rutaceae)	*C. x aurantium, C. aurantifolia, C. jambhiri, C. limon, C. reticulata; hybrids of: C. paradisi, C. reshni, C. sinensis, C. trifoliata, C. volkameriana*	polyembryony (clonal)	scion architecture, size control/dwarfing, fruit quality, rapid growth, polyembryony, disease/pest resistance; abiotic tolerance: drought, cold, salt, flooding	71,445,353	[166-168,225]
Mango *Mangifera indica* L. (Anacardiaceae)	*M. indica, M. casturi (trials only)*	polyembryony (clonal), seed	size control/dwarfing, graft compatibility, polyembryony; abiotic tolerance: calcareous soil, salt	43,300,070[3]	[226-228]
Tangerine, mandarin *Citrus reticulata* Blanco (Rutaceae)	*C. x aurantium, C. aurantifolia, C. jambhiri, C. limon, C. reticulata; hybrids of: C. paradisi, C. reshni, C. sinensis, C. trifoliata, C. volkameriana*	polyembryony (clonal)	size control/dwarfing, rapid growth, scion architecture, polyembryony, disease/pest resistance; abiotic tolerance: cold, drought, salt, flooding,	28,678,214	[166,167, 225]

Pear *Pyrus communis* L. (Rosaceae)	*Amelanchier spp., Crataegus spp., Cydonia oblonga, P. amygdaliformis, P. betulifolia, P. calleryana, P. caucasica, P. communis, P. cordata, P. elaeagnifolia, P. kawakamii, P. nivalis, P. pashia, P. pyrifolia, P. syriaca, P. ussuriensis, P. xerophila, and Sorbus spp., hybrids of: C. oblonga, P. bretschneideri, P. elaeagnifolia, P. heterophylla, P. longipes, P. nivalis, P. pyrifolia, P. sinaica, P. ussuriensis*	seed, clonal	size control/dwarfing, precocity, productivity, yield, fruit quality, fruit size, ease of clonal propagation, disease resistance, graft compatibility; abiotic tolerance: cold, iron and calcium chlorosis	25,203,754	[207,208, 229]
Peach *Prunus persica* (L.) Batsch (Rosaceae)	*P. cerasifera, P. davidiana, P. dulcis, P. ferganensis, P. insititia, P. kansuensis, P. mira, P. persica, P. pumila, P. salicina, P. spinosa; hybrids of P. angustifolia, P. besseyi, P. cerasifera, P. davidiana, P. dulcis, P. persica, P. salicina, P. spinosa*	seed, clonal	size control/dwarfing, ease of vegetative propagation, graft compatibility, disease/pest resistance, abiotic tolerance: drought, cold, anaerobic soil conditions, flooding, iron chlorosis, calcareous and compact soils	21,638,953	[230, 231]
Olive *Olea europaea* L. (Oleaceae)	*O. europaea*	seed, clonal	size control/dwarfing, rooting ability, graft compatibility, disease resistance; abiotic tolerance: drought, salt	20,396,700	[188, 189]
Lemon and lime *Citrus limon* (L.) Osbeck, *C. aurantifolia* (Cristm.) (Rutaceae)	*C. x aurantium, C. aurantifolia, C. jambhiri, C. limon, C. reticulata; hybrids of: C. paradisi, C. reshni, C. sinensis, C. trifoliata, C. volkameriana*	polyembryony (clonal)	scion architecture, size control/dwarfing, rapid growth, polyembryony, disease/pest resistance; abiotic tolerance: cold, drought, salt, flooding	15,191,482	[166-168,225]
Papaya *Carica papaya* L.[4] (Caricaceae)	*C. papaya*	?	fruit quality	12,420,585	[161]

Plum and sloe *Prunus domestica* L. (*P. spinosa* L. *P. x cerasifera* Ehrh.) (Rosaceae)	hybrids of: *P. americana, P. armeniaca, P. besseyi, P. cerasifera, P. domestica, P. dulcis, P. hortulana, P. insititia, P. munsoniana, P. persica, P. pumila, P. salicina, P. spinosa, P. tomentosa*	seed, clonal	scion vigor and architecture, size control/dwarfing, precocity, graft compatibility, ease of clonal propagation, nutrient uptake, disease/pest resistance; abiotic tolerance: cold, calcareous soils, drought, flooding	11,528,337	[199]
Coffee *Coffea arabica* L., *C. canephora* var. robusta (L. Linden) A. Chev.(Rubiaceae)	*C. canephora, C. liberica, C. liberica var. dewevrei*	seed	fruit quality, growth, production, pest resistance, drought tolerance	8,920,840	[22,169,232]
Grapefruit *Citrus paradisi* Macfad.(Rutaceae)	*C. x aurantium, C. aurantifolia, C. jambhiri, C. limon, C. reticulata; hybrids of: C. paradisi, C. reshni, C. sinensis, C. trifoliata, C. volkameriana*	polyembryony (clonal)	scion architecture, size control/dwarfing, rapid growth, polyembryony, disease/pest resistance, abiotic tolerance: cold, drought, salt, flooding	8,453,446	[166-168,225]
Tea *Camellia sinensis* L. (Theaceae)	*C. sinensis, C. irrawadiensis, C. taliensis*	clonal	high production, drought tolerance	5,345,523	[160]
Avocado *Persea americana* Mill. (Lauraceae)	*P. americana*	clonal, (seed)	precocity, disease resistance, salt tolerance	4,717,102	[192]
Persimmon *Diospyros kaki* L.f.(Ebenaceae)	*D. rhombifolia* (as interstock), *D. virginiana*	seed	size control/dwarfing, graft compatibility	4,637,357	[172,233]
Cocoa *Theobroma cacao* L. (Malvaceae)	*T. cacao*	clonal	size control/dwarfing, cultivation density, disease resistance	4,585,552	[212, 213]
Cashew nut *Anacardium occidentale* L. (Anacardiaceae)	*A. occidentale*	seed	size control/dwarfing, precocity	4,439,960	[234]

Apricot *Prunus armeniaca* L. (Rosaceae)	*P. armeniaca, P. cerasifera, P. domestica, P. mume, P. persica*, interspecific hybrids thereof	seed, (clonal)	scion vigor, fruit size, yield, tree longevity, precocity, rootstock vigor, graft compatibility, disease/pest resistance; abiotic tolerance: salt, cold	4,111,076	[197,235, 236]
Walnut *Juglans regia* L. (Juglandaceae)	*J. hindsii, J. major, J. mandshurica, J. microcarpa, J. nigra;* hybrids of *J. nigra* and *J. hindsii, Pterocarya stenoptera*	seed, (clonal)	disease resistance; abiotic tolerance: salt, acidic soils	3,458,046	[178, 179]

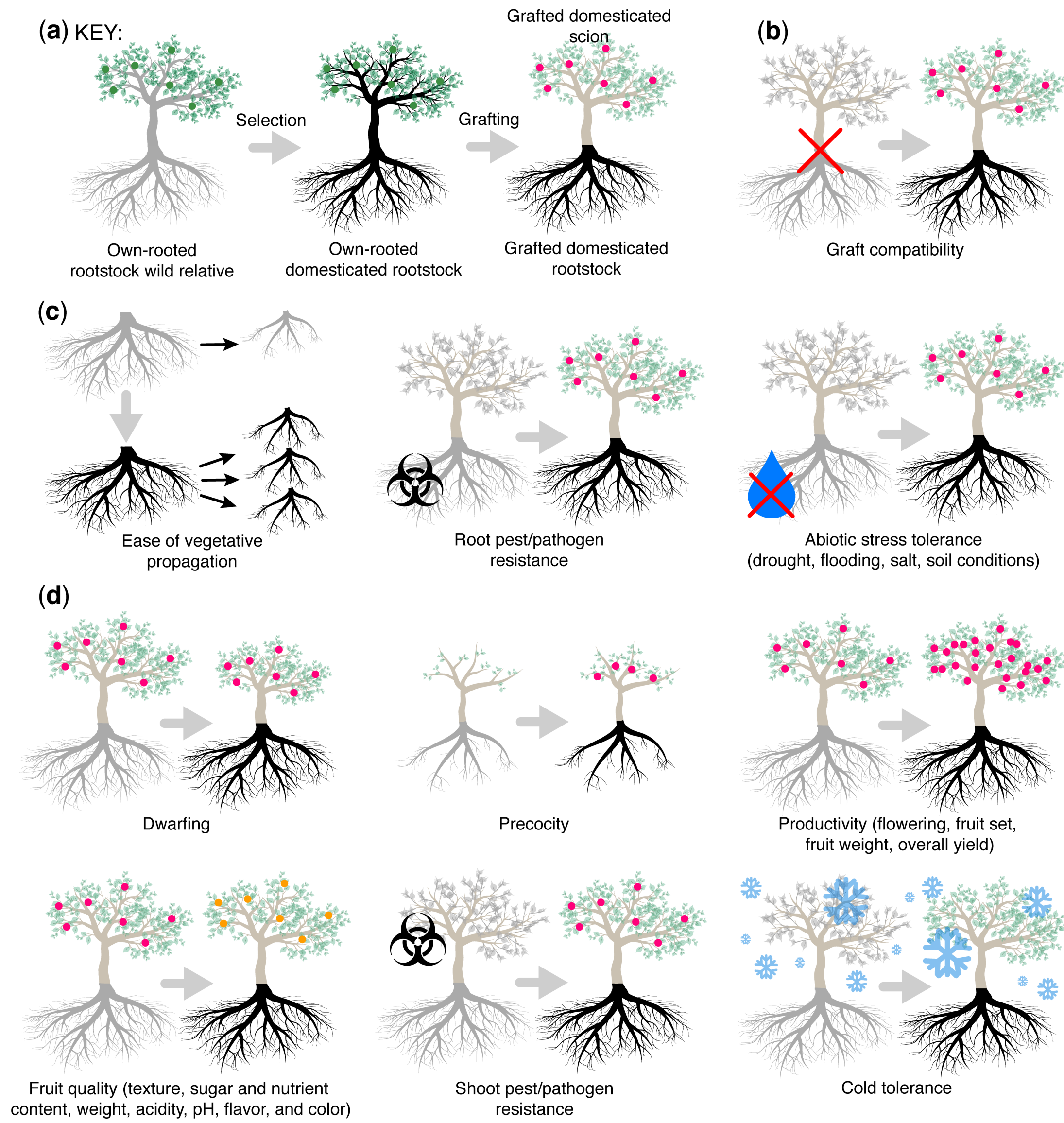
[1]clonal = asexually produced, seed = sexually produced, polyembryonic = from clonal embryos
[2][26]
[3] Estimated tonnage for *Mangifera indica, Garcinia mangostana,* and *Psidium guajava* combined
Carica papaya is an herbaceous perennial, but is included here because cultivation and production practices are similar to that of woody perennials

Figure Captions

Figure 3.1, Key Figure. Primary targets of rootstock selection. Rootstocks used in perennial agriculture (A) have been selected from a pool of wild germplasm and bred for (B) their ability to graft to cultivated scions, (C) the root phenotype, and (D) their ability to impact the phenotype of the grafted scion.

Figures

Figure 3.1



Appendices Captions

3.1. Trends Box
3.2. Outstanding Questions Box

Appendices

Appendix 3.1.

Trends

As concerns mount about food secur- ity in a changing climate, attention is refocusing on perennial crops as important components of sustainable agriculture.

In many economically important woody perennial crops (e.g., many Rosaceae, Citrus, and grapes), a fruit-bearing shoot (scion) is grafted to a root system (rootstock) that is genetically distinct from the scion.

Rootstocks are selected for rooting and grafting capacity, abiotic and biotic stress tolerance, and their ability to beneficially alter scion phenotypes.

Relatively little is known about the diversity of rootstocks used for any given crop, the geographic origins or current distribution of cultivated root- stocks, or their domestication.

A common scion can be grafted to segregating rootstock populations to produce a genetic map of both the traits of the rootstocks themselves and their effects on scion phenotype.

Appendix 3.2.

Outstanding Questions

For a given woody perennial crop, what is the domestication history of its rootstocks? Which wild species contributed to the germplasm of cultivated rootstocks and what is their geographic distribution in nature?

What are the genetic underpinnings of phenotypic variation observed in the rootstock itself (e.g., root architecture, abiotic and biotic stress tolerance) and graft-transmissible effects on the scion? Are there scion-modulated traits in the rootstock and, if so, what is their genetic basis?

How are the genetic and phenotypic interactions between rootstock and scion affected by the environment?

What are the molecular signals (e.g., transcripts, sRNAs, proteins, metabolites, hormones) underlying graft-transmissible phenotypes? How far can large portions of DNA traverse the graft junction?

Is Darwin's concept of graft hybridization explained by epigenetics? Does grafting induce heritable epigenetic changes that alter important agronomic traits?

To what extent does the soil microbiome impact rootstock function and scion phenotype? Does the rootstock influence scion microbiome composition?

CHAPTER IV

A FRUITFUL PHYLOGENY: RAD-SEQ REVEALS THE EVOLUTIONARY RELATIONSHIPS OF CULTIVATED AND WILD SPECIES IN THE MANGO GENUS *(MANGIFERA,* ANACARDIACEAE)

Abstract

The field of population genomics has long capitalized on domesticated species, which provide tractable systems in which to study evolutionary phenomena, but few studies use comparative phylogenetic methods to understand the evolution of domesticated lineages. We use restriction site associated DNA sequencing (RADseq) to estimate the phylogenetic relationships of *Mangifera* (Anacardiaceae), a genus of tropical trees that includes mango, *M. indica,* and multiple other cultivated species. We also explore the effects of intraspecific resampling and different bioinformatic parameters that are considered to be important for RADseq, such as clustering threshold and the permissible amount of missing data, on the resulting dataset and downstream topology. We present the first multilocus phylogenetic hypothesis for *Mangifera,* which reveals that the genus, as traditionally circumscribed, is not monophyletic. Additionally, we find that the five major clades we recover can be characterized by differences in fruit morphology, and each contains both cultivated and wild species. We find that the level of missing data allowed within a dataset has the greatest impact on the number of loci recovered and on topology, while clustering threshold has less of an impact. Together, our results demonstrate that RADseq is an effective tool for phylogenetic study of non-model systems and that *Mangifera* represents a unique system in which to study the evolution of closely related cultivated tree species.

Introduction

Domesticated species were foundational to Darwin's formulation of the theory of evolution by natural selection (Darwin 1859). With over 2,500 species domesticated from 160 families across the plant kingdom (Meyer et al. 2012), crops provide systems in which to address questions occurring at both shallow and deep timescales. However, while the field of population genetics continues to take advantage of crop systems to study evolutionary processes (e.g., Arnold 2004; Kovach et al. 2007; Purugganan and Fuller 2009; Meyer and Purugganan 2013; Olsen and Wendel 2013; Washburn et al. 2016), the field of phylogenetics has yet to capitalize on domesticated systems in the same manner. Rather than using domesticated systems for comparative evolutionary analyses, most phylogenetic studies of crops use phylogenetic tools to inform crop breeding or provide a foundational understanding of the origin of domesticated species (e.g., Weese and Bohs 2010; Chomicki and Renner 2014; Hawkins et al. 2015; Wong et al. 2015; Rendón-Anaya et al. 2017). These efforts are critically important to crop improvement and should not be discounted, but they leave tantalizingly rich questions about the process of domestication across parallel lineages and at deeper timescales unanswered.

Fundamentally, domestication is a case of co-evolution and mutualism between humans and crop species that results in a spectrum of levels of intensity of cultivation and breeding (e.g., Clement 1999; Zeder 2006; Pickersgill 2007). The domestication syndrome, a suite of characters associated with and acquired by plants during the process of domestication, can be considered a complex phenotype (Washburn et al. 2016). Recently, phylogenetic studies elucidated the stepwise fashion of other complex

phenotypes, like the evolution of CAM photosynthesis (Hancock and Edwards 2014) and could be similarly beneficial in understanding the process of crop domestication at the molecular and genetic levels. Domesticated species are particularly amenable to studies of stepwise evolution because they often have extant wild progenitor populations, as well as both semi-domesticated and wild relatives, yet few studies investigate the evolutionary relationships between these wild, semi-domesticated, and domesticated congeners (e.g., Sanjur et al. 2002; Wang et al. 2017). With recent research highlighting the different trajectories of perennial and annual species under domestication (Miller and Gross 2011; Gaut et al. 2015), there is a clear need for new, diverse systems in which to explore the process of crop domestication in long-lived species. Here, we provide a phylogenetic hypothesis for a genus of tropical fruit trees, taking a key step toward comparative domestication studies in perennial crop systems.

The genus *Mangifera* includes 69 species of tropical trees (Kostermans and Bompard 1993), making it the third largest genus in the poison ivy family Anacardiaceae (Pell et al. 2011). The cultivated mango, *Mangifera indica*, is by far the most well known species in the genus. The mango was likely domesticated in India some 4,000 years ago (Mukherjee 1949), and today mangoes are one of the most important tropical fruit crops in the world (FAO 2003; FAOSTAT 2013). The native range of *Mangifera* stretches from India eastward to the Solomon Islands, with the highest diversity present in the Malay Peninsula and the Malesian islands of Borneo and Sumatra (Kostermans and Bompard 1993). Of the 69 species of *Mangifera*, all produce fleshy fruits that range from red, egg-sized drupes that are consumed by hornbills and other birds (e.g., *M. griffithii*) to mango-like yellow fruits that are a favorite of primates including orangutans (e.g., *M.*

pentandra), to large, green- or brown-skinned fruits that are eaten by elephants and rhinos (e.g., *M. pajang*) (Phillipps and Phillipps 2016). In addition to being a favorite of iconic Malesian megafauna, 27 species of *Mangifera* are consumed by humans. These species fall along the wild-domesticated continuum (Kostermans and Bompard 1993) that includes wild-harvested, incipiently domesticated, semi-domesticated, and domesticated species.

Within the family Anacardiaceae, *Mangifera* has traditionally been placed as sister to the small Southeast Asian genus *Bouea*, which contains 3 species (Pell et al. 2011). Recent phylogenetic analysis placed *Mangifera* and *Bouea* in clade IV of the tribe Anacardioideae, with close relatives including two Southeast Asian genera, *Gluta* and *Swintonia,* the African genus *Fegimanra,* and the Neotropical genus *Anacardium* (Weeks et al. 2014). The majority of *Mangifera* species are large trees, reaching up to 30 m in height, and growing scattered throughout the lowland evergreen tropical forests of South and Southeast Asia (Kostermans and Bompard 1993). Most species flower irregularly, and a few species, such as *M. lagenifera,* are said to flower only once every 5-10 years (Kostermans and Bompard 1993). The combination of *Mangifera* species' life history and ecology makes it difficult to locate, identify, collect, and study them in the wild. As a result, these species are underrepresented in both living and herbarium collections, and while much research has sought to improve our understanding of cultivated mango (e.g., Surapaneni et al. 2013; Ravishankar et al. 2015; Sherman et al. 2015; Singh 2016; Kuhn et al. 2017), relatively little research has investigated any elements of other *Mangifera* species. Notably, recent works highlight the economic botany of some *Mangifera* species including *M. sylvatica, M. pentandra, M. caloneura, and M. linearifolia* (Ueda et al.

2011, 2015, 2016; Akhter et al. 2016; Baul et al. 2016). The most recent and most comprehensive taxonomic treatment of *Mangifera* by Kostermans and Bompard (1993) used morphological characters to propose an infrageneric classification for the genus with two subgenera and a total of six sections (Fig. 4.1). Previous efforts to produce a hypothesis of phylogenetic relationships used AFLPs and locus-by-locus sequencing in as many as 19 species, but none included representatives of each of the proposed infrageneric groups, and all failed to provide well-resolved and well-supported phylogenies (Eiadthong et al. 1999, 2000; Yonemori et al. 2002; Yamanaka et al. 2006; Fitmawati and Hartana 2010; Hidayat et al. 2011; Suparman et al. 2013; Dinesh et al. 2015; Fitmawati et al. 2017).

RADseq In Phylogenetics

Restriction-site associated DNA sequencing (RADseq) was first developed as a tool for acquiring reduced representation population genomic data from non-model organisms (Miller et al. 2007; Baird et al. 2008; Davey et al. 2010). To this end, it has been used to investigate population structure, adaptation, demographics, hybridization and introgression, and biogeography (e.g., Emerson et al. 2010; Hohenlohe et al. 2010, 2011; Hess et al. 2012; The Heliconius Genome Consortium 2012; Corander et al. 2013; Combosch and Vollmer 2015). More recently, RADseq has proven to be an innovative and effective tool for resolving the phylogenetic relationships of previously intractable groups of closely related species, particularly recent radiations (e.g., Wagner et al. 2013; Wang et al. 2013; Eaton and Ree 2013; Jones et al. 2013; Nadeau et al. 2013; Cruaud et al. 2014; Hipp et al. 2014; DaCosta and Sorenson 2016; Paun et al. 2016; Díaz-Arce et al.

2016; Herrera and Shank 2016; Massatti et al. 2016; Tripp et al. 2017; Vargas et al. 2017; Fernández-Mazuecos et al. 2017). Yet RADseq seems to be effective in still deeper time-scales: in silico tests suggest it is effective in resolving infrageneric relationships in *Drosophila* up to 60 Ma (Rubin et al. 2012; Cariou et al. 2013) and in situ studies used RADseq to successfully resolve relationships in 30-55 Ma clades (Leaché et al. 2015; Liu et al. 2015; Eaton et al. 2017; McVay et al. 2017).

The transition of RADseq from population genomic to phylogenomic applications has not occurred without some criticisms, which have been discussed in detail by multiple authors (Takahashi et al. 2014; Leaché et al. 2015; Ree and Hipp 2015; Huang and Knowles 2016; Rivers et al. 2016; Eaton et al. 2017; Leaché and Oaks 2017), and are briefly presented here. One of the primary concerns with RADseq is the amount of missing data present in a dataset and its effect on downstream analyses (Leaché et al. 2015; Huang and Knowles 2016; Eaton et al. 2017; Leaché and Oaks 2017; Tripp et al. 2017). Allele dropout, or the lack of data at a particular locus for one or more individuals, occurs frequently in RADseq datasets, both randomly as a result of low sequence coverage and systematically across phylogenetic distance as a result of the accumulation of mutations in restriction enzyme sites (Leaché et al. 2015; Eaton et al. 2017). Counter-intuitively, removing missing data from a RADseq dataset has been shown to decrease the phylogenetic informativeness of the dataset (Leaché et al. 2015; Huang and Knowles 2016; Eaton et al. 2017; Tripp et al. 2017).

Another controversial issue for RADseq and similar methods, like genotyping by sequencing (GBS), is orthology estimation. Following sequencing, millions of raw reads from each individual must be clustered into loci according to sequence identity, then

these loci must be clustered across individuals to identify regions that will be considered orthologous. If the clustering threshold is too low, there is a risk that paralogous sequences will be clustered together (referred to as undersplitting), whereas if the threshold is too high, orthologous sequences will be treated as distinct loci (oversplitting). There is no objective way to select a clustering threshold (Ree and Hipp 2015) though some guidelines have been proposed for population-level analysis (Paris et al. 2017). Therefore, the effect of clustering threshold on downstream analyses is still a topic of some debate.

To date, RADseq has primarily been used for phylogenomic analysis in relatively small taxonomic groups, often including fewer than 20 taxa and 50 samples total. One recent exception is the work of Singhal et al. (2017), who used RADseq to investigate diversity from over 500 individuals across 76 taxa of *Ctenotus* lizards. As RADseq becomes a more popular phylogenomic tool, it is important to understand how taxon re-sampling and the overall number of individuals included in an analysis interact with bioinformatic parameters to impact downstream analyses.

Here, we present the first multilocus phylogenetic hypothesis of the tropical perennial fruit genus, *Mangifera*, and demonstrate its promise as a novel system in which to study domestication. We also explore how the common RADseq bioinformatic parameters, including the number of samples in a dataset, the clustering threshold, and the amount of missing data impact the resulting dataset (including the number of loci and parsimony informative sites retained) and affect subsequent tree topology. Our phylogenetic hypothesis of *Mangifera* provides a framework for future investigations of character evolution, biogeography, and domestication events in this fruitful genus.

Materials and Methods

Sampling

Leaves of 280 individuals of *Mangifera* and outgroup taxa were collected from herbarium or living specimens at botanic gardens, private collections, and forest plots in Singapore, Malaysia, Indonesia, and the United States. Tissue was frozen at -80ºC or dried in silica and stored at 4ºC. When possible, living specimens were vouchered and deposited as herbarium specimens. Nearly all specimens were collected in sterile condition, and most accessions had no previous fertile vouchers to verify identifications. Therefore, identification was often determined by leaf characters such as venation patterns, which are distinctive in many taxa (Kostermans and Bompard 1993). Some samples were removed from the final dataset because of poor sequencing results or to reduce duplicate sampling within species. In total, 201 individuals representing approximately 36 taxa were used for phylogenetic analysis (Table 4.1, Appendix 4.1).

Library preparation

The ddRADseq libraries were prepared according to the protocol of Peterson et al. (2012). Genomic DNA was extracted from samples using a Qiagen DNEasy plant mini kit or a modified CTAB protocol (Doyle and Doyle 1990). Following initial restriction enzyme trials, for each individual, 300-1000 ng high molecular weight genomic DNA was digested with NlaIII and MluCI (New England Biolabs). Prior to pooling groups of 8 individuals together, custom designed barcoded adapters were ligated onto each sample (Appendix 4.2). To target 350 bp inserts, size selection of sublibraries was performed on a Pippin Prep (Sage Science) using an external marker with a tight size selection at 425

bp. Sublibraries were amplified using short-cycle PCR in six separate reactions to reduce PCR bias. For each sublibrary, one of 12 unique DNA indices was added during PCR (Appendix 4.3). Amplified sublibraries were quality-checked with an Agilent Bioanalzyer DNA High Sensitivity Chip. If overamplification of a sublibrary was observed, Pippin Prep size selection was performed to remove non-target DNA. Libraries of 96 pooled individuals were sequenced in individual lanes of 150 bp rapid runs on Illumina HiSeq 2500 at the University of Southern California's Genome and Cytometry Core.

Bioinformatics

Bioinformatic analysis was performed on Florida International University's high performance computing cluster (FIU HPCC). Raw fastq files for each sublibrary were analyzed using the program FASTQC v.0.11.4 (Andrews 2010) to check for overall quality.

The bioinformatic pipeline ipyRAD v.0.6.11 (Eaton 2014) was used to process raw reads and call single nucleotide polymorphisms (SNPs). The pipeline was run under default conditions with the following exceptions: one mismatched base was allowed within the barcode sequence, the *adapter filtering* parameter was set to 2, and the maximum read depth for loci, *maxdepth,* was set to 1000. To explore the effect of clustering threshold on the dataset recovered and on tree topology, de novo clustering of reads was performed at three different thresholds (0.85, 0.90, 0.95), using the same value for both within and between sample clustering. To determine how the amount of missing data in a dataset affects both the composition of the dataset and downstream topology,

each dataset was filtered by the minimum number of individuals required to have data at a locus (*min_samples_locus,* or *m*) at four different values of *m:* 4 individuals (the minimum number required to obtain phylogenetic information, Eaton et al. 2017), and 20%, 50%, and 90% of the total number of individuals in a dataset. Percentages, rather than specific numbers, were selected in order to obtain subsets (see below) with comparable levels of missing information. These *m20%, m50%,* and *m90%* values constrain the maximum percent of missing data to 80%, 50%, and 10% (but see Results for actual levels of missing data). For each set of parameters, a supermatrix of full-length reads (including SNPs and indels) and a matrix of unlinked SNPs was produced. The described analysis was performed for subsets of 201, 98, and 41 individuals *(phyloBIG, phyloMED, phyloSM*, respectively), which differed in the number of replicate samples included within species (Table 4.1). The resulting 36 unique datasets of full-length reads were analyzed to examine the total number of loci recovered, the percent of missing data in a dataset, the adjusted average number of parsimony informative sites per informative site (adjusted for invariable loci), and the adjusted average number of variable sites per informative locus (adjusted for invariable loci).

The genome of the mango is relatively small (haploid genome 439 Mb) (Arumuganathan and Earle 1991), and all individuals that have been tested (6 species, including 25 cultivars of *M. indica*) are of the same ploidy level, with 2n=40 chromosomes (Mukherjee 1950a, 1950b, 1953, 1957). While some cytological evidence suggests that *M. indica* may be of (neo)allopolyploid origin (Mukherjee 1950b), this conclusion has been disputed (Viruel et al. 2005; Iyer and Schnell 2009; Arias et al.

2012). We did not find significant evidence of polyploidy in our dataset, and therefore, in the present study we treat *Mangifera* as diploid.

Maximum Likelihood (RAxML)

Simulation studies show that maximum likelihood (ML) analysis of large concatenated RADseq matrices provide robust species trees that are consistent with those produced using other markers (Rivers et al. 2016), but using full sequences rather than SNPs alone is preferable for obtaining accurate branch lengths and topologies (Leaché et al. 2015). Therefore, supermatrices of concatenated full-length loci for 30 of the 36 datasets were analyzed in a ML framework with RAxML (Stamatakis 2014) using the BFGS (Broyden-Fletcher-Goldfarb-Shanno) method to optimize GTR rate parameters, and executing 1,000 (or 100 for the three largest datasets, *phyloBIG_c85/90/95_m20%*) bootstrap replicates followed by a thorough maximum likelihood search. Datasets were analyzed on the RAxML-HPC2 on XSEDE Workflow available on the CIPRES portal (Miller et al. 2010). Six datasets, *phyloBIG_c85/90/95_m4* and *phyloMED_c85/90/95_m4*, were prohibitively large, with over 50,000 loci each, and therefore were not analyzed. Phylogenetic trees were visualized and inspected in FigTree v. 1.4.3 (Rambaut 2006) and support values were verified using the R package ape v. 4.1 (Paradis et al. 2004) to ensure correct placement after rooting (Czech et al. 2017).

To examine topological differences between ML phylogenies produced from 30 different datasets, weighted normalized Robinson Foulds distances (Robinson and Foulds 1981) and branch score distances (Kuhner and Felsenstein 1994) were calculated in the R

package phangorn v. 2.20 (Schliep 2011) and visualized using multidimensional scaling (MDS) in the R package ape v. 4.1 (Paradis et al. 2004).

SVD Quartets

Supermatrices of unlinked SNPs (one per locus) were analyzed using SVD Quartets, a computationally efficient method that infers the relationships between quartets of taxa under the coalescent model (Chifman and Kubatko 2014). For the *phyloSM_c85_m4/8/21* datasets, all possible quartets were analyzed with 100 nonparametric bootstrap replicates. Bootstrap support was mapped onto the best quartet tree using sumtrees v.4.0.0 (Sukumaran and Holder 2010, 2015).

Results

Sequencing Results & Data Processing

A total of 454,840,461 raw reads were obtained for 280 individuals across three lanes of sequencing, with 244,672,938 reads for the 201 individuals analyzed here (average reads per individual: 1,217,278; standard deviation: 551,095; for individual results see Appendix 4.1). FastQC analysis of each sublibrary indicated high per-base sequence quality across the entire 150 bp length.

Effect of Bioinformatic Parameters

For each of the 36 different datasets, we calculated five summary statistics: 1) the total number of loci recovered, 2) the percent of SNPs that are parsimony informative, 3) the average number of variable sites per informative locus, 4) the average number of

parsimony informative sites per informative locus, and 5) the percent of missing loci in a dataset (Table 4.2, Fig. 4.2). Changes in the subset of individuals analyzed (*phyloBIG/MED/SM*) and the parameter settings (*c85/90/95, m4/20/50/90%*) greatly influenced the size and amount of missing data in resulting datasets. The number of loci recovered ranged from 208 *(phyloSM_c85_m90%)* to 143,628 *(phyloBIG_c95_m4)* while the amount of missing data ranged from a low of 5.31% *(phyloBIG_c90_m90%)* to a high of 92.71% *(phyloBIG_c90_m4)*. The combination of the subset of individuals and parameter settings also affected the average number of variants per locus (3.50-11.73, *phyloBIG_c90_m4* and *phyloBIG_c90_m50%* respectively), the percentage of variants that were parsimony informative (31.56-62.45%, *phyloSM_c85_m20%* and *phyloBIG_c90_m4* respectively), and the number of parsimony informative sites per locus (1.52-6.99, *phyloSM_c90_m90%* and *phyloBIG_c90_m50%,* respectively).

Of the three parameters tested, the minimum number of individuals required to have data for a locus (*m*) had the greatest impact on the number of reads recovered and the percent of missing data: the nine *m4* datasets had the most loci recovered and the highest percentages of missing data, followed by the nine *m20%* datasets, the nine *m50%* datasets, and the nine *m90%* datasets. The average variants per locus and average parsimony informative sites per locus were generally lowest in the *m90%* and *m4* datasets and highest in the *m20%* and *m50%* datasets. The datasets with the highest percentages of variants that were parsimony informative were *phyloBIG* datasets with *m20/50/90%*, while the lowest were *phyloSM* datasets with *m4/20/50%.* The *phyloSM_c85* supermatrices analyzed with SVD Quartets were comprised of unlinked SNPs and were 20,140 (*m4*), 7,967 (*m20%*), and 1,738 (*m50%*) sites long.

Parameter Effects on ML Topology

For the topologies of *phyloSM/MED/BIG* subsets, MDS plots of weighted normalized Robinson Foulds (RF) distances (Robinson and Foulds 1981) and branch score distances (Kuhner and Felsenstein 1994) are presented in Figure 4.3. Datasets of similar *m* values cluster together, indicating that the *m* parameter, which controls the amount of missing data in a dataset, has a particularly strong influence on topology. The impact of *m* on topology was confirmed by visual analysis of the phylogenies, which showed obvious differences in topology between trees produced by the *m90%* datasets compared to all others. Across all datasets, there was relatively little effect of clustering threshold (*c*) on topology as inferred by RF distance. Visual inspection of the datasets found no clear differences in bootstrap support of topologies for the *m4/20/50%* datasets, which were generally higher than support values for trees produced with *m90%*. The relatively poor resolution and low bootstrap support of *m90%* topologies is not surprising in light of the fact that these datasets include far fewer loci than the *m4/m20%/m50%* datasets.

Phylogeny of Mangifera

The ML analyses of *Mangifera* and *Bouea* species consistently recover five clades with high bootstrap support (>90%, Fig. 4.4, Appendix 4.4). These five clades loosely correspond to the infrageneric classification proposed by Kostermans and Bompard (1993). Clade I (dark green) includes many species previously included in subgenus Mangifera section Mangifera: *M. indica, M. lalijiwa, M. laurina, M. casturi, M. zeylanica, M. pentandra, M. altissima,* and *M. gedebe,* along with multiple unidentified

taxa. Clade II (light green) includes *M. odorata, M. foetida, M. pajang, M. macrocarpa* and one unidentified taxon, most of which were previously classified as subgenus Limus section Perennes. Clade III (yellow) shares some taxa with subgenus Mangifera section Rawa, and includes *M. quadrifida, M. gracilipes, M. magnifica, M. monandra, M. griffithii, M. rufocostata,* and *M. subsessilifolia.* Clade IV contains two of the three species of *Bouea,* which has traditionally been defined as sister to *Mangifera.* Clade V is represented by *M. caesia* and *M. superba*, which were previously placed in subgenus Limus section Deciduae. Within the five clades, species-level relationships are generally well supported, but vary among the different datasets.

SVD Quartets

As expected, SVD Quartets provided greater resolution and higher support values as the number of SNPs in a dataset increased (Appendix 4.4). The 50% majority rules trees inferred by SVD Quartets for *phyloSM_c85_m4/20/50%* recovered clades that roughly correspond to those identified by ML analysis, but fail to resolve the relationships among the clades (Appendix 4.5). The results of SVD Quartets analysis for *phyloSM_c85_m4/20%* datasets provide support for the polyphyly of *Mangifera* (Appendix 4.5).

Discussion

RADseq phylogenomics

Here, we present one of the largest and temporally deepest-scale RADseq phylogenetic studies to date and find ML inference of a concatenated supermatrix of

RADseq loci to be effective in resolving phylogenetic relationships in a formerly intractable group of plants. Previous work has used RADseq to produce phylogenies from clades of plants with crown ages ranging from 30-50 Ma, including Oaks (McVay et al. 2017), *Morella* (Liu et al. 2015), and *Viburnum* (Eaton et al. 2017). The deepest outgroup included here, *Anacardium,* is estimated to have diverged from other taxa around the end of the Paleocene, 55 Ma, and *Gluta* is estimated to have diverged 40 Ma (Weeks et al. 2014).

In congruence with previous studies, we find that reducing the missing data in a dataset drastically reduces the number of loci included in the dataset and also excludes a many variable and parsimony informative sites (e.g., Leaché et al. 2015; Huang and Knowles 2016; Eaton et al. 2017; Tripp et al. 2017). Therefore, in downstream phylogenetic analyses, RADseq datasets with low levels of missing data *(m90)* failed to produce resolved topologies and well-supported branches. In theory, a dataset with maximal phylogenetic information is obtained using the *m4* parameter (Eaton et al. 2017), but in practice, we found these datasets were roughly 2.5-15 times larger than comparable *m20* datasets, and some were prohibitively large for phylogenetic analysis of concatenated full-length reads using currently available software. From the *m4* datasets that we were able to analyze, our results indicate that the increase in the total number of loci, while computationally costly, does not necessarily translate to better resolution or branch support in phylogenetic trees. The *minind (m)* parameter had a much greater impact on the number of loci retained and the overall topology of the tree than did the clustering *(c)* parameter. Datasets using different clustering thresholds produced

relatively similar topologies, indicating that the accuracy of orthology estimation may not be of great concern in RADseq phylogenomics.

While the impacts of the amount of missing data and clustering threshold in RADseq phylogenomics have been explored in previous studies, here, we incorporated an additional variable - the number of individuals included in a dataset. As species-level phylogenomics becomes a more popular endeavor, we expect studies with intraspecific re-sampling to become increasingly common. Our datasets that varied by the number of replicate individuals within a species generally recovered similar topologies, with one notable exception (see discussion of *M. odorata* below). However, summary statistics indicate that the datasets themselves differed substantially. For any given combination of clustering threshold and *minind* value, datasets that contained more individuals retained more loci, and while the loci had fewer variable sites on average, they had slightly more parsimony informative sites. Overall, this indicates that the nature of the loci retained in datasets with more or less intraspecific re-sampling differs in some way. Certainly, our results implore a more thorough examination of how the characteristics of loci change as additional intraspecific samples are included in RADseq datasets, and would benefit from study in a system with a well-annotated genome sequence to explore the patterns we observed here.

Mangifera taxonomy and systematics

The first multilocus phylogenetic hypothesis for *Mangifera* indicates that the genus*,* as traditionally circumscribed, is paraphyletic. Clade V, which includes *M. caesia* (8 samples) and *M. superba* (4 samples), is consistently recovered as sister to the clade of

Bouea (Clade IV) and *Mangifera* sensu stricto (s.s., Clades I-III). On the basis of descriptions of *Mangifera* species (Kostermans and Bompard 1993), we presume that an additional three taxa, *M. decandra, M. lagenifera,* and *M. kemanga* should be included in Clade V. Both *M. caesia* and *M. kemanga* are well known, cultivated species in Malaysia and Indonesia (Morton 1987; Kostermans and Bompard 1993), and *M. caesia* has previously been included in phylogenetic analyses (Yamanaka et al. 2006; Hidayat et al. 2011) though never alongside enough outgroup taxa to test the monophyly of *Mangifera*. However, the taxa in Clade V share synapomorphic morphological characters (erect purple flowers, prominent bud scales, brown or green pear-shaped drupes with white to pink or purple flesh, (Kostermans and Bompard 1993) that support their status as a novel genus. The present study is the one of the first to identify a novel genus using a RADseq phylogeny.

The relationships among the three clades of *Mangifera* s.s. are well supported, though some species-level relationships, particularly those within Clade I, change depending on the dataset analyzed. Most notably, the placement of *M. odorata* differs between *phyloSM* phylogenies, where it is recovered as sister to *M. foetida*, and *phyloMED/BIG* phylogenies, which place it as sister to Clade I (Appendix 4.4). Kuwini mango, *M. odorata,* is a cultivated species that has never been found in the wild (Kostermans and Bompard 1993) and has long been thought to be a hybrid between the cultivated *M. foetida* (horse mango) and common mango, *M. indica* (Hou 1978). In 2002, the hybrid origin of *M. odorata* was confirmed by amplified fragment length polymorphism analysis (Teo et al. 2002, also see Ch. V). We therefore deduce the unstable placement of *M. odorata* within the phylogeny is the result of different loci

being retained in the datasets for different subsets of individuals and reflects the species' hybrid origin. Our results suggest that the number of replicate samples within a species and the number of samples included in a dataset can impact the placement of hybrid lineages, though we are not aware of other studies that have found similar results.

The phylogeny for *Mangifera* indicates that cultivated mango, *M. indica*, appears to be a member of a closely related clade of taxa that includes *M. lalijiwa, M. laurina, M. casturi, M. pentandra,* and *M. zeylanica,* as well as many unidentified accessions. Of these, *M. laurina* and *M. lalijiwa* seem to be most closely related to mango, but the relationship among these species is not well supported and is unstable across datasets produced using different clustering and *minind* parameters. One possible cause for the poor resolution of relationships within Clade I is that there may be high levels of interspecific gene flow occurring between closely related species within this clade. Hybridization is common in plants, and even more so in outcrossing perennial tree species like *Mangifera*. Given that *M. indica* is widely cultivated, it is likely that the species has experienced some level of gene flow with congeners. Additionally, taxa that were not sampled here, especially those from India, Myanmar, and Thailand, are also likely to be closely related to mango, and inclusion of these species may help to resolve relationships within Clade I.

Along with *M. indica*, a few other taxa are of particular interest within the genus *Mangifera* and deserve additional attention in future studies. Sister to all other taxa in Clade I, *M. gedebe* is the most widely distributed species in the genus, with a range that spans from Myanmar to the Solomon Islands (Kostermans and Bompard 1993). The seeds of *M. gedebe* are labyrinthine, exhibiting a reticulate endosperm that allows them to

float and be dispersed by water (Kostermans and Bompard 1993). Since *M. gedebe* is found in both coastal and inland waterways across a broad geographic range, it is possible that it is a complex of multiple species, as was suggested by Hou (1978). Additionally, the species *M. quadrifida* has a convoluted taxonomic history, and was most recently described as having two varieties: *M. quadrifida* var. *quadrifida* and *M. quadrifida* var. *longipetiolata.* Here, we find *M. quadrifida* to be paraphyletic, justifying the earlier identification of *M. quadrifida* var. *longipetiolata* as a distinct species, *M. longipetiolata*. Overall, our results indicate that more thorough study and careful taxonomic revision of the genus *Mangifera* is required in order to aid efforts to understand the diversity and evolutionary history of the genus.

Evolutionary insights

Our multilocus phylogenetic hypothesis for the genus *Mangifera* provides insight into a genus that is rife with opportunity for the study of comparative evolution across the domestication continuum. *Mangifera* s.s., which includes an estimated 64 species, is the third largest genus in the family Anacardiaceae after *Searsia* and *Semecarpus* (Pell et al. 2011), both of which are estimated to be significantly older (~42 and 32 Ma vs. 25 Ma, Weeks et al. 2014), raising questions about the origins of the genus and its rapid speciation. Presently, limited taxon sampling, incomplete morphological descriptions, and underrepresentation in herbarium collections preclude a thorough analysis of the biogeography of *Mangifera* or the evolution of particular traits of interest. However, given that Clades IV and V are restricted to Malesia, which is also the center of diversity for *Mangifera* s.s. (Kostermans and Bompard 1993), and that the recent estimated

divergence times of *Mangifera* s.s. and *Bouea* (25 Ma, Weeks et al. 2014), it seems probable that *Mangifera* originated in Malesia, rather than in India, as was recently proposed (Singh et al. 2016).

The phylogeny of *Mangifera* uncovers a promising system in which to study, in multiple closely related species, the process of perennial crop evolution across the domestication continuum. In each of the five major lineages recovered here (*Bouea, Mangifera* s.s. Clades I-III, Clade V), there exist species that are wild and those that are, to varying extents, cultivated and domesticated (Fig. 4.5). The study of parallel domestication events in closely related species is relatively uncommon, but a few recent works demonstrate that these studies can provide important insights into the evolution of crop species at all stages of domestication. Velasco et al. (2016) compared two closely related taxa, peach *(Prunus persica)* and almond *(Prunus dulcis)*, to explore the genomic signatures of domestication in these perennial systems and found that fruit morphology likely diverged prior to the species' domestication. Other studies have used SNP markers to explore the relationships between wild and semi-domesticated species to infer crop origins, including in carrot *(Daucus spp.,* Arbizu et al. 2016) and squashes *(Cucurbita* spp., Kates et al. 2017). A particularly relevant recent study by Wang et al. (2017) used wild, semi-domesticated, and cultivated citrus to explore the evolution of asexual reproduction by nucellar polyembryony, a rare trait in flowering plants that is horticulturally significant and also reported in at least four *Mangifera* species (Kostermans and Bompard 1993, Bompard 2009).

In addition to being a novel system of perennial domestication, the three genera included in the present study (*Bouea, Mangifera* s.s., Clade V) represent an interesting

clade in which to study frugivory. Species in all three genera produce fleshy drupes, unlike their closest relatives *Gluta* and *Swintonia* (Pell et al. 2011), and are among the largest simple fruits (i.e., harvested and dispersed in one piece) in Malesia (Corlett 1998). Within the phylogeny presented here, each of the five main clades can be distinguished by fruit characteristics. Clade V contains species with pear-shaped fruits that are generally large and brown or green in color, and are dispersed by rhinos and elephants (Phillipps and Phillipps 2016). Out of the five clades, *Bouea* species have the smallest drupes, which turn yellow to orange upon ripening (Pell et al. 2011). The three clades of *Mangifera* also have distinctive fruit characters, with species in Clade I (which includes *M. indica*) predominantly characterized by medium to large fruits that turn yellow/orange/red or stay green upon ripening. Fruits in Clade II are large and typically stay green or turn brown upon ripening and are probably dispersed by large mammals (Phillipps and Phillipps 2016). Species in clade III have fruits that tend to turn red, red/brown or dark purple/black upon ripening and are reportedly dispersed by hornbills (Phillipps and Phillipps 2016). The differences in fruit characteristics between clades suggest an important role for frugivory in the relatively recent and rapid speciation of *Mangifera.* While the role of frugivores in shaping fruit-trait evolution remains contentious, recent work indicates that the fruit size (but not color) is impacted by the presence of frugivore seed dispersers in Malesia (Brodie 2017). Given the imperiled nature of large frugivores in Southeast Asia, including Asian elephants, rhinos, hornbills, and orangutans, species of *Mangifera* that depend on these large dispersers may be in jeopardy.

Beyond the long-term evolutionary implications of the loss of primary seed dispersers, *Mangifera* species are directly threatened by deforestation, even more so because they are often selectively targeted as high-quality timber (Kostermans and Bompard 1993). The IUCN Redlist (IUCN 2012) assessment of 45 *Mangifera* species lists two as extinct in the wild, one as critically endangered, and nine as endangered, with twelve additional species considered vulnerable, three near threatened, and ten as data deficient, including (wild populations of) *M. indica*. Only eight species are assessed as being of least concern (IUCN 2012). While there is a resurgent effort to conserve wild relatives of important crop species (Meilleur and Hodgkin 2004; Maxted and Kell 2009; Ford-Lloyd et al. 2011; Hunter and Heywood 2011; Maxted et al. 2012), *Mangifera* is among the least well-represented genera in ex situ collections (Castañeda-Álvarez et al. 2016), and a great deal more work is required to ensure the long-term survival of this unique and important genus of tropical fruit trees.

Literature Cited

Akhter S., McDonald M.A., Marriott R. 2016. *Mangifera sylvatica* (Wild Mango): A new cocoa butter alternative. Sci. Rep. 6:32050.

Andrews S. 2010 FastQC: a quality control tool for high throughput sequence data. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Arbizu C.I., Ellison S.L., Senalik D., Simon P.W., Spooner D.M. 2016. Genotyping-by-sequencing provides the discriminating power to investigate the subspecies of *Daucus carota* (Apiaceae). BMC Evol. Biol. 16:234.

Arias R.S., Borrone J.W., Tondo C.L., Kuhn D.N., Irish B.M., Schnell R.J. 2012. Ch. 7 Genomics of Tropical Fruit Tree Crops. In: Schnell R.J., Priyadarshan P.M., editors. Genomics of Tree Crops. New York, NY: Springer New York. p. 209–239.

Arnold M.L. 2004. Natural hybridization and the evolution of domesticated, pest and disease organisms. Mol. Ecol. 13:997–1007.

Arumuganathan K., Earle E.D. 1991. Nuclear DNA content of some important plant species. Plant Mol. Biol. Report. 9:415–415.

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:e3376.

Baul T.K., Jahedul Alam M., Nath T.K. 2016. *Mangifera sylvatica* Roxb. in the Forests of South-Eastern Bangladesh: A Potential Underutilised Tree for Small-Scale Forestry. Small-scale For. 15:149–158.

Brodie J.F. 2017. Evolutionary cascades induced by large frugivores. Proc. Natl. Acad. Sci. 114:11998–12002.

Bompard J.M. 2009. Taxonomy and Systematics. In: Litz R.E., editor. The Mango: Botany, Production and Uses. Cambridge: CAB International. p. 19-41.

Cariou M., Duret L., Charlat S. 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecol. Evol. 3:846–52.

Castañeda-Álvarez N.P., Khoury C.K., Achicanoy H.A., Bernau V., Dempewolf H., Eastwood R.J., Guarino L., Harker R.H., Jarvis A., Maxted N., Müller J. V., Ramirez-Villegas J., Sosa C.C., Struik P.C., Vincent H., Toll J. 2016. Global conservation priorities for crop wild relatives. Nat. Plants. 2:16022.

Chifman J., Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics. 30:3317–3324.

Chomicki G., Renner S.S. 2014. Watermelon origin solved with molecular phylogenetics including Linnaean material: another example of museomics. New Phytol. 205:526–532.

Clement C.R. 1999. 1492 and the loss of Amazonian crop genetic resources. I. The relation between domestication and human population decline. Econ. Bot. 53:188–202.

Combosch D.J., Vollmer S. V. 2015. Trans-Pacific RAD-Seq population genomics confirms introgressive hybridization in Eastern Pacific *Pocillopora* corals. Mol. Phylogenet. Evol. 88:154–162.

Corander J., Majander K.K., Cheng L., Merilä J. 2013. High degree of cryptic population differentiation in the Baltic Sea herring *Clupea harengus*. Mol. Ecol. 22:2931–2940.

Corlett R.T. 1998. Frugivory and seed dispersal by vertebrates in the Oriental (Indomalaysian) Region. Biol. Rev. 73:413–448.

Cruaud A., Gautier M., Galan M., Foucaud J., Sauné L., Genson G., Dubois E., Nidelet S., Deuve T., Rasplus J.Y. 2014. Empirical assessment of RAD sequencing for interspecific phylogeny. Mol. Biol. Evol. 31:1272–1274.

Czech L., Huerta-Cepas J., Stamatakis A. 2017. A critical review on the use of support values in tree viewers and bioinformatics toolkits. Mol. Biol. Evol. 34:1535–1542.

DaCosta J.M., Sorenson M.D. 2016. DdRAD-seq phylogenetics based on nucleotide, indel, and presence-absence polymorphisms: Analyses of two avian genera with contrasting histories. Mol. Phylogenet. Evol. 94:122–135.

Darwin C. 1859. The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Racies in the Struggle for Life. New York: Random House.

Davey J.W., Davey J.L., Blaxter M.L., Blaxter M.W. 2010. RADSeq: next-generation population genetics. Brief. Funct. Genomics. 9:416–23.

Díaz-Arce N., Arrizabalaga H., Murua H., Irigoien X., Rodríguez-Ezpeleta N. 2016. RAD-seq derived genome-wide nuclear markers resolve the phylogeny of tunas. Mol. Phylogenet. Evol. 102:202–207.

Dinesh M.R., Ravishankar K.V., Nischita P. 2015. Exploration , Characterization and Phylogenetic Studies in Wild *Mangifera indica* Relatives. Am. J. Plant Sci. 6:2151–2160.

Doyle J.J., Doyle J.L. 1990. Isolation of plant DNA from fresh tissue. Focus (Madison). 12:13–15.

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

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.

Eaton D.A.R.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.

Eiadthong W., Yonemori K., Kanzaki S., Sugiura A., Utsunomiya N., Subhadrabandhu S. 2000. Amplified Fragment Length Polymorphism Analysis for Studying Genetic Relationships among *Mangifera* Species in Thailand. J. Am. Hortic. Soc. 125:160–164.

Eiadthong W., Yonemori K., Sugiura A., Utsunomiya N., Subhadrabandhu S. 1999. Analysis of phylogenetic relationships in *Mangifera* by restriction site analysis of an amplified region of cpDNA. Sci. Hortic. (Amsterdam). 80:145–155.

Emerson K.J., Merz C.R., Catchen J.M., Hohenlohe P.A., Cresko W.A., Bradshaw W.E., Holzapfel C.M. 2010. Resolving postglacial phylogeography using high-throughput sequencing. Proc. Natl. Acad. Sci. 107:16196–16200.

FAO. 2003. Tropical Fruits. In: Medium-term prospects for agricultural commodities. Food and Agriculture Organization of the United Nations, Rome, Italy. p. 1–6.

FAOSTAT. 2013. Food and Agriculture Organization of the United Nations Statistics Division. Available from: http://www.fao.org/faostat/.

Fernández-Mazuecos M., Mellers G., Vigalondo B., Sáez L., Vargas P., Glover B.J. 2017. Resolving Recent Plant Radiations: Power and Robustness of Genotyping-by-Sequencing. Syst. Biol. 0:1–19.

Fitmawati, Harahap S.P., Sofiyanti N. 2017. Short Communication: Phylogenetic analysis of mango (*Mangifera*) in Northern Sumatra based on gene sequences of cpDNA trnL-F intergenic spacer. Biodiversitas. 18:715–719.

Fitmawati, Hartana A. 2010. Phylogenetic Study of *Mangifera laurina* and its Related Species Using cpDNA trnL-F Spacer Markers. 17:9–14.

Ford-Lloyd B. V., Schmidt M., Armstrong S.J., Barazani O., Engels J., Hadas R., Hammer K., Kell S.P., Kang D., Khoshbakht K., Li Y., Long C., Lu B.-R., Ma K., Nguyen V.T., Qiu L., Ge S., Wei W., Zhang Z., Maxted N. 2011. Crop Wild Relatives—Undervalued, Underutilized and under Threat? Bioscience. 61:559–565.

Gaut B.S., Díez C.M., Morrell P.L. 2015. Genomics and the Contrasting Dynamics of Annual and Perennial Domestication. Trends Genet. 31:709–719.

Hancock L., Edwards E.J. 2014. Phylogeny and the inference of evolutionary trajectories. J. Exp. Bot. 65:3491–3498.

Hawkins J.S., Ramachandran D., Henderson A., Freeman J., Carlise M., Harris A., Willison-Headley Z. 2015. Phylogenetic reconstruction using four low-copy nuclear loci strongly supports a polyphyletic origin of the genus *Sorghum*. Ann. Bot. 116:291–299.

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.

Hess J.E., Campbell N.R., Close D.A., Docker M.F., Narum S.R. 2012. Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. Mol. Ecol.:2898–2916.

Hidayat T., Pancoro A.D.I., Kusumawaty D. 2011. Utility of matK Gene to Assess Evolutionary Relationships of Genus *Mangifera* (Anacardiaceae) in Indonesia and Thailand. Biotropia (Bogor). 18:74–80.

Hipp A.L., Eaton D.A.R.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:e93975.

Hohenlohe P.A., Amish S.J., Catchen J.M., Allendorf F.W., Luikart G. 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Mol. Ecol. Resour. 11:117–22.

Hohenlohe P.A., Phillips P.C., Cresko W.A. 2010. Using Population Genomics to Detect Selection in Natural Populations: Key Concepts and Methodological Considerations. Int. J. Plant Sci. 171:1059–1071.

Hou D. 1978. Anacardiaceae. Flora Malesiana Vol 8 Part 3. Alphen Aan Den Rijn, The Netherlands: Sijthoff & Noordhoff International Publishers. p. 395–548.

Huang H., Knowles L.L. 2016. Unforeseen consequences of excluding missing data from next-generation sequences: Simulation study of rad sequences. Syst. Biol. 65:357–365.

Hunter D., Heywood V. [Eds.] 2011. Crop Wild Relatives: A Manual of in situ Conservation. Washington, DC: Earthscan. 440 p.

IUCN. 2012. World Conservation Monitoring Centre 1998. *IUCN Red List of Threatened Species. Version 2012.2.* Available at: www.iucnredlist.org.

Iyer C.P.A., Schnell R.J. 2009. Breeding and Genetics. In: Litz R.E., editor. The Mango: Botany, Production and Uses. Cambridge: CAB International. p. 68–83.

Jones J.C., Fan S., Franchini P., Schartl M., Meyer A. 2013. The evolutionary history of *Xiphophorus* fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. Mol. Ecol. 22:2986–3001.

Kates H.R., Soltis P.S., Soltis D.E. 2017. Evolutionary and domestication history of *Cucurbita* (pumpkin and squash) species inferred from 44 nuclear loci. Mol. Phylogenet. Evol. 111:98–109.

Kostermans A.J.G.H., Bompard J.M. 1993. The Mangoes: Their Botany, Nomenclature, Horticulture, and Utilization. San Diego, CA: Academic Press. 233 p.

Kovach M.J., Sweeney M.T., McCouch S.R. 2007. New insights into the history of rice domestication. Trends Genet. 23:578–587.

Kuhn D.N., Bally I.S.E., Dillon N.L., Innes D., Groh A.M., Rahaman J., Ophir R., Cohen Y., Sherman A. 2017. Genetic Map of Mango: A Tool for Mango Breeding. Front. Plant Sci. 8:577.

Kuhner M.K., Felsenstein J. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol Biol Evol. 11:459–468.

Leaché A.D., Banbury B.L., Felsenstein J., De Oca A.N.M., Stamatakis A. 2015. Short tree, long tree, right tree, wrong tree: New acquisition bias corrections for inferring SNP phylogenies. Syst. Biol. 64:1032–1047.

Leaché A.D., Oaks J.R. 2017. The utility of single nucleotide polymorphism (SNP) data in phylogenetics. Annu. Rev. Ecol. Evol. Syst. 48:69–84.

Liu L., Jin X., Chen N., Li X., Li P., Fu C. 2015. Phylogeny of *Morella rubra* and its relatives (Myricaceae) and genetic resources of Chinese bayberry using RAD sequencing. PLoS One. 10:e0139840.

Massatti R., Reznicek A.A., Knowles L.L. 2016. Utilizing RADseq data for phylogenetic analysis of challenging taxonomic groups: A case study in *Carex* sect. *Racemosae*. Am. J. Bot. 103:337–347.

Maxted N., Kell S., Ford-Lloyd B., Dulloo E., Toledo Á. 2012. Toward the Sytematic Conservation of Global Crop Wild Relative Diversity. Crop Sci. 52:774–785.

Maxted N., Kell S.P. 2009. Establishment of a global network for the in situ conservation of crop wild relatives: Status and needs. FAO Commission on Genetic Resources for Food and Agriculture, Rome, Itlay. 266 pp.

McVay J.D., Hipp A.L., Manos P.S. 2017. A genetic legacy of introgression confounds phylogeny and biogeography in oaks. Proc. R. Soc. London B. 284:20170300.

Meilleur B.A., Hodgkin T. 2004. In situ conservation of crop wild relatives: status and trends. Biodivers. Conserv. 13:663–684.

Meyer R.S., DuVal A.E., Jensen H.R. 2012. Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. New Phytol. 196:29–48.

Meyer R.S., Purugganan M.D. 2013. Evolution of crop species: genetics of domestication and diversification. Nat. Rev. Genet. 14:840–52.

Miller A.J., Gross B.L. 2011. From forest to field: perennial fruit crop domestication. Am. J. Bot. 98:1389–414.

Miller M., Pfeiffer W., Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gatew. Comput. Environ. Work. p. 1–8.

Miller M.R., Dunham J.P., Amores A., Cresko W.A., Johnson E.A. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 17:240–8.

Morton J. 1987. Mango. Fruits of Warm Climates. p. 221–239.

Mukherjee S.K. 1949. The mango and its wild relatives. Sci. Cult. 26:5–9.

Mukherjee S.K. 1950a. Cytological investigation of the mango (*Mangifera indica* L.) and the allied Indian species. Proc. Nat. Inst. Sci., India. 16:287–303.

Mukherjee S.K. 1950b. Mango: its allopolyploid nature. Nature. 166:196–197.

Mukherjee S.K. 1953. Origin, distribution and phylogenetic affiinity of the species of *Mangifera* L. Econ. Bot. 7:130–162.

Mukherjee S.K. 1957. Cytology of Some Malayan Species of *Mangifera*. Cytologia (Tokyo). 22:239–241.

Nadeau N.J., Martin S.H., Kozak K.M., Salazar C., Dasmahapatra K.K., Davey J.W., Baxter S.W., Blaxter M.L., Mallet J., Jiggins C.D. 2013. Genome-wide patterns of divergence and gene flow across a butterfly radiation. Mol. Ecol. 22:814–26.

Olsen K.M., Wendel J.F. 2013. Crop plants as models for understanding plant adaptation and diversification. Front. Plant Sci. 4:290.

Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics. 20:289–290.

Paris J.R., Stevens J.R., Catchen J.M. 2017. Lost in parameter space: A road map for stacks. Methods Ecol. Evol. 8:1360–1373.

Paun O., Turner B., Trucchi E., Munzinger J., Chase M.W., Samuel R. 2016. Processes Driving the Adaptive Radiation of a Tropical Tree (*Diospyros*, Ebenaceae) in New Caledonia, a Biodiversity Hotspot. Syst. Biol. 65:212–227.

Pell S.K., Mitchell J.D., Miller A.J., Lobova T.A. 2011. Anacardiaceae. In: Kubitzki K., editor. The Families and Genera of Vascular Plants. Vol. X. Berlin, Heidelberg: Springer-Verlag. p. 7–50.

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.

Phillipps Q., Phillipps K. 2016. Mammals of Borneo. Princeton, NJ: Princeton University Press. p. 58–59.

Pickersgill B. 2007. Domestication of plants in the Americas: insights from Mendelian and molecular genetics. Ann. Bot. 100:925–40.

Purugganan M.D., Fuller D.Q. 2009. The nature of selection during plant domestication. Nature. 457:843–8.

Rambaut, A. 2006. FigTree: tree figure drawing tool. Available from: http://tree.bio.ed.ac.uk/software/figtree/.

Ravishankar K. V, Bommisetty P., Bajpaj A., Srivastava N., Mani B.H., Vasugi C., Rajan S., Dinesh M.R. 2015. Genetic diversity and population structure analysis of mango (*Mangifera indica*) cultivars assessed by microsatellite markers. Trees. 29:775–783.

Ree R.H., Hipp A.L. 2015. Inferring phylogenetic history from restriction site associated DNA (RADseq). In: Hörandl E., Appelhans M.S., editors. Next-Generation Sequencing in Plant Systematics. p. 181–204.

Rendón-Anaya M., Montero-Vargas J.M., Saburido-Álvarez S., Vlasova A., Capella-Gutierrez S., Ordaz-Ortiz J.J., Aguilar O.M., Vianello-Brondani R.P., Santalla M., Delaye L., Gabaldón T., Gepts P., Winkler R., Guigó R., Delgado-Salinas A., Herrera-Estrella A. 2017. Genomic history of the origin and domestication of common bean unveils its closest sister species. Genome Biol. 18:60.

Rivers D.M., Darwell C.T., Althoff D.M. 2016. Phylogenetic analysis of RAD-seq data: examining the influence of gene genealogy conflict on analysis of concatenated data. Cladistics. 32:672–681.

Robinson D.F., Foulds L.R. 1981. Comparison of phylogenetic trees. Math. Biosci. 53:131–147.

Rubin B.E.R., Ree R.H., Moreau C.S. 2012. Inferring phylogenies from RAD sequence data. PLoS One. 7:e33394.

Sanjur O.I., Piperno D.R., Andres T.C., Wessel-Beaver L. 2002. Phylogenetic relationships among domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene: Implications for crop plant evolution and areas of origin. Proc. Natl. Acad. Sci. U.S.A. 99:535–40.

Schliep K.P. 2011. phangorn: Phylogenetic analysis in R. Bioinformatics. 27:592–593.

Sherman A., Rubinstein M., Eshed R., Benita M., Ish-Shalom M., Sharabi-Schwager M., Rozen A., Saada D., Cohen Y., Ophir R. 2015. Mango (*Mangifera indica* L.) germplasm diversity based on single nucleotide polymorphisms derived from the transcriptome. BMC Plant Biol. 15:277.

Singh N.K., Mahato A.K., Jayaswal P.K., Singh A., Singh S., Singh N., Rai V., Mithra S.V.A., Gaikwad K., Sharma N., Lal S., Srivastava M., Prakash J., Kalindindi U., Singh S.K., Singh A.K., Khan K., Mishra R.K., Rajan S., Bajpai A., Sandhyra B.S., Nischita P., Ravishankar K.V., Dinesh M.R., Kumar N., Jaiswal S., Iquebal M.A., Kumar D., Rai A., Sharma R.S. 2016. Origin, Diversity and Genome Sequence of Mango (*Mangifera indica* L.). Indian J. Hist. Sci. 51:355–368.

Singhal S., Huang H., Title P.O., Donnellan S.C., Holmes I., Rabosky D.L. 2017. Genetic diversity is largely unpredictable but scales with museum occurrences in a species-rich clade of Australian lizards. Proc. R. Soc. B Biol. Sci. 284:20162588.

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

Sukumaran J., Holder M.T. 2010. DendroPy: A Python library for phylogenetic computing. Bioinformatics. 26:1569–1571.

Sukumaran J., Holder M.T. 2015. SumTrees: Phylogenetic Tree Summarization. Available from: https://github.com/jeetsukumaran/DendroPy.

Suparman S., Pancoro A., Hidayat T. 2013. Phylogenetic Analisys of *Mangifera* Based on rbcL Sequences, Chloroplast DNA. Sci. Pap. Ser. B, Hortic. LVII:235–240.

Surapaneni M., Vemireddy L.R., Begum H., Purushotham Reddy B., Neetasri C., Nagaraju J., Anwar S.Y., Siddiq E.A. 2013. Population structure and genetic analysis of different utility types of mango (*Mangifera indica* L.) germplasm of Andhra Pradesh state of India using microsatellite markers. Plant Syst. Evol. 299:1215–1229.

Takahashi T., Nagata N., Sota T. 2014. Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. Mol. Phylogenet. Evol. 80:77–81.

Teo L.L., Kiew R., Set O., Lee S.K., Gan Y.Y. 2002. Hybrid status of kuwini, *Mangifera odorata* Griff. (Anacardiaceae) verified by amplified fragment length polymorphism. Mol. Ecol. 11:1465–9.

The Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature. 487:94–8.

Tripp E.A., Tsai Y.H.E., Zhuang Y., Dexter K.G. 2017. RADseq dataset with 90% missing data fully resolves recent radiation of *Petalidium* (Acanthaceae) in the ultra-arid deserts of Namibia. Ecol. Evol. 7:7920–7936.

Ueda Y., Higuchi H., Nawata E. 2011. Wild Mangoes in Mainland Southeast Asia Their Local Names, Uses and Growing Environments. Trop. Agr. Dev. 55:55–67.

Ueda Y., Higuchi H., Nawata E. 2015. Kalon Mango, a Superior Fruit from *Mangifera pentandra* Hook. f. in Northeast Thailand: An Ethnobotanical Study of Indigenous Relationships between People and Plants. Trop. Agr. Dev. 59:89–100.

Ueda Y., Higuchi H., Nawata E. 2016. Wild Mangoes in North Thailand: An Ethnobotanical Study of Local Names and Uses. Trop. Agr. Dev. 60:93–102.

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.

Velasco D., Hough J., Aradhya M., Ross-Ibarra J. 2016. Evolutionary Genomics of Peach and Almond Domestication. G3-Genes Genom. Genet. 7:3985–3993.

Viruel M.A., Escribano P., Barbieri M., Ferri M., Hormaza J.I. 2005. Fingerprinting, embryo type and geographic differentiation in mango (*Mangifera indica* L., Anacardiaceae) with microsatellites. Mol. Breed. 15:383–393.

Wagner C.E., Keller I., Wittwer S., Selz O.M., Mwaiko S., Greuter L., Sivasundar A., Seehausen O. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Mol. Ecol. 22:787–98.

Wang X., Xu Y., Zhang S., Cao L., Huang Y., Cheng J., Wu G., Tian S., Chen C., Liu Y., Yu H., Yang X., Lan H., Wang N., Wang L., Xu J., Jiang X., Xie Z., Tan M., Larkin R.M., Chen L.L., Ma B.G., Ruan Y., Deng X., Xu Q. 2017. Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. Nat. Genet. 49:765–772.

Wang X.Q., Zhao L., Eaton D.A.R., Li D.Z., Guo Z.H. 2013. Identification of SNP markers for inferring phylogeny in temperate bamboos (Poaceae: Bambusoideae) using RAD sequencing. Mol. Ecol. Resour. 13:938–945.

Washburn J.D., Bird K.A., Conant G.C., Pires J.C. 2016. Convergent Evolution and the Origin of Complex Phenotypes in the Age of Systems Biology. Int. J. Plant Sci. 177:305–318.

Weeks A., Zapata F., Pell S.K., Daly D.C., Mitchell J.D., Fine P.V.A. 2014. To move or to evolve: Contrasting patterns of intercontinental connectivity and climatic niche evolution in “Terebinthaceae” (Anacardiaceae and Burseraceae). Front. Genet. 5:409.

Weese T.L., Bohs L. 2010. Eggplant origins: Out of Africa, into the Orient. Taxon. 59:49–56.

Wong M.M.L., Gujaria-Verma N., Ramsay L., Yuan H.Y., Caron C., Diapari M., Vandenberg A., Bett K.E. 2015. Classification and characterization of species within the genus *Lens* using genotyping-by-sequencing (GBS). PLoS One. 10: e0122025.

Yamanaka N., Hasran M., Xu D.H., Tsunematsu H., Idris S., Ban T. 2006. Genetic relationship and diversity of four *Mangifera* species revealed through AFLP analysis. Genet. Resour. Crop Evol. 53:949–954.

Yonemori K., Honsho C., Kanzaki S., Eiadthong W., Sugiura A. 2002. Phylogenetic relationships of *Mangifera* species revealed by ITS sequences of nuclear ribosomal DNA and a possibility of their hybrid origin. Plant Syst. Evol. 231:59–75.

Zeder M. a. 2006. Central questions in the domestication of plants and animals. Evol. Anthropol. Issues, News, Rev. 15:105–117.

Tables

Table 4.1. Sampling for phylogenetic analysis. For each proposed subgenus and section, the number of species included and total number of species included in the group is given in parentheses. The number of replicates included in each of the three subsets (*phyloBIG, phyloMED, phyloSM*) is provided.

Subgenus/Section (# sampled/total # in group)	Taxon	Number of Individuals		
		phylo BIG	phylo MED	phylo SM
Limus/Deciduae (6/11)	*Mangifera caesia*	8	3	1
Limus/Deciduae (6/11)	*Mangifera foetida*	21	8	1
Limus/Deciduae (6/11)	*Mangifera macrocarpa*	2	2	1
Limus/Deciduae (6/11)	*Mangifera odorata*	17	5	1
Limus/Deciduae (6/11)	*Mangifera pajang*	4	4	1
Limus/Deciduae (6/11)	*Mangifera superba*	4	1	1
Mangifera/Euantherae (1/3)	*Mangifera pentandra*	2	2	1
Mangifera/Mangifera (11/34)	*Mangifera altissima*	1	1	1
Mangifera/Mangifera (11/34)	*Mangifera casturi*	5	4	1
Mangifera/Mangifera (11/34)	*Mangifera indica*	69	15	2
Mangifera/Mangifera (11/34)	*Mangifera lalijiwa*	2	2	1
Mangifera/Mangifera (11/34)	*Mangifera laurina*	6	2	2
Mangifera/Mangifera (11/34)	*Mangifera magnifica*	2	2	1
Mangifera/Mangifera (11/34)	*Mangifera monandra*	1	1	1
Mangifera/Mangifera (11/34)	*Mangifera quadrifida*	8	8	2
Mangifera/Mangifera (11/34)	*Mangifera rufocostata*	1	1	1
Mangifera/Mangifera (11/34)	*Mangifera similis*	2	2	1
Mangifera/Mangifera (11/34)	*Mangifera zeylanica*	2	1	1
Mangifera/Marchandora (1/1)	*Mangifera gedebe*	4	3	1
Mangifera/Rawa (2/9)	*Mangifera gracilipes*	1	1	1
Mangifera/Rawa (2/9)	*Mangifera griffithii*	9	5	1
Mangifera/sp. (22)	*Mangifera* sp.	22	1	1
Mangifera/Unknown/Unknown (1/11)	*Mangifera subsessilifolia*	3	3	1
Outgroup	*Anacardium occidentale*	1	1	1
Outgroup	*Bouea macrophylla*	1	1	1
Outgroup	*Bouea oppositifolia*	2	1	1
Outgroup	*Gluta malayana*	1	1	1
		201	98	41

Table 4.2. Summary statistics for (A) *phyloBIG,* (B) *phyloMED,* and (C) *phyloSM* phylogenetic datasets (201, 98, and 41 individuals, respectively. The total number of loci recovered, percent of loci that are parsimony informative (PI), adjusted average variants per locus, adjusted average PI sites per locus, and percent of missing loci are given. Values are shaded in a red to green color scale, with red indicating less desirable characteristics (e.g., lower numbers of loci, higher percentages of missing data) and green indicating more desirable characteristics.

A

Dataset	Total Loci	Percent of Var that are PI	Adj. Avg. Var/locus	Adj. Avg. PI/locus	Percent Missing Loci
phyloBIG_c85_m4	89887	45.76%	8.38	3.83	92.52%
phyloBIG_c90_m4	107117	46.70%	8.32	3.89	92.71%
phyloBIG_c95_m4	143628	47.93%	6.74	3.23	92.66%
phyloBIG_c85_m20%	6494	58.99%	11.53	6.80	59.29%
phyloBIG_c90_m20%	7269	59.63%	11.73	6.99	59.82%
phyloBIG_c95_m20%	9950	59.74%	11.20	6.69	60.42%
phyloBIG_c85_m50%	1765	60.02%	10.69	6.42	29.44%
phyloBIG_c90_m50%	1901	60.08%	10.85	6.52	29.40%
phyloBIG_c95_m50%	2491	60.59%	11.19	6.78	30.06%
phyloBIG_c85_m90%	217	62.45%	3.50	2.18	5.40%
phyloBIG_c90_m90%	245	60.14%	4.05	2.43	5.31%
phyloBIG_c95_m90%	263	58.32%	3.59	2.09	5.61%

B

Dataset	Total Loci	Percent of Var that are PI	Adj. Avg. Var/locus	Adj. Avg. PI/locus	Percent Missing Loci
phyloMED_c85_m4	52440	39.80%	8.91	3.55	88.09%
phyloMED_c90_m4	61455	40.80%	8.90	3.63	88.39%
phyloMED_c95_m4	79280	41.69%	7.22	3.01	88.49%
phyloMED_c85_m20%	6866	50.46%	11.36	5.73	59.09%
phyloMED_c90_m20%	7574	51.05%	11.60	5.92	59.47%
phyloMED_c95_m20%	9505	51.50%	10.95	5.64	59.35%
phyloMED_c85_m50%	1890	52.54%	10.99	5.77	29.83%
phyloMED_c90_m50%	2035	52.63%	10.83	5.70	29.82%
phyloMED_c95_m50%	2576	53.31%	11.13	5.94	30.05%
phyloMED_c85_m90%	241	55.83%	3.64	2.03	5.89%
phyloMED_c90_m90%	268	54.71%	4.14	2.26	5.79%
phyloMED_c95_m90%	285	53.97%	3.70	1.99	6.07%

Table 4.2. (continued)

C

Dataset	Total Loci	Percent of Var that are PI	Adj. Avg. Var/locus	Adj. Avg. PI/locus	Percent Missing Loci
phyloSM_c85_m4	21582	31.56%	9.30	2.94	77.41%
phyloSM_c90_m4	24199	32.19%	9.25	2.98	77.70%
phyloSM_c95_m4	26771	31.57%	7.24	2.29	77.47%
phyloSM_c85_m20%	8526	37.11%	10.30	3.82	61.49%
phyloSM_c90_m20%	9384	37.70%	10.43	3.93	61.77%
phyloSM_c95_m20%	10461	36.76%	8.85	3.25	61.31%
phyloSM_c85_m50%	2027	38.46%	10.52	4.05	30.39%
phyloSM_c90_m50%	2172	38.97%	10.67	4.16	30.25%
phyloSM_c95_m50%	2542	38.22%	9.77	3.73	30.53%
phyloSM_c85_m90%	208	45.86%	4.02	1.84	6.04%
phyloSM_c90_m90%	235	42.23%	4.41	1.86	5.95%
phyloSM_c95_m90%	230	42.63%	3.56	1.52	5.97%

Figure Captions

Figure 4.1. Taxonomy of *Mangifera* as proposed by Kostermans and Bompard (1993). Two subgenera and a total of six sections were delimited based on morphological characteristics. The number of species proposed to be in each group is shown in parentheses, with 11 described species of unknown placement.

Figure 4.2. Summary of (A) number of loci (B) Percent missing loci (C) Average Variants per locus, (D) Average Parsimony informative sites per locus, and (E) Percent of SNPs that are Parsimony informative across *phyloBIG, phyloMED,* and *phyloSM* datasets clustered at three different thresholds (85%, 90%, 95%) and for four levels of *min_individuals* (4, 20%, 50%, 90%).

Figure 4.3. Multidimensional scaling of weighted Robinson-Foulds Distance and Branch Score Distance between maximum likelihood topologies produced for A) *phyloBIG* B) *phyloMED* and C) *phyloSM* datasets.

Figure 4.4. Maximum likelihood phylogenetic hypothesis of *Mangifera* based on the *phyloSM_m20_c85* dataset, with support from 1000 bootstrap replicates shown on branches. Highlighting indicates the five main clades recovered; Clade I (dark green) includes *M. indica, M. lalijiwa, M. laurina, M. casturi, M. zeylanica, M. pentandra, M. altissima, M. gedebe,* and multiple unidentified taxa; Clade II (light green) includes *M. odorata, M. foetida, M. pajang, M. macrocarpa* and one unidentified taxon; clade III (yellow) includes *M. quadrifida, M. gracilipes, M. magnifica, M. monandra, M. griffithii, M. rufocostata,* and *M. subsessilifolia*; Clade IV (orange) contains *Bouea* species, and Clade V (red) includes two taxa previously included in *Mangifera*, *M. caesia* and *M. superba.* The names of species that are domesticated to some extent (including wild-collected, incipiently domesticated, cultivated, and highly domesticated) are shown in blue.

Figures

Figure 4.1.

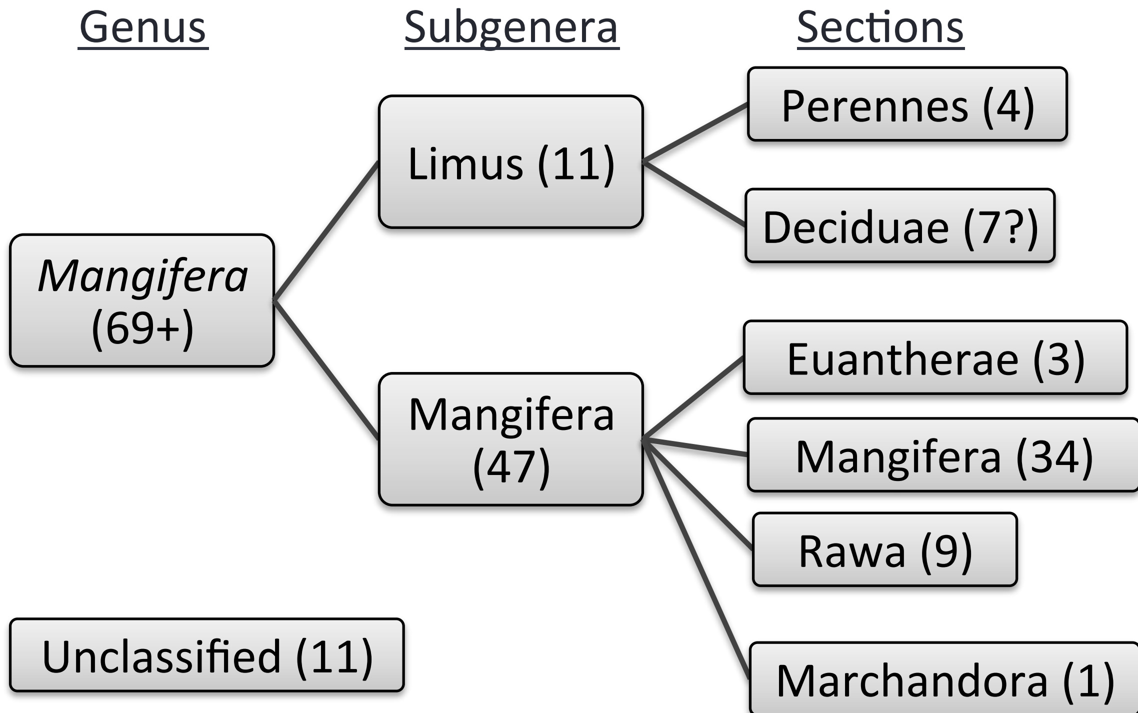


Figure 4.2.

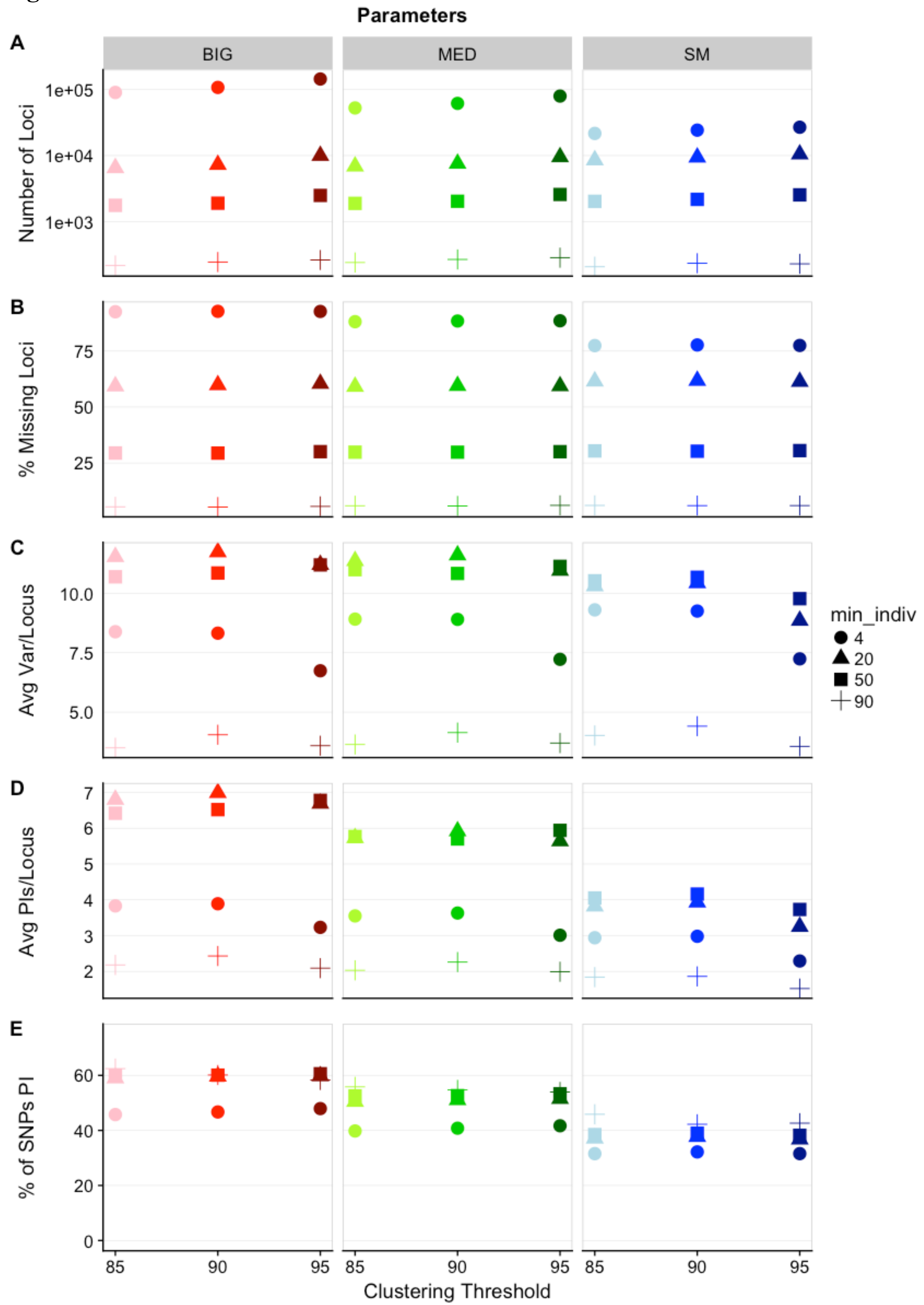


Figure 4.3.

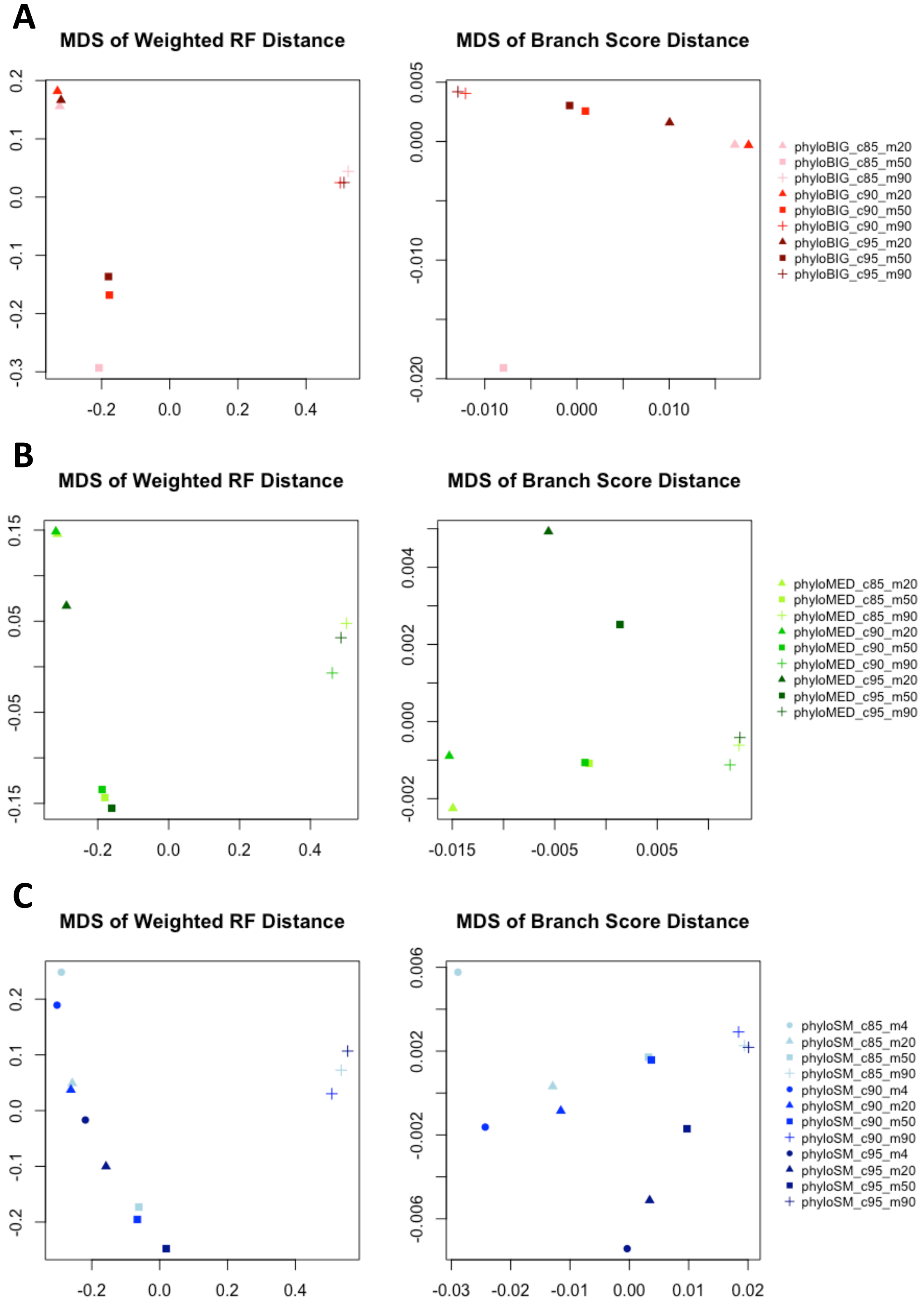
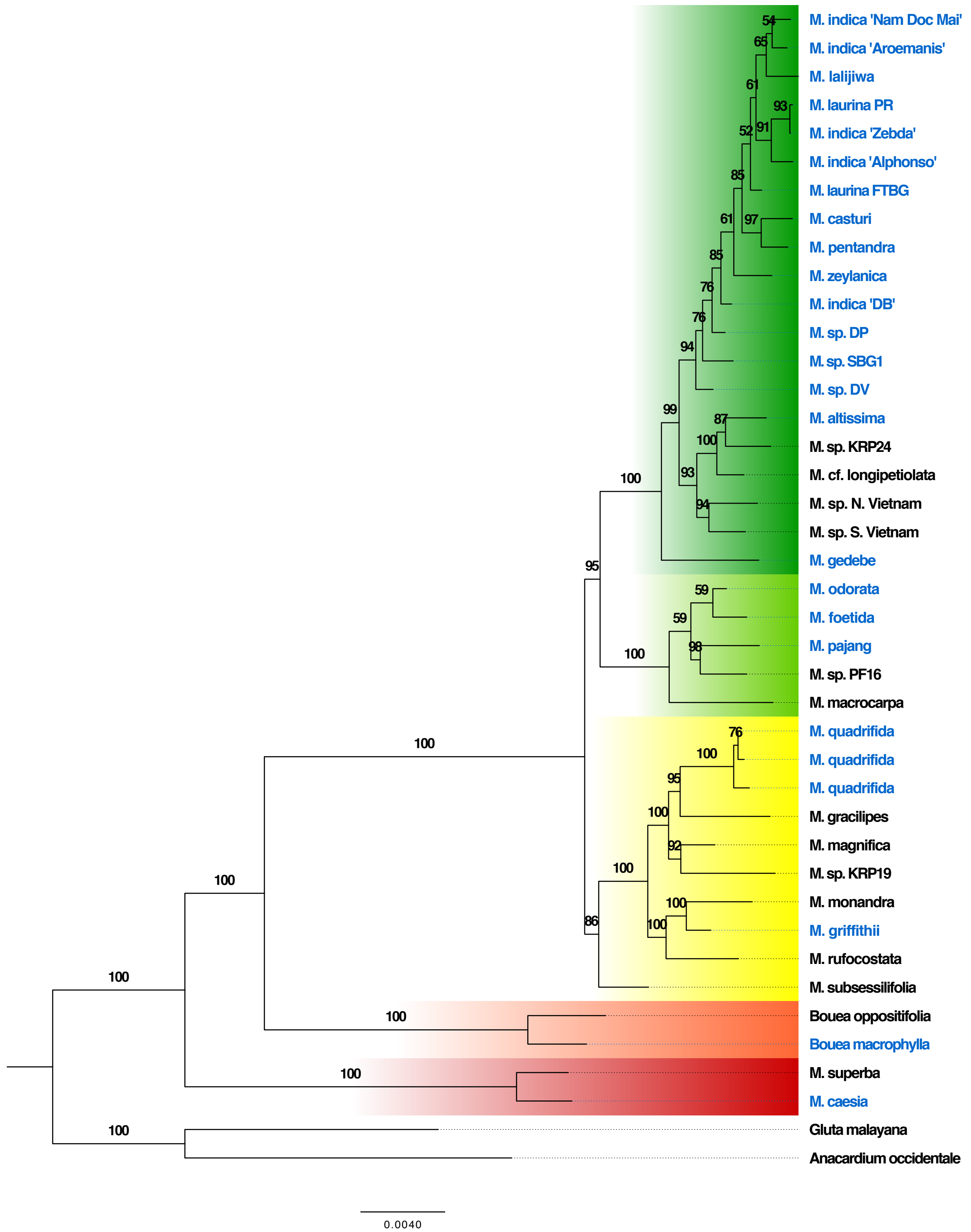


Figure 4.4.



Appendices Captions

Appendix 4.1. Information for samples included in this study, including sample name, Species ID (which may differ from that of the sample name), and collection location (FTBG = Fairchild Tropical Botanic Garden, FTG = Fairchild Tropical Garden Herbarium, SBG = Singapore Botanic Garden, KRP = Purwodadi Botanic Garden, KRB = Bogor Botanic Garden, PF = Pasoh Forest, PA = Pasoh Arboretum, FS = Miami Dade Fruit and Spice Park, GBTB = Gardens by the Bay, FRIM = Forestry Research Institute of Malaysia, JFL/CL/SKP/DV = private individuals). Information on provenance is provided where available. Number of raw reads for each sample is given and inclusion in *phyloBIG/MED/SM* is given as presence/absence (1/0).

Appendix 4.2. Oligonucleotide sequences for eight sets of barcoded adapters used for ddRADseq. Adapters were ordered as single-stranded oligonucleotides, and the forward and reverse strands are given as P1.1 and P1.2, respectively. The unique barcode sequence is shown in lowercase within the full sequence and provided in a separate column for clarity.

Appendix 4.3. Oligonucleotide sequences for twelve indexed PCR primers. One primer (P1) was universal and used for all samples, while the second primer (P2) included a unique index sequence.

Appendix 4.4-01–4.4-30. Maximum likelihood phylogenies for 30 datasets included in this study (see figure for dataset name). Individual sample names are colored to correspond to clades outlined in Fig. 4. (dark green = clade I, light green = clade II, yellow = clade III, orange = clade IV, red = clade V, black = outgroup) with bootstrap support for 100 (4.01-4.09) or 1000 (4.10-4.30) replicates.

Appendix 4.5-01–4.5-03. Quartet-based 50% majority rules phylogenies estimated with SVD Quartets for *phyloSM_c85/90/95_m4* datasets. Individual sample names are colored to correspond to clades outlined in Fig. 4. (dark green = clade I, light green = clade II, yellow = clade III, orange = clade IV, red = clade V, black = outgroup) with bootstrap support from 100 replicates.

Appendices

Appendix 4.1.

Sample Name	Species ID	Collected From	Accession Number	Specimen Number	Provenance	Lane/ Sublibrary	Raw Reads	Dataset		
								BIG	*MED*	*SM*
11A_MI_154_ Madame_Francis	*M. indica*	FTBG	2004-1064*A	–	–	1/11	942123	1	0	0
11C_MI_84_ Thai_Everbearing	*M. indica*	FTBG	2004-1177*A	–	–	1/11	993848	1	0	0
11G_MI_27_Pairi	*M. indica*	FTBG	2004-1081*A	–	–	1/11	1063491	1	0	0
11H_MI_86_M_casturi_982186A	*M. casturi*	FTBG	982186*A	–	Kalimantan	1/11	1335965	1	1	0
12_AMI_127_ Bullocks_Heart	*M. indica*	FTBG	2004-1230*A	–	–	1/12	468764	1	1	0
12_BMI_123_Gaylour	*M. indica*	FTBG	2004-1180	–	–	1/12	478989	1	0	0
12_CMI_71_Siamese	*M. indica*	FTBG	2004-1076	–	–	1/12	257760	1	0	0
12_DMI_90_ M_rubropetala	*M. sp.*	FTBG	2003-1730A	–	Kalimantan	1/12	313415	1	0	0
12_EMI_100_M_pajang	*M. foetida*	FTBG	2012-2354*A	–	Brunei	1/12	244582	1	1	0
12_FMI_151_ M_laurina6PR	*M. laurina*	FTBG	2013-0552 A	–	–	1/12	523064	1	1	1
12_GMI_75_ Braham_kai_mau	*M. indica*	FTBG	2004-1186	–	–	1/12	596576	1	0	0
12_HMI_33_Poh_Gedong	*M. indica*	FTBG	2010-0398A	–	–	1/12	297073	1	0	0
13B_SBG_4_M_pentandra	*M. pentandra*	SBG	20122617*A	EW 150	–	1/13	2100410	1	1	0
13F_MI_81_Mallika	*M. indica*	FTBG	2003-1722*A	–	–	3/13	1919662	1	1	0
13G_KRP_29_M_sp	*M. sp.* Complex 1	KRP	XVI.D.II.11	EW 139	–	3/13	1824176	1	1	0
13H_GBTB_2_M_caesia	Genus novum sp. *caesia*	GBTB	NA	EW 145	–	3/13	286956	1	1	0
14A_MI_92_ M_odorata_Row_E	*M. odorata*	FTBG	2008-1293	–	Malaya	3/14	1891440	1	0	0
14B_KRP_4_ M_foetida_cv_Pakel	*M. odorata*	KRP	IX.C.25	EW 117	Gedong Kuning	3/14	1260649	1	0	0
14C_KRP_31_M_indica_ cv_Gandik_luyung	*M. indica* complex	KRP	IX.B.24A	EW 141	–	3/14	612705	1	0	0
14E_MI_10_ Tommy_Atkins	*M. indica*	FTBG	2003-1734*A	–	–	3/14	1588941	1	0	0
14G_JFL_501_ M_caesia_wanji	Genus novum sp. *caesia*	JFL	NA	–	–	3/14	1496010	1	1	0
14H_KRP_9_M_indica	*M. sp.* Complex 1	KRP	XVI.E.21	EW 122	Sumba, NTT	3/14	1504680	1	0	0
15B_MI_58_zeylanica	*M. zeylanica*	FTBG	2012-2376 A	–	–	3/15	1375976	1	0	0
15C_MI_16_Joe_Long	*M. odorata*	FTBG	2004-1197	–	–	3/15	1546871	1	0	0
15D_FTG_1_ M_sp_Vietnam	*M. sp.* Complex 1	FTG	FTG 00154069	–	Vietnam	3/15	101792	1	0	0
15E_KRP_33_ M_indica_cv_Madu	*M. indica* complex	KRP	IX.B.14c	EW 143	–	3/15	1237635	1	0	0
15F_MI_103_M_foetida	*M. foetida*	FTBG	2014-0266*A	–	–	3/15	1420235	1	0	0
15G_PF_26_M_griffithii	*M. griffithii*	PF	124822	EW 232	Peninsular Malaysia	3/15	894100	1	0	0

16C_MI_40_Cac	*M. indica*	FTBG	2006-1295*A	–	–	3/16	1555717	1	0	0
16E_JFL_504_M_lalijiwa	*M. indica* complex	JFL	NA	–	–	3/16	1627664	1	1	1
16F_FRIM_11_ M_cf_odorata	*M. odorata*	FRIM	Y02-1473, 33020193	EW 188	–	3/16	1189156	1	1	0
16G_SKP_59_M_sp	*M. sp.* Complex 1	SKP	NA	SKP 1044	Nha Trang, Vietnam	3/16	1421892	1	1	1
16H_PA_3_M_quadrifida	*M. sp.* Complex 1	PA	746	EW 197	Peninsular Malaysia	3/16	1443282	1	1	0
17B_SBG_12_M_caesia	Genus novum sp. *caesia*	SBG	19970923*A	EW 159	–	3/17	1716599	1	1	1
17E_SKP_510_M_foetida	*M. sp.* Complex 1	SKP	NA	SKP 1097	Cuc Phuong NP, Vietnam	3/17	1159927	1	1	1
17F_MI_50_ Butterfly_Hainan	*M. indica*	FTBG	2010-0391*A	–	–	3/17	2005251	1	1	0
17G_KRP_30_ M_indica_cv_Gurih	*M. indica* complex	KRP	IX.B.21a	EW 140	Java	3/17	473154	1	0	0
3A_MI_37_Golek	*M. indica*	FTBG	2006-1292*A	–	–	1/3	729910	1	0	0
3B_MI_117_Mabrouka	*M. indica*	FTBG	2003-1713*A	–	–	1/3	561681	1	0	0
3C_MI_20_Carabao	*M. indica*	FTBG	2005-1791*A	–	–	1/3	480442	1	1	0
3D_MI_35_Himsagar	*M. indica*	FTBG	2006-1278*A	–	–	1/3	586464	1	0	0
3E_MI_32_Gedong_Ginco	*M. indica*	FTBG	2004-1215*A	–	–	1/3	440891	1	1	0
3F_MI_69_Langra_Benarsi	*M. indica*	FTBG	2005-1945*A	–	–	1/3	798799	1	0	0
6A_MI_102_Chok_Anon	*M. indica*	FTBG	2006-1301*A	–	–	1/6	672634	1	0	0
6B_MI_158_Baptiste	*M. indica*	FTBG	2006-1309*A	–	–	1/6	511517	1	0	0
6C_MI_11_Nam_Doc_Mai	*M. indica*	FTBG	2003-1724*A	–	–	1/6	603488	1	1	1
6F_MI_148_Gilas	*M. indica*	FTBG	2010-0366*A	–	–	1/6	737869	1	0	0
6G_MI_68_ Pohn_Sawadee	*M. indica*	FTBG	2004-1088*A	–	–	1/6	767434	1	0	0
6H_MI_72_ M_quadrifida_RowA	*M. quadrifida*	FTBG	2012-2379A	–	–	1/6	1056687	1	1	0
9C_MI_130_ M_mempelam	*M. laurina*	FTBG	2012-2371*B	–	–	1/9	719950	1	0	0
9F_MI_156_Depih_Pasir	*M. sp.*	FTBG	2012-2369*A	–	–	1/9	735616	1	1	1
9H_MI_34_Rumanii	*M. indica*	FTBG	2005-1742*A	–	–	1/9	1189430	1	0	0
AA_FS2_M_zeylanica	*M. zeylanica*	FS	NA	–	–	A/2	1389807	1	1	1
AB_FRIM6_M_quadrifida	*M. quadrifida* complex	FRIM	T04-1467	EW 183	–	A/2	251955	1	1	1
AC_FRIM8_M_odorata	*M. odorata*	FRIM	33020171	EW 815	–	A/2	1287058	1	1	0
AD_FRIM10_M_odorata	*M. odorata*	FRIM	W05-1614, 33020185	EW 187	–	A/2	1282316	1	0	0
AE_GBTB_3_M_indica	*M. indica* complex	GBTB	NA	–	–	A/2	1619145	1	0	0
AF_K1_M_foetida	*M. foetida*	FRIM	NA	EW 223	Peninsular Malaysia	A/2	1104520	1	0	0
AG_K2_M_odorata	*M. odorata*	FRIM	NA	EW 224	Peninsular Malaysia	A/2	1483656	1	0	0
AH_KRB_1_M_sp	*M. foetida*	KRB	XIX.F.2.51	–	W. Java	A/2	1901836	1	0	0

BA_FS1_M_odorata	*M. indica* complex	FS	NA	–	–	B/2	987343	1	1	0
BB_K3_M_indica	*M. sp.* Complex 2	FRIM	NA	EW 225	Peninsular Malaysia	B/2	814416	1	0	0
BC_K4_M_foetida	*M. foetida*	FRIM	NA	EW 226	Peninsular Malaysia	B/2	1107182	1	1	0
BD_KRB_5_ M_macrocarpa	*M. macrocarpa*	KRB	VI.B.125	–	Java	B/2	985244	1	1	1
BE_KRB_28_M_sp	*M. pajang*	KRB	V.II.E.181	–	E. Kalimantan	B/2	1199722	1	1	0
BF_KRP1_ M_foetida_pakel	*M. sp.* Complex 1	KRP	IX.B.17	EW 114	Semarang (C. Java)	B/2	684994	1	1	0
BG_KRP3_ M_foetida_pakel	*M. foetida*	KRP	IX.C.27	EW 116	Semarang (C. Java)	B/2	1091284	1	1	0
BH_KRP8_M_foetida	*M. foetida*	KRP	IX.C.11	EW 121	Blitar	B/2	1165602	1	0	0
CA_KRP_5_ M_odorata_cv_Kuweni	*M. odorata*	KRP	IX.C.37a	EW 118	Semarang (C. Java)	C/2	609735	1	1	0
CB_MI_91_Myatrynat	*M. indica*	FTBG	2006-1293*A	–	–	C/2	631319	1	0	0
CD_KRP_6_M_foetida_ cv_Pakel_lumut	*M. indica* complex	KRP	IX.C.9b	EW 119	Semarang (C. Java)	C/2	514920	1	0	0
CF_MI_42_Swethintha	*M. indica*	FTBG	2004-1211*A	–	–	C/2	537252	1	0	0
CG_MI_77_Amrapali	*M. indica*	FTBG	2012-2391*A	–	–	C/2	599107	1	0	0
CH_KRP_7_M_odorata	*M. odorata*	KRP	IX.C.10a	EW 120	Bantul	C/2	713827	1	1	1
DA_MI_114_Zebda	*M. laurina*	FTBG	2005-1788*A	–	–	D/2	887852	1	1	1
DB_MI_24_Turpentine	*M. indica*	FTBG	2003-1738*A	–	–	D/2	799605	1	0	0
DC_KRP_26_M_similis	*M. quadrifida* complex	KRP	XVI.D.II.16	–	E. Kalimantan	D/2	678691	1	1	1
DD_MI_97_Diab	*M. indica*	FTBG	2004-1061*A	–	–	D/2	812452	1	0	0
DF_MI_80_ Nam_Tam_Teem	*M. indica*	FTBG	2004-1228*A	–	–	D/2	702325	1	0	0
EA_KRB_4_M_gedebe	*M. gedebe*	KRB	VI.D.5	–	Sumatra, Lampung	E/2	1210992	1	1	1
EB_MI_74_Pam_kai_mia	*M. indica*	FTBG	2012-2402*A	–	–	E/2	1315873	1	0	0
EC_MI_44_M_aquea	*M. laurina*	FTBG	NA	–	–	E/2	1534535	1	0	0
ED_SBG2014_10_ M_magnifica	*M. quadrifida* complex	SBG	20110755*A	–	–	E/2	1323694	1	1	0
EF_MI_95_Cairo	*M. indica*	FTBG	2004-1219*A	–	–	E/2	1072510	1	0	0
EG_MI_85_M_laurina_1_B4	*M. sp.*	FTBG	2013-0555A	–	–	E/2	1466982	1	1	1
FA_MI_76_Chao_savoy	*M. indica*	FTBG	2003-1712*A	–	–	F/2	1760475	1	0	0
FD_KRP_28_M_casturi	*M. casturi*	KRP	XVI.D.II.14	EW 138	S. Kalimantan	F/2	1650119	1	1	1
FF_MI51_Royal_Special	*M. indica*	FTBG	2004-1087*A	–	–	F/2	1311009	1	0	0
FG_MI_53_Sindhri	*M. indica*	FTBG	2004-1055*A	–	–	F/2	1876470	1	0	0
GA_MI_47_Rawa	*M. quadrifida*	FTBG	2012-2356	–	–	G/2	1172739	1	1	0
GB_KRP_11_M_sp	*M. casturi*	KRP	XVI.E.22	EW 134	Maluku	G/2	1107713	1	0	0
GD_KRB_7_M_applanata	Genus novum sp. *caesia*	KRB	VI.B.108a	–	Kalimantan	G/2	947622	1	0	0

GF_MI_64_Poh_Pakel	*M. lalijiwa*	FTBG	2010-0397*A	–	–	G/2	928187	1	0	0
GH_MI_57_Jumbo_Kesar	*M. indica*	FTBG	2009-0822*A	–	–	G/2	1219797	1	0	0
HA_MI_88_ Pancahdarakalasa	*M. indica*	FTBG	2004-1181*A	–	–	H/2	1561762	1	0	0
HC_MI_63_Dusheri	*M. indica*	FTBG	2005-1765*A	–	–	H/2	1599887	1	1	0
HD_KRP_12_M_sp	*M. gedebe*	KRP	XVI.E.48	EW 125	E. Kalimantan	H/2	1498756	1	1	0
HE_MI_39_Pyu_Pyu_Kalay	*M. indica*	FTBG	2009-0816*A	–	–	H/2	1942535	1	1	0
HF_MI_79_ Aslul_Mukararara	*M. indica*	FTBG	2008-1289*A	–	–	H/2	1222409	1	0	0
HH_MI_88_M_pentandra	*M. sp.*	FTBG	2012-2368*A	–	–	H/2	1712812	1	1	0
IA_MI_12_Tong_Dam	*M. indica*	FTBG	2003-1707*A	–	–	I/2	1032280	1	0	0
IB_KRB_6_M_oblongifolia	*M. pajang*	KRB	VI.B.151	–	E. Kalimantan	I/2	778298	1	1	0
IC_MI_59_M_lalijiwa_G	*M. lalijiwa*	FTBG	2004-1213*A	–	–	I/2	1167782	1	1	0
ID_MI_14_Alphonso	*M. indica*	FTBG	2004-1053*A	–	–	I/2	1518263	1	1	1
IE_MI_87_Depih_Biasa	*M. sp.*	FTBG	2012-2373*A	–	–	I/2	1159995	1	1	1
IF_MI_43_Saigon	*M. indica*	FTBG	2006-1276*A	–	–	I/2	832868	1	1	0
IG_MI_4_Phimsen_Mun	*M. indica*	FTBG	2003-1753*A	–	–	I/2	990893	1	0	0
IH_MI_36_ Alampur_Baneshan	*M. indica*	FTBG	2010-0370*A	–	–	I/2	1092326	1	0	0
JB_FRIM_13_M_gedebe	*M. gedebe*	FRIM	33020263	EW 190	–	J/2	831842	1	1	0
JC_MI_55_Frenz_odorata	*M. odorata*	FTBG	2010-0365*A	–	–	J/2	1317838	1	0	0
JD_MI_3_Cambodiana	*M. indica*	FTBG	2005-1753*A	–	–	J/2	1023296	1	0	0
JE_MI_61_Praya_Savoy	*M. indica*	FTBG	2006-1289*A	–	–	J/2	1142384	1	0	0
JF_MI_116_ Tyler_Premiere	*M. indica*	FTBG	2004-1218*A	–	–	J/2	754210	1	1	0
JG_MI_54_Sig_Siput	*M. indica*	FTBG	2005-1746*A	–	–	J/2	893846	1	0	0
JH_MI_70_M_rampagni	*M. odorata*	FTBG	2001-0889*A	–	Sarawak	J/2	1097061	1	0	0
KA_CL_1_M_altissima	*M. sp.* Complex 1	CL	NA	–	–	K/2	1311074	1	1	1
KB_PA_1_M_griffithii	*M. griffithii*	PA	585	EW 196	Peninsular Malaysia	K/2	1467741	1	0	0
KE_KRP_19_M_sp	*M. sp.* 'KRP1'	KRP	XVI.E.52	EW 132	N. Sulawesi	K/2	1433693	1	1	1
KG_FRIM_17_M_foetida	*M. foetida*	FRIM	33020262	EW 194	–	K/2	1423614	1	0	0
LB_KRP_32_ M_indica_kepodang	*M. indica* complex	KRP	IX.B.18A	EW 142	–	L/2	972685	1	0	0
LC_PA_6_M_foetida	*M. foetida*	PA	NA	EW 200	Peninsular Malaysia	L/2	1062234	1	1	0
LD_PA_5_M_quadrifida	*M. sp.* Complex 1	PA	134	EW 199	Peninsular Malaysia	L/2	997042	1	1	1
LE_CL_3_M_monandra	*M. monandra*	CL	NA	–	–	L/2	1165246	1	1	1
LF_SBG_19A_ M_quadrifida	*M. quadrifida* complex	SBG	20110756*A	EW 162	–	L/2	813211	1	1	1
LG_PA_7_M_quadrifida	*M. longipetiolata*	PA	350	EW 221	Peninsular Malaysia	L/2	1046816	1	1	0
LH_PF_17_M_sp	*M. sp.* Complex 1	PF	NA	EW 217	Peninsular Malaysia	L/2	511627	1	1	0

MB_PF_6_M_sp_Griffithii	*M. griffithii*	PF	64758	EW 206	Peninsular Malaysia	M/3	498864	1	1	0
MC_MI_101_ M_quadrifida	*M. quadrifida*	FTBG	2012-2356*A	–	–	M/3	2765794	1	1	0
MD_KRP_13_M_sp	*M. odorata*	KRP	XVI.E.44	EW 126	S. Kalimantan	M/3	1808845	1	0	0
MG_KRP_10_M_longipes	*M. sp.* Complex 1	KRP	XVI.E.24	EW 123	Sumba, NTT	M/3	1736083	1	1	0
MH_FRIM_5_M_odorata	*M. odorata*	FRIM	33020097	EW 182	–	M/3	2014983	1	0	0
NA_MI_56_ Hindi_Besanara	*M. indica*	FTBG	2004-1233*A	–	–	N/3	1511170	1	0	0
NB_SBG_9_M_pajang	*M. pajang*	SBG	20030635*B	EW 156	–	N/3	1306837	1	1	0
ND_KRP_20_ M_cf_kemanga	Genus novum sp. *caesia*	KRP	XVI.F.I.7	–	E. Kalimantan	N/3	682941	1	0	0
NE_SBG_2014_3_ M_caesia	Genus novum sp. *caesia*	SBG	00/05062*A	–	–	N/3	1201478	1	0	0
NF_KRP_24_M_sp	*M. sp.* Complex 1	KRP	XVI.D.II.2	–	Sulawesi	N/3	1145006	1	1	1
NH_PF_2_M_superba	Genus novum sp. *superba*	PF	74350	EW 202	Peninsular Malaysia	N/3	968760	1	0	0
OA_Duval_01_ M_pelipisan	*M. sp.* Complex 1	DV	NA	–	–	O/3	1153813	1	1	1
OC_PA_8_M_griffithii	*M. griffithii*	PA	173	EW 222	Peninsular Malaysia	O/3	1086868	1	0	0
OD_PF_25_G_malayana	*Gluta malayana*	PF	16708	EW 231	Peninsular Malaysia	O/3	1328115	1	1	1
OF_SBG_13_ B_oppositifolia	*Bouea oppositifolia*	SBG	20105419*B	EW 160	–	O/3	761593	1	1	1
OG_SBG_34_M_odorata	*M. odorata*	SBG	19970994*A	EW 176	–	O/3	1278585	1	0	0
PA_MI_89_Aeromanis	*M. indica*	FTBG	2004-1189*A	–	–	P/3	1549107	1	1	1
PB_FRIM_7_M_caesia	Genus novum sp. *caesia*	FRIM	33020187	EW 184	–	P/3	1568198	1	0	0
PD_KRB_31_M_sp	*M. foetida*	KRB	VII.E.179	–	N. Sulawesi	P/3	1543712	1	1	0
PE_PF_22_M_magnifica	*M. magnifica*	PF	122925	EW 228	Peninsular Malaysia	P/3	1546703	1	1	0
PF_PF_18_ M_subsessilifolia	*M. subsessilifolia*	PF	NA	EW 218	Peninsular Malaysia	P/3	523956	1	1	0
QB_KRB_8_M_griffithii	*M. casturi*	KRB	VII.E.170a	–	S. Kalimantan	Q/3	1223203	1	1	0
QC_FS_3_A_occidentale	*Anacardium occidentale*	FS	NA	–	–	Q/3	1410021	1	1	1
QD_MI_9_Totapuri	*M. indica*	FTBG	2004-1222*A	–	–	Q/3	1574796	1	0	0
QH_PF_1_M_sp	*M. griffithii*	PF	NA	EW 201	Peninsular Malaysia	Q/3	1636681	1	1	0
RB_SBG_10_M_pajang	*M. pajang*	SBG	NA	EW 157	–	R/3	1169736	1	1	1
RC_Mcaes1_M_caesia	*M. caesia*	FTBG	2012-2361*A	–	Sumatra	R/3	1381216	1	0	0
RD_PF_15_M_sp	*M. quadrifida* complex	PF	151873	EW 214	Peninsular Malaysia	R/3	1078866	1	1	0
RE_MI_31_Kaeo_Luemkon	*M. indica*	FTBG	2007-1056*A	–	–	R/3	1396821	1	0	0

RF_SBG_24b_M_griffithii	*M. griffithii*	SBG	NA	EW 168	–	R/3	1066482	1	1	1
RH_MI_15_Ivory	*M. indica*	FTBG	2003-1723*A	–	–	R/3	1658796	1	0	0
SA_GBTB_1_M_griffithii	*M. griffithii*	GBTB	NA	EW 146	–	S/3	1252858	1	1	0
SB_MI_78_M_A_carle	*M. casturi*	FTBG	2007-1060*A	–	–	S/3	1612855	1	1	0
SC_KRP_27_M_indica	*M. sp.* Complex 1	KRP	XVI.D.II.8	EW 137	Sulawesi	S/3	1594169	1	1	0
SE_PF_16_M_sp	*M. sp.* 'limus'	PF	NA	EW 216	Peninsular Malaysia	S/3	1929107	1	1	1
SF_SBG_25_M_foetida	*M. foetida*	SBG	NA	EW 169	–	S/3	1367395	1	0	0
SG_SBG_16_ B_oppositifolia	*Bouea oppositifolia*	SBG	20105419*D	–	–	S/3	1067279	1	0	0
SH_PF_13_M_superba	Genus novum sp. *superba*	PF	151471	EW 213	Peninsular Malaysia	S/3	1379844	1	0	0
TA_PF_24_ M_subsessilifolia	*M. subsessilifolia*	PF	26703	EW 230	Peninsular Malaysia	T/1	281205	1	1	1
TB_KRP_2_M_minor	*M. indica* complex	KRP	IX.B.32	EW 115	Sulawesi Tengah	T/1	1587805	1	1	0
TC_KRB_2_M_similis	*M. quadrifida* complex	KRB	VI.D.8a	–	Bangka I., S. Sumatra	T/1	1755500	1	1	0
TD_KRP_22_M_sp	*M. laurina*	KRP	XVI.F.I.19a	EW 134	–	T/1	1830319	1	0	0
TE_PF_23_ M_subsessilifolia	*M. subsessilifolia*	PF	66366	EW 229	Peninsular Malaysia	T/1	351204	1	1	0
TF_KRB_30_ M_rufocostata	*M. rufocostata*	KRB	VII.E.178	–	E. Kalimantan	T/1	2007780	1	1	1
TG_MI_165_M_griffithii	*M. griffithii*	FTBG	2012-2372*A	–	–	T/1	2155304	1	0	0
TH_PF_3_M_magnifica	*M. magnifica*	PF	NA	EW 203	Peninsular Malaysia	T/1	321271	1	1	1
UA_SBG_35_M_foetida	*M. foetida*	SBG	00/6994*A	EW 177	–	U/1	2400886	1	1	0
UB_SBG_30_M_foetida	*M. foetida*	SBG	20060155*F	EW 172	–	U/1	2399019	1	0	0
UC_KRP_17_ M_macrocarpa	*M. macrocarpa*	KRP	XVI.E.56	EW 130	W. Kalimantan	U/1	2654829	1	1	0
UD_SBG_22_M_foetida	*M. foetida*	SBG	NA	EW 166	–	U/1	2849945	1	0	0
UE_SBG_28_M_foetida	*M. foetida*	SBG	003135*A	EW 171	–	U/1	2537820	1	1	1
UF_PF_27_M_superba	Genus novum sp. *superba*	PF	144915	EW 233	Peninsular Malaysia	U/1	556437	1	1	1
UG_PF_10_M_gracilipes	*M. gracilipes*	PF	214421	EW 210	Peninsular Malaysia	U/1	2893557	1	1	1
UH_SBG_32_M_foetida	*M. foetida*	SBG	20091887*A	EW 174	–	U/1	2343765	1	0	0
WA_MI_45_ M_PR_Martex	*M. laurina*	FTBG	2012-2413*A	–	–	W/1	1917441	1	0	0
WF_MI_52_Maha_Chanok	*M. indica*	FTBG	2012-2399*A	–	–	W/1	2338767	1	0	0
XA_SBG_20_M_gedebe	*M. gedebe*	SBG	NA	EW 164	–	X/1	1489679	1	0	0
XB_SBG_6_M_quadrifida	*M. quadrifida* complex	SBG	NA	EW 152	–	X/1	1559839	1	1	0
XC_SBG_8_B_macrophylla	*Bouea macrophylla*	SBG	19970946*A	EW 155	–	X/1	1161155	1	1	1
XD_SBG_27_M_foetida	*M. foetida*	SBG	003135*A	EW 170	–	X/1	1443188	1	0	0

XE_SBG_23_M_odorata	*M. odorata*	SBG	NA	EW 167	–	X/1	1468589	1	1	0
XF_FRIM_3_M_indica	*M. indica* complex	FRIM	NA	EW 180	–	X/1	1676509	1	0	0
XG_FRIM_2_M_foetida	*M. foetida*	FRIM	NA	EW 179	–	X/1	1479415	1	0	0
XH_SBG_31_M_odorata	*M. indica* complex	SBG	NA	EW 173	–	X/1	1514572	1	1	0
YB_MI_48_Imam_Pasand	*M. indica*	FTBG	2004-1187*A	–	–	Y/1	1222060	1	0	0
YC_MI_65_Ratna	*M. indica*	FTBG	2012-2403*B	–	–	Y/1	1833551	1	0	0
YH_MI_73_Cowasji_patel	*M. indica*	FTBG	2004-1079*A	–	–	Y/1	3144639	1	0	0
ZA_SBG_7a_M_kemanga	Genus novum sp. *superba*	SBG	20060155*b	EW 153	–	Z/1	825693	1	0	0
ZB_FRIM_1_M_foetida	*M. foetida*	FRIM	NA	EW 178	–	Z/1	729331	1	1	0
ZC_SBG_33_M_odorata	*M. odorata*	SBG	20090008*A	EW 175	–	Z/1	1040003	1	0	0
ZD_SBG_2_M-foetida	*M. foetida*	SBG	200903556 *C	EW 148	–	Z/1	576639	1	0	0
ZE_SBG_21_M_foetida	*M. foetida*	SBG	NA	EW 165	–	Z/1	996406	1	0	0
ZF_SBG_3_M-odorata	*M. odorata*	SBG	20093557*A	EW 149	–	Z/1	856966	1	0	0
ZG_SBG_26_M_pentandra	*M. pentandra*	SBG	20117045*A	–	–	Z/1	823561	1	1	1
ZH_SBG_1_M_sp	*M. sp.* Complex 1	SBG	NA	EW 147	–	Z/1	985203	1	1	1

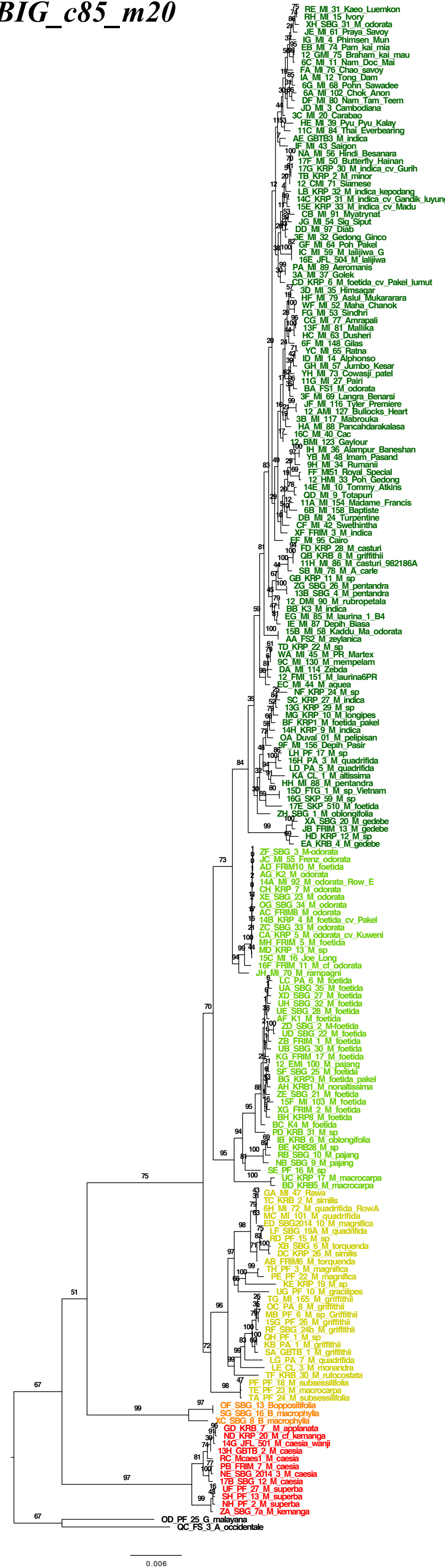
Appendix 4.2.

Adapter Name	Oligonucleotide Sequence	Barcode Sequence
HBG_adapt1_flex_P1.1	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT cca gag tgt CA T*G	CCAGAGTGT
HBG_adapt1_flex_P1.2	/5Phos/aca ctc tgg AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT	
HBG_adapt2_flex_P1.1	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT tga gcg act CA T*G	TGAGCGACT
HBG_adapt2_flex_P1.2	/5Phos/agt cgc tca AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT	
HBG_adapt3_flex_P1.1	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT tgg tct ctg CA T*G	TGGTCTCTG
HBG_adapt3_flex_P1.2	/5Phos/cag aga cca AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT	
HBG_adapt4_flex_P1.1	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT gta atc cag CAT*G	GTAATCCAG
HBG_adapt4_flex_P1.2	/5Phos/ctg gat tac AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT	
HBG_adapt5_flex_P1.1	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT gaa tgc gtc CAT*G	GAATGCGTC
HBG_adapt5_flex_P1.2	/5Phos/gac gca ttc AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT	
HBG_adapt6_flex_P1.1	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT atc agt gac CAT*G	ATCAGTGAC
HBG_adapt6_flex_P1.2	/5Phos/gtc act gat AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT	
HBG_adapt7_flex_P1.1	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT cac cga cta CAT*G	CACCGACTA
HBG_adapt7_flex_P1.2	/5Phos/tag tcg gtg AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT	
HBG_adapt8_flex_P1.1	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT gac gcg tga CAT*G	GACGCGTGA
HBG_adapt8_flex_P1.2	/5Phos/tca cgc gtc AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT	

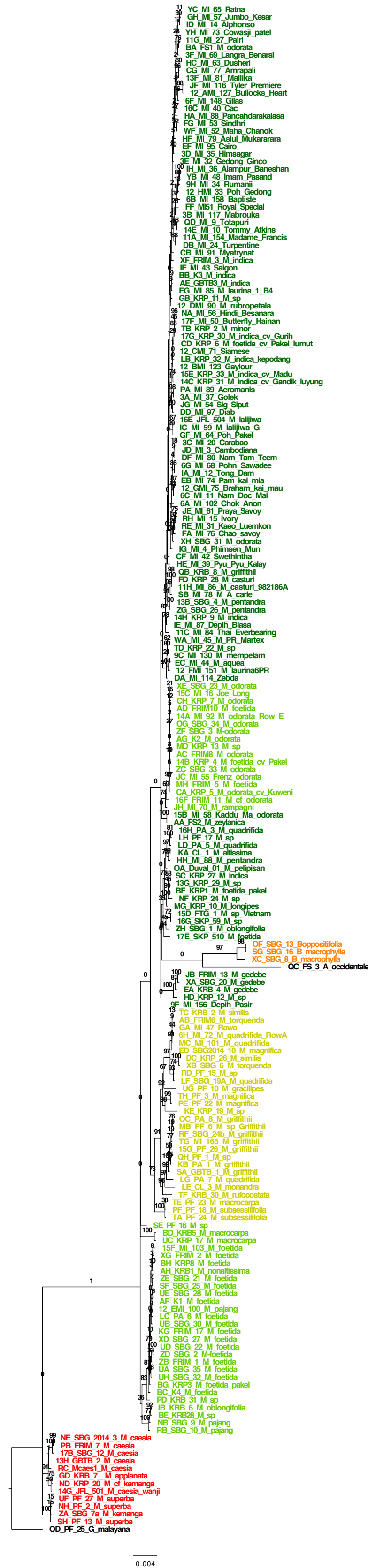
Appendix 4.3

PCR Primer	PCR Index	Index Sequence	Oligonucleotide sequence
PCR1	–	–	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC* G
PCR2	1	ATCACG	CAA GCA GAA GAC GGC ATA CGA GAT CGT GAT GTG ACT GGA GTT CAG ACG TGT G*C
PCR2	3	TTAGGC	CAA GCA GAA GAC GGC ATA CGA GAT GCC TAA GTG ACT GGA GTT CAG ACG TGT G*C
PCR2	7	CAGATC	CAA GCA GAA GAC GGC ATA CGA GAT GAT CTG GTG ACT GGA GTT CAG ACG TGT G*C
PCR2	12	CTTGTA	CAA GCA GAA GAC GGC ATA CGA GAT TAC AAG GTG ACT GGA GTT CAG ACG TGT G*C
PCR2	13	AGTCAA	CAA GCA GAA GAC GGC ATA CGA GAT TTG ACT GTG ACT GGA GTT CAG ACG TGT G*C
PCR2	16	CCGTCC	CAA GCA GAA GAC GGC ATA CGA GAT GGA CGG GTG ACT GGA GTT CAG ACG TGT G*C
PCR2	21	GTTTCG	CAA GCA GAA GAC GGC ATA CGA GAT CGA AAC GTG ACT GGA GTT CAG ACG TGT G*C
PCR2	24	GGTAGC	CAA GCA GAA GAC GGC ATA CGA GAT GCT ACC GTG ACT GGA GTT CAG ACG TGT G*C
PCR2	29	CAACTA	CAA GCA GAA GAC GGC ATA CGA GAT TAG TTG GTG ACT GGA GTT CAG ACG TGT G*C
PCR2	37	ATTCCG	CAA GCA GAA GAC GGC ATA CGA GAT ATT CCG GTG ACT GGA GTT CAG ACG TGT G*C
PCR2	42	TAATCG	CAA GCA GAA GAC GGC ATA CGA GAT CGA TTA GTG ACT GGA GTT CAG ACG TGT G*C
PCR2	43	TACAGC	CAA GCA GAA GAC GGC ATA CGA GAT GCT GTA GTG ACT GGA GTT CAG ACG TGT G*C

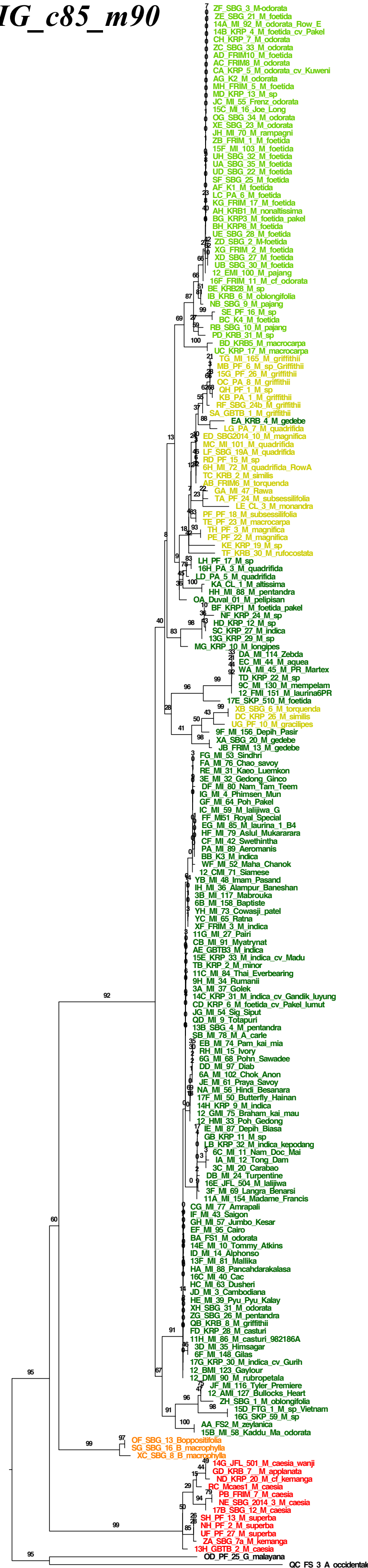
Appendix 4.4-01. *phyloBIG_c85_m20*



Appendix 4.4-02. *phyloBIG_c85_m50*

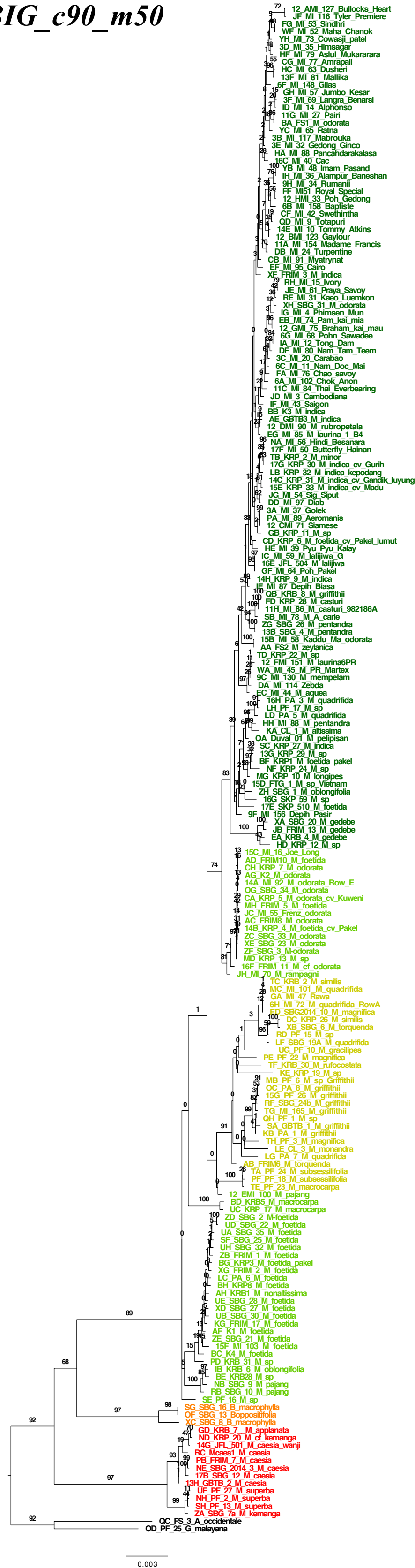


Appendix 4.4-03. *phyloBIG_c85_m90*



Appendix 4.4-04. *phyloBIG_c90_m20*

Appendix 4.4-05. *phyloBIG_c90_m50*



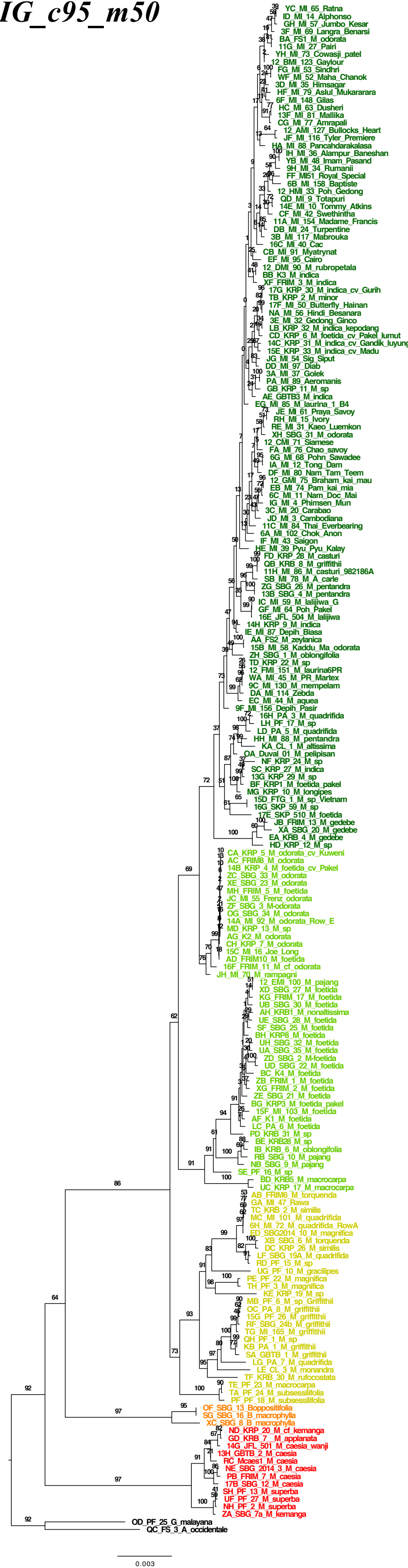
Appendix 4.4-06. *phyloBIG_c90_m90*

ZB_FRIM_1_M_foetida
15F_MI_103_M_foetida
UE_SBG_28_M_foetida
ZE_SBG_21_M_foetida
XE_SBG_23_M_odorata
AD_FRIM10_M_foetida
MD_KRP_13_M_sp
14B_KRP_4_M_foetida_cv_Pakel
CH_KRP_7_M_odorata
14A_MI_92_M_odorata_Row_E
AC_FRIM8_M_odorata
MH_FRIM_5_M_foetida
15C_MI_16_Joe_Long
JC_MI_55_Frenz_odorata
OG_SBG_34_M_odorata
ZC_SBG_33_M_odorata
ZF_SBG_3_M-odorata
AG_K2_M_odorata
CA_KRP_5_M_odorata_cv_Kuweni
BG_KRP3_M_foetida_pakel
LC_PA_6_M_foetida
BH_KRP8_M_foetida
KG_FRIM_17_M_foetida
AF_K1_M_foetida
AH_KRB1_M_nonaltissima
UD_SBG_22_M_foetida
UH_SBG_32_M_foetida
UB_SBG_30_M_foetida
UA_SBG_35_M_foetida
SF_SBG_25_M_foetida
XG_FRIM_2_M_foetida
ZD_SBG_2_M-foetida
XD_SBG_27_M_foetida
JH_MI_70_M_rampagni
12_EMI_100_M_pajang
16F_FRIM_11_M_cf_odorata
IB_KRB_6_M_oblongifolia
BE_KRB28_M_sp
NB_SBG_9_M_pajang
BC_K4_M_foetida
SE_PF_16_M_sp
PD_KRB_31_M_sp
RB_SBG_10_M_pajang
UC_KRP_17_M_macrocarpa
BD_KRB5_M_macrocarpa
TG_MI_165_M_griffithii
15G_PF_26_M_griffithii
QH_PF_1_M_sp
KB_PA_1_M_griffithii
MB_PF_6_M_sp_Griffithii
OC_PA_8_M_griffithii
RF_SBG_24b_M_griffithii
SA_GBTB_1_M_griffithii
LG_PA_7_M_quadrifida
EA_KRB_4_M_gedebe
RD_PF_15_M_sp
LF_SBG_19A_M_quadrifida
GA_MI_47_Rawa
ED_SBG2014_10_M_magnifica
6H_MI_72_M_quadrifida_RowA
MC_MI_101_M_quadrifida
TC_KRB_2_M_similis
AB_FRIM6_M_torquenda
TF_KRB_30_M_rufocostata
TH_PF_3_M_magnifica
PE_PF_22_M_magnifica
KE_KRP_19_M_sp
PF_PF_18_M_subsessilifolia
TA_PF_24_M_subsessilifolia
TE_PF_23_M_macrocarpa
16H_PA_3_M_quadrifida
LH_PF_17_M_sp
LD_PA_5_M_quadrifida
HH_MI_88_M_pentandra
KA_CL_1_M_altissima
OA_Duval_01_M_pelipisan
EC_MI_44_M_aquea
DA_MI_114_Zebda
TD_KRP_22_M_sp
WA_MI_45_M_PR_Martex
9C_MI_130_M_mempelam
12_FMI_151_M_laurina6PR
17E_SKP_510_M_foetida
XB_SBG_6_M_torquenda
DC_KRP_26_M_similis
9F_MI_156_Depih_Pasir
XA_SBG_20_M_gedebe
JB_FRIM_13_M_gedebe
BF_KRP1_M_foetida_pakel
HD_KRP_12_M_sp
NF_KRP_24_M_sp
13G_KRP_29_M_sp
SC_KRP_27_M_indica
MG_KRP_10_M_longipes
LE_CL_3_M_monandra
FG_MI_53_Sindhri
FA_MI_76_Chao_savoy
3E_MI_32_Gedong_Ginco
6B_MI_158_Baptiste
DF_MI_80_Nam_Tam_Teem
3C_MI_20_Carabao
RE_MI_31_Kaeo_Luemkon
12_DMI_90_M_rubropetala
12_BMI_123_Gaylour
12_CMI_71_Siamese
CG_MI_77_Amrapali
WF_MI_52_Maha_Chanok
DB_MI_24_Turpentine
BB_K3_M_indica
EG_MI_85_M_laurina_1_B4
HF_MI_79_Aslul_Mukararara
3D_MI_35_Himsagar
6F_MI_148_Gilas
DD_MI_97_Diab
YB_MI_48_Imam_Pasand
IH_MI_36_Alampur_Baneshan
XF_FRIM_3_M_indica
YC_MI_65_Ratna
YH_MI_73_Cowasji_patel
17G_KRP_30_M_indica_cv_Gurih
FF_MI51_Royal_Special
CF_MI_42_Swethintha
PA_MI_89_Aeromanis
3B_MI_117_Mabrouka
6G_MI_68_Pohn_Sawadee
RH_MI_15_Ivory
IA_MI_12_Tong_Dam
EB_MI_74_Pam_kai_mia
14H_KRP_9_M_indica
NA_MI_56_Hindi_Besanara
17F_MI_50_Butterfly_Hainan
6A_MI_102_Chok_Anon
CD_KRP_6_M_foetida_cv_Pakel_lumut
14C_KRP_31_M_indica_cv_Gandik_luyung
11C_MI_84_Thai_Everbearing
CB_MI_91_Myatrynat
GB_KRP_11_M_sp
3A_MI_37_Golek
6C_MI_11_Nam_Doc_Mai
LB_KRP_32_M_indica_kepodang
15E_KRP_33_M_indica_cv_Madu
TB_KRP_2_M_minor
JG_MI_54_Sig_Siput
AE_GBTB3_M_indica
9H_MI_34_Rumanii
QD_MI_9_Totapuri
11A_MI_154_Madame_Francis
12_HMI_33_Poh_Gedong
16E_JFL_504_M_lalijiwa
IC_MI_59_M_lalijiwa_G
ZG_SBG_26_M_pentandra
IG_MI_4_Phimsen_Mun
11H_MI_86_M_casturi_982186A
XH_SBG_31_M_odorata
SB_MI_78_M_A_carle
FD_KRP_28_M_casturi
EF_MI_95_Cairo
IF_MI_43_Saigon
GH_MI_57_Jumbo_Kesar
ID_MI_14_Alphonso
HA_MI_88_Pancahdarakalasa
14E_MI_10_Tommy_Atkins
GF_MI_64_Poh_Pakel
13B_SBG_4_M_pentandra
QB_KRB_8_M_griffithii
JD_MI_3_Cambodiana
JE_MI_61_Praya_Savoy
11G_MI_27_Pairi
12_GMI_75_Braham_kai_mau
16C_MI_40_Cac
IE_MI_87_Depih_Biasa
13F_MI_81_Mallika
3F_MI_69_Langra_Benarsi
HC_MI_63_Dusheri
BA_FS1_M_odorata
HE_MI_39_Pyu_Pyu_Kalay
JF_MI_116_Tyler_Premiere
12_AMI_127_Bullocks_Heart
16G_SKP_59_M_sp
15D_FTG_1_M_sp_Vietnam
ZH_SBG_1_M_oblongifolia
15B_MI_58_Kaddu_Ma_odorata
AA_FS2_M_zeylanica
UG_PF_10_M_gracilipes
OF_SBG_13_Boppositifolia
SG_SBG_16_B_macrophylla
XC_SBG_8_B_macrophylla
14G_JFL_501_M_caesia_wanji
GD_KRB_7__M_applanata
ND_KRP_20_M_cf_kemanga
RC_Mcaes1_M_caesia
NE_SBG_2014_3_M_caesia
PB_FRIM_7_M_caesia
17B_SBG_12_M_caesia
NH_PF_2_M_superba
UF_PF_27_M_superba
SH_PF_13_M_superba
ZA_SBG_7a_M_kemanga
13H_GBTB_2_M_caesia
OD_PF_25_G_malayana
QC_FS_3_A_occidentale

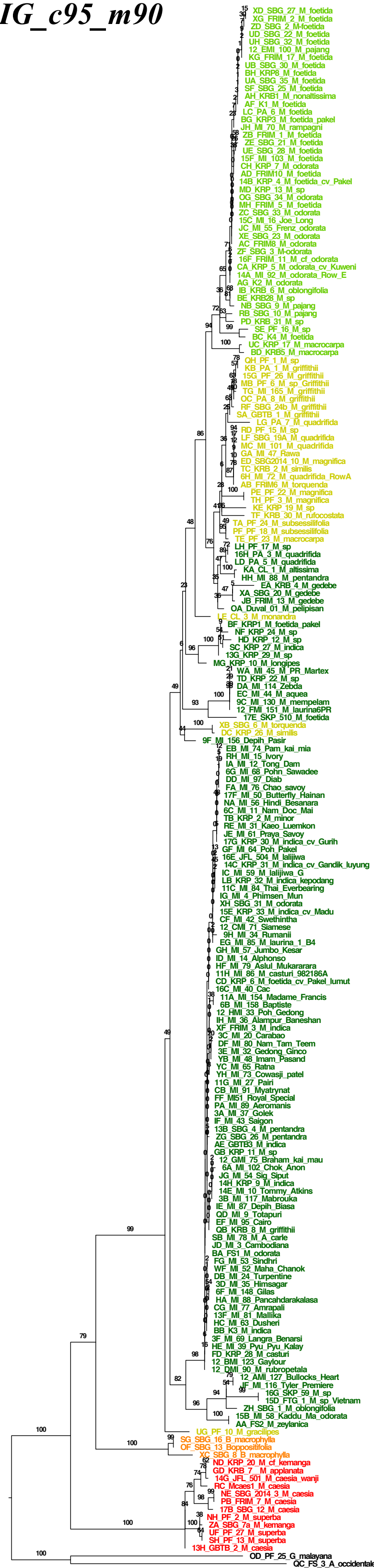
9.0E-4

Appendix 4.4-07. *phyloBIG_c95_m20*

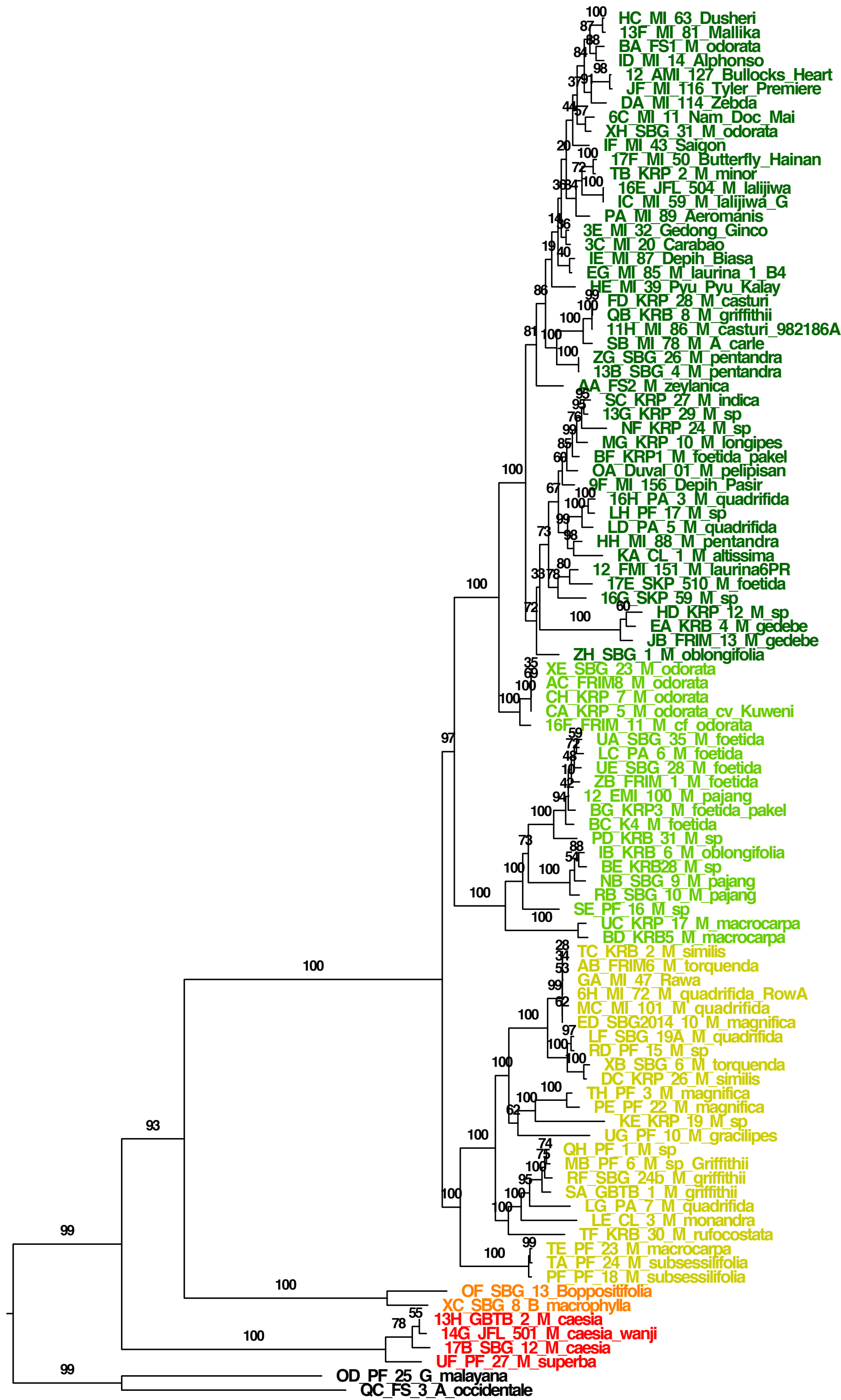
Appendix 4.4-08. *phyloBIG_c95_m50*



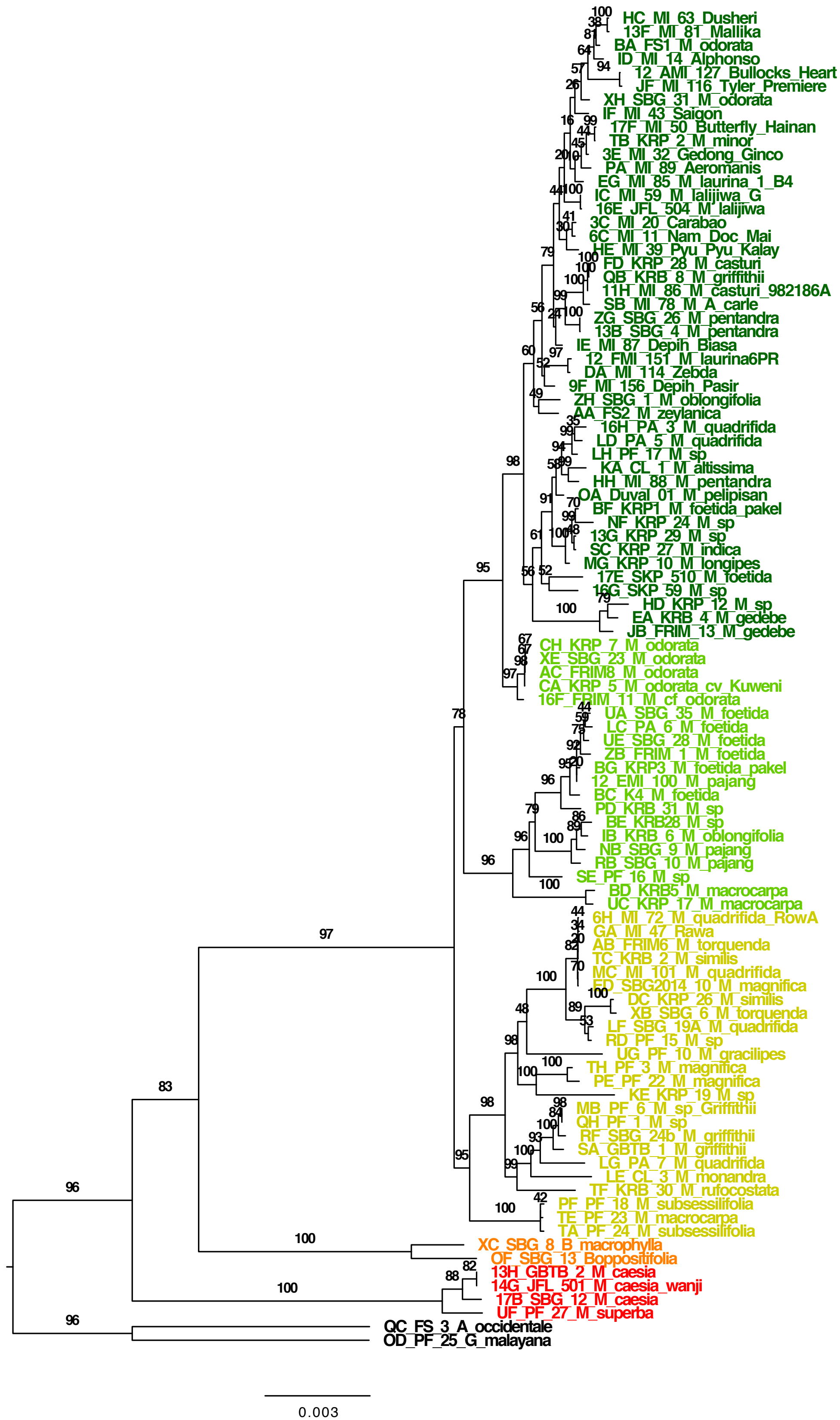
Appendix 4.4-09. *phyloBIG_c95_m90*



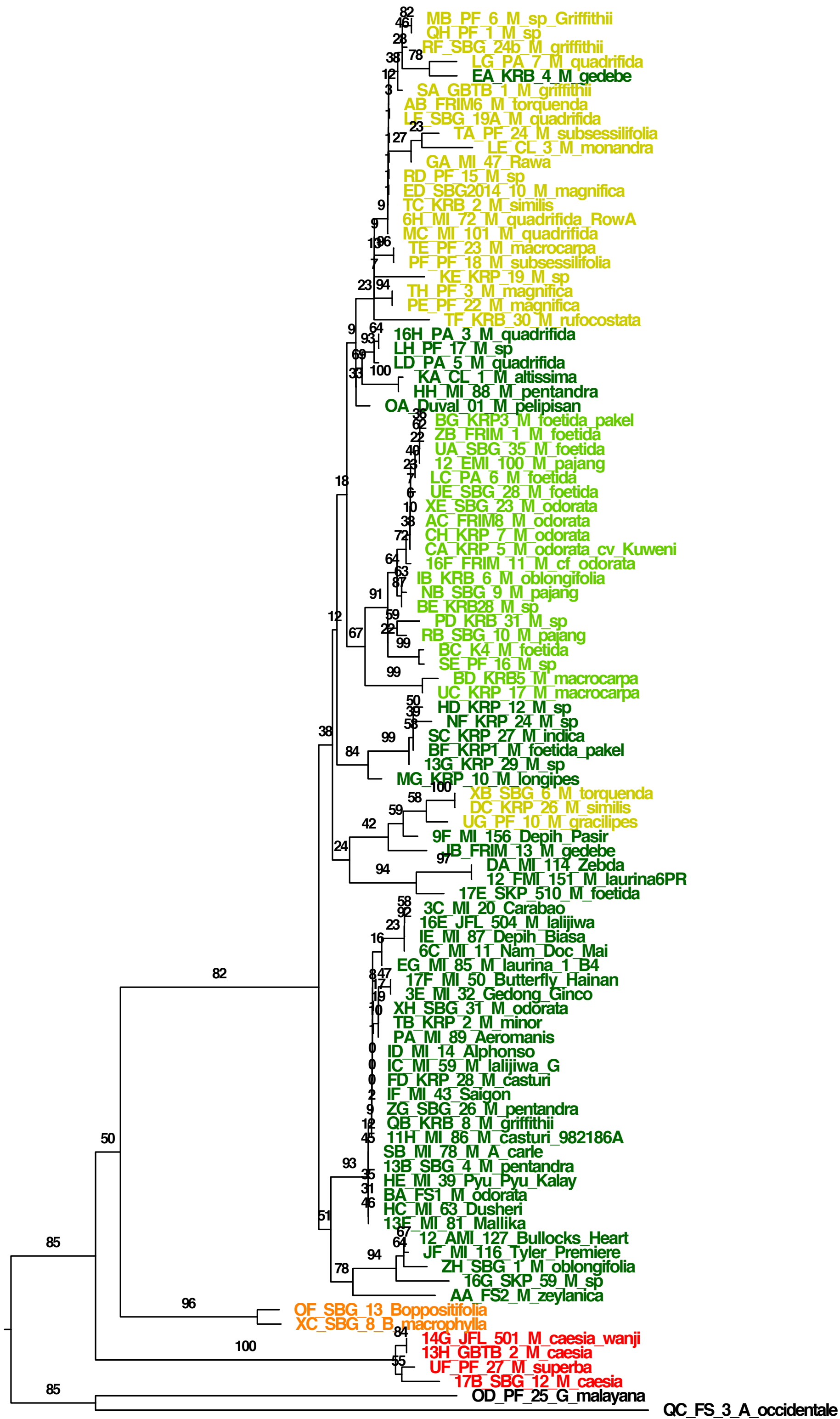
Appendix 4.4-10. ***phyloMED_c85_m20***



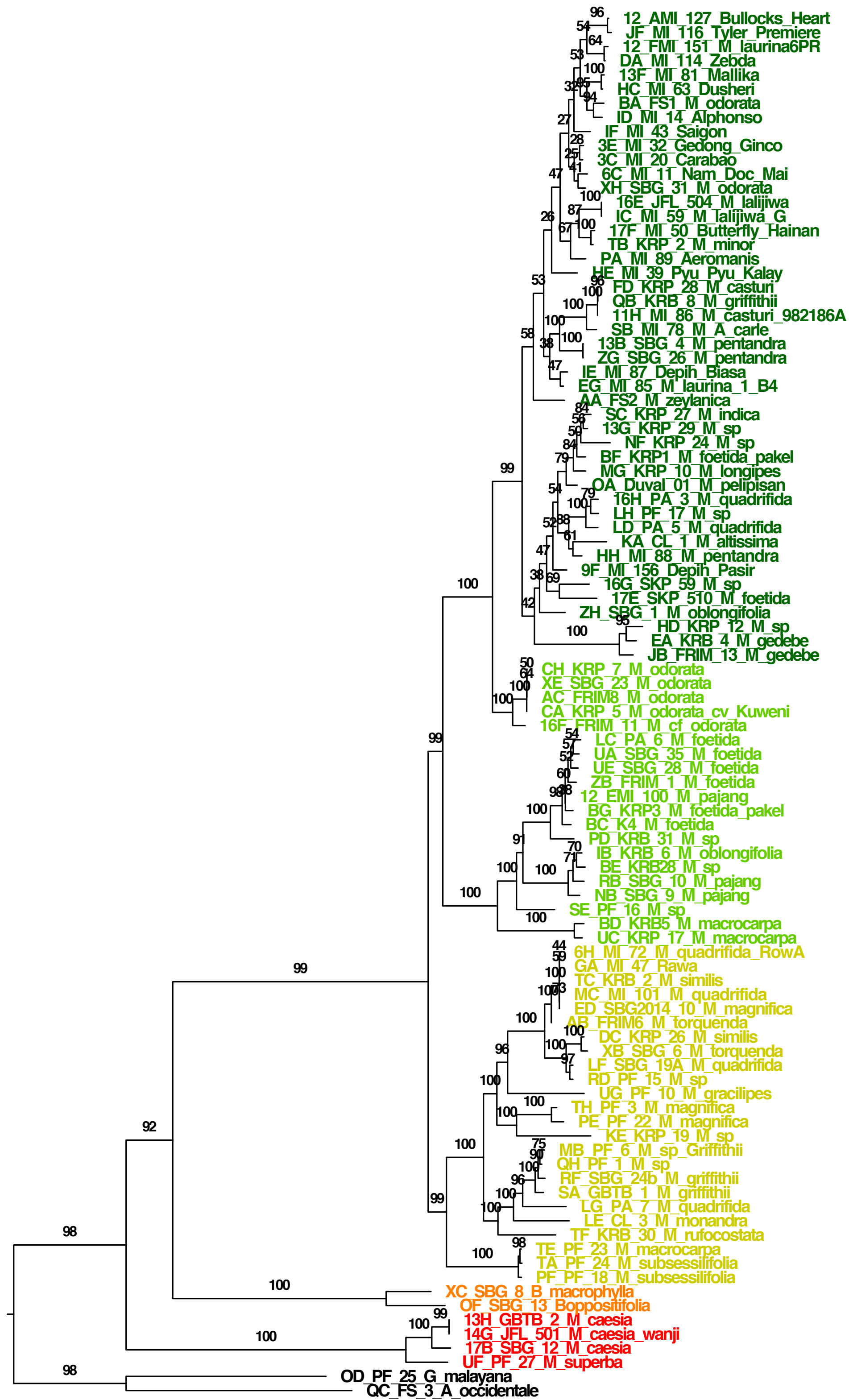
Appendix 4.4-11. ***phyloMED_c85_m50***



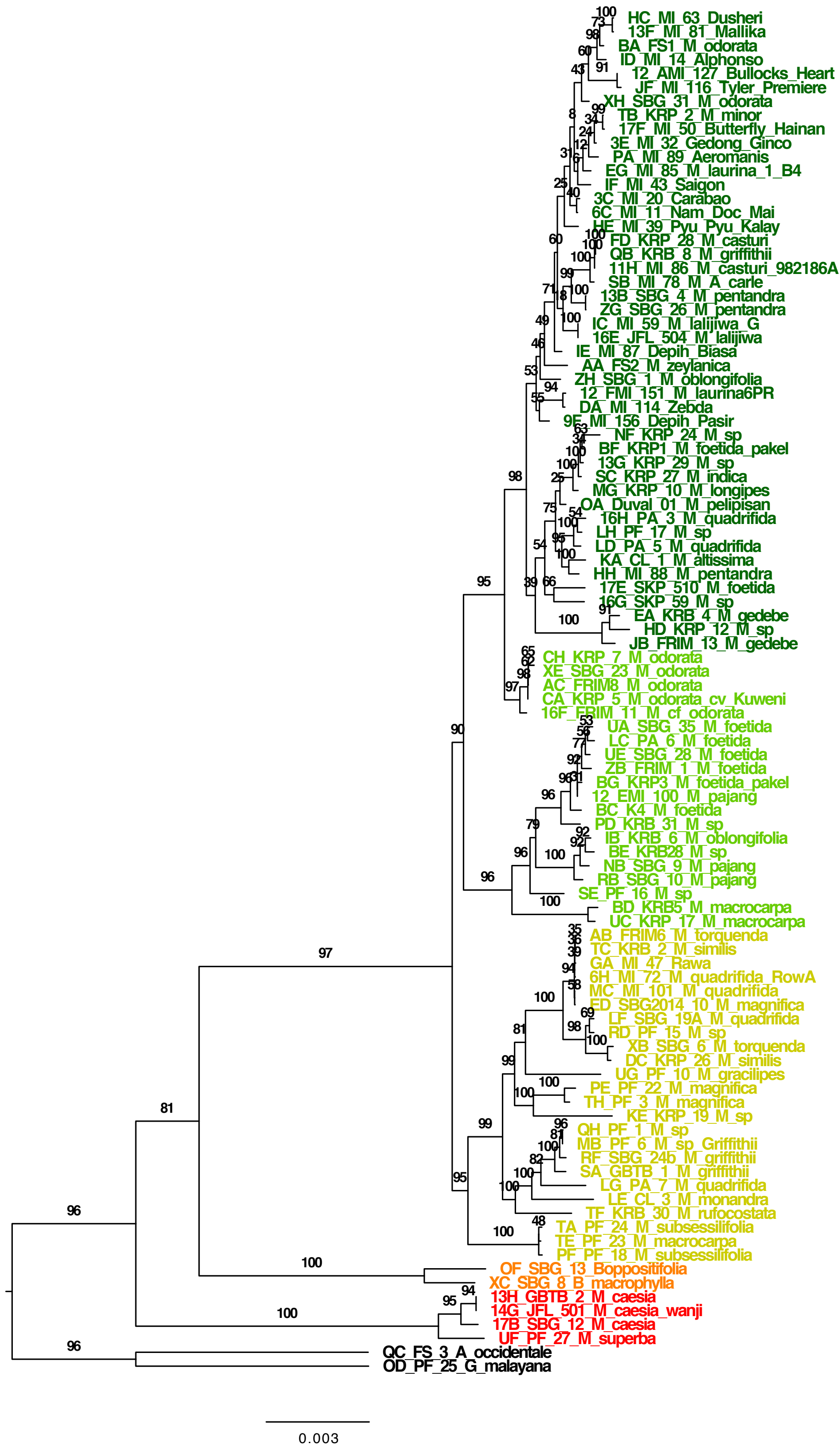
Appendix 4.4-12. ***phyloMED_c85_m90***



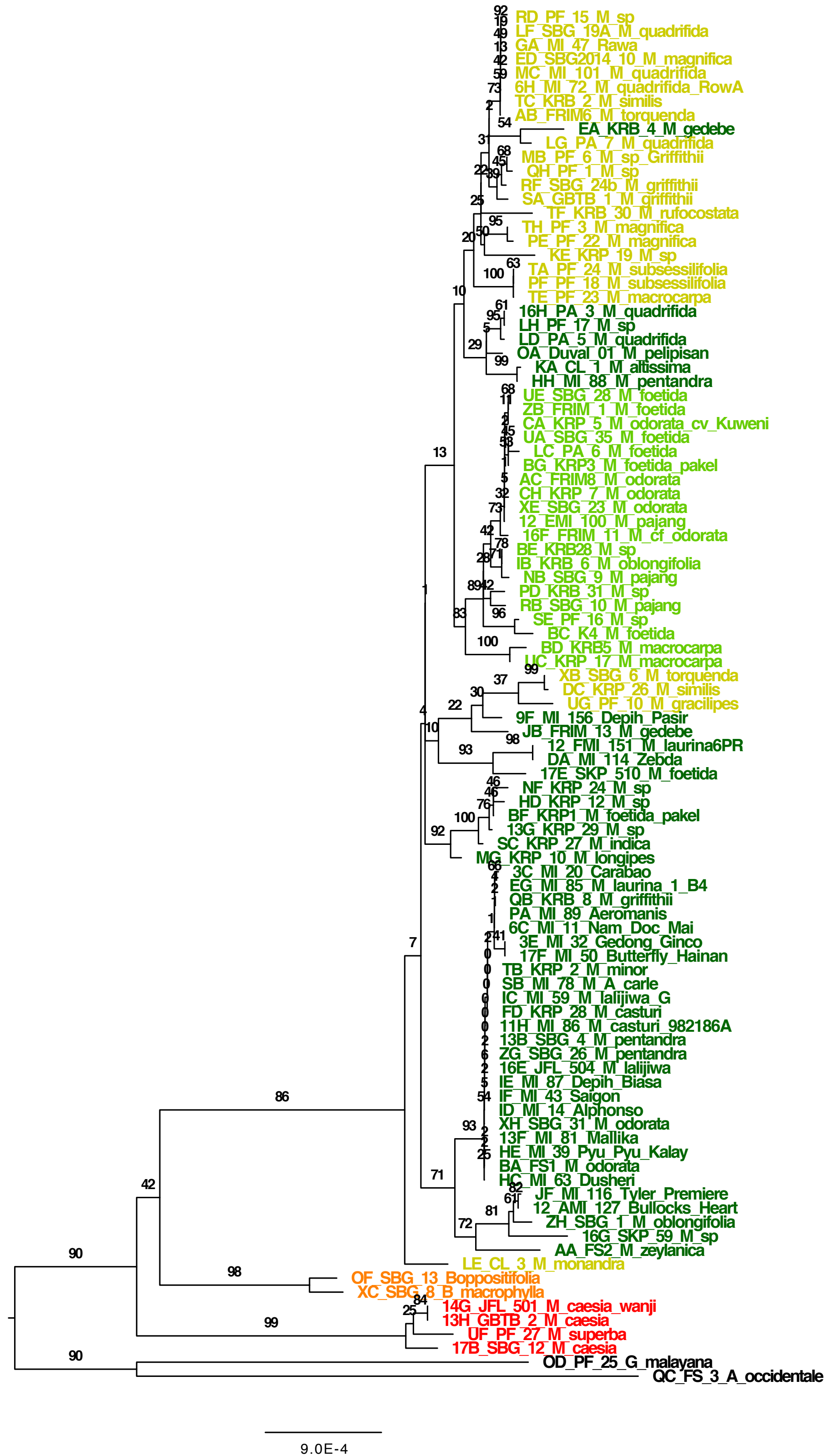
Appendix 4.4-13. ***phyloMED_c90_m20***



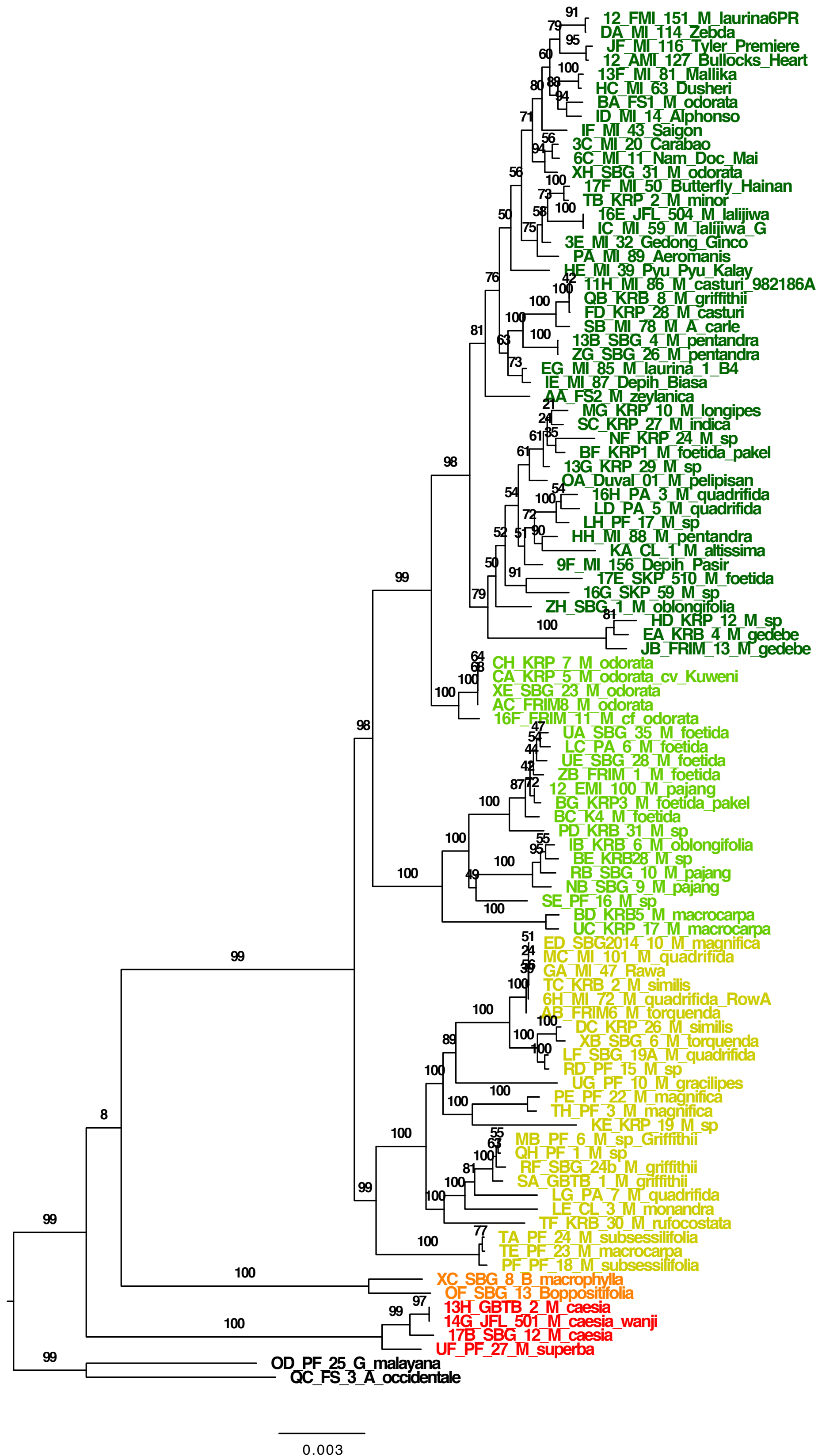
Appendix 4.4-14. *phyloMED_c90_m50*



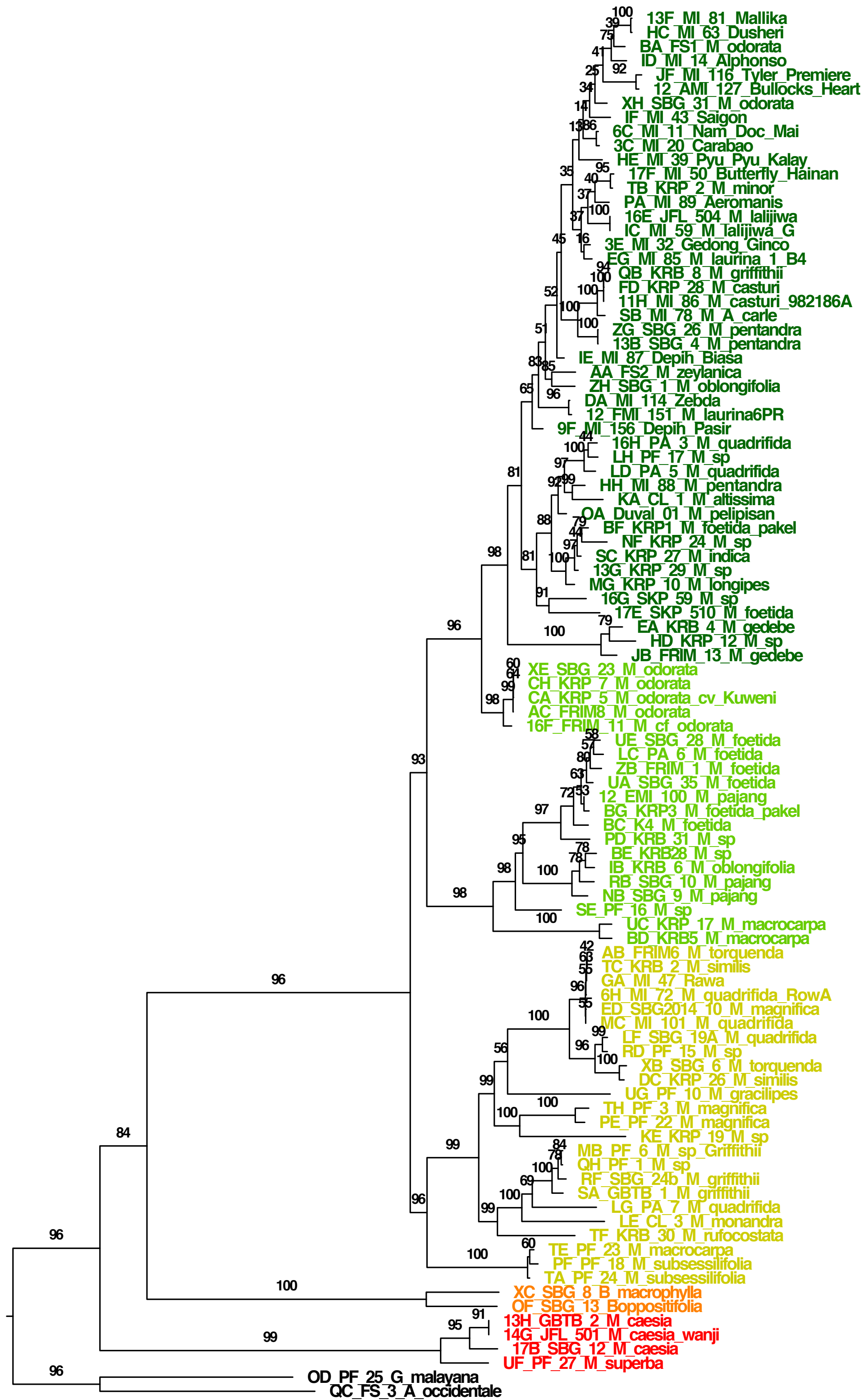
Appendix 4.4-15. ***phyloMED_c90_m90***



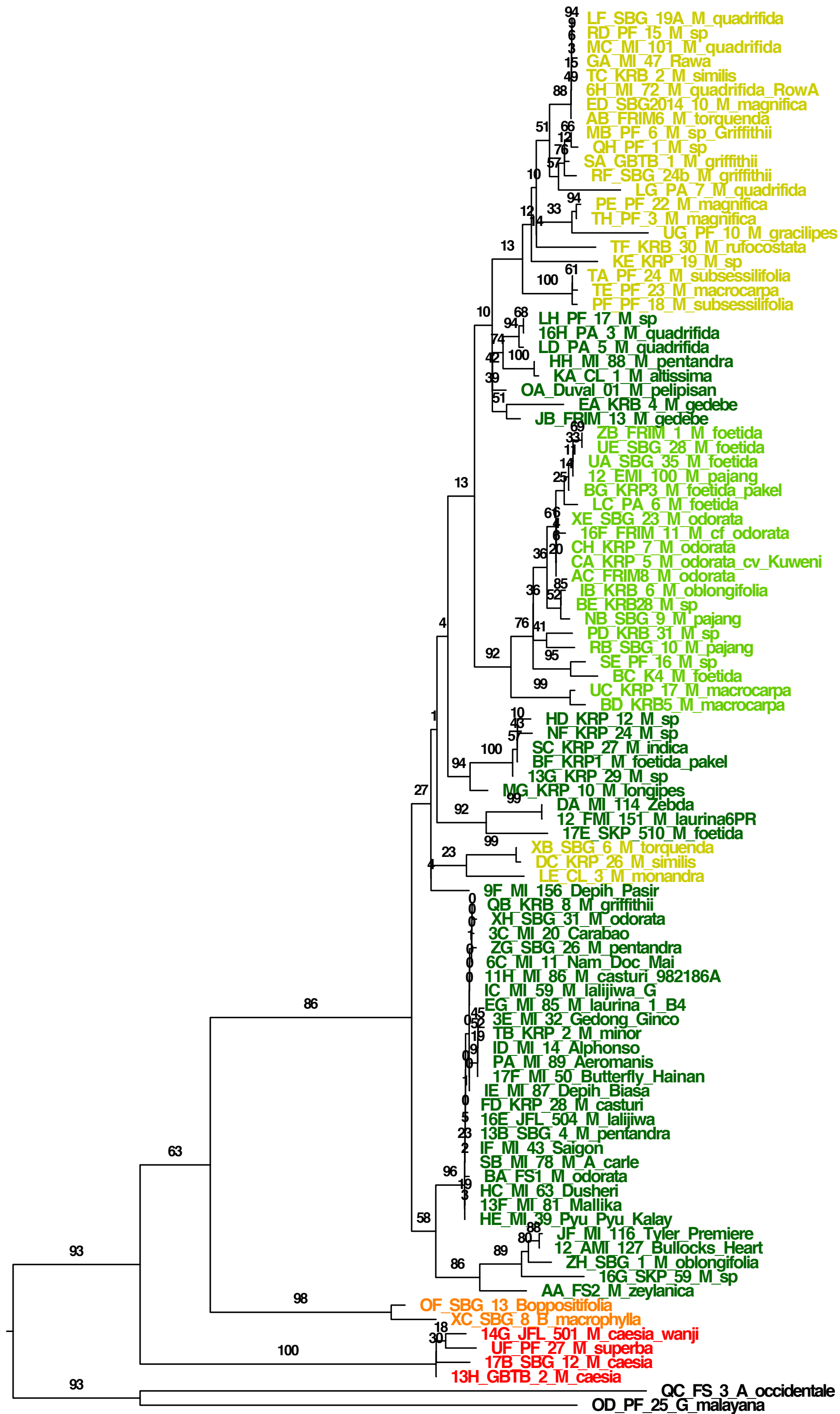
Appendix 4.4-16. *phyloMED_c95_m20*



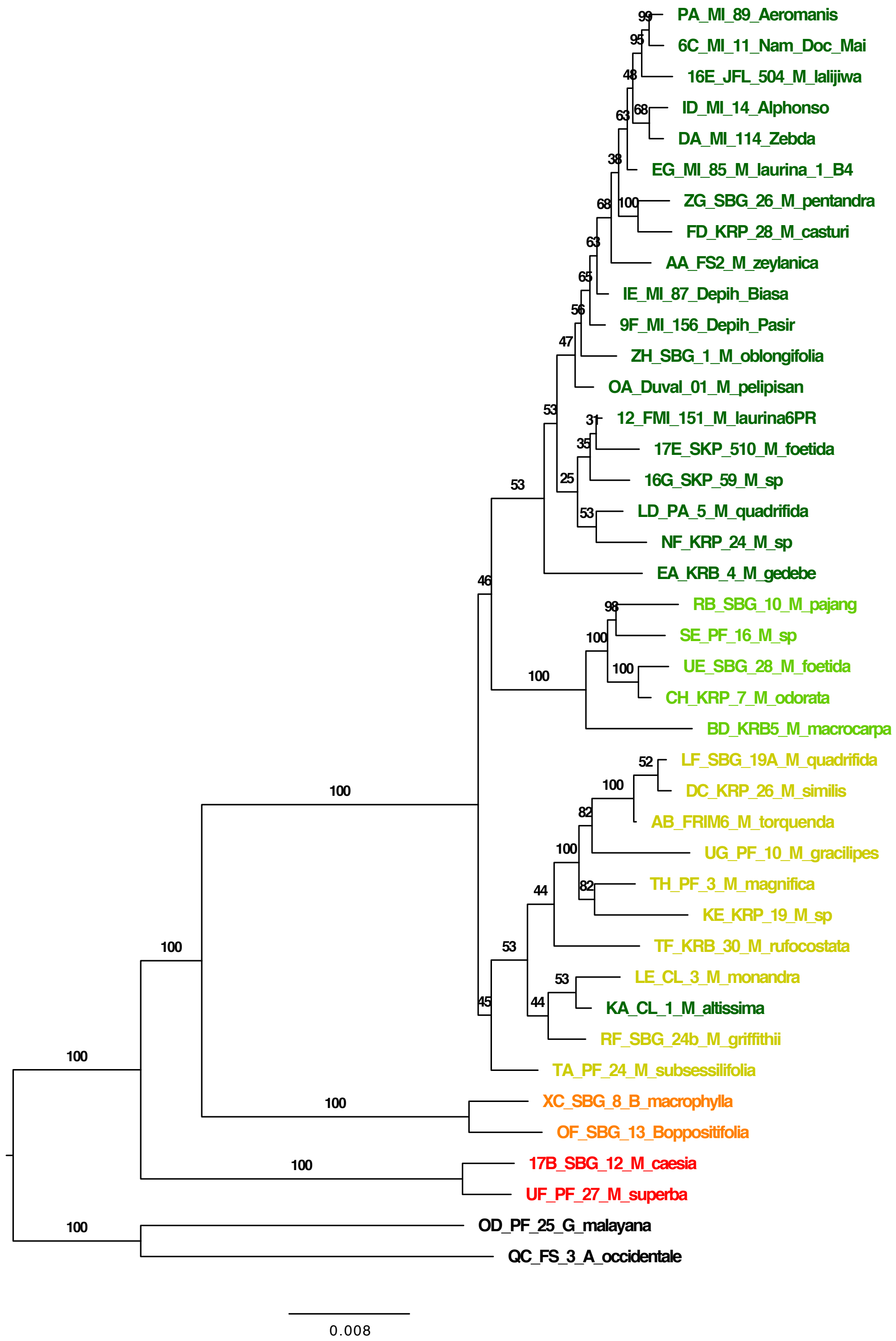
Appendix 4.4-17. ***phyloMED_c95_m50***



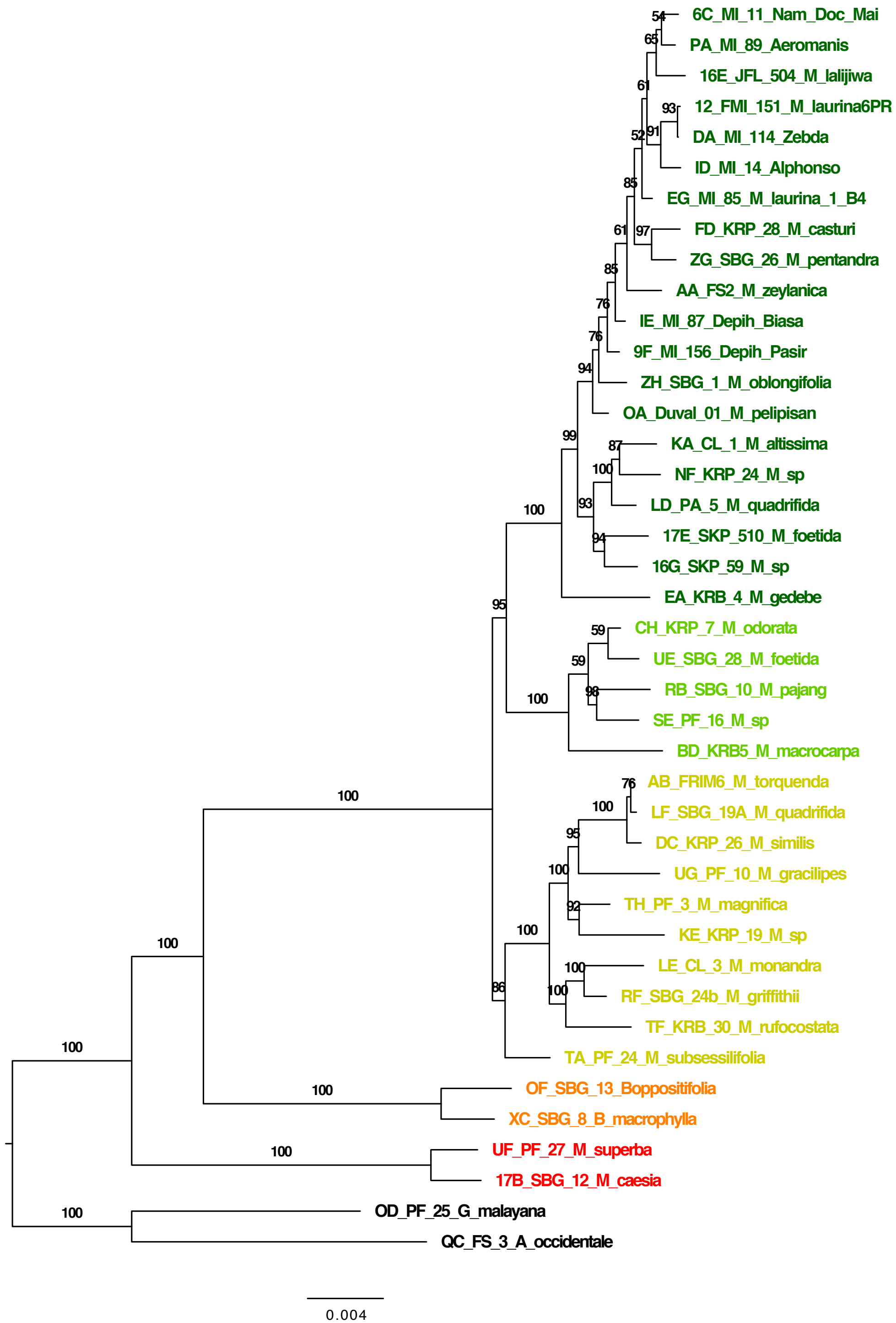
Appendix 4.4-18. ***phyloMED_c95_m90***



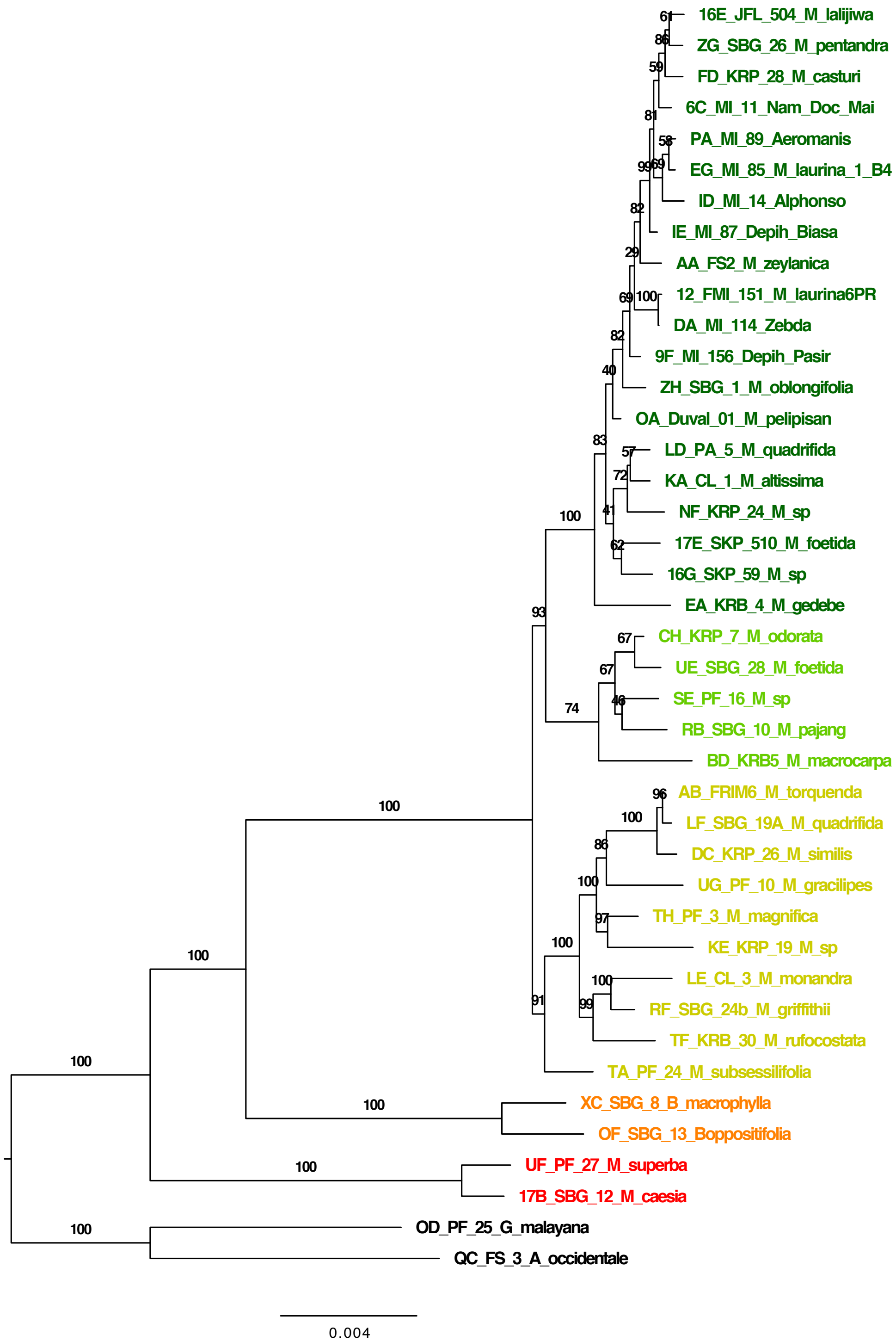
Appendix 4.4-19. ***phyloSM_c85_m4***



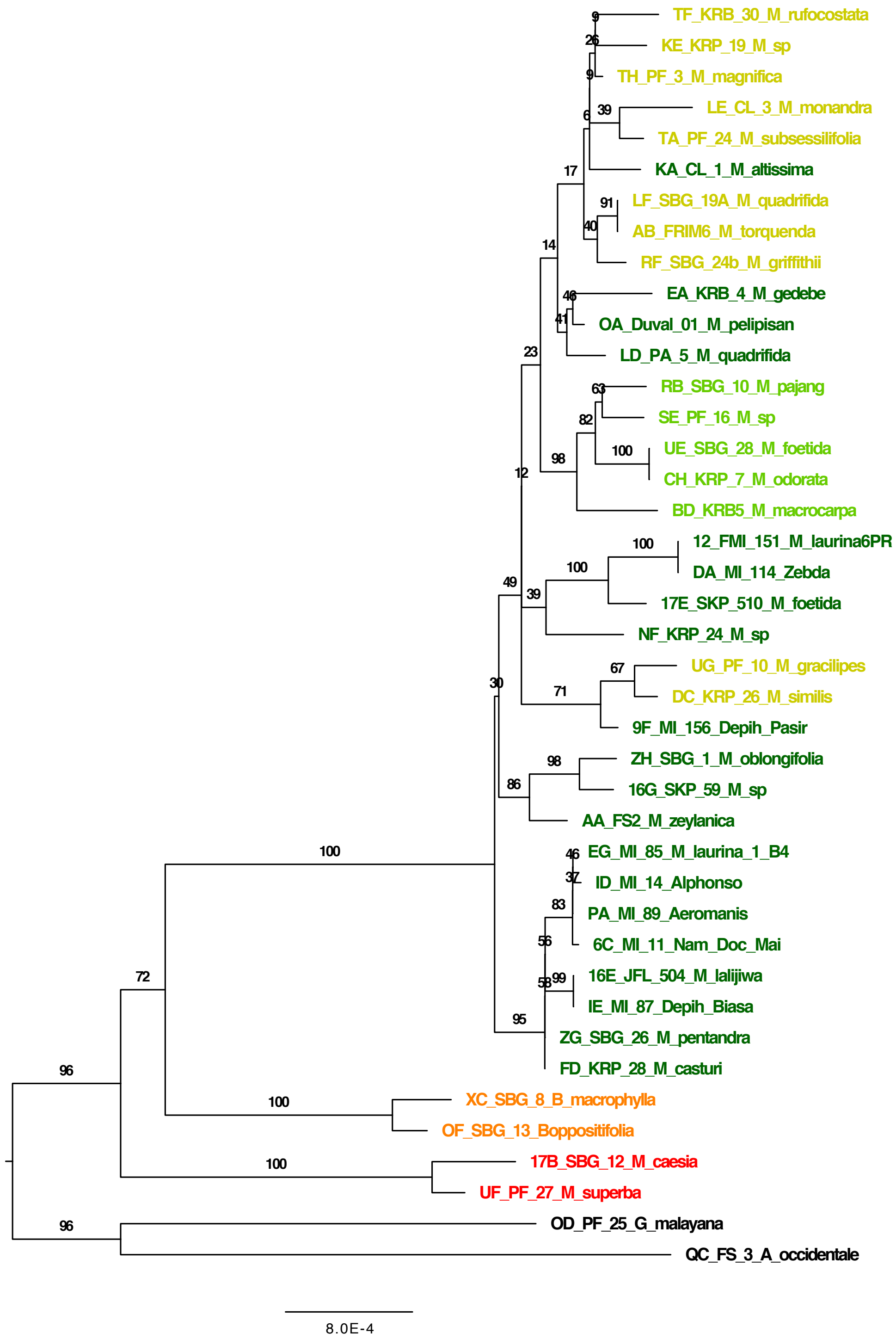
Appendix 4.4-20. ***phyloSM_c85_m20***



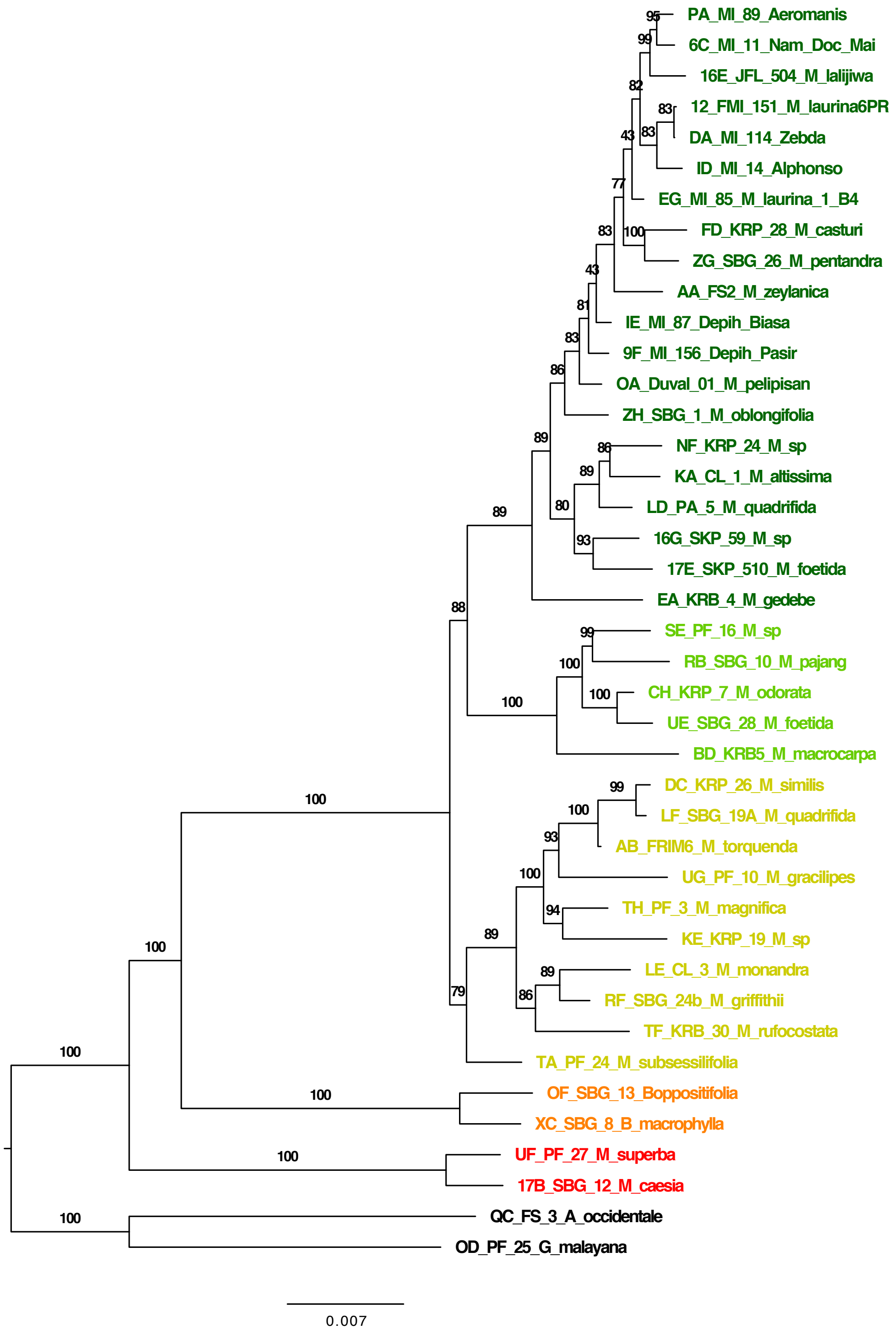
Appendix 4.4-21. ***phyloSM_c85_m50***



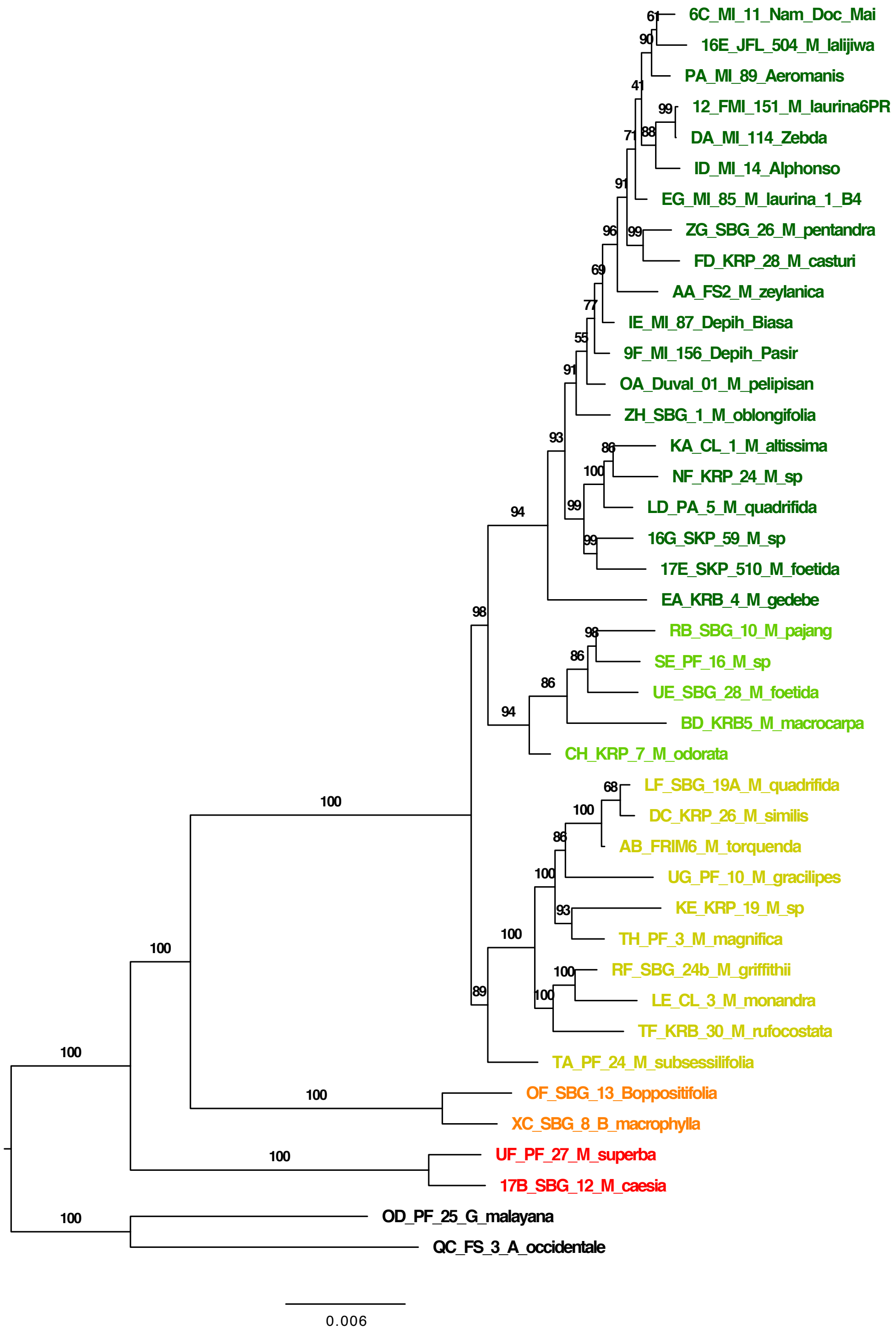
Appendix 4.4-22. ***phyloSM_c85_m90***



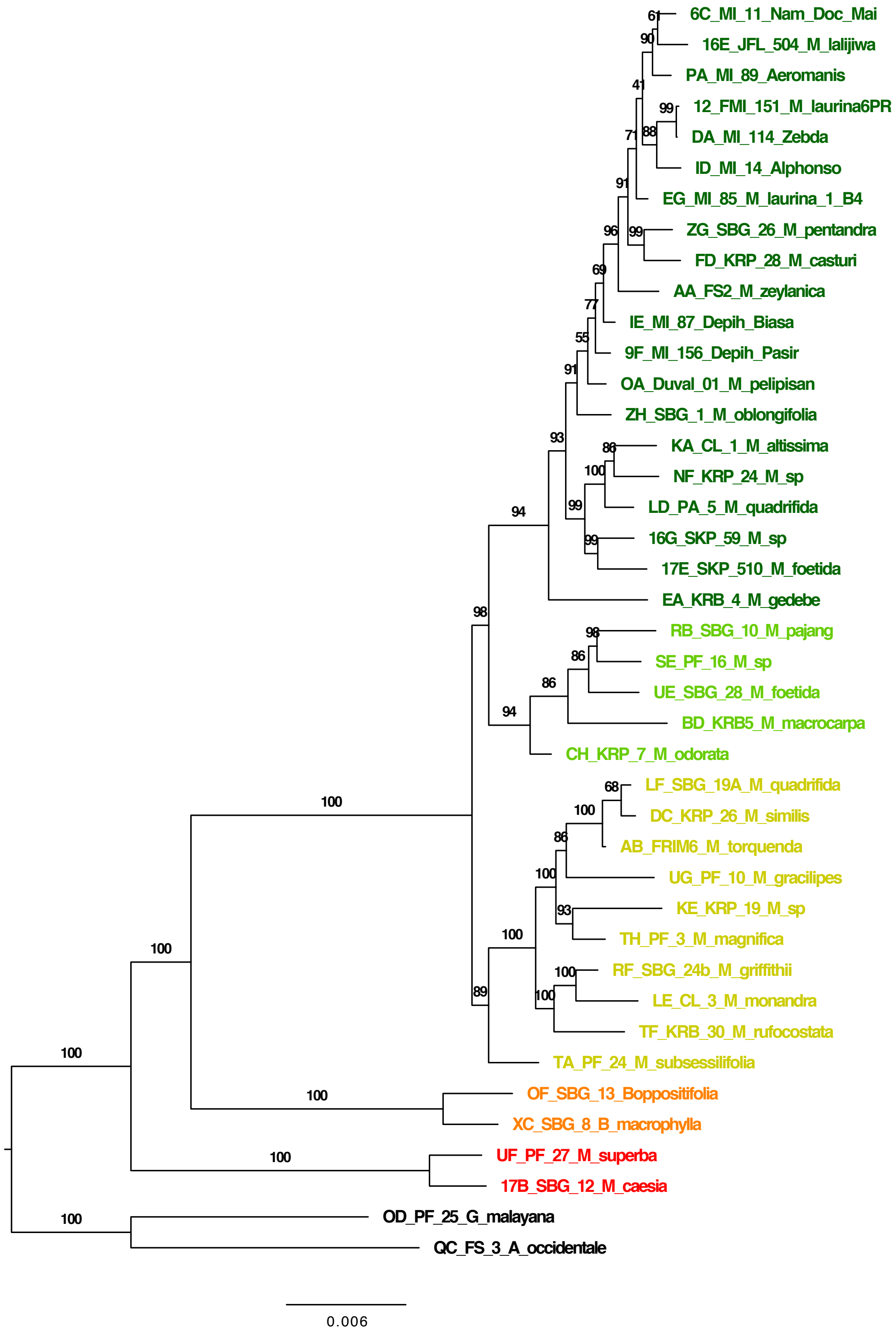
Appendix 4.4-23. ***phyloSM_c90_m4***



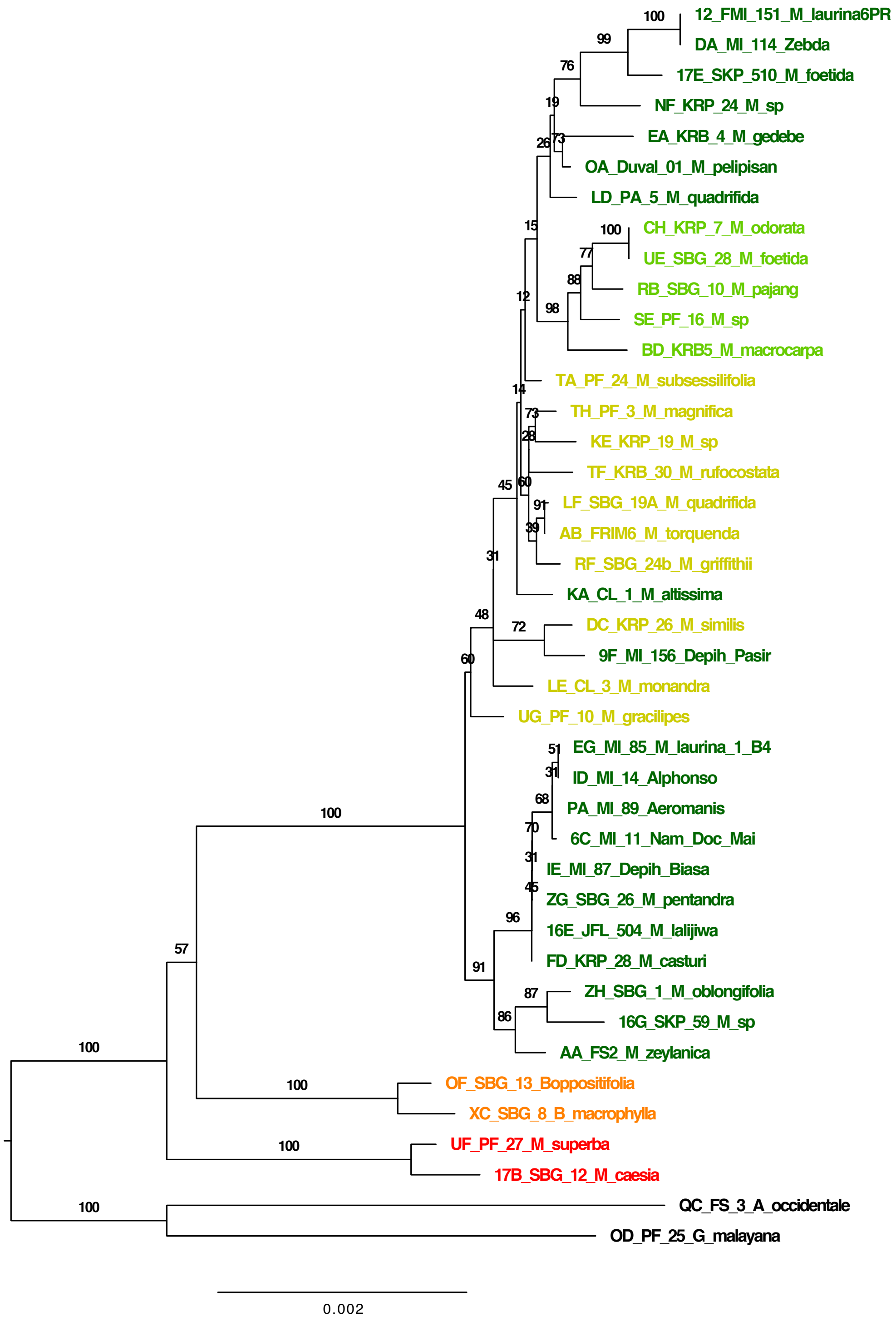
Appendix 4.4-24. ***phyloSM_c90_m20***



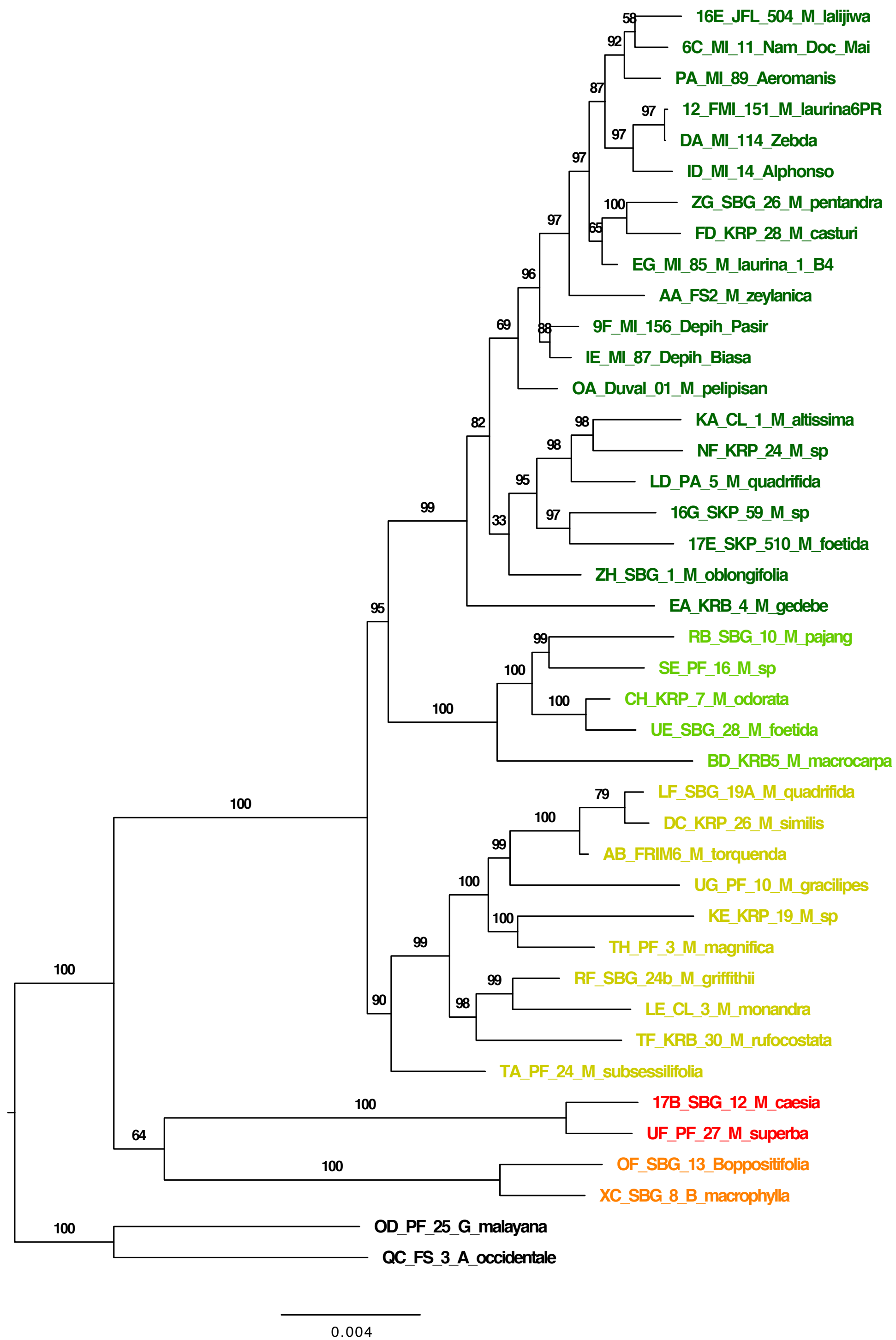
Appendix 4.4-25. ***phyloSM_c90_m50***



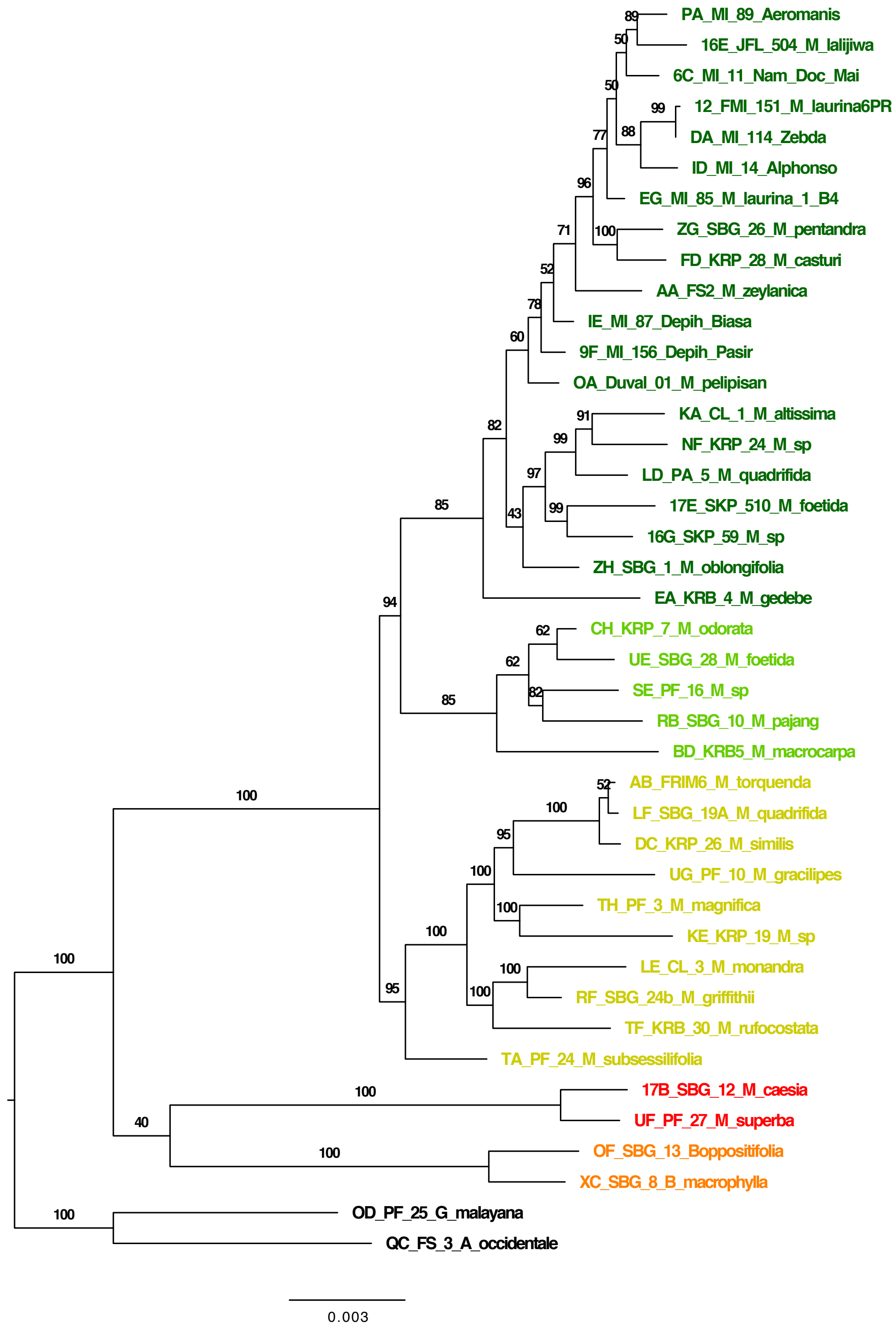
Appendix 4.4-26. *phyloSM_c90_m90*



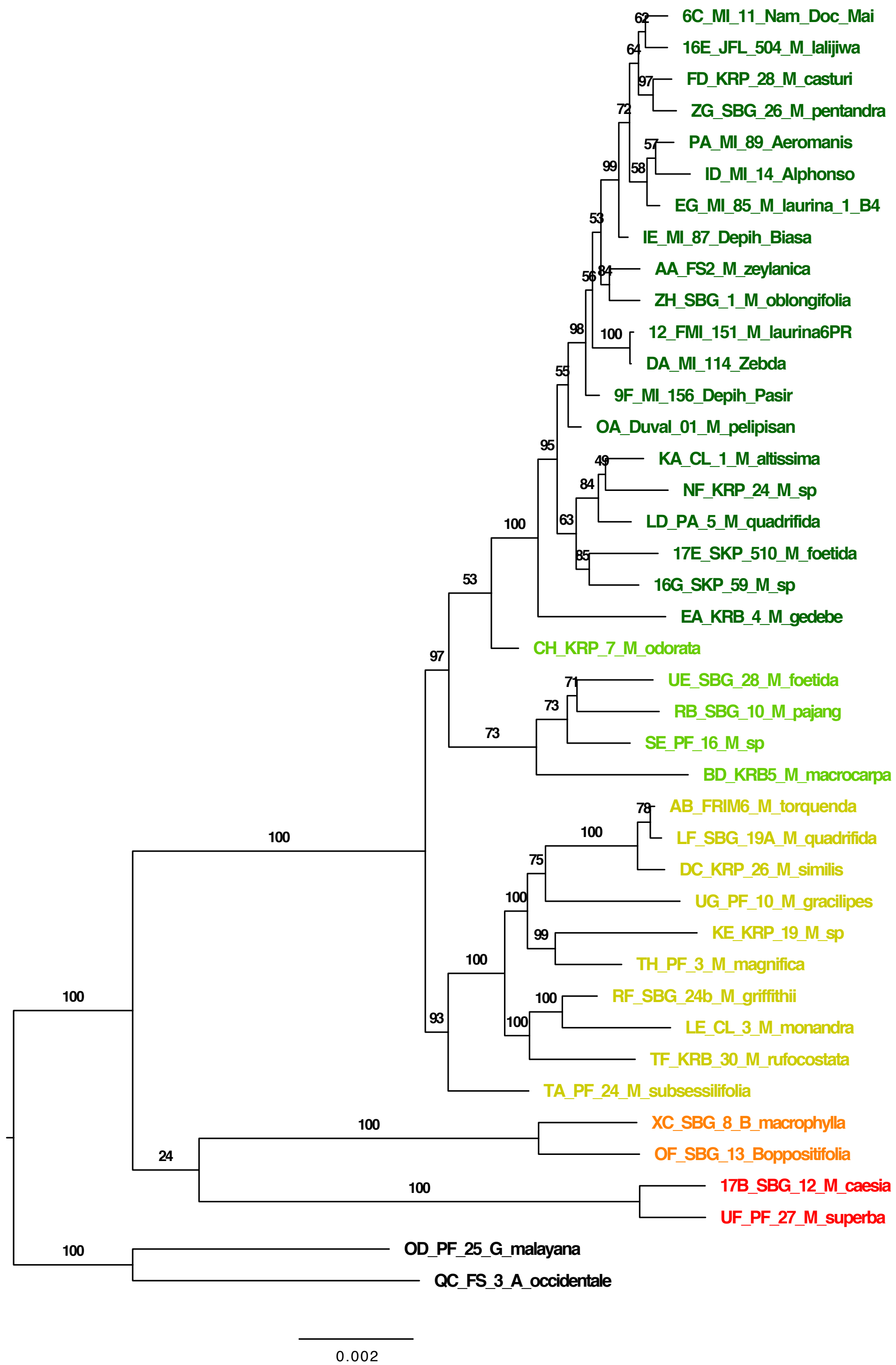
Appendix 4.4-27. ***phyloSM_c95_m4***



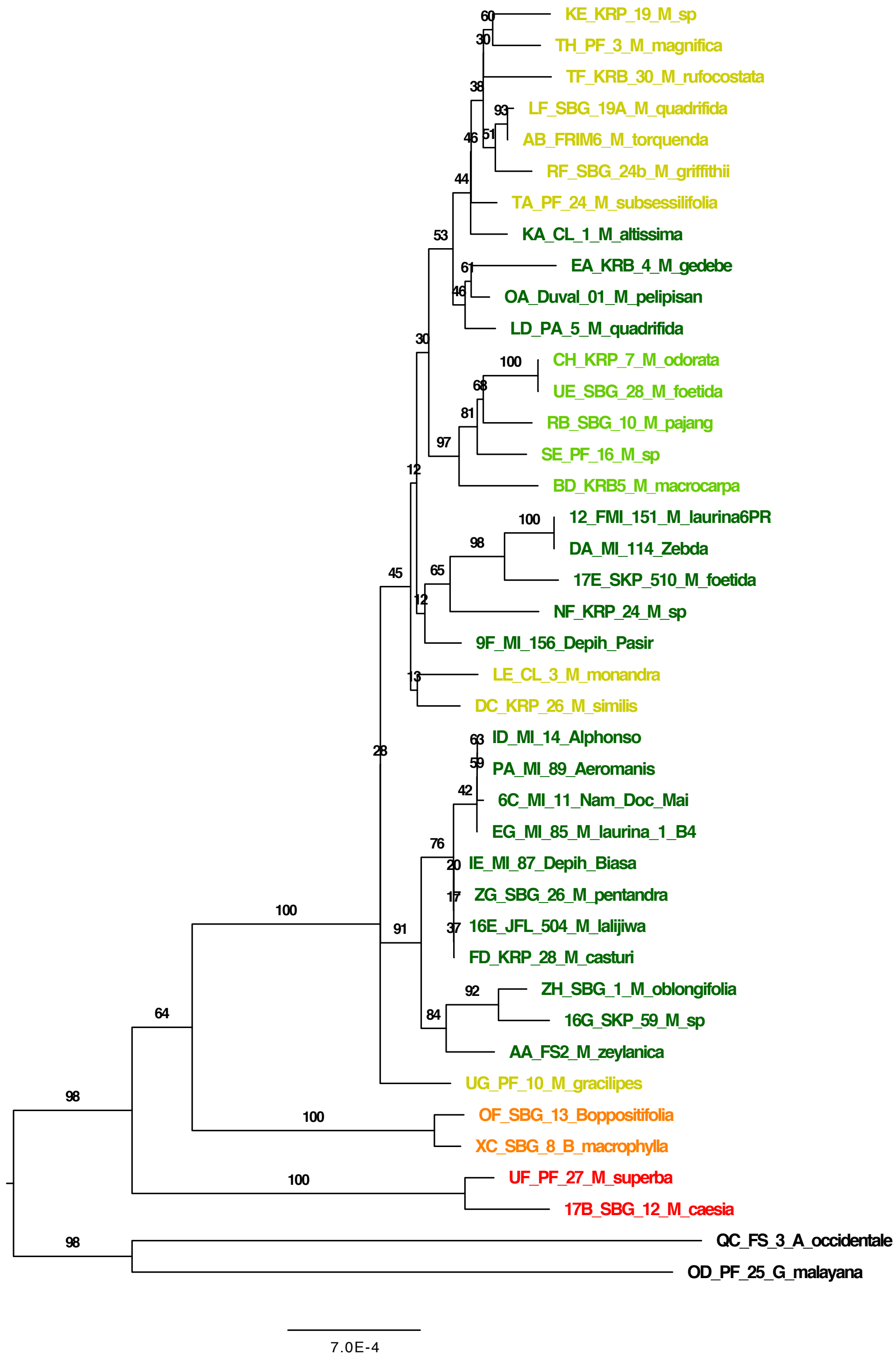
Appendix 4.4-28. ***phyloSM_c95_m20***



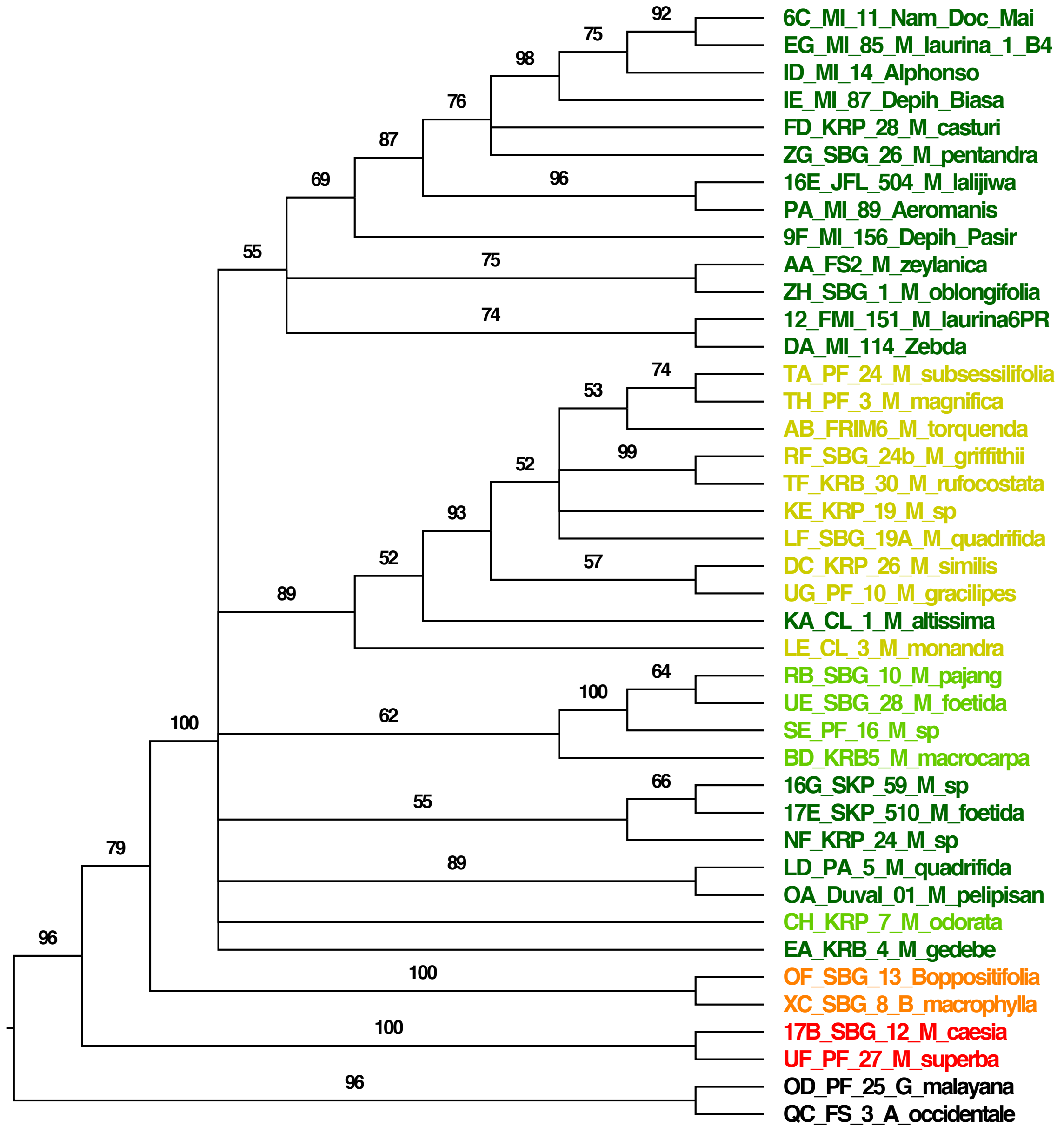
Appendix 4.4-29. ***phyloSM_c95_m50***



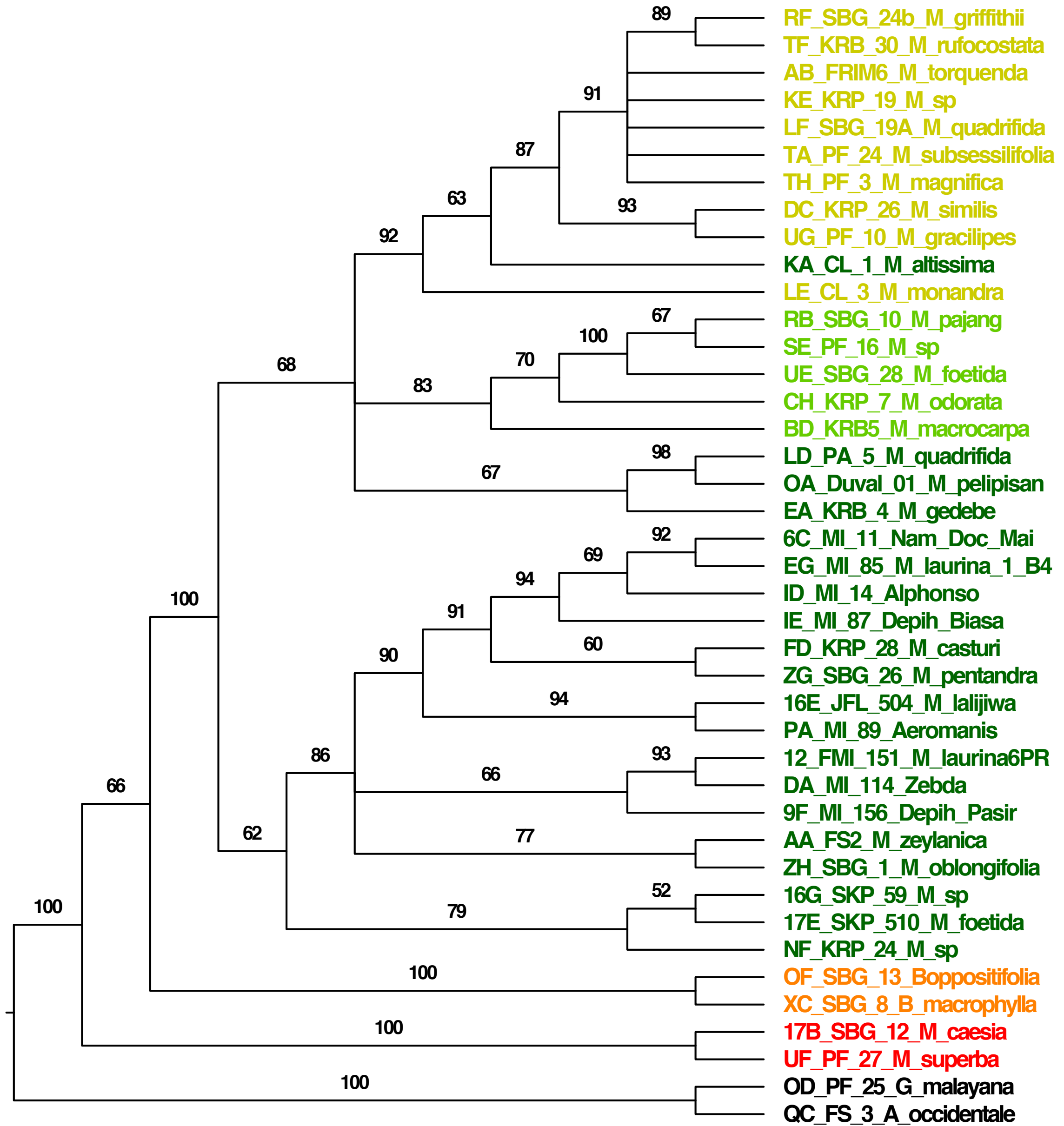
Appendix 4.4-30. ***phyloSM_c95_m90***



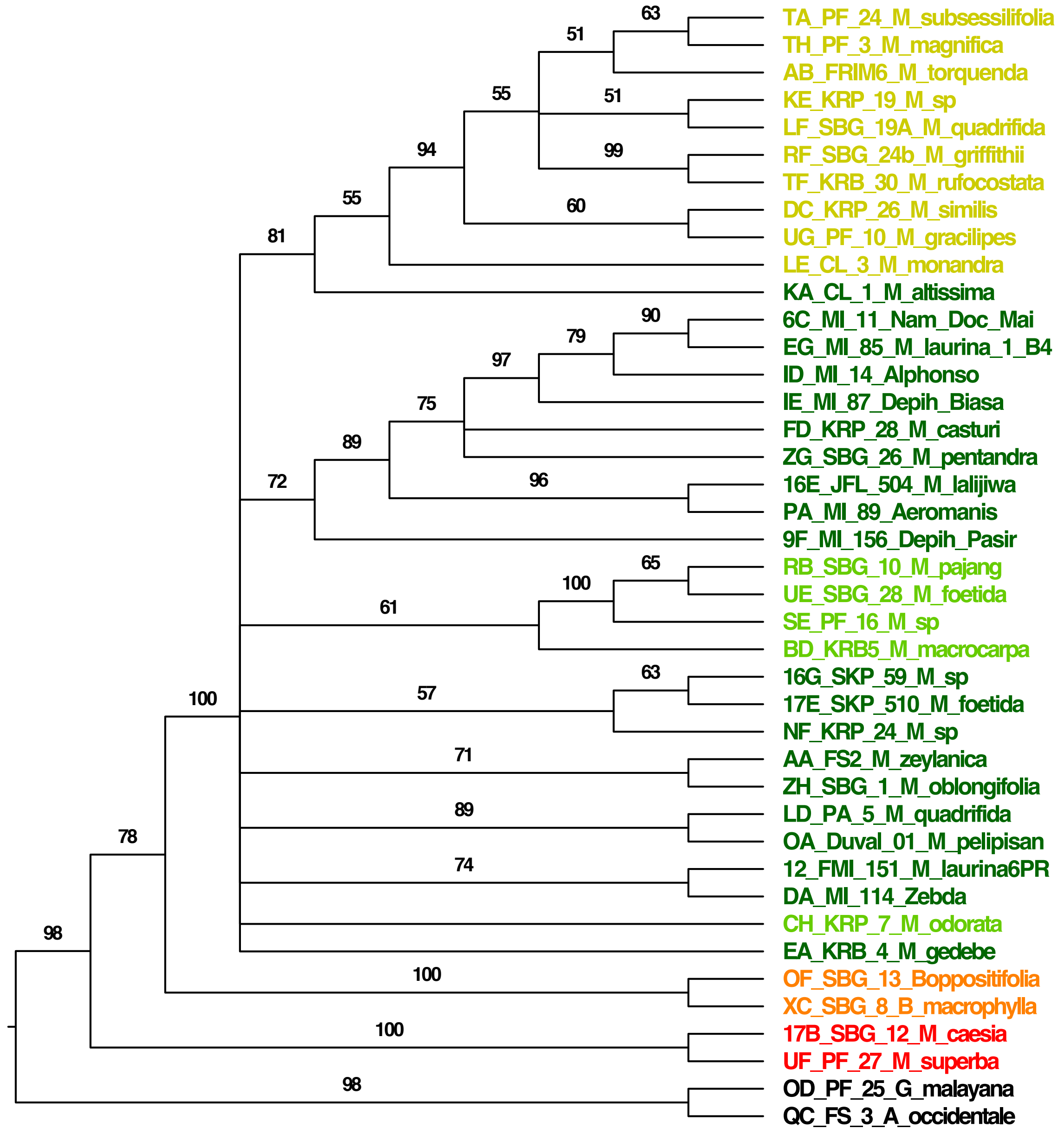
Appendix 4.5-01. ***phyloSM_c85_m4***



Appendix 4.5-02. ***phyloSM_c90_m4***



Appendix 4.5-03. ***phyloSM_c95_m4***



CHAPTER V

POPULATION GENOMIC ANALYSIS OF CULTIVATED MANGO (*MANGIFERA INDICA* L.) SUGGESTS A COMPLEX HISTORY OF DOMESTICATION

Abstract

Humans have domesticated diverse species from across the plant kingdom, yet much of our foundational knowledge of domestication has come from studies investigating relatively few of the most important annual food crops. In annual species, domestication typically involves a series of genetic bottlenecks resulting in low genetic diversity in crops, yet evidence indicates that some perennial species are more robust to the effects of domestication bottlenecks and maintain relatively high levels of diversity. Here, we examine the impacts of domestication on genetic diversity in a tropical perennial fruit species, mango (*Mangifera indica*). We generated genomic SNP data from 108 mango cultivars of known origin to test for a domestication-associated bottleneck and examine the geographic distribution of genetic diversity within cultivate *M. indica*. We find no significant loss of diversity associated with the mango's introduction into new regions of the world. However, our results show that mango cultivars from Southeast Asia contain unique genetic diversity not present in cultivars from other regions of the world, suggesting mango may have a more complex history of domestication than previously supposed. Our work has direct implications for mango breeding and genebank management, and also builds on recent efforts to understand how woody perennial crops respond to domestication.

Introduction

Over the past 12,000 years, humans have domesticated thousands of species from across the plant kingdom (Meyer *et al.* 2012; Meyer & Purugganan 2013; Gaut *et al.* 2015). The process of crop domestication is a special case of co-evolution that gradually increases plant-human interdependence and results in various levels of intensity of cultivation and breeding (Clement 1999; Zeder 2006; Pickersgill 2007). As such, the domestication process provides tractable systems in which to study convergent evolution, gene flow, adaptation, diversification, and genome evolution (e.g., Arnold 2004; Kovach *et al.* 2007; Purugganan & Fuller 2009; Meyer & Purugganan 2013; Olsen & Wendel 2013; The International Peach Genome Initiative 2013; Washburn *et al.* 2016). Understanding how these evolutionary forces impact crop genetic diversity and characterizing the standing genetic variation within cultivated germplasm is key to crop improvement efforts (e.g., Iqbal *et al.* 2001; Burke *et al.* 2002; Esquinas-Alcázar 2005; Doebley *et al.* 2006; Pickersgill 2007; Gross & Olsen 2010; Miller & Gross 2011; Kassa *et al.* 2012). However, our current understanding of plant domestication is founded on studies of highly domesticated annual staples like cereals and grain legumes (e.g., Singh *et al.* 1991; Wang *et al.* 1999; Matsuoka *et al.* 2002; Li *et al.* 2006; Londo *et al.* 2006; Huang *et al.* 2012; Hufford *et al.* 2013; Saintenac *et al.* 2013; von Wettberg *et al.* 2018) and, consequently, there remain many gaps in our understanding of the broader context of domestication – across a wide span of taxonomic and geographic diversity, among species that have undergone different degrees of domestication, and among species with different life history strategies (Miller & Gross 2011; Meyer *et al.* 2012).

One of the central dogmas of domestication is that crops undergo an often-severe decrease in genetic diversity in response to three key bottleneck (or founder) events (Ladizinsky 1985; Cooper *et al.* 2001; Doebley *et al.* 2006; van de Wouw *et al.* 2010; Miller & Gross 2011). During the initial stages of cultivation, as important traits are selected for or against, crops generally undergo a 'domestication bottleneck' (Cooper *et al.* 2001; van de Wouw *et al.* 2010). Compounding the primary loss of diversity, many crops experience a secondary 'dispersal bottleneck' when they are introduced into new regions (Cooper *et al.* 2001; van de Wouw *et al.* 2010). Soybean, for example, was subjected to an intense introduction bottleneck when it was introduced from Asia into North America (Hyten *et al.* 2006). The concept of a dispersal bottleneck is connected to Vavilov's premise of crop 'centers of origin', which posits that the geographical origin of a crop contains the greatest variation of morphological types (Vavilov 1987), thereby implying a loss of diversity as a crop is dispersed. Studies have also shown that as breeding and cultivation intensify, some crops suffer a tertiary 'improvement bottleneck' (Cooper *et al.* 2001; van de Wouw *et al.* 2010). The drastic reductions in diversity incurred during these three bottleneck events can negatively impact a crop's ability to adapt to novel environments, pests and diseases (e.g., Abbo *et al.* 2003; Esquinas-Alcázar 2005). However, the relative impacts of each bottleneck vary both within and among crops depending in large part on the biology of the species itself.

Perennial crop species have recently received increased attention highlighting their relatively different trajectories under domestication compared to annuals (Miller & Gross 2011; Gaut *et al.* 2015). In general, woody perennials tend to retain greater levels of genetic diversity under cultivation than do annuals (Miller & Gross 2011). For

example, a recent genome-wide analysis of peach *(Prunus dulcis)* and its close relative almond *(P. persica)* showed no evidence of genetic bottlenecks associated with domestication in either species (Velasco *et al.* 2016), and similar results have been found for grape *(Vitis vinifera,* Myles *et al.* 2011) and apple *(Malus* x*domestica,* Gross *et al.* 2014). The relatively weak bottleneck observed in many perennial species is largely a result of characteristics common to the perennial life history: a long generation time and the predominance of self-incompatibility (Miller & Gross 2011). The former means that perennial crops have experienced fewer generations of selection under domestication than their annual counterparts (Pickersgill 2007), while the latter explains how perennials maintain high levels of heterozygosity despite the fact that their per-unit-of-time mutation rates are far slower than in annual species (Savolainen & Pyhäjärvi 2007). In addition, clonal propagation techniques common in woody perennial cultivation allow any individual – including F1 hybrids, triploids, and sterile or seedless parthenocarpic individuals – to be preserved for posterity, effectively halting the domestication process in that clone and potentially limiting the loss of genetic diversity in perennial species (Zohary, 2004; Miller & Gross 2011). However, not all perennial crops retain high levels of diversity: the tropical species coffee *(Coffea arabica)* and cacao *(Theobroma cacao)* have both experienced significant losses of diversity during domestication (Anthony *et al.* 2002; Aerts *et al.* 2013; Yang *et al.* 2013).

The King of Fruits

The mango, *Mangifera indica* L. (Anacardiaceae) is a perennial fruit tree that has been cultivated on the Indian subcontinent for an estimated 4,000 years, where it is called

'The King of Fruits' (Mukherjee 1949). This timeline places the domestication of mango contemporaneously with that of citron *(Citrus medica)*, walnut *(Juglans regia)*, peach *(Prunus persica)*, sweet orange *(Citrus* x*sinensis)*, lychee *(Litchi chinensis)*, lemon *(Citrus limon)*, and jujube *(Ziziphus jujuba)*, and prior to that of the other domesticated species in the poison ivy family: pistachio *(Pistacia vera)*, cashew *(Anacardium occidentale)*, Peruvian peppertree *(Schinus molle)*, and jocote *(Spondias purpurea)* (Meyer *et al.* 2012). Unpruned, mango trees can reach over 30 meters in height and live for more than a century, producing tons of fruit throughout their lifetime.

On the basis of historical documents and artifacts, *M. indica* is thought to have been cultivated on the Indian subcontinent for thousands of years before it was introduced elsewhere (Mukherjee 1949; Fig. 5.1), and it is presumed that India represents the center of origin of mango (Vavilov 1987; Kostermans & Bompard 1993). Buddhist monks were likely the first to introduce mango outside its original range of cultivation during their trips to Southeast Asia in the 4th and 5th centuries (Mukherjee 1949). The mango began its westward journey much later, when Persian traders brought the tree to East Africa in the 9th or 10th centuries (Mukherjee 1949). In the 16th century, as global botanical trade continued to grow, the Portuguese likely reintroduced the mango into East Africa from their territory in Goa (Mukherjee 1949). The Portuguese would continue to facilitate the mango's range expansion, transporting it to West Africa, and then to Brazil sometime around 1700 (Popenoe 1920; Mukherjee 1949). From there, the mango spread throughout the Caribbean, reaching Barbados in 1742 and Jamaica by 1782 (Popenoe 1920; Mukherjee 1949). As a Spanish colony, Mexico had a unique history of introductions, with mangoes arriving from the Caribbean as well as directly from the

Philippines, which was also under Spanish rule at the time (Popenoe 1920; Mukherjee 1949). It was not until 1833 that the first mango reached the shores of Florida (Popenoe 1920). In the 1900s, mango became the subject of intensive breeding programs in South Florida, which produced many of today's most important commercial cultivars including 'Tommy Atkins', 'Haden', 'Keitt', and 'Kent' (Knight *et al.* 2009). For this reason, South Florida has been termed a secondary center of domestication for mango (Knight & Schnell 1994). Today, the mango is one of the world's most important fruits and is grown in tropical and subtropical climates across the world (FAO 2003, FAOSTAT 2013).

Mysteries and Molecules

Despite its importance as a global food crop and its immense cultural importance in many regions of the world, the mango's origin is still somewhat mysterious. Current ranges of wild *M. indica* (cultivated and wild mango are considered to be the same species) are not well characterized, and the IUCN's red list currently categorizes the species as 'data deficient' (IUCN 2012). Most authors agree that wild *M. indica* originated in the region of Indo-Burma, primarily in Northeastern India, Bangladesh, Bhutan, and Nepal, perhaps extending south into Myanmar and northern Thailand, and presuppose a single domestication event for cultivated *M. indica* (DeCandolle 1884; Mukherjee 1972; Vavilov 1987; Mukherjee & Litz 2009; Singh *et al.* 2016). However, some other authors contend that morphological differentiation within cultivated *M. indica* signifies a more complex history of domestication (see discussion; Bompard 2009; Iyer & Schnell 2009).

Phylogeographic studies help us elucidate the origins of crops and the reveal the impacts of domestication on these species (e.g., Olsen & Schaal 1999; Salamini *et al.*

2002; Londo *et al.* 2006; Gunn *et al.* 2011; Kassa *et al.* 2012; Loor Solorzano *et al.* 2012). While a lack of accessible wild *M. indica* populations precludes investigations of a primary bottleneck associated with the initial domestication of mango, the recent and well-documented history of mango's human-mediated migration into new regions of the world provides an opportunity to determine whether the species experienced a genetic bottleneck during successive founder events. Although many previous studies have provided insight into the diversity and genetic structure of mango cultivars within specific regions, including Kenya (Sennhenn *et al.* 2013), Myanmar (Hirano *et al.* 2012), China (Luo *et al.* 2011), Colombia (Diaz-Matallana *et al.* 2009), Brazil (Dos Santos Ribeiro *et al.* 2012), Iran (Shamili *et al.* 2012), and especially India (Ravishankar *et al.* 2000; Kumar *et al.* 2001; Karihaloo *et al.* 2003; Damodaran et al. 2012; Vasugi *et al.* 2012; Surapaneni *et al.* 2013; Ravishankar *et al.* 2015), only a handful have examined mango cultivars originating across a broad geographic range. Works by Schnell *et al.* (2006) and Dillon *et al.* (2013), both of which used microsatellite markers, found Southeast Asian mango cultivars to be the most differentiated from other populations, while Sherman *et al.* (2015) found population structure between Asian and Western mango cultivars. Here, we use SNP markers from double digest restriction site associated DNA sequencing (ddRADseq, Peterson *et al.* 2012) to explore geographic patterns of diversity in mango cultivars within genebank collections that originated from different geographic regions. As a reduced representation genomic technique, RADseq identifies SNPs from across the genome (Miller *et al.* 2007, Baird *et al.* 2008), and has proven to be a useful tool for investigating population structure and phylogeography in non-model organisms, including crop species (e.g., Xu *et al.* 2014; Atchison *et al.* 2016; Pan *et al.*

2016; Singh *et al.* 2016; Gao *et al.* 2017; Stetter *et al.* 2017). We aim to **a)** determine the geographic distribution of genetic diversity in mango, **b)** test whether India represents a 'center of diversity' for mango, **c)** quantify the genetic bottleneck mango underwent during its migration to Africa and the Americas, and **d)** provide insight into the origin of cultivated mango. Our work has a threefold impact, informing the management practices for germplasm resources, providing a better understanding of the history and genetic impacts of domestication in mango, and adding to the growing body of literature that seeks to understand how perennial crops evolve under domestication.

Methods

Sampling

To explore the geographic distribution of genetic diversity in mango, we selected 113 cultivars from mango genebanks in South Florida (Fairchild Tropical Botanic Garden [FTBG], U.S. Department of Agriculture's Subtropical Horticulture Research Station [USDA]) that originated in eight different geographic regions: India, Indochina (Myanmar, Thailand, Cambodia, Laos, Vietnam), Malesia (Malaysia, Indonesia, the Philippines), Africa (limited germplasm required pooling of all African samples), South America, Mexico, the Caribbean (Cuba, Jamaica, Haiti, the Dominican Republic), and Florida (Appendix 5.1). We attempted to sample the most diverse and characteristic mangoes from each region, emphasizing historical cultivars whenever possible. Additionally, we collected leaves of 54 samples of unidentified cultivars of *M. indica* and closely related *Mangifera* species from FTBG, Miami-Dade Fruit and Spice Park (MDFS), Singapore Botanic Garden (SBG), Gardens by the Bay (Singapore), Purwodadi

Botanic Garden (KRP, East Java, Indonesia), Bogor Botanic Garden (KRB, West Java, Indonesia), the Forestry Research Institute of Malaysia (FRIM, Kepong Malaysia), Pasoh Forest Arboretum (PA, Simpang Pertang, Malaysia), and Pasoh Forest Reserve (PF, Simpang Pertang, Malaysia) (Table 4.1). Leaf samples were stored at -80ºC or dried in silica and stored at 4ºC. DNA was extracted from each sample using the DNEasy plant mini kit (Qiagen) or a modified CTAB protocol (Doyle & Doyle 1990).

RADseq library preparation and locus assembly

Three ddRADseq libraries were prepared following the protocol of Peterson *et al.* (2012). The 167 samples for this study were combined with 113 additional samples that were analyzed elsewhere (see Chapters III and V). High molecular weight DNA (300-1000 ng) was digested with NlaIII and MluCI (New England Biolabs). Custom-designed oligonucleotides containing unique barcode sequences (Chapter IV, Appendix 4.2) were ligated onto each individual prior to pooling eight samples into 12 separate sublibraries per lane (36 sublibraries across three lanes total). Pippin Prep (Sage Science) was used to size select 350 bp inserts (tight size selection, 425 bp, external marker). Short-cycle PCR was performed in sextuplicate to amplify and add a unique index to sublibraries (Chapter IV, Appendix 4.3), which were then quality-checked on an Agilent Bioanalyzer DNA High Sensitivity Chip. For libraries where overamplification was observed, non-target DNA was removed by size-selection on Pippin Prep, with a subsequent Bioanalyzer quality-check. Each of the three libraries was sequenced at The University of Southern California's Genome and Cytometry Core in a rapid run of Illumina HiSeq 2500 as a single lane of 150 bp single end reads.

The program FASTQC v.0.11.4 (Andrews 2010) was used to check overall quality of raw fastq files for each sublibrary. After demultiplexing reads within each sublibrary based on the individual barcode (see below), nine individuals from this study were excluded based on very low sequencing coverage, bringing the total number of individuals analyzed to 158 (dataset *FULL*), with 108 samples from mango cultivars (dataset *CULT*) and 50 samples from closely related species or unidentified accessions (Table 5.1).

Raw reads were processed using the ipyRAD bioinformatic pipeline (Eaton 2014) on Florida International University's high performance computing cluster (FIU HPCC) using default parameters except for: *maxdepth* = 1000, *max_barcodes_mismatch* = 1, *filter_adapters* = 2, *min_samples_locus* = 4, and *clust_threshold* = 0.95 using de novo clustering. The clustering threshold was set to 0.95 to account for previous reports of high heterozygosity within mango (Sherman *et al.* 2015; Singh 2016; Kuhn *et al.* 2017) and because we included closely related *Mangifera* species in some datasets. For population genetic analysis, ipyRAD was used to produce a file containing a single randomly selected SNP from each locus for downstream analyses. We performed filtering (ipyRAD step 7) independently for the complete dataset (*FULL*) and the subset of 108 mango cultivars (*CULT*). We produced an additional output file in ipyRAD of the full dataset for UPGMA clustering that consisted of all SNPs with the parameter *min_samples_locus* = 33 (*FULL_33*, Table 5.1). For the *FULL* and *CULT* datasets, we used a custom python script to remove loci that contained more than 10% missing data and to ensure that no individuals had >50% missing data across all loci.

Population structure and admixture

A dendrogram of uncorrected distances using unweighted pair-group method with arithmetic mean (UPGMA) was built for dataset *FULL_33* in PAUP v. 4.0a build 153 (Swofford 2002). The dendrogram was visualized in FigTree v. 1.4.3 (Rambaut 2006) and rooted with the species *M. gedebe*, which has been previously shown to be sister to all other species within the clade containing *M. indica* (see Chapter III). Additionally, principal component analysis (PCA) was used to visualize population structure within mango cultivars (*CULT*) in the R package adegenet (Jombart 2008; Jombart & Ahmed 2011), with two obvious outliers (African cultivars 'Zebda' and 'Tyler Premiere') removed from the plot.

To detect population structure and admixture within mango cultivars (*CULT*) and the full dataset (*FULL*), *K*-means clustering was conducted in the Bayesian software STRUCTURE v. 2.3.4 (Pritchard *et al.* 2000; Falush *et al.* 2003; Hubisz *et al.* 2009). For each dataset, lambda was estimated by averaging the mean value of lambda with $K = 1$ across 10 independent runs of 100,000 iterations with a 10,000 step burn-in period. Using the estimated value of lambda for each dataset, 10 runs of 100,000 iterations followed by a 10,000 step burn in were performed for $K = 1$ to 10. The optimal value of *K* was determined using StructureHarvester v. 0.6.94 (Earl & vonHoldt 2012) according to the ΔK method of Evanno *et al.* (2005). Results were summarized with CLUMPP v. 1.1.2 (Jakobsson & Rosenberg 2007) using the greedy option (M=2) for K=1-7 or the LargeKGreedy option (M=3) for K=8-10, with G' similarity and 1,000 random permutations. The results were visualized using DISTRUCT v. 1.1 (Rosenberg 2004), and individuals were labeled by subpopulation and region.

Population differentiation

The combined evidence from UPGMA, PCA, and STRUCTURE analyses indicated that two individuals labeled as African mango cultivars in the FTBG collection were probably misidentified and in fact belong to a closely related *Mangifera* species. Therefore, these cultivars ('Zebda' and 'Tyler Premiere') were removed for subsequent analyses of mango cultivar diversity, and loci with a minor allele frequency <0.0001 were filtered out using the R package poppr (Kamvar et al. 2014), creating a dataset of 220 SNPs for 106 cultivars (*CULT_106*, Table 5.1).

To examine population differentiation within 106 mango cultivars (*CULT_106*), we performed hierarchical analysis of molecular variance (AMOVA, Excoffier *et al.* 1992, Michalakis & Excoffier 1996) in the software GENODIVE v. 2.0b27 (Meirmans & Van Tienderen 2004) under an infinite allele model and with 999 permutations to test for significant differences. Prior to the AMOVA, missing data were filled in with randomly drawn alleles determined by the overall allele frequencies. Separate AMOVAs were performed on three distinct sets of hierarchical divisions of diversity: among individual, population, and region; among individual, population, and continent; and among individual, region, and continent (for population, region, and continent classifications, see Table 5.1). To look for significant genetic differentiation between mango cultivars originating from distinct geographic regions, pairwise values of population differentiation (F_{ST} of Weir & Cockerham 1984) between all populations, regions, and continents were calculated in GENODIVE v. 2.0b27 (Meirmans & Van Tienderen 2004) and significance was evaluated with a correction for multiple tests.

Indices of genetic diversity

Common measures of genetic diversity were calculated for the eight populations of mango cultivars (India, Indochina, Malesia, Africa, the Americas, the Caribbean, Mexico, Florida) in *CULT_106*. Observed heterozygosity (H_O), gene diversity (H_E, the expected heterozygosity within subpopulations assuming Hardy-Weinberg Equilibrium), and the inbreeding coefficient (F_{IS}) were calculated in the R package hierfstat (Goudet 2005). Additional packages were used to calculate allelic richness (PopGenReport, Adamack & Gruber 2014), nucleotide diversity (π) (pegas, Paradis 2010), private alleles (poppr, Kamvar et al. 2014), and percent polymorphism (adegenet, Jombart 2008; Jombart & Ahmed 2011).

Results

Sequencing and assembly

Across three lanes of sequencing, we obtained 454,840,461 raw reads for all 280 individuals and 201,811,265 for the 158 individuals included in this study (average 1,277,286, standard deviation 541,376, Appendix 5.1). The FastQC results indicated that reads were of high quality across the entire 150 bp length. After filtering, the complete dataset (*FULL*), and the subset for mango cultivars (*CULT*) recovered 612 and 281 unlinked SNPs, respectively. For the *FULL_33* dataset, which included all variable sites across all 158 individuals and allowed for a higher level of missing data, 126,653 SNPs were identified.

Population Structure

The UPGMA dendrogram provides information about genetic structure at both intraspecific and interspecific levels (Fig. 5.2). Individual samples of *M. pentandra, M. casturi, M. gedebe,* and *M. zeylanica* clustered within their respective species, in agreement with previous phylogenetic analysis (see Ch. III, this work). The dendrogram also reveals a group of samples of uncertain identity that includes the African cultivar 'Zebda' and three samples previously identified as *M. laurina* ('PR_Martex', 'Mempelam', and 'Aquea'). Within the main group of *M. indica* cultivars, the dendrogram reveals two distinct clusters. The first predominantly consists of Southeast Asian cultivars and is subdivided into clusters of Indochinese and Malesian cultivars (including many Malesian samples that were of uncertain identity). Notably, there are a few individuals from other regions within the Southeast Asian group, including 'Joellen' (Florida), 'Hindi Besanara' (Africa), and 'Diab' (Africa). Additionally, two Mexican cultivars ('Ataulfo' and 'Manila') cluster with the lone Philippine cultivar ('Carabao'), corroborating the historical documentation that indicates some Mexican mango germplasm was introduced directly from the Philippines. Three samples identified as *M. lalijiwa* ('M_lalijiwa', 'Poh Pakel', 'M_lalijiwa_G') are also grouped within the Malesian subcluster. The second of the two primary *M. indica* groups contains cultivars from all regions of the world except Malesia, including five samples from Indochina ('Saigon', 'Swethintha', 'Myatrynat', 'Cac', 'Maha_Chanok'). Within the group, there is little evidence of clustering associated with geography, though one subcluster contains only individuals from the Caribbean, South America, Africa and Mexico (along with one unidentified sample). Of note, six cultivars from Florida cluster closely together, including the economically important

'Tommy_Atkins' and 'Keitt'. Potentially because of the higher levels of missing data allowed, two samples ('Pyu Pyu Kalay' Indochina and 'Royal Special' India) did not cluster in either of the two major groups, and a third ('Tyler Premiere' Africa) was recovered as a solitary individual nested between the two *M. indica* clusters.

Analysis of the full dataset (*FULL*) in the program STRUCTURE indicated $K = 2$ as the optimal number of populations using the ΔK method of Evanno (Fig. 5.3a, Appendix 5.2), though we found additional informative structure for $K = 3$. For the *FULL* dataset, mango cultivars from Florida, the Caribbean, South America, Africa (with the exception of two individuals), and India show high levels of shared ancestry from a single group and only a few individuals indicate low levels of admixture with a secondary group. In contrast, almost all cultivars from Indochina and Malesia show high levels of admixture with the second group. Admixed ancestry from groups one and two was also found in *M. casturi, M. pentandra, and M. lalijiwa*. Both *M. gedebe* and *M. laurina* are assigned to group three with little evidence of admixture. A few individuals, including three cultivars from Africa, both samples of *M. zeylanica,* and multiple unidentified samples from Florida and Malesia, were inferred to be of admixed ancestry between groups one and three or between all three groups. Unsurprisingly, a variety of ancestry is assigned to individuals from the unidentified accessions in Florida and Malesia, with individuals assigned to group one, group three, or showing admixture between two or more of the populations. Of note, no individuals are inferred to have >60% ancestry from group two.

Population structure of the subset of mango cultivars (*CULT*) was first examined with the program STRUCTURE, which found K of 4 to be optimal using the ΔK method

of Evanno (Fig. 5.4, Appendix 5.3). These results show mango cultivars from Southeast Asia have different ancestry compared to cultivars from other regions. In general, cultivars from Florida, the Caribbean, South America, Mexico, Africa (with the exception of two individuals), and India are inferred to be of admixed ancestry from groups one and two, while Southeast Asian cultivars are assigned to group three, with some admixture from groups one and two. Three African cultivars show admixture with a fourth population that is not found at high levels in other cultivars. Indicative of the ongoing exchange of germplasm across the world, all populations include some individuals that deviate from the overall pattern for that population. For analysis of *CULT_106* (106 mango cultivars) using principal components, the first principal component explained 8.37% of the variance while the second explained 5.79% (Fig. 5.5). The PCA clustered cultivars from India with those from Florida, the Caribbean, South America, Africa, and Mexico. The majority of mango cultivars from Malesia and Indochina form a distinct cluster. Together, the results of clustering analyses indicate that Southeast Asian cultivars contain unique genetic diversity compared to cultivars from other regions of the world.

Genetic Diversity and Population Differentiation

Measures of genetic diversity were calculated for the eight populations of mango cultivars (Table 5.2). In general, levels of diversity were similar across all populations. Levels of observed heterozygosity (H_O) were highest for Malesia and Mexico (0.1576 and 0.1545, respectively) and lowest for Indochina (0.1322). Mexico also had the highest levels of gene diversity (H_E 0.1594) while Florida had the lowest (0.1330). Values for the inbreeding coefficient F_{IS} ranged from 0.0646 (Africa) to -0.0851 (Malesia), indicating

relatively low levels of inbreeding in mango cultivars. Values of allelic richness differed very little between populations, with the highest level found in the Mexican population (1.1998) and the lowest found in the Malesian population (1.1603). We observed the highest nucleotide diversity in the African population (0.0405) while the Floridian population had the lowest (0.0196). Percent polymorphism varied from 70.00% in the Indochinese population to 26.82% in the Malesian population. The number of private alleles was highest in the Indian and Indochinese populations (21 and 26, respectively).

Despite similar levels of genetic diversity across populations, many pairs of populations were significantly differentiated from one another by pairwise calculations of F_{ST} (Table 5.3). The Floridian population was singularly significantly different from all other populations. Both the Indochinese and Indian populations were significantly different from all other populations except that of Malesia. However, hierarchical AMOVA showed very low levels of population differentiation, with the majority of variation (92.3-92.6%) found within individuals (Table 5.4). Significant differences were observed among populations within continents and among regions within a continent (p = 0.001), as well as among regions (p = 0.013).

Discussion

Here, we analyzed mango cultivars and closely related *Mangifera* species to describe phylogeographic patterns of diversity, explicitly test whether India represents a 'center of diversity' for mango, and quantify the genetic bottleneck that mango underwent as it was introduced into new regions of the world. Collectively, our results provide

insight into the origin and domestication of one of the world's most important perennial fruit crops.

Insights into the diversity of cultivated mango

Traditionally, crops are thought to have a center of diversity near where they were originally domesticated (Vavilov 1987) and experience a loss of this baseline diversity as the result of introduction bottlenecks (Cooper *et al.* 2001; van de Wouw *et al.* 2010). However, relatively few studies have sought to quantify the introduction bottlenecks experienced by perennial species during domestication or test for centers of origin for these species. While most scholars believe that mango was domesticated in India, here, we find no evidence that mango has a center of diversity in India or that it has experienced a loss of diversity during its introduction into new regions of the world. Instead, we find that most metrics of genetic diversity are similar across all regions, with percent polymorphism and the number of private alleles indicating that India and Southeast Asia contain higher diversity compared to other regions. This finding does not necessarily preclude India as a center of origin for mango, but perhaps indicates that, like some other perennial species, the mango has been robust to the effects introduction bottlenecks during its domestication.

In the early 1900s, mango cultivation and breeding programs intensified in the Americas, especially in South Florida, which went on to produce many of today's most commercially important cultivars. The novel characteristics of these cultivars and their success in the global market led South Florida to be dubbed a secondary center of domestication (Knight & Schnell 1994), though previous molecular work has shown this

to be unfounded (Schnell *et al.* 2006). We find Floridian mangoes do not have greater genetic diversity than those from other regions, providing further evidence that Florida is not a secondary center of diversity. In fact, UPGMA clustering indicates many of the Floridian cultivars appear to be closely related to one another, including the three most commercially important Floridian cultivars included in this study, Tommy Atkins, Kent, and Keitt. This finding highlights an important concern in perennial crop cultivation: the loss of diversity at the population level, rather than the individual level. Although most perennial species have high within-individual heterozygosity, they are clonally propagated and therefore commercial orchards have virtually no population-level diversity, putting them at risk for disease outbreaks (Gross *et al.* 2012). The lack of diversity in commercial orchards is exacerbated when the most important commercial cultivars come from a narrow genetic base, as is the case for the three Floridian cultivars.

Simulation studies have shown metrics of diversity calculated from RADseq datasets may be inflated because of allele dropout and high levels of missing data (Gautier *et al.* 2012; Arnold *et al.* 2013), and we therefore restricted the amount of missing data in our dataset. Contrary to these expectations, our estimates of gene diversity in mango were considerably lower than those from the only other comparable report. Sherman et al. (2015) estimated gene diversity from transcriptome-derived SNP markers to have a median value of 0.28-0.43, roughly 2-4 times higher than the average (and median) values calculated here. The explanation for this discrepancy is not immediately clear, however, more recent empirical work indicates that missing data may not inflate diversity indices in empirical datasets as much as was initially proposed (Hodel *et al.* 2017). One possibility for the differences in gene diversity between studies

is that low sequence coverage and low tolerance for missing data in the present study made it so that only highly conserved regions within the genome were examined, these regions being less likely to have polymorphisms (Huang & Knowles 2016). As we progress toward a high-quality sequence of the mango genome (Singh *et al.* 2016; D. Kuhn, pers. comm.) better estimations of heterozygosity across the genome of mango will be possible.

Phylogeography of the mango

Our analysis of genetic structure within cultivated mango germplasm consistently recovers two distinct groups of mango cultivars, corresponding to individuals from Southeast Asia (Indochina and Malesia) and those from other regions of the world. The UPGMA analysis recovers substructure within the Southeast Asian group, with Indochinese and Malesian cultivars clustering separately, though this distinction is not found in STRUCTURE analysis or PCA. In support of historical documentation that suggests Mexico received introductions of mango from directly from the Philippines, two of the five Mexican cultivars cluster closely with the lone Philippine cultivar. While the Philippines is considered part of Malesia, the group of Mexican and Philippine cultivars clusters with Indochinese cultivars rather than Malesian cultivars. In all three analyses (UPGMA, STRUCTURE, PCA) mango cultivars from India, Florida, Africa, the Caribbean, South America, and three Mexican individuals cluster together, with relatively little population structure observed within this group.

Although the amount of diversity in each of the mango populations we analyzed was relatively similar, we find clear evidence from three clustering methods (UPGMA,

PCA, STRUCTURE) that Southeast Asian cultivars contain unique genetic diversity. The result is in agreement with some previous molecular studies (Schnell et al. 2006; Dillon et al. 2013) and also corresponds to observations of morphological diversity in mango. On the basis of a suite of fruit characteristics, mango cultivars are categorized as either Indian or Indochinese types (Crane & Campbell 1994). Indian cultivars tend to have an apparent color change when ripe, turning orange or red, and are rounded with fibrous, strong-flavored flesh. They also generally have a seed that is monoembryonic, producing a single seedling. In contrast, Indochinese cultivars tend to turn yellow or remain green when ripe, display a prominent "nose" or "beak", and have flesh that is less fibrous and mild in flavor. Indochinese cultivars also typically have polyembryonic seeds, containing a single zygotic embryo and multiple embryos derived from the maternal nucellar tissue (Mukherjee & Litz 2009). Nucellar embryony is a rare trait in angiosperms, though the phenomenon has been observed in at least two other species of *Mangifera, (M. laurina, M. casturi;* Kostermans & Bompard 1993) and is found in another cultivated genus within the order Sapindales, *Citrus* (Wang *et al.* 2017).

The unique diversity found in Southeast Asian mango cultivars suggests that mango may follow one or two other trends seen in perennial crops: multiple domestications and interspecific hybridization with congeneric species (Miller & Gross 2011). Both of these phenomena are common in the course of perennial fruit crop domestication, a process that likely occurs on a broader geographic scale and over a longer period of time than it does in annual species (Miller & Gross 2011). Perennial fruit crops that are known to have multiple origins include breadfruit *(Artocarpus altilis)*, pecan *(Carya illinoinensis)*, hazelnut *(Coryus avellana)*, coconut *(Cocos nucifera)*, olive

(Olea europaea), apricot *(Prunus armeniaca)*, peach *(Prunus persica)*, pear *(Pyrus communis)*, red raspberry *(Rubus idaeus),* blackberry *(Rubus spp.)*, and jocote *(Spondias purpurea)* (Miller & Gross 2011). The list of perennial fruit crops that are the result of hybridization events between congeneric species is much longer, but includes sweet orange *(Citrus sinensis)*, fig *(Ficus carica)*, walnut *(Juglans regia)*, avocado *(Persea americana)*, and grape *(Vitis vinifera)* (Miller & Gross 2011). In the case of mango, our results indicate that some of the genetic diversity present in modern-day mangoes may not have originated in India. Bompard (2009) previously proposed that, despite archaeological and linguistic evidence, *M. indica* might have been domesticated independently in India and Indochina. However, mounting evidence of interspecific hybridization events within the genus *Mangifera* (Ch. VI) suggests that the novel diversity seen in Indochinese cultivars may be the result of genetic introgression. Teasing apart the seemingly complex history of domestication in mango requires more thorough sampling of wild *M. indica,* Indian, Indochinese, and Malesian mango cultivars and landraces, and closely related *Mangifera* species.

Remaining gaps and future goals

While we did not observe a center of diversity in India or Florida or a loss of diversity associated with the mango's dispersal into Africa and the Americas, this line of inquiry deserves additional attention. Given that population structure has been observed within Indian mango germplasm (Ravishankar *et al.* 2000; Kumar *et al.* 2001; Karihaloo *et al.* 2003; Damodaran *et al.* 2012; Vasugi *et al.* 2012; Surapaneni *et al.* 2013; Ravishankar *et al.* 2015; Singh *et al.* 2016), we made an effort to include a diverse subset

of Indian cultivars in our analysis, however it is possible that the individuals included here do not fully encompass the diversity present in India. Additionally, sampling from within Africa was restricted because of the limited number of African genebank accessions. Future efforts should be made to address the lack of African germplasm in U.S. accessions and refine our understanding of the phylogeography of mango in Africa.

Here, we tested whether mango incurred a dispersal bottleneck by comparing cultivars from different regions of the world. However, the question of whether mango underwent a loss of diversity during the initial phases of domestication cannot be answered without including samples from the mango's wild progenitors, though future analysis using coalescent simulations of demography may help shed light on this issue. For a number of reasons, it may be difficult to locate and identify the mango's wild progenitor populations. Although wild populations of *M. indica* have been reported in regions of Northeastern India, Nepal, Bhutan, Bangladesh, and Myanmar (Kostermans & Bompard 1993), to our knowledge, these populations have not been recently surveyed and have never been rigorously studied in a genetic framework. As a result of intensifying land use requirements in this region of the world, it is possible that many of populations of wild *M. indica* have been extirpated. Additionally, whether the individuals in this region truly represent wild *M. indica* or whether they are naturalized offspring of previously cultivated individuals may be difficult to determine. Naturalized mango trees are frequently observed in the Neotropics, and, to the casual observer, appear to be wild (Bompard, 2009). Further complicating this problem is the fact that many closely related *Mangifera* species bear remarkable resemblance to cultivated mango, and common names of these species are often translated to "wild mango" (Kostermans & Bompard

1993). With continuing deforestation in the region, a thorough investigation of all remaining putative populations of wild *M. indica* and closely related species is an increasingly urgent need.

Literature Cited

Abbo S, Berger J, Turner NC (2003) Evolution of cultivated chickpea: four bottlenecks limit diversity and constrain adaptation. *Functional Plant Biology*, **30**(10), 1081–1087.

Adamack AT, Gruber B (2014) PopGenReport: Simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution*, **5**(4), 384–387.

Aerts R, Berecha G, Gijbels P, *et al.* (2013) Genetic variation and risks of introgression in the wild *Coffea arabica* gene pool in south-western Ethiopian montane rainforests. *Evolutionary Applications*, **6**(2), 243–252.

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Anthony F, Combes MC, Astorga C, *et al.* (2002) The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. *Theoretical and Applied Genetics*, **104**(5), 894–900.

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

Arnold ML (2004) Natural hybridization and the evolution of domesticated, pest and disease organisms. *Molecular Ecology*, **13**(5), 997–1007.

Atchison GW, Nevado B, Eastwood RJ, *et al.* (2016) Lost crops of the incas: Origins of domestication of the Andean pulse crop Tarwi, *Lupinus mutabilis*. *American Journal of Botany*, **103**(9), 1592–1606.

Baird NA, Etter PD, Atwood TS, *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS One*, **3**(10), e3376.

Bompard JM (2009) Taxonomy and Systematics. In R. E. Litz (Ed.), *The Mango: Botany, Production and Uses* (2nd ed.). Cambridge, MA: CAB International.

Burke JM, Tang S, Knapp SJ, Rieseberg LH (2002) Genetic Analysis of Sunflower Domestication. *Genetics*, **161**, 1257–1267.

Clement CR (1999) 1492 and the loss of Amazonian crop genetic resources. I. The relation between domestication and human population decline. *Economic Botany*, **53**(2), 188–202.

Cooper HD, Spillane C, Hodgkin T (2001) *Broadening the genetic base of crop production (Livre numérique Google)*. *Organization*.

Crane JH, Campbell CW (1994) *The Mango*. *Fact Sheet HS-2*.

Damodaran T, Kannan R, Ahmed I, *et al.* (2012) Assessing genetic relationships among mango (*Mangifera indica* L.) accessions of Andaman Islands using inter simple sequence repeat markers. *New Zealand Journal of Crop and Horticultural Science*, **40**(4), 229–240.

DeCandolle (1884) *Origin of Cultivated Plants*. Kegan Paul, Trench & Company.

Diaz-Matallana M, Schuler-Garcia I, Ruiz-Garcia M, Hodson de Jaramillo E (2009) Analysis of diversity among six populations of Colombian mango (*Mangifera indica* L. cvar. Hilacha) using RAPDs markers. *Electronic Journal of Biotechnology*, **12**(3).

Dillon NL, Bally ISE, Wright CL, *et al.* (2013) Genetic diversity of the Australian National Mango Genebank. *Scientia Horticulturae*, **150**, 213–226.

Doebley JF, Gaut BS, Smith BD (2006) The Molecular Genetics of Crop Domestication. *Cell*, **127**, 1309–1321.

Dos Santos Ribeiro ICN, Lima Neto FP, Santos CAF (2012) Allelic database and accession divergence of a Brazilian mango collection based on microsatellite markers. *Genetics and Molecular Research : GMR*, **11**(4), 4564–74.

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus*, **12**, 13–15.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**(2), 359–361.

Eaton DAR (2014) PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**(13), 1844–1849.

Esquinas-Alcázar J (2005) Science and society: protecting crop genetic diversity for food security: political, ethical and technical challenges. *Nature Reviews. Genetics*, **6**(12), 946–53.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, **14**(8), 2611–2620.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application., **491**, 479–491.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**(4), 1567–87.

FAO (2003) Tropical Fruits. In *Medium-term prospects for agricultural commodities*. Food and Agriculture Organization of the United Nations, Rome, Italy. (pp. 1–6).

FAOSTAT (2013). Food and Agriculture Organization of the United Nations Statistics Division. Available from: http://www.fao.org/faostat/.

Gao Y, Yin S, Wu L, *et al.* (2017) Genetic diversity and structure of wild and cultivated *Amorphophallus paeoniifolius* populations in southwestern China as revealed by RAD-seq. *Scientific Reports*, **7**(1), 2–11.

Gaut BS, Díez CM, Morrell PL (2015) Genomics and the Contrasting Dynamics of Annual and Perennial Domestication. *Trends in Genetics*, **31**(12), 709–719.

Gautier M, Gharbi K, Cezard T, *et al.* (2012) The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology*, 3165–3178.

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes,* **5**(1), 184-186.

Gross BL, Henk AD, Forsline PL, *et al.* (2012) Identification of interspecific hybrids among domesticated apple and its wild relatives. *Tree Genetics & Genomes*, **8**(6), 1223–1235.

Gross BL, Henk AD, Richards CM, *et al.* (2014) Genetic diversity in *Malus* x*domestica* (Rosaceae) through time in response to domestication. *American Journal of Botany*, **101**(10), 1770–1779.

Gross BL, Olsen KM (2010) Genetic perspectives on crop domestication. *Trends in Plant Science*, **15**(9), 529–37.

Gunn BF, Baudouin L, Olsen KM (2011) Independent origins of cultivated coconut (*Cocos nucifera* L.) in the old world tropics. *PloS One*, **6**(6), e21143.

Hirano R, Oo TH, Watanabe KN (2012) Myanmar mango landraces reveal genetic uniqueness over common cultivars from Florida, India, and Southeast Asia. *Genome*, **53**, 321–330.

Hodel RGJ, Chen S, Payton AC, *et al.* (2017) Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: comparing microsatellites and RAD-Seq and investigating loci filtering. *Scientific Reports*, **7**(1), 17598.

Huang H, Knowles LL (2016) Unforeseen consequences of excluding missing data from next-generation sequences: Simulation study of RAD sequences. *Systematic Biology*, **65**(3), 357–365.

Huang P, Molina J, Flowers JM, *et al.* (2012) Phylogeography of Asian wild rice, *Oryza rufipogon*: a genome-wide view. *Molecular Ecology*, **21**(18), 4593–604.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**(5), 1322–32.

Hufford MB, Lubinksy P, Pyhäjärvi T, *et al.* (2013) The genomic signature of crop-wild introgression in maize. *PLoS Genetics*, **9**(5), e1003477.

Hyten DL, Song Q, Zhu Y, *et al.* (2006) Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(45), 16666–71.

Iqbal MJ, Reddy OUK, El-Zik KM, Pepper AE (2001) A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. *TAG Theoretical and Applied Genetics*, **103**(4), 547–554.

IUCN (2012) World Conservation Monitoring Centre 1998. Retrieved from www.iucnredlist.org

Iyer CPA, Schnell RJ (2009) Breeding and Genetics. In R. E. Litz (Ed.), *The Mango: Botany, Production and Uses* (2nd ed., pp. 68–83). Cambridge: CAB International.

Jakobsson M, Rosenberg NA (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**(14), 1801–1806.

Jombart T (2008) Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**(11), 1403–1405.

Jombart T, Ahmed I (2011) adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**(21), 3070–3071.

Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ,* **2**, e281.

Karihaloo JL, Dwivedi YK, Sunil A, *et al.* (2003) Analysis of genetic diversity of Indian mango cultivars using RAPD markers. *Journal of Horticultural Science and Biotechnology*, **78**(3), 285–289.

Kassa MT, Penmetsa RV, Carrasquilla-Garcia N, *et al.* (2012) Genetic patterns of domestication in pigeonpea (*Cajanus cajan* (L.) Millsp.) and wild *Cajanus* relatives. *PloS One*, **7**(6), e39563.

Knight RJ, Campbell RJ, Maguire I (2009) Important Mango Cultivars and their Descriptions. In R. E. Litz (Ed.), *The Mango: Botany, Production and Uses* (2nd ed., pp. 42–66). Cambridge, MA: CAB International.

Knight RJ, Schnell RJ (1994) Mango Introduction in Florida and the 'Haden' Cultivar's Significance to the Modern Industry. *Economic Botany*, **48**(2), 139–145.

Kostermans AJGH, Bompard JM (1993) *The Mangoes: Their Botany, Nomenclature, Horticulture, and Utilization*. San Diego, CA: Academic Press.

Kovach MJ, Sweeney MT, McCouch SR (2007) New insights into the history of rice domestication. *Trends in Genetics*, **23**(11), 578–587.

Kuhn DN, Bally ISE, Dillon NL, et al. Sherman A (2017) Genetic Map of Mango: A Tool for Mango Breeding. *Frontiers in Plant Science*, **8**(April), 1–11.

Kumar NVH, Narayanaswamy P, Prasad DT, *et al.* (2001) Estimation of genetic diversity of commercial mango (*Mangifera indica* L.) cultivars using RAPD markers. *Journal of Horticultural Science & Biotechnology*, **76**(5), 529–533.

Ladizinsky G (1985) Founder effect in crop-plant evolution. *Economic Botany*, **39**(2), 191–199.

Li C, Zhou A, Sang T (2006) Rice domestication by reducing shattering. *Science*, **311**(5769), 1936–9.

Londo JP, Chiang Y, Hung K, *et al.* (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proceedings of the National Academy of Sciences*, **103**(25), 9578–9583.

Loor Solorzano RG, Fouet O, Lemainque A, *et al.* (2012) Insight into the wild origin, migration and domestication history of the fine flavour nacional *Theobroma cacao* L. variety from Ecuador. *PloS One*, **7**(11), e48438.

Luo C, He XH, Chen H *et al.* (2011) Genetic diversity of mango cultivars estimated using SCoT and ISSR markers. *Biochemical Systematics and Ecology,* **39**(4-6), 676-684.

Matsuoka Y, Vigouroux Y, Goodman MM, *et al.* (2002) A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences of the United States of America*, **99**(9), 6080–4.

Meirmans PG, Van Tienderen PH (2004) Genotype and Genodive: Two Programs for the Analysis of Genetic Diversity of Asexual Organisms. *Molecular Ecology Notes*, **4**(4), 792–794.

Meyer RS, DuVal AE, Jensen HR (2012) Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *The New Phytologist*, **196**(1), 29–48.

Meyer RS, Purugganan MD (2013) Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics*, **14**(12), 840–52.

Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics*, **142**(3), 1061–1064.

Miller AJ, Gross BL (2011) From forest to field: perennial fruit crop domestication. *American Journal of Botany* **98**(9), 1389–1414.

Miller MR, Dunham JP, Amores A, *et al.* (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, **17**(2), 240–8.

Mukherjee SK (1949) The mango and its wild relatives. *Science and Culture*, **26**, 5–9.

Mukherjee SK (1972) Origin of Mango (*Mangifera indica*). *Economic Botany*, **26**(3), 260–264.

Mukherjee SK, Litz RE (2009) Introduction: Botany and Importance. In R. E. Litz (Ed.), *The Mango: Botany, Production and Uses* (2nd ed., pp. 1–18). Wallingford, UK: CAB International.

Myles S, Boyko AR, Owens CL, *et al.* (2011) Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences*, **108**(9), 3530–3535.

Olsen KM, Schaal BA (1999) Evidence on the origin of cassava: Phylogeography of *Manihot esculenta*. *Proceedings of the National Academy of Sciences*, **96**(May), 5586–5591.

Olsen KM, Wendel JF (2013) A Bountiful Harvest: Genomic Insights into Crop Domestication Phenotypes. *Annual Review of Plant Biology*, **64**(1), 47–70.

Pan Y, Wang X, Sun G, *et al.* (2016) Application of RAD sequencing for evaluating the genetic diversity of domesticated panax notoginseng (Araliaceae). *PLoS ONE*, **11**(11), 1–17.

Paradis E (2010) Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, **26**(3), 419–420.

Peterson BK, Weber JN, Kay EH, *et al.* (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non model species. *PloS One*, **7**(5), 1–11.

Pickersgill B (2007) Domestication of plants in the Americas: insights from Mendelian and molecular genetics. *Annals of Botany*, **100**(5), 925–40.

Popenoe W (1920) The Mango. In *Manual of Tropical and Subtropical Fruits* (pp. 79–145). New York: The Macmillian Company.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**(2), 945–59.

Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. *Nature*, **457**(7231), 843–8.

Rambaut, A (2006) FigTree: tree figure drawing tool. Available from: http://tree.bio.ed.ac.uk/software/figtree/.

Ravishankar KV, Anand L, Dinesh MR (2000) Assessment of genetic relatedness among mango cultivars of India using RAPD markers. *The Journal of Horticultural Science and Biotechnology*, **75**(2), 198–201.

Ravishankar KV, Bommisetty P, Bajpaj A, *et al.* (2015) Genetic diversity and population structure analysis of mango (*Mangifera indica*) cultivars assessed by microsatellite markers. *Trees*, **29**(3), 775–783.

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Saintenac C, Jiang D, Wang S, Akhunov E (2013) Sequence-based mapping of the polyploid wheat genome. *G3: Genes, Genomes, Genetics*, **3**(7), 1105–14.

Salamini F, Ozkan H, Brandolini A, *et al.* (2002) Genetics and geography of wild cereal domestication in the near east. *Nature Reviews Genetics*, **3**(6), 429–41.

Savolainen O, Pyhäjärvi T (2007) Genomic diversity in forest trees. *Current Opinion in Plant Biology*, **10**(2), 162–7.

Schnell RJ, Brown JS, Olano CT, *et al.* (2006) Mango Genetic Diversity Analysis and Pedigree Inferences for Florida Cultivars Using Microsatellite Markers. *Journal of the American Horticultural Society*, **131**(2), 214–224.

Sennhenn A, Prinz K, Gebauer J, *et al.* (2013) Identification of mango (*Mangifera indica* L.) landraces from Eastern and Central Kenya using a morphological and molecular approach. *Genetic Resources and Crop Evolution*, **61**(1), 7–22.

Shamili M, Fatahi R, Hormaza JI (2012) Characterization and evaluation of genetic diversity of Iranian mango (*Mangifera indica* L., Anacardiaceae) genotypes using microsatellites. *Scientia Horticulturae*, **148**, 230–234.

Sherman A, Rubinstein M, Eshed R, *et al.* (2015) Mango (*Mangifera indica* L.) germplasm diversity based on single nucleotide polymorphisms derived from the transcriptome. *BMC Plant Biology*, **15**(1), 277.

Singh NK, Mahato AK, Jayaswal PK, *et al.* (2016) Origin, Diversity and Genome Sequence of Mango (*Mangifera indica* L.). *Indian Journal of History of Science*, **51**(2), 355–368.

Singh SP, Gepts P, Debouck DG (1991) Races of Common Bean. *Economic Botany*, **45**(3), 379–396.

Stetter MG, Müller T, Schmid KJ (2017) Genomic and phenotypic evidence for an incomplete domestication of South American grain amaranth (*Amaranthus caudatus*). *Molecular Ecology*, **26**(3), 871–886.

Surapaneni M, Vemireddy LR, Begum H, *et al.* (2013) Population structure and genetic analysis of different utility types of mango (*Mangifera indica* L.) germplasm of Andhra Pradesh state of India using microsatellite markers. *Plant Systematics and Evolution*, **299**(7), 1215–1229.

Swofford DL (2002) PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods). Sunderland, Massachusetts: Sinauer Associates.

The International Peach Genome Initiative (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics*, **45**(5), 487–494.

van de Wouw M, Kik C, van Hintum T, *et al.* (2010) Genetic erosion in crops: concept, research results and challenges. *Plant Genetic Resources*, **8**(1), 1–15.

Vasugi C, Dinesh MR, Sekar K, *et al.* (2012) Genetic diversity in unique indigenous mango accessions (Appemidi) of the Western Ghats for certain fruit characteristics. *Current Science*, **103**(2), 199–207.

Vavilov NI (1987) *Origin and Geography of Cultivated Plants*. (D. Löve, Trans.). Cambridge: Cambridge University Press.

Velasco D, Hough J, Aradhya M, Ross-Ibarra J (2016) Evolutionary Genomics of Peach and Almond Domestication. *Genes, Genomes, Genetics*, **6**(12), 3985-3993.

von Wettberg EJ, Chang PL, Greenspan A, *et al.* (2018) Ecology and community genomics of an important crop wild relative as a prelude to agricultural innovation. *Nature Communications,* in press.

Wang R, Stec A, Hey J, *et al.*(1999) The limits of selection during maize domestication. *Nature*, **398**(March), 236–239.

Wang X, Xu Y, Zhang S, *et al.* (2017) Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nature Genetics*, **49**(5), 765–772.

Washburn JD, Bird KA, Conant GC, Pires JC (2016) Convergent Evolution and the Origin of Complex Phenotypes in the Age of Systems Biology. *International Journal of Plant Sciences*, **177**(4), 305–318.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, **38**(6), 1358–1370.

Xu P, Xu S, Wu X, *et al.* (2014) Population genomic analyses from low-coverage RAD-Seq data: A case study on the non-model cucurbit bottle gourd. *Plant Journal*, **77**(3), 430–442.

Yang JY, Scascitelli M, Motilal LA, *et al.* (2013) Complex origin of Trinitario-type *Theobroma cacao* (Malvaceae) from Trinidad and Tobago revealed using plastid genomics. *Tree Genetics & Genomes*, **9**(3), 829–840.

Zeder MA (2006) Central questions in the domestication of plants and animals. *Evolutionary Anthropology: Issues, News, and Reviews*, **15**(3), 105–117.

Zohary D (2004) Unconscious Selection and the Evolution of Domesticated Plants. *Economic Botany*, **58**(1), 5–10.

Tables

Table 5.1. Samples included in datasets *FULL*, *FULL_33*, *CULT*, and *CULT_106*.

Species	Population	Region	Continent	Individuals in Dataset			
				FULL	*FULL_33*	*CULT*	*CULT_106*
M. gedebe	--	--	--	4	4	0	0
M. casturi	--	--	--	4	4	0	0
M. lalijiwa	--	--	--	2	2	0	0
M. laurina	--	--	--	3	3	0	0
M. pentandra	--	--	--	2	2	0	0
***M.* spp.**	Florida_sp	--	--	9	9	0	0
***M.* spp.**	Malesia_sp	--	--	23	23	0	0
M. zeylanica	--	--	--	2	2	0	0
M. indica	Africa	Africa	Africa	13	13	13	11
M. indica	S America	Americas	Americas	8	8	8	8
M. indica	Caribbean	Caribbean	Americas	19	19	19	19
M. indica	Florida	Florida	Americas	20	20	20	20
M. indica	India	India	India	19	19	19	19
M. indica	Indochina	SE Asia	Asia	21	21	21	21
M. indica	Malesia	SE Asia	Asia	3	3	0	0
M. indica	Mexico	Americas	Americas	5	5	0	0
Total				**158**	**158**	**108**	**106**
# SNPS in dataset:				**612**	**126,653**	**281**	**220**

Table 5.2. Measures of diversity for 106 mango cultivars from eight populations calculated from SNP loci (CULT_106). For each column, warmer colors reflect lower values. *Ho*: observed heterozygosity; *He*: heterozygosity within populations, aka 'gene diversity'; *Fis*: inbreeding coefficient; Ar: allelic richness; *π:* nucleotide diversity; %Poly: percent polymorphic; Ap: private alleles.

Population	*Ho*	*He*	*Fis*	Ar	π	%Poly	Ap
Africa	0.1366	0.1488	0.0646	1.1885	0.0405	52.27%	8
Caribbean	0.1443	0.1404	0.0022	1.1843	0.0293	55.00%	6
Florida	0.1454	0.1330	-0.0701	1.1755	0.0196	50.91%	3
India	0.1370	0.1420	0.0303	1.1813	0.0328	64.55%	21
Indochina	0.1322	0.1384	0.0366	1.1767	0.0324	70.00%	26
Malesia	0.1576	0.1357	-0.0851	1.1603	0.0226	26.82%	1
Mexico	0.1545	0.1594	0.0077	1.1998	0.0390	43.18%	1
SAmerica	0.1338	0.1401	0.0286	1.1796	0.0329	48.18%	2

Table 5.3. Pairwise differentiation between eight populations of 106 mango cultivars (CULT_106) calculated from 220 SNP loci. Values of F_{ST} are given above the diagonal, and p-values are given below the diagonal (light grey).

	India	Africa	SAmerica	Mexico	Caribbean	Florida	Indochina	Malesia
India	--	0.046	0.054	0.052	0.048	0.046	0.086	0.021
Africa	*0.001	--	0.011	0.031	0.032	0.059	0.067	-0.002
SAmerica	*0.008	0.163	--	0.009	0.007	0.057	0.1	0.036
Mexico	*0.011	*0.038	0.322	--	0.023	0.061	0.075	0.007
Caribbean	*0.001	*0.007	0.226	0.073	--	0.064	0.117	0.051
Florida	*0.001	*0.001	*0.001	*0.001	*0.001	--	0.109	0.072
Indochina	*0.001	*0.001	*0.001	*0.001	*0.001	*0.001	--	-0.024
Malesia	0.16	0.445	0.084	0.406	*0.028	*0.005	0.864	--

Table 5.4. Hierarchical analysis of molecular variance among individuals (a) within populations nested in regions, (b) within populations nested in continents, (c) within regions nested in continents calculated from dataset CULT_106.

a)

Source of Variation	%var	Nested in	F-stat	F-value	P-value
w/in Ind	0.926	--	F_it	0.074	--
Among Ind	0.009	Pop	F_is	0.009	0.153
Among Pop	-0.002	Reg	F_sc	-0.002	0.673
Among Reg	0.067	--	F_ct	0.067	**0.013

b)

Source of Variation	%var	Nested in	F-stat	F-value	P-value
w/in Ind	0.923	--	F_it	0.077	--
Among Ind	0.009	Pop	F_is	0.009	0.15
Among Pop	0.053	Cont	F_sc	0.054	**0.001
Among Cont	0.014	--	F_ct	0.014	0.236

c)

Source of Variation	%var	Nested in	F-stat	F-value	P-value
Within Ind	0.924	--	F_it	0.076	--
Among Ind	0.009	Reg	F_is	0.009	0.181
Among Reg	0.058	Cont	F_sc	0.059	**0.001
Among Cont	0.01	--	F_ct	0.01	0.326

** $\alpha \leq 0.001$

Figure Captions

Figure 5.1. Map of the human-mediated migration of the mango. Colors represent the geographic populations of mango cultivars analyzed in this study and correspond to labels used throughout the results. The mango is thought to have originated and been domesticated in India, Nepal, Bangladesh, and Bhutan (red). It was first dispersed into Indochina (blue) and Malesia (green), then into East and West Africa (purple), South America (Brazil, orange), the Caribbean (pink), Mexico (yellow), and Florida (brown).

Figure 5.2. UPGMA dendrogram of mango cultivars from eight populations, closely related wild *Mangifera* species, and unidentified accessions (*FULL_33*). Individuals are color-coded according to their region of origin (map inset). Unidentified accessions and *Mangifera* species are labeled in grey.

Figure 5.3. Inferred population structure for the full dataset (*FULL*) as visualized with the software distruct with (a) two and (b) three ancestral populations. Each vertical bar represents a single individual that is assigned ancestry to one or more of the four ancestral populations as indicated by the colors. Groups are labeled by region (for *M. indica*) or as sp (species) or unk (unknown) (top) and population (for *M. indica)*, species name, or MSP (Malesian species) or FSP (Floridian species) (bottom).

Figure 5.4. Inferred population structure for mango cultivars (dataset *CULT*) as visualized with the software distruct with four ancestral populations. Each vertical bar represents a single individual that is assigned ancestry to one or more of the four populations (as indicated by the colors). Groups are labeled by region (top) and population (bottom).

Figure 5.5. Principal component analysis of mango cultivars (dataset *CULT*) from eight populations (map inset). Axes are labeled with the percent of variation explained by the corresponding principal component. Two individuals identified as outliers in UPGMA and STRUCTURE analyses were not plotted here.

Figures

Figure 5.1.

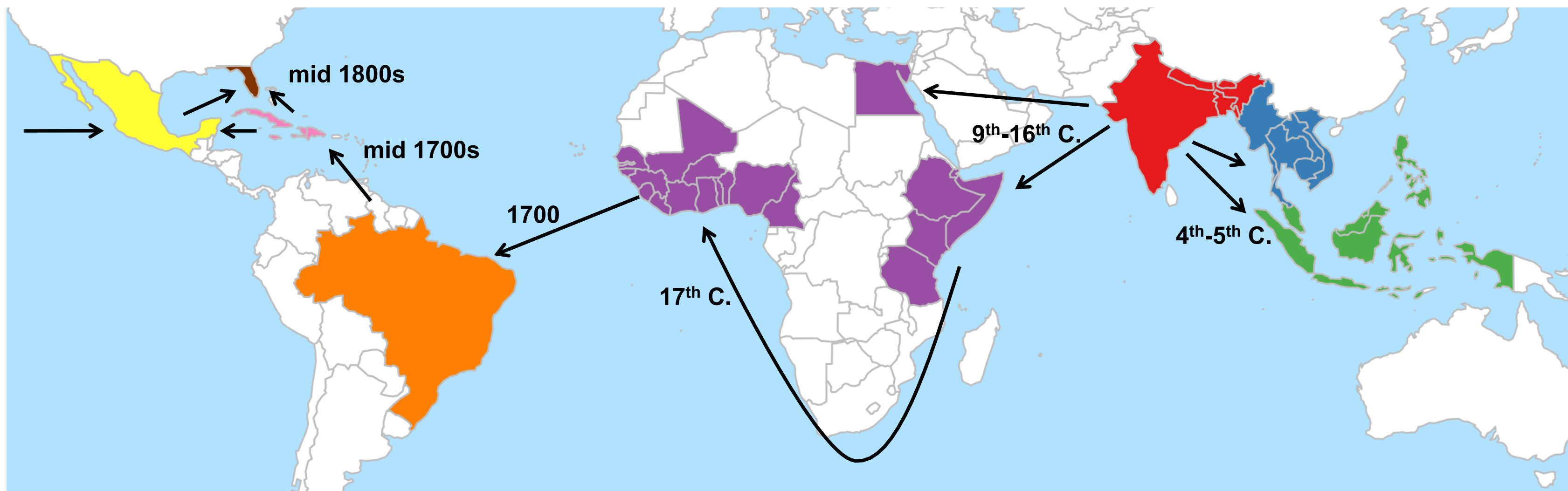


Figure 5.2.

ZH_SBG_1_M_oblongifolia
EC_MI_44_M_aquea
HH_MI_88_M_pentandra
DA_MI_114_Zebda
TD_KRP_22_M_sp
9C_MI_130_M_mempelam
WA_MI_45_M_PR_Martex
13B_SBG_4_M_pentandra
ZG_SBG_26_M_pentandra
SB_MI_78_M_A_carle
FD_KRP_28_M_casturi
11H_MI_86_M_casturi_982186A
QB_KRB_8_M_griffithii
15B_MI_58_Kaddu_Ma_odorata
AA_FS2_M_zeylanica
HE_MI_39_Pyu_Pyu_Kalay
FF_MI51_Royal_Special
11C_MI_84_Thai_Everbearing
JD_MI_3_Cambodiana
LA_MI_145_Joellen
6D_MI_104_Ataulfo
3C_MI_20_Carabao
DG_MI_105_Manilita
6A_MI_102_Chok_Anon
DF_MI_80_Nam_Tam_Teem
IA_MI_12_Tong_Dam
FA_MI_76_Chao_savoy
IG_MI_4_Phimsen_Mun
6C_MI_11_Nam_Doc_Mai
EB_MI_74_Pam_kai_mia
JE_MI_61_Praya_Savoy
RE_MI_31_Kaeo_Luemkon
6G_MI_68_Pohn_Sawadee
RH_MI_15_Ivory
BB_K3_M_indica
XF_FRIM_3_M_indica
9F_MI_156_Depih_Pasir
EG_MI_85_M_laurina_1_B4
IE_MI_87_Depih_Biasa
16E_JFL_504_M_lalijiwa
GF_MI_64_Poh_Pakel
IC_MI_59_M_lalijiwa_G
AE_GBTB3_M_indica
GB_KRP_11_M_sp
3A_MI_37_Golek
PA_MI_89_Aeromanis
DD_MI_97_Diab
JG_MI_54_Sig_Siput
15E_KRP_33_M_indica_cv_Madu
TB_KRP_2_M_minor
17F_MI_50_Butterfly_Hainan
NA_MI_56_Hindi_Besanara
CD_KRP_6_M_foetida_cv_Pakel_lumut
3E_MI_32_Gedong_Ginco
14C_KRP_31_M_indica_cv_Gandik_luyung
LB_KRP_32_M_indica_kepodang
JF_MI_116_Tyler_Premiere
FH_MI_83_Tenom
9A_MI_112_Mamita
BA_FS1_M_odorata
11D_MI_22_Julie
WC_MI_5_Graham
FC_MI_13_Bicochuelo
11F_MI_109_Malindi
6B_MI_158_Baptiste
EE_MI_26_Espada
RA_MI_30_Extrema
OH_MI_137_Peach
HB_MI_163_Sabre
YG_MI_146_Piva
9E_MI_133_Banilejo
11B_MI_140_Azucar
WB_MI_98_Vallenato
9D_MI_29_Rosa
KF_MI_149_Toledo
DB_MI_24_Turpentine
QG_MI_1_Prieto
11A_MI_154_Madame_Francis
KC_MI_131_Oro
EH_MI_17_East_Indian
16A_MI_28_Number_11
17C_MI_96_Long
9H_MI_34_Rumanii
9G_MI_120_Pettigrew
IH_MI_36_Alampur_Baneshan
YB_MI_48_Imam_Pasand
CB_MI_91_Myatrynat
CF_MI_42_Swethintha
GC_MI_23_Lanceitilla
DE_MI_113_Martin
QD_MI_9_Totapuri
FB_MI_111_Cheno
YE_MI_128_Sensation
OB_MI_126_Irwin
YD_MI_115_Lippens
CC_MI_106_Esmeralda
NG_MI_144_Diplomatico
WD_MI_132_Torbet
PG_MI_138_San_Felipe
YF_MI_122_Hodson
XH_SBG_31_M_odorata
3G_MI_129_Kent
3B_MI_117_Mabrouka
3H_MI_119_Nelpetite
GE_MI_6_Keitt
11E_MI_160_Parvin
14E_MI_10_Tommy_Atkins
YA_MI_124_Smith
GG_MI_93_Valencia_Pride
WH_MI_135_Alice
DH_MI_107_Mesk
EF_MI_95_Cairo
6F_MI_148_Gilas
13F_MI_81_Mallika
CG_MI_77_Amrapali
HC_MI_63_Dusheri
HF_MI_79_Aslul_Mukararara
3D_MI_35_Himsagar
CE_MI_25_Pascual
HA_MI_88_Pancahdarakalasa
FG_MI_53_Sindhri
WF_MI_52_Maha_Chanok
6E_MI_121_Fairchild
IF_MI_43_Saigon
ME_MI_155_Manila
QA_MI_18_Glenn
16C_MI_40_Cac
YC_MI_65_Ratna
GH_MI_57_Jumbo_Kesar
HG_MI_162_Corazon
ID_MI_14_Alphonso
3F_MI_69_Langra_Benarsi
SD_MI_147_Harris
YH_MI_73_Cowasji_patel
FE_MI_99_Ewais
WG_MI_125_Cogshall
WE_MI_136_Manga_Blanca
13E_MI_110_Baileys_Marvel
9B_MI_152_Pruter
11G_MI_27_Pairi
JA_MI_21_Bombay
EA_KRB_4_M_gedebe
HD_KRP_12_M_sp
JB_FRIM_13_M_gedebe
XA_SBG_20_M_gedebe
17E_SKP_510_M_foetida
KA_CL_1_M_altissima
16G_SKP_59_M_sp
16H_PA_3_M_quadrifida
LD_PA_5_M_quadrifida
LH_PF_17_M_sp
NF_KRP_24_M_sp
MG_KRP_10_M_longipes
OA_Duval_01_M_pelipisan
14H_KRP_9_M_indica
SC_KRP_27_M_indica
13G_KRP_29_M_sp
BF_KRP1_M_foetida_pakel

0.0070

Figure 5.3a.

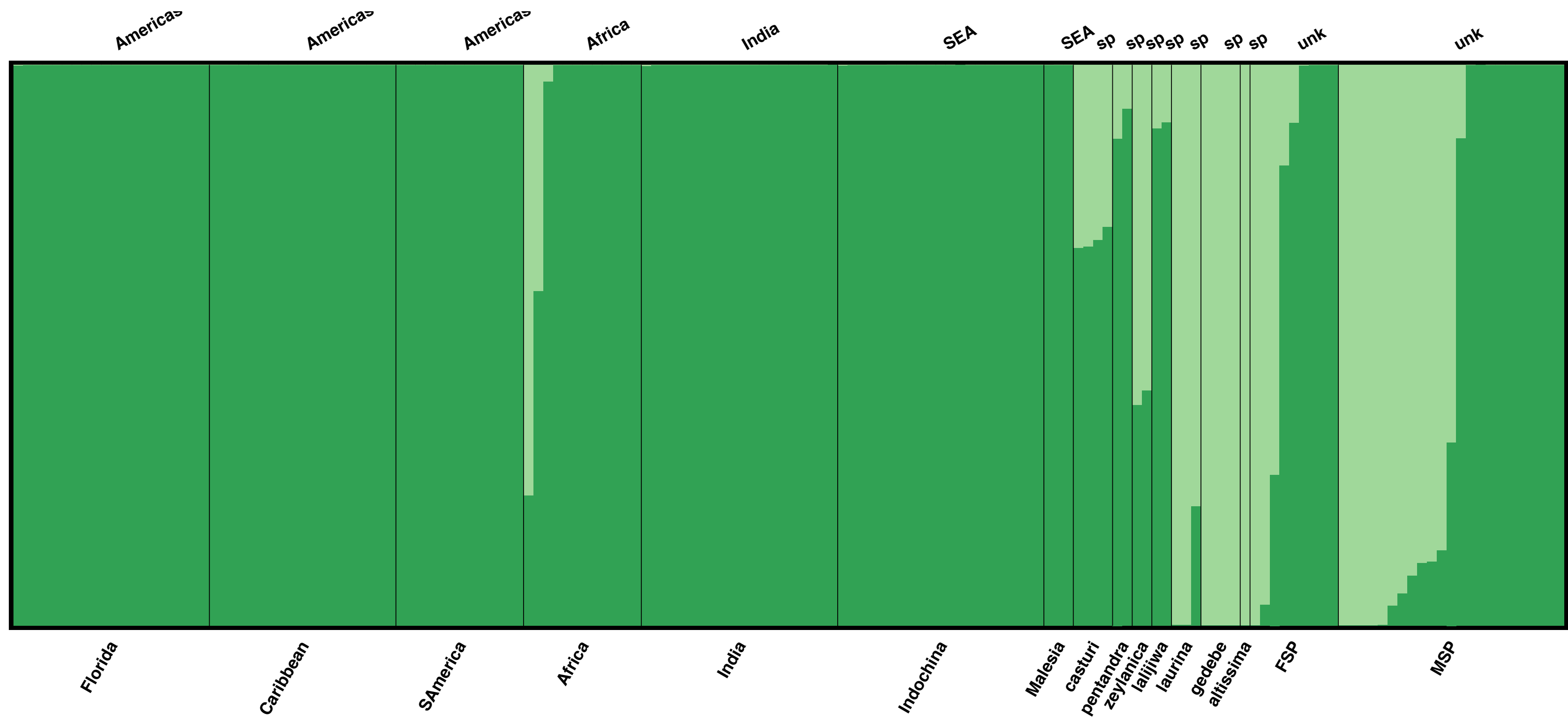


Figure 5.3b.

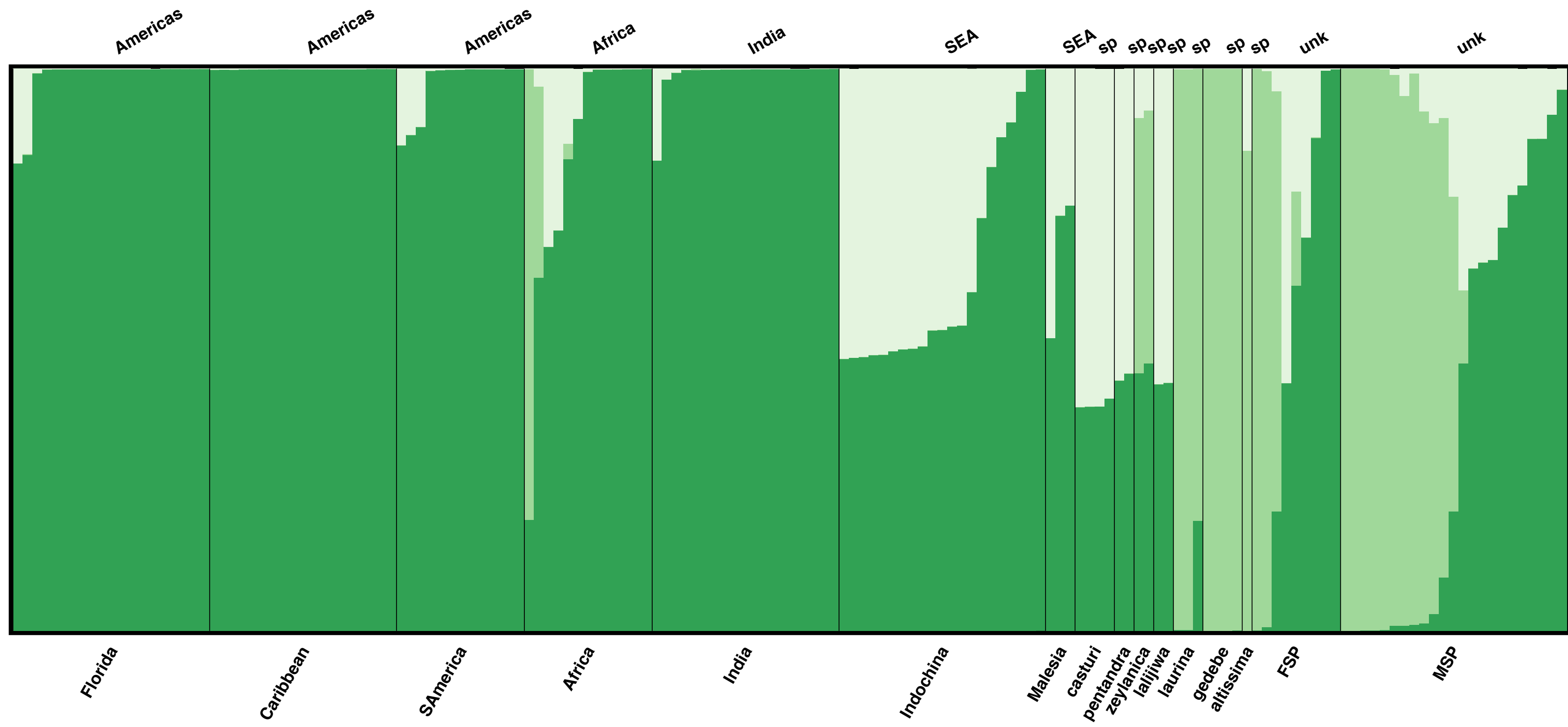


Figure 5.4.

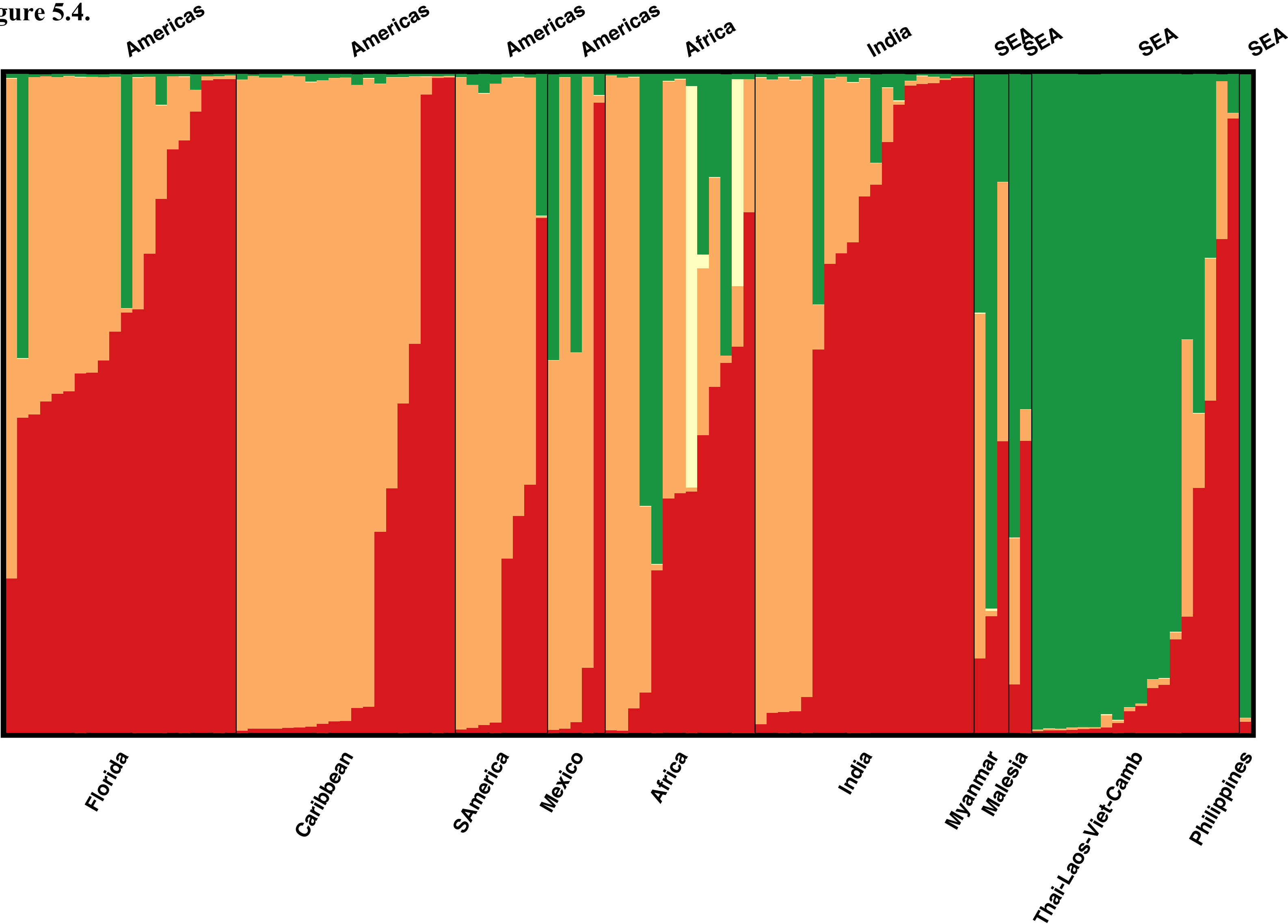
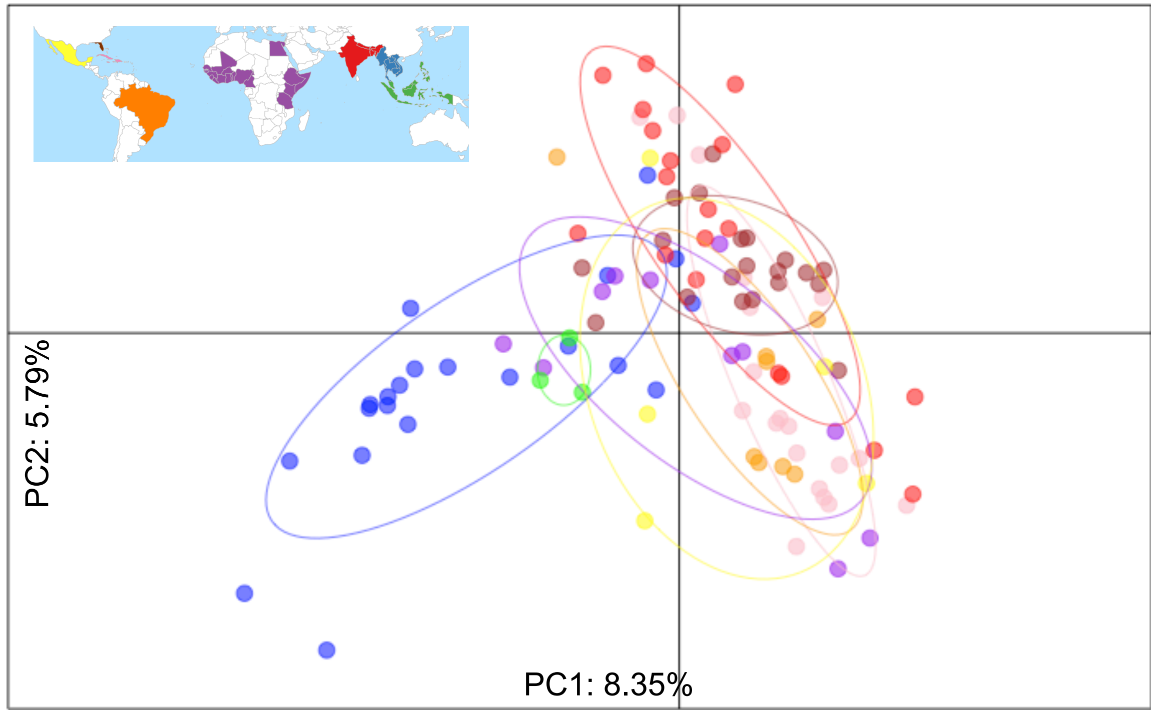


Figure 5.5.



Appendices Captions

Appendix 5.1. Information for the 158 samples included in this study. Sample ID consists of the individual ddRAD Sample ID, the study collection number and putative identification (cultivar or species name). The collection location (FTBG = Fairchild Tropical Botanic Garden, FTG = Fairchild Tropical Garden Herbarium, SBG = Singapore Botanic Garden, KRP = Purwodadi Botanic Garden, KRB = Bogor Botanic Garden, USDA = USDA Subtropical Horticulture Research Station at Chapman Field, FSP = Miami Dade Fruit and Spice Park, GBTB = Gardens by the Bay, FRIM = Forestry Research Institute of Malaysia, PA = Pasoh Forest Arboretum, PF = Pasoh Forest Research Station) and accession number within the respective collection are provided. The ddRAD library, sublibrary, and individual sample ID are given, as well as the number of raw reads for each individual. Each *M. indica* sample was classified within a Continent, Region, Population, and Subpopulation. Individuals of unknown provenance (Malesian species = MSP, Floridian species = FSP) or samples of *Mangifera* species were categorized accordingly. Nine individuals that were removed from the study as a result of low sequence coverage are not shown.

Appendix 5.2. Plot of ΔK metric as described by Evanno et al. (2005) for the full dataset *(FULL)*. The optimal number of ancestral populations is deemed to be that which has the highest value of ΔK.

Appendix 5.3. Plot of ΔK metric as described by Evanno et al. (2005) for 108 mango cultivars (*CULT*). The optimal number of ancestral populations is deemed to be that which has the highest value of ΔK.

Appendices

Appendix 5.1

Sample ID	Collected From	Accession Number	ddRAD Library	ddRAD Sublibrary	Raw Reads	Continent	Region	Population	Subpopulation
11A_MI_154_Madame_Francis	FTBG	2004-1064*A	1	11	942123	New	Americas	Caribbean	Caribbean
11B_MI_140_Azucar	FTBG	2005-1756*A	1	11	745859	New	Americas	Americas	SAmerica
11C_MI_84_Thai_Everbearing	FTBG	2004-1177*A	1	11	993848	Old	SEA	Indochina	Thai-Viet-Camb
11D_MI_22_Julie	FTBG	2003-1721*A	1	11	828538	New	Americas	Caribbean	Caribbean
11E_MI_160_Parvin	FTBG	2004-1072*A	1	11	986003	New	Americas	Florida	Florida
11F_MI_109_Malindi	FTBG	2007-1232*A	1	11	1014668	Old	Africa	Africa	Africa
11G_MI_27_Pairi	FTBG	2004-1081*A	1	11	1063491	Old	India	India	India
11H_MI_86_M_casturi_982186A	FTBG	982186*A	1	11	1335965	casturi	casturi	casturi	casturi
13B_SBG_4_M_pentandra	SBG	2012-2617*A	3	13	2100410	pentandra	pentandra	pentandra	pentandra
13E_MI_110_Baileys_Marvel	FTBG	2004-1074*A	3	13	2346160	New	Americas	Florida	Florida
13F_MI_81_Mallika	FTBG	2003-1722*A	3	13	1919662	Old	India	India	India
13G_KRP_29_M_sp	KRP	XVI.D.II.11	3	13	1824176	MSP	MSP	MSP	MSP
14C_KRP_31_M_indica_cv_Gandik_luyung	KRP	IX.B.24A	3	14	612705	MSP	MSP	MSP	MSP
14E_MI_10_Tommy_Atkins	FTBG	2003-1734*A	3	14	1588941	New	Americas	Florida	Florida
14H_KRP_9_M_indica	KRP	XVI.E.21	3	14	1504680	MSP	MSP	MSP	MSP
15B_MI_58_Kaddu_Ma_odorata	FTBG	2012-2376*A	3	15	1375976	zeylanica	zeylanica	zeylanica	zeylanica
15E_KRP_33_M_indica_cv_Madu	KRP	IX.B.14c	3	15	1237635	MSP	MSP	MSP	MSP
16A_MI_28_Number_11	FTBG	2004-1084*A	3	16	1770080	New	Americas	Caribbean	Caribbean
16C_MI_40_Cac	FTBG	2006-1295*A	3	16	1555717	Old	SEA	Indochina	Thai-Viet-Camb
16E_JFL_504_M_lalijiwa	Other	NA	3	16	1627664	lalijiwa	lalijiwa	lalijiwa	lalijiwa
16G_SKP_59_M_sp	Other	SKP 1044	3	16	1421892	MSP	MSP	MSP	MSP
16H_PA_3_M_quadrifida	PA	746	3	16	1443282	MSP	MSP	MSP	MSP

17C_MI_96_Long	FTBG	2004-1080*A	3	17	2044610	New	Americas	Caribbean	Caribbean
17E_SKP_510_M_foetida	Other	SKP 1097	3	17	1159927	MSP	MSP	MSP	MSP
17F_MI_50_Butterfly_Hainan	FTBG	2010-0391*A	3	17	2005251	FSP	FSP	FSP	FSP
3A_MI_37_Golek	FTBG	2006-1292*A	1	3	729910	Old	SEA	Indochina	Thai-Viet-Camb
3B_MI_117_Mabrouka	FTBG	2003-1713*A	1	3	561681	Old	Africa	Africa	Africa
3C_MI_20_Carabao	FTBG	2005-1791*A	1	3	480442	Old	SEA	Malesia	Philippines
3D_MI_35_Himsagar	FTBG	2006-1278*A	1	3	586464	Old	India	India	India
3E_MI_32_Gedong_Ginco	FTBG	2004-1215*A	1	3	440891	Old	SEA	Malesia	Indonesia
3F_MI_69_Langra_Benarsi	FTBG	2005-1945*A	1	3	798799	Old	Africa	Africa	Africa
3G_MI_129_Kent	FTBG	2003-1735*A	1	3	395813	New	Americas	Florida	Florida
3H_MI_119_Nelpetite	FTBG	2004-1199*A	1	3	617784	Old	Africa	Africa	Africa
6A_MI_102_Chok_Anon	FTBG	2006-1301*A	1	6	672634	Old	SEA	Indochina	Thai-Viet-Camb
6B_MI_158_Baptiste	FTBG	2006-1309*A	1	6	511517	New	Americas	Caribbean	Caribbean
6C_MI_11_Nam_Doc_Mai	FTBG	2003-1724*A	1	6	603488	Old	SEA	Indochina	Thai-Viet-Camb
6D_MI_104_Ataulfo	FTBG	2003-1710*A	1	6	625516	New	Americas	Mexico	Mexico
6E_MI_121_Fairchild	FTBG	2003-1719*A	1	6	1089530	New	Americas	Americas	SAmerica
6F_MI_148_Gilas	FTBG	2010-0366*A	1	6	737869	Old	India	India	India
6G_MI_68_Pohn_Sawadee	FTBG	2004-1088*A	1	6	767434	Old	SEA	Indochina	Thai-Viet-Camb
9A_MI_112_Mamita	FTBG	2004-1203*A	1	9	759933	New	Americas	Caribbean	Caribbean
9B_MI_152_Pruter	FTBG	2004-1221*A	1	9	605033	New	Americas	Florida	Florida
9C_MI_130_M_mempelam	FTBG	2012-2371*B	1	9	719950	laurina	laurina	laurina	laurina
9D_MI_29_Rosa	FTBG	2003-1711*A	1	9	765321	New	Americas	Americas	SAmerica
9E_MI_133_Banilejo	FTBG	2008-1292*A	1	9	919915	New	Americas	Caribbean	Caribbean
9F_MI_156_Depih_Pasir	FTBG	2012-2369*A	1	9	735616	FSP	FSP	FSP	FSP
9G_MI_120_Pettigrew	FTBG	2005-1766*A	1	9	958812	New	Americas	Florida	Florida
9H_MI_34_Rumanii	FTBG	2005-1742*A	1	9	1189430	Old	India	India	India
AA_FS2_M_zeylanica	FSP	NA	2	A	1389807	zeylanica	zeylanica	zeylanica	zeylanica

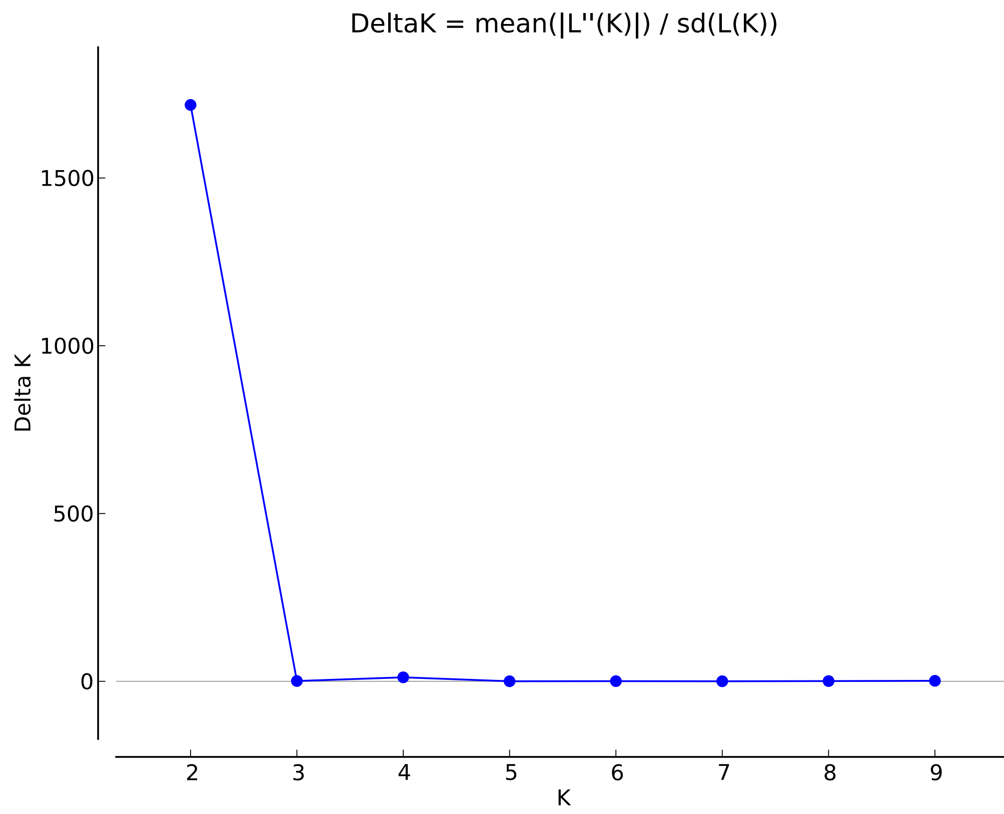
AE_GBTB3_M_indica	GBTB	NA	2	A	1619145	MSP	MSP	MSP	MSP
BA_FS1_M_odorata	FSP	NA	2	B	987343	FSP	FSP	FSP	FSP
BB_K3_M_indica	FRIM	NA	2	B	814416	MSP	MSP	MSP	MSP
BF_KRP1_M_foetida_pakel	KRP	IX.B.17	2	B	684994	MSP	MSP	MSP	MSP
CB_MI_91_Myatrynat	FTBG	2006-1293*A	2	C	631319	Old	SEA	Indochina	Myanmar
CC_MI_106_Esmeralda	FTBG	2003-1747*A	2	C	525923	New	Americas	Mexico	Mexico
CD_KRP_6_M_foetida_cv_Pakel_ lumut	KRP	IX.C.9b	2	C	514920	MSP	MSP	MSP	MSP
CE_MI_25_Pascual	FTBG	2003-1743*A	2	C	757128	New	Americas	Caribbean	Caribbean
CF_MI_42_Swethintha	FTBG	2004-1211*A	2	C	537252	Old	SEA	Indochina	Myanmar
CG_MI_77_Amrapali	FTBG	2012-2391*A	2	C	599107	Old	India	India	India
DA_MI_114_Zebda	FTBG	2005-1788*A	2	D	887852	Old	Africa	Africa	Africa
DB_MI_24_Turpentine	FTBG	2003-1738*A	2	D	799605	New	Americas	Caribbean	Caribbean
DD_MI_97_Diab	FTBG	2004-1061*A	2	D	812452	Old	Africa	Africa	Africa
DE_MI_113_Martin	FTBG	2005-1782*A	2	D	1043022	New	Americas	Florida	Florida
DF_MI_80_Nam_Tam_Teem	FTBG	2004-1228*A	2	D	702325	Old	SEA	Indochina	Thai-Viet-Camb
DG_MI_105_Manilita	FTBG	2004-1054*A	2	D	792009	New	Americas	Mexico	Mexico
DH_MI_107_Mesk	FTBG	2003-1715*A	2	D	754290	Old	Africa	Africa	Africa
EA_KRB_4_M_gedebe	KRB	VI.D.5	2	E	1210992	gedebe	gedebe	gedebe	gedebe
EB_MI_74_Pam_kai_mia	FTBG	2012-2402*A	2	E	1315873	Old	SEA	Indochina	Thai-Viet-Camb
EC_MI_44_M_aquea	FTBG	NA	2	E	1534535	laurina	laurina	laurina	laurina
EE_MI_26_Espada	FTBG	2004-1086*A	2	E	1826917	New	Americas	Americas	SAmerica
EF_MI_95_Cairo	FTBG	2004-1219*A	2	E	1072510	Old	Africa	Africa	Africa
EG_MI_85_M_laurina_1_B4	FTBG	2013-0555*A	2	E	1466982	FSP	FSP	FSP	FSP
EH_MI_17_East_Indian	FTBG	2004-1073*A	2	E	1550385	New	Americas	Caribbean	Caribbean
FA_MI_76_Chao_savoy	FTBG	2003-1712*A	2	F	1760475	Old	SEA	Indochina	Thai-Viet-Camb
FB_MI_111_Chene	FTBG	2003-1736*A	2	F	1728268	Old	Africa	Africa	Africa
FC_MI_13_Biscochuelo	FTBG	2004-1193*A	2	F	1492636	New	Americas	Caribbean	Caribbean
FD_KRP_28_M_casturi	KRP	XVI.D.II.14	2	F	1650119	casturi	casturi	casturi	casturi

FE_MI_99_Ewais	FTBG	2004-1220*A	2	F	2265290	Old	Africa	Africa	Africa
FF_MI51_Royal_Special	FTBG	2004-1087*A	2	F	1311009	Old	India	India	India
FG_MI_53_Sindhri	FTBG	2004-1055*A	2	F	1876470	Old	India	India	India
FH_MI_83_Tenom	FTBG	2012-2407*A	2	F	1674109	FSP	FSP	FSP	FSP
GB_KRP_11_M_sp	KRP	XVI.E.22	2	G	1107713	MSP	MSP	MSP	MSP
GC_MI_23_Lancetilla	FTBG	2003-1726*A	2	G	1221555	New	Americas	Americas	SAmerica
GE_MI_6_Keitt	FTBG	2003-1732*A	2	G	1715767	New	Americas	Florida	Florida
GF_MI_64_Poh_Pakel	FTBG	2010-0397*A	2	G	928187	FSP	FSP	FSP	FSP
GG_MI_93_Valencia_Pride	FTBG	2004-1243*A	2	G	971353	New	Americas	Florida	Florida
GH_MI_57_Jumbo_Kesar	FTBG	2009-0822*A	2	G	1219797	Old	India	India	India
HA_MI_88_Pancahdarakalasa	FTBG	2004-1181*A	2	H	1561762	Old	India	India	India
HB_MI_163_Sabre	USDA	NA	2	H	1421046	Old	Africa	Africa	Africa
HC_MI_63_Dusheri	FTBG	2005-1765*A	2	H	1599887	Old	India	India	India
HD_KRP_12_M_sp	KRP	XVI.E.48	2	H	1498756	gedebe	gedebe	gedebe	gedebe
HE_MI_39_Pyu_Pyu_Kalay	FTBG	2009-0816*A	2	H	1942535	Old	SEA	Indochina	Myanmar
HF_MI_79_Aslul_Mukararara	FTBG	2008-1289*A	2	H	1222409	Old	India	India	India
HG_MI_162_Corazon	FTBG	2012-2385*A	2	H	1742916	New	Americas	Caribbean	Caribbean
HH_MI_88_M_pentandra	FTBG	2012-2368*A	2	H	1712812	FSP	FSP	FSP	FSP
IA_MI_12_Tong_Dam	FTBG	2003-1707*A	2	I	1032280	Old	SEA	Indochina	Thai-Viet-Camb
IC_MI_59_M_lalijiwa_G	FTBG	2004-1213*A	2	I	1167782	lalijiwa	lalijiwa	lalijiwa	lalijiwa
ID_MI_14_Alphonso	FTBG	2004-1053*A	2	I	1518263	Old	India	India	India
IE_MI_87_Depih_Biasa	FTBG	2012-2373*A	2	I	1159995	FSP	FSP	FSP	FSP
IF_MI_43_Saigon	FTBG	2006-1276*A	2	I	832868	Old	SEA	Indochina	Thai-Viet-Camb
IG_MI_4_Phimsen_Mun	FTBG	2003-1753*A	2	I	990893	Old	SEA	Indochina	Thai-Viet-Camb
IH_MI_36_Alampur_Baneshan	FTBG	2010-0370*A	2	I	1092326	Old	India	India	India
JA_MI_21_Bombay	FTBG	2005-1789*A	2	J	1073891	New	Americas	Caribbean	Caribbean
JB_FRIM_13_M_gedebe	FRIM	33020263	2	J	831842	gedebe	gedebe	gedebe	gedebe
JD_MI_3_Cambodiana	FTBG	2005-1753*A	2	J	1023296	Old	SEA	Indochina	Thai-Viet-Camb

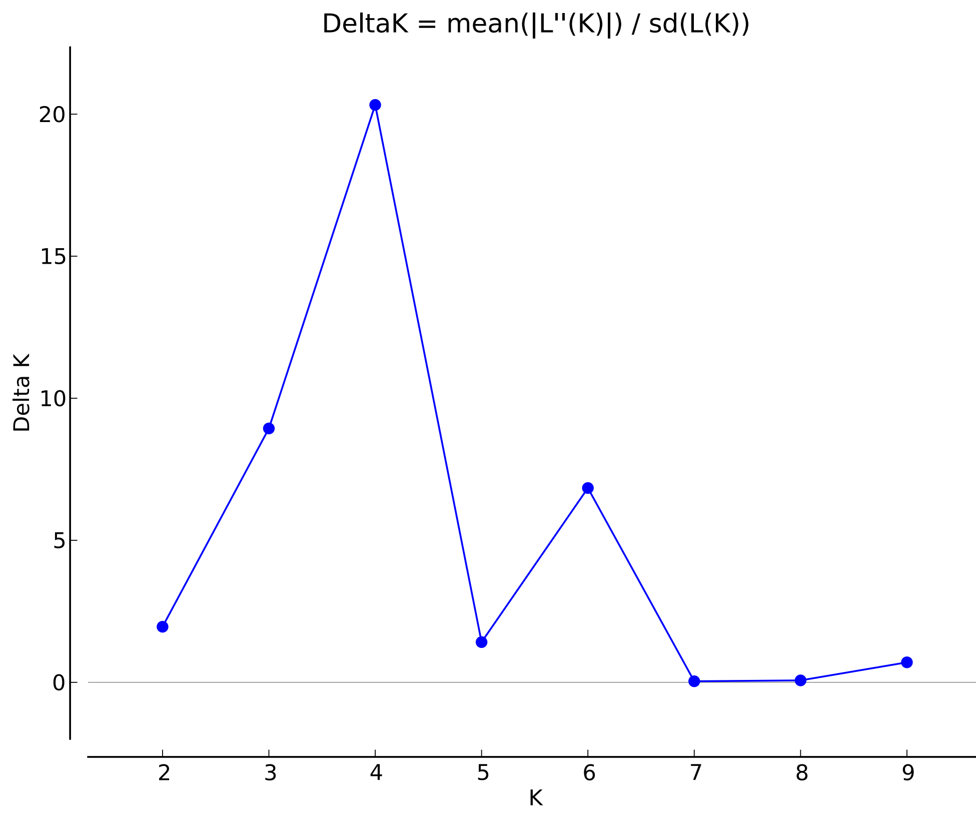
JE_MI_61_Praya_Savoy	FTBG	2006-1289*A	2	J	1142384	Old	SEA	Indochina	Thai-Viet-Camb
JF_MI_116_Tyler_Premiere	FTBG	2004-1218*A	2	J	754210	Old	Africa	Africa	Africa
JG_MI_54_Sig_Siput	FTBG	2005-1746*A	2	J	893846	Old	SEA	Indochina	Thai-Viet-Camb
KA_CL_1_M_altissima	Other	NA	2	K	1311074	altissima	altissima	altissima	altissima
KC_MI_131_Oro	FTBG	2004-1075*A	2	K	1441712	New	Americas	Mexico	Mexico
KF_MI_149_Toledo	FTBG	2004-1236*A	2	K	1190421	New	Americas	Caribbean	Caribbean
LA_MI_145_Joellen	FTBG	2004-1240*A	2	L	1141491	New	Americas	Florida	Florida
LB_KRP_32_M_indica_kepodang	KRP	IX.B.18A	2	L	972685	MSP	MSP	MSP	MSP
LD_PA_5_M_quadrifida	PA	134	2	L	997042	MSP	MSP	MSP	MSP
LH_PF_17_M_sp	PF	NA	2	L	511627	MSP	MSP	MSP	MSP
ME_MI_155_Manila	FTBG	2005-1787*A	3	M	2352090	New	Americas	Mexico	Mexico
MG_KRP_10_M_longipes	KRP	XVI.E.24	3	M	1736083	MSP	MSP	MSP	MSP
NA_MI_56_Hindi_Besanara	FTBG	2004-1233*A	3	N	1511170	Old	India	India	India
NF_KRP_24_M_sp	KRP	XVI.D.II.2	3	N	1145006	MSP	MSP	MSP	MSP
NG_MI_144_Diplomatico	FTBG	2004-1189*A	3	N	1295275	New	Americas	Americas	SAmerica
OA_Duval_01_M_pelipisan	Other	NA	3	O	1153813	FSP	FSP	FSP	FSP
OB_MI_126_Irwin	FTBG	2005-1783*A	3	O	1277059	New	Americas	Florida	Florida
OH_MI_137_Peach	FTBG	2006-1298*A	3	O	1502683	New	Americas	Caribbean	Caribbean
PA_MI_89_Aeromanis	FTBG	2005-1740*A	3	P	1549107	Old	SEA	Malesia	Indonesia
PG_MI_138_San_Felipe	FTBG	2004-1184*A	3	P	1548724	New	Americas	Caribbean	Caribbean
QA_MI_18_Glenn	FTBG	2003-1749*A	3	Q	1461121	New	Americas	Florida	Florida
QB_KRB_8_M_griffithii	KRB	VII.E.170A	3	Q	1223203	casturi	casturi	casturi	casturi
QD_MI_9_Totapuri	FTBG	2004-1222*A	3	Q	1574796	Old	India	India	India
QG_MI_1_Prieto	FTBG	2004-1221*A	3	Q	1469444	New	Americas	Caribbean	Caribbean
RA_MI_30_Extrema	FTBG	2005-1747*A	3	R	1520233	New	Americas	Americas	SAmerica
RE_MI_31_Kaeo_Luemkon	FTBG	2007-1056*A	3	R	1396821	Old	SEA	Indochina	Thai-Viet-Camb
RH_MI_15_Ivory	FTBG	2003-1723*A	3	R	1658796	Old	SEA	Indochina	Thai-Viet-Camb
SB_MI_78_M_A_carle	FTBG	2007-1060*A	3	S	1612855	casturi	casturi	casturi	casturi

SC_KRP_27_M_indica	KRP	XVI.D.II.8	3	S	1594169	MSP	MSP	MSP	MSP
SD_MI_147_Harris	FTBG	2005-1739*A	3	S	1497220	New	Americas	Florida	Florida
TB_KRP_2_M_minor	KRP	IX.B.32	1	T	1587805	MSP	MSP	MSP	MSP
TD_KRP_22_M_sp	KRP	XVI.F.I.19a	1	T	1830319	MSP	MSP	MSP	MSP
WA_MI_45_M_PR_Martex	FTBG	2012-2413*A	1	W	1917441	laurina	laurina	laurina	laurina
WB_MI_98_Vallenato	FTBG	2004-1225*A	1	W	1521493	New	Americas	Americas	SAmerica
WC_MI_5_Graham	FTBG	2004-1065*A	1	W	1320415	New	Americas	Caribbean	Caribbean
WD_MI_132_Torbet	FTBG	2004-1059*A	1	W	863785	New	Americas	Florida	Florida
WE_MI_136_Manga_Blanca	FTBG	2012-2400*A	1	W	2199792	New	Americas	Caribbean	Caribbean
WF_MI_52_Maha_Chanok	FTBG	2012-2399*A	1	W	2338767	Old	SEA	Indochina	Thai-Viet-Camb
WG_MI_125_Cogshall	FTBG	2003-1720*A	1	W	598447	New	Americas	Florida	Florida
WH_MI_135_Alice	FTBG	2006-1308*A	1	W	932156	New	Americas	Florida	Florida
XA_SBG_20_M_gedebe	SBG	NA	1	X	1489679	gedebe	gedebe	gedebe	gedebe
XF_FRIM_3_M_indica	FRIM	NA	1	X	1676509	MSP	MSP	MSP	MSP
XH_SBG_31_M_odorata	SBG	NA	1	X	1514572	MSP	MSP	MSP	MSP
YA_MI_124_Smith	FTBG	2005-1764*A	1	Y	3870665	New	Americas	Florida	Florida
YB_MI_48_Imam_Pasand	FTBG	2004-1187*A	1	Y	1222060	Old	India	India	India
YC_MI_65_Ratna	FTBG	2012-2403*B	1	Y	1833551	Old	India	India	India
YD_MI_115_Lippens	FTBG	2003-1709*A	1	Y	2700472	New	Americas	Florida	Florida
YE_MI_128_Sensation	FTBG	2005-1743*A	1	Y	2184330	New	Americas	Florida	Florida
YF_MI_122_Hodson	FTBG	2004-1078*A	1	Y	2025757	New	Americas	Florida	Florida
YG_MI_146_Piva	FTBG	2003-1729*A	1	Y	1891232	Old	Africa	Africa	Africa
YH_MI_73_Cowasji_patel	FTBG	2004-1079*A	1	Y	3144639	Old	India	India	India
ZG_SBG_26_M_pentandra	SBG	2011-7045*A	1	Z	823561	pentandra	pentandra	pentandra	pentandra
ZH_SBG_1_M_oblongifolia	SBG	NA	1	Z	985203	MSP	MSP	MSP	MSP

Appendix 5.2.



Appendix 5.3.



CHAPTER VI
INTERSPECIFIC HYBRIDIZATION IN *MANGIFERA*

Abstract

Premise: Homoploid hybridization is known to play an important role in the evolution of plants, including many crop species, but can have different outcomes including introgression between parental taxa and the formation of new evolutionary lineages. We investigate the occurrence and consequences of hybridization between the economically important tree crop *Mangifera indica* (mango) and two congeneric species in Southeast Asia.

Methods: A total of 90 samples of the hybrid *M. odorata* and its parental taxa, *M. indica* (mango) and *M. foetida*, along with 65 samples of a newly proposed hybrid, *M. casturi* and its putative parental taxa, *M. indica* and *M. quadrifida,* were sampled and genotyped using restriction site associated DNA sequencing. For each hybrid, we assessed population structure and admixture and indices of genetic diversity, including multilocus linkage disequilibrium.

Key Results: We found no evidence of introgression between *M. foetida* and *M. indica* cultivars from Southeast Asia, but find support for a hybrid origin of *M. casturi.* Both hybrids show low levels of intraspecific genetic diversity and individuals have high genetic identity and significant multilocus linkage disequilibrium.

Conclusions: For both *M. odorata* and *M. casturi,* our results are consistent with hybrid lineages that have formed only a few times and have since been maintained clonally. While grafting may play a role in the continued propagation of these hybrids, we suggest that the ability of *M. odorata* and *M. casturi* to reproduce asexually through nucellar polyembryony has allowed the hybrids to persist independently of grafting.

Introduction

Hybridization has long been considered a creative force in the evolutionary process and is known to have played a particularly important role in plant evolution and speciation (e.g., Stebbins, 1950; Anderson and Stebbins, 1954; Arnold, 1992; Rieseberg, 1997; Mallet, 2007; Soltis and Soltis, 2009; Abbott et al., 2013; Yakimowski and Rieseberg, 2014). Broadly defined as the interbreeding of individuals from different genetic populations, hybridization occurs via multiple mechanisms and results in varied outcomes (e.g., Barton and G.M., 1985; Mallet, 2007; Soltis and Soltis, 2009; Abbott et al., 2013). In plants, interspecific hybridization can occur with or without whole genome duplication, resulting in homoploid or allopolyploid hybrids, respectively. Typically, allopolyploids are reproductively isolated from their parental taxa as a result of chromosomal incompatibilities, and are therefore generally considered cases of saltational evolution (Mallet, 2007; Soltis and Soltis, 2009). However, the consequences of homoploid hybridization are more complex. In some cases, low fitness of homoploid hybrids is thought to contribute to reproductive isolation of parental taxa (Servedio and Noor, 2003), while in other cases, hybrid intermediates may provide a bridge for introgression (gene flow) between parental taxa (e.g., Anderson, 1949; Arnold, 1992; Arnold et al., 2012). Alternatively, homoploid hybrids may form lineages (or species) that are ecologically distinct and which are or may become reproductively isolated from parental species (Soltis and Soltis, 2009; Yakimowski and Rieseberg, 2014).

Although hybridization is a widespread phenomenon in plants, it is known to be more common in certain circumstances. For instance, long-lived plants, which tend to be self-incompatible, are more likely to hybridize than annuals (Ellstrand et al., 1996; Petit

and Hampe, 2006; Miller and Gross, 2011). Closely related allopatric species that are brought into secondary contact either by natural dissolution of geographic barriers or by human-mediated introductions are also more likely to hybridize, as they may not have developed reproductive isolation (Coyne and Orr, 2004; Harrison and Larson, 2014). Therefore, it comes as no surprise that many domesticated plants, especially perennials, are of confirmed or putative hybrid ancestry (Warschefsky et al., 2014; Chapter II, Appendix 2.1), including apple (*Malus* x*domestica,* Coart et al., 2006; Cornille et al., 2012), potato (*Solanum tuberosum,* Rodríguez et al., 2010), banana (*Musa acuminata,* De Langhe et al., 2010), and many citrus species (*Citrus* spp. Wu et al., 2018). Such hybrid crops may be the inadvertent result of cultivation in proximity to congeneric species or the outcome of intentional crosses.

The genus *Mangifera* includes approximately 69 tropical tree species, all of which are native to South and Southeast Asia (Kostermans and Bompard, 1993). The most well known species of *Mangifera, M. indica* (mango), was likely domesticated around 4,000 years ago in India (Mukherjee, 1949). The first introduction of *M. indica* outside its original center of domestication likely occurred during the 4th and 5th centuries, when Buddhist monks from India traveled to Southeast Asia (Mukherjee, 1949). There, it came into contact with approximately 35 other species of *Mangifera* native to the region, many of which are edible and a few of which are regionally cultivated in small orchards and backyard gardens (Kostermans and Bompard, 1993). Though only *M. indica* has been investigated, *Mangifera* species are assumed to be self-incompatible and therefore obligately outcrossing (Kostermans and Bompard, 1993; Mukherjee and Litz, 2009). It follows that the cultivation of such outcrossing congers in close proximity to one another

would lead to interspecific hybridization, a phenomenon that has been documented in the genus (Kostermans and Bompard, 1993; Teo et al., 2002; Bompard, 2009).

Today, *M. indica* is cultivated in subtropical and tropical regions around the world and is one of the most economically important tropical fruit species (FAO, 2003; FAOSTAT, 2013). On the basis of fruit morphology, cultivars of *M. indica* are separated into two types, Indian and Southeast Asian (Crane and Campbell 1994). Molecular research demonstrates that these cultivar types form distinct genetic groups (Schnell et al., 2005; Dillon et al., 2013), and that Southeast Asian cultivars contain genetic diversity that is not present in Indian populations (Chapter IV). However, potential causes of the genetic and morphological differentiation observed in *M. indica* have not been explored. While one theory is that Indian and Southeast Asian cultivar types are the result of independent domestication events (Bompard, 2009), an alternative explanation is that the novel diversity observed in Southeast Asian *M. indica* was introduced by introgressive hybridization with a congeneric species.

The kuwini mango, *M. odorata*, was first described by Griffith in 1854 (Griffith, 1854). Only known from cultivation in Southeast Asia (Kostermans and Bompard, 1993; Teo et al., 2002), *M. odorata* had long been thought to be a hybrid of *M. indica* and *M. foetida* (horse mango), another cultivated species, on the basis of morphological characters (Hou, 1978). In 2002, Teo et al. confirmed the hybrid status of *M. odorata* using amplified fragment length polymorphisms (AFLPs) and found that *M. odorata* has greater genetic affinity to *M. foetida* than to *M. indica*. The species was later given taxonomic status as a hybrid, *M.* x*odorata* Griff, (pro sp.) (Kiew, 2002). However, no further research has explored the dynamics of hybridization between *M. indica* and *M.*

odorata or tested whether the novel diversity observed in Southeast Asian cultivars of *M. indica* is the result of introgression from *M. foetida.*

Mangifera casturi, (kastooree mango) was recently described by Kostermans and Bompard (1993). Only known from cultivation in Kalimantan, the Indonesian region of Borneo, *M. casturi* is classified as extinct in the wild (IUCN, 2012). The species' characteristic dark purple to black fruit and orange flesh led the authors to suggest it was most closely related to another Bornean species with similar fruit coloration, *M. quadrifida.* However, recent phylogenetic analysis indicates that despite its fruit morphology, *M. casturi* is not closely allied to *M. quadrifida,* but is instead a close relative of *M. indica* (Chapter IV, Fig. 4.4). Considering this new phylogenetic information, it is feasible that *M. casturi* may be a hybrid of *M. indica* and *M. quadrifida,* an idea not previously proposed.

Here, we explore the occurrence and consequences of hybridization in two separate instances in *Mangifera*. Analyzing SNP markers obtained by RAD sequencing, we: 1) determine whether the novel genetic diversity observed in Southeast Asian *M. indica* cultivars can be attributed to introgression from *M. foetida* via *M. odorata*; 2) look for evidence of a hybrid origin of *M. casturi*; and 3) characterize the genetic diversity and population structure of *M. odorata* and *M. casturi*.

Methods

Sampling

Leaf samples were collected from 280 specimens in living collections in Singapore, Indonesia, Malaysia, and the United States. Fresh leaves were stored at -80ºC

or dried in silica and stored at 4ºC. Genomic DNA was extracted using a Qiagen DNEasy mini extraction kit or by a modified CTAB protocol (Doyle and Doyle, 1990). Samples were analyzed as two independent datasets: 90 samples were analyzed to examine hybridization in *M. odorata* (Table 6.1), and 65 individuals were analyzed to examine hybridization in *M. casturi* (Table 6.2). Samples not analyzed in this chapter were analyzed for Chapters IV and V.

Library preparation

Double-digest restriction site-associated DNA sequencing (ddRADseq) libraries were prepared according to the protocol of Peterson et al. (2012). Briefly, 300-1000 ng genomic DNA was digested with NlaIII and MluCI before ligating on custom-designed adapters that contained one of eight unique barcode sequences (Chapter IV, Appendix 4.2). Groups of eight individuals with different barcode sequences were pooled together to form a sublibrary prior to size selection using a Pippin Prep (Sage Science) that targeted 350 bp inserts (tight size selection at 425 bp using an external marker). To amplify sublibraries and add a unique PCR index (Chapter IV, Appendix 4.3), short cycle PCR was performed, using six separate reactions to avoid PCR bias. The quality of amplified sublibraries was checked on an Agilent Bioanalyzer (DNA high sensitivity chip). For any sublibraries where overamplification occurred, size selection on Pippin Prep was performed again to remove non-target DNA. Following size selection, 12 sublibraries consisting of 96 individuals total were pooled in equimolar amounts. Each of the three libraries was sequenced at the University of Southern California's Genome and Cytometry Core in a single lane of rapid run on an Illumina HiSeq 2500.

Bioinformatics

Raw fastq files were quality checked using the software FASTQC v.0.11.4 (Andrews, 2010). The ipyRAD bioinformatic pipeline (Eaton, 2014) was used to analyze raw reads under default parameters except for the following: a single mismatched base was allowed within the barcode sequence, adapter filtering was set to option 2, maximum read depth for loci was set to 1000. The 280 individuals that were sequenced were subset into the *M. odorata* and *M. casturi* datasets (90 and 65 individuals, respectively) in the ipyRAD workflow (remaining samples were analyzed in Chapters IV and V). Clustering within and between individuals was performed at 85% identity. For this study, the minimum number of individuals required to have data at a given locus was set to filter out loci with >10% missing data for the *M. odorata* dataset and >50% missing data for the *M. casturi* dataset.

Population structure and admixture

Population structure and admixture of the *M. odorata* and *M. casturi* datasets were analyzed independently using the Bayesian K-means program STRUCTURE v. 2.3.4 (Pritchard et al., 2000; Falush et al., 2003; Hubisz et al., 2009). First, an average value of lambda for K=1 was estimated across 10 runs of 100,000 iterations with 10,000 steps of burn-in. Population structure was estimated for K of 1 to 10, using the estimated value of lambda. For each value of K, 10 runs of 100,000 iterations with 10,000 steps of burn-in were completed. The optimal value of K was determined according to the deltaK parameter of Evanno et al. (2005) in the program STRUCTURE HARVESTER v. 0.6.94 (Earl and vonHoldt, 2012). For each value of K, the program CLUMPP v. 1.1.2

(Jakobsson and Rosenberg, 2007) was used to summarize results across the 10 runs performed using the greedy option (M=2) for K=1-7 or the LargeKGreedy option (M=3) for K=8na -10 with G' similarity and 1,000 random permutations. Summarized results were visualized using the program DISTRUCT v. 1.1 (Rosenberg, 2004).

Population structure was also assessed independently for the *M. odorata* and *M. casturi* datasets using principal component analysis (PCA) in the R package adegenet (Jombart, 2008; Jombart and Ahmed, 2011). For the *M. odorata* dataset, two individuals identified as outliers were removed from PCA analysis and downstream analyses, resulting in a dataset of 88 individuals. The R package INTROGRESS (Gompert and Buerkle, 2010) was used to calculate hybrid indices (Buerkle, 2005) and interspecific heterozygosity for the *M. odorata* and *M. casturi* datasets. Additionally, a neighbor network analysis was performed on the *M. odorata* and *M. casturi* datasets using the program SPLITSTREE (Huson and Bryant, 2006) under default parameters.

Population Genetics

Multiple indices of genetic diversity were calculated for hybrid and parental species in the *M. odorata* and *M. casturi* datasets. Percent polymorphism, observed heterozygosity, and gene diversity were calculated using the R package adegenet (Jombart, 2008; Jombart and Ahmed, 2011). Allelic richness and the number of private alleles (SNPs) in each species were calculated in the R packages PopGenReport (Adamack and Gruber, 2014) and poppr (Kamvar et al., 2014), respectively. The number of private alleles was also calculated as a percent of the total number of SNP loci analyzed within each dataset. For *M. odorata* and *M. casturi,* the index *rbarD* (Agapow

and Burt, 2001), a standardized form the index of association (Brown et al., 1980) that assesses multilocus linkage disequilibrium, was calculated in poppr using 999 permutations to test for significance.

Results

Sequencing Results

For the entire 280 individuals sequenced across three lanes, we obtained 454,840,461 raw reads, with 116,110,642 for the 90 individuals included in the *M. odorata* dataset (average: 1,290.118, SD: 558,342) and 80,460,503 for the 65 individuals in the *M. casturi* dataset (average: 1,237,854, SD: 550,028). FastQC results indicated reads were of high quality across the entire 150 bp length. After filtering for missing data, we recovered 481 unlinked SNPs from the *M. odorata* dataset and 3,609 unlinked SNPs from the *M. casturi* dataset.

Mangifera odorata

For the *M. odorata* dataset, analysis of population structure in the Bayesian software STRUCTURE found $K = 2$ to be the optimal number of populations according to the ΔK method of Evanno (Fig. 6.1, Appendix 6.1). All samples of the parental species *M. foetida* were assigned to a single population, with high levels of identity (99.1-100.0%). The Indian and Southeast Asian populations of *M. indica* were assigned to a second population, with individual identity ranging from 99.9-100%. No differentiation was observed between Indian and Southeast Asian populations for $K = 2$. All 17 individuals of *M. odorata* showed high levels of admixture, with 58.73-74.01% of

ancestry assigned to the *M. foetida* population and 25.99-41.27% of ancestry assigned to the *M. indica* population. Because the ΔK method of Evanno is known to identify only the highest existing level of population structure, larger values of K were examined, but were not found to show any additional meaningful structure.

Analysis of the *M. odorata* dataset using principal components shows three clusters, which correspond to *M. indica, M. odorata,* and *M. foetida* (Fig. 6.2). The first principal component (PC1) accounts for 37.68% of the variation observed in the data and separates the three species, with *M. odorata* recovered as intermediate to *M. foetida* and *M. indica*. The second principal component (PC2) accounts for 6.32% of the variation present in the dataset and further distinguishes *M. odorata* from *M. foetida* while also stratifying Southeast Asian and Indian populations of *M. indica*.

Of the 17 *M. odorata* individuals examined, 16 have a very similar hybrid index (0.32-0.36, where 0 is *M. foetida* and 1 is *M. indica*) and similar values of interspecific heterozygosity (23.29-25.51%) (Fig. 6.3). The remaining individual is genetically distinct, with an estimated hybrid index of 0.44 and interspecific heterozygosity of 17.86%. Network analysis of the *M. odorata* dataset places all samples of *M. odorata* directly on the branch separating *M. indica* and *M. foetida,* though the branch length between *M. odorata* and *M. foetida* is shorter than that between *M. odorata* and *M. indica* (Fig. 6.4).

Indices of genetic diversity for the *M. odorata* dataset (Table 6.3) show that the hybrid individuals have higher levels of observed heterozygosity (H_O = 0.3005) than *M. indica* (H_O = 0.1072) and *M. foetida* (H_O = 0. 0487). Gene diversity is highest in *M. odorata* (H_S = 0.1638), followed by *M. indica* (H_S = 0.1186) and *M. foetida* (H_S = 0.0644).

A negative inbreeding coefficient for *M. odorata* (F_{IS} = -0.6230) indicates an excess of outcrossing, while *M. indica* and *M. foetida* have values indicating slight inbreeding (F_{IS} = 0.0840 and 0.3115, respectively). *Mangifera indica* has the lowest percent polymorphism (25.32%), while *M. foetida* has the highest (57.29%) and intermediate levels are seen in *M. odorata* (43.22%). Similar values of nucleotide diversity are observed in *M. indica* and *M. foetida* (0.0309 and 0.0246, respectively) while *M. odorata* has very low nucleotide diversity (0.0016). For *M. odorata,* values of *rbarD*, a measure of multilocus linkage disequilibrium, are significantly different from zero (*rbarD* = 0.3984, p = 0.001, Appendix 6.2). The number and percent of private alleles across all loci is highest for *M. indica* (173, 44.25%), lower in *M. foetida* (46, 11.76%), and very low in *M. odorata* (1, 0.26%).

Mangifera casturi

Analysis of the *M. casturi* dataset in the program STRUCTURE found K = 3, 5, and 8 to be equally optimal according to the ΔK method of Evanno (Fig. 6.5, Appendix 6.3). Patterns of genetic structure are generally similar across the three values of K. For K = 3, *M. indica* samples are assigned high identity to a single population (78.38-84.63%) with moderate identity from a secondary population (13.43-16.92%) and low levels of identity from a third population (3.06-5.47%). Samples from *M. quadrifida* show a distinct pattern, with highest identity coming from the third population (50.82-60.44%), and moderate levels from both the second (18.00-21.50%) and first (18.05-30.42%) populations. The samples of *M. casturi* are intermediate to the putative parental populations: assignment to the first population ranges from 58.02-59.77%, with

assignment to the second population at 18.57-19.83% and assignment to the third population at 21.63-22.39%. As the number of populations increases from three to five to eight, the population assignments for *M. indica* and *M. quadrifida* become more distinct, while *M. casturi* continues to have assignments that are intermediate to the putative parental species and shows no unique ancestry.

For the *M. casturi* dataset, principal component analysis recovers four clusters of individuals, with one cluster corresponding to *M. indica,* a second corresponding to *M. casturi*, and two clusters representing *M. quadrifida* (Fig. 6.6). The first principal component accounts for 27.20% of the variation in the dataset and separates *M. casturi* and *M. quadrifida* from *M. indica*. The second principal component accounts for 9.73% of the variation in the dataset and separates *M. casturi* and the two populations of *M. quadrifida*.

Analysis of the four samples of *M. casturi* using the INTROGRESS software indicates that the samples are very similar, with hybrid indices (where 0 is *M. quadrifida* and 1 is *M. indica*) between 0.60-0.62 and interspecific heterozygosity of 25.3-26.3% (Fig. 6.7). Network analysis of the *M. casturi* dataset in the program SPLITSTREE places *M. casturi* directly intermediate to *M. indica* and *M. quadrifida*. However, the neighbor network identifies two populations of *M. quadrifida,* one of which is also found to be an intermediate (Fig. 6.8). As a result of the population differentiation of *M. quadrifida* observed in the PCA and SPLITSTREE analyses, we attempted to analyze the *M. casturi* dataset independently for each of the *M. quadrifida* populations, but the small sample sizes of the individual *M. quadrifida* populations prohibited robust analysis.

Genetic diversity indices of the *M. casturi* dataset (Table 6.4) show similar patterns to those observed in the *M. odorata* dataset. Individuals of *M. casturi* have higher observed heterozygosity (H_O = 0. 2869) than the putative parental taxa (*M. indica* H_O = 0.0939, *M. foetida* H_O = 0.1473). Gene diversity was lowest in *M. indica* (H_S = 0.1112), and similar in *M. casturi* and *M. quadrifida* (H_S = 0.1580 and 0.1377, respectively). A strongly negative inbreeding coefficient for *M. casturi* (F_{IS} = -0.7801) indicates an excess of outcrossing in this species, while values closer to zero were calculated for *M. indica* and *M. quadrifida* (F_{IS} = 0.1213 and -0.0394, respectively). The highest percent polymorphism is observed in *M. indica* (53.75%), while *M. quadrifida* and *M. casturi* have similar values (37.60% and 32.70%, respectively). Similar values of nucleotide diversity are observed in *M. indica* and *M. quadrifida* (0.0321 and 0.0331, respectively) while the value for *M. casturi* is much lower (0.0045). For both *M. odorata* and *M. quadrifida,* values of *rbarD*, a measure of multilocus linkage disequilibrium, are significantly different from zero (*rbarD* = 0.2522 and 0.3011, p = 0.001, Appendix 6.4). The number and percent of private alleles across all loci is highest for *M. indica* (1453, 40.26%), lower in *M. quadrifida* (937, 25.96%), and lowest in *M. casturi* (62, 1.72%).

Discussion

In this study, we use RADseq to explore the consequences of hybridization between a widely cultivated fruit tree, *M. indica,* and two different congeners in Southeast Asia. We find support for the previously demonstrated hybrid origin of *M. odorata* and, for the first time, evidence that *M. casturi* is also of hybrid origin. Both *M. odorata* and *M. casturi* lack unique genetic diversity compared to their parental taxa, as

indicated by the low number of private alleles present in hybrid populations and by population structure and network analyses. However, neither *M. odorata* nor *M. casturi* is directly intermediate to its respective parental taxa, as would be expected for first generation (F1) hybrids. Instead, both hybrids have levels of interspecific heterozygosity near 25%, indicating that they are likely the result of a backcross with one parental species. In the case of *M. odorata*, our data support the results of Teo et al. (2002), who found it to be more closely related to *M. foetida*. Therefore, *M. odorata* is probably the product of a backcross between an F1 hybrid and *M. foetida.* On the other hand, *M. casturi* has greater genetic affinity to *M. indica* than to *M. quadrifida,* and is most likely the result of an F1 hybrid backcross with *M. indica.*

Although *M. odorata* is of hybrid origin, we find no evidence of introgression between *M. indica* and *M. foetida* via the hybrid intermediate *M. odorata.* On the contrary, the genetic similarity of 16 of the 17 individuals of *M. odorata,* as evidenced by significant multilocus linkage disequilibrium and similar measures of admixture (interspecific heterozygosity, hybrid index, population structure assignment) combined with strongly negative inbreeding coefficients suggest that *M. odorata* is the result of a limited number of hybridization events. A previous study of diversity in 11 landraces of *M. odorata* using microsatellite markers also found it to have a very narrow genetic base (Yamanaka et al., 2006). We find patterns of diversity in *M. casturi* to be similar to *M. odorata,* with significant multilocus linkage disequilibrium, and individuals having very similar measures of admixture. Notably, in the case of *M. casturi* we find some evidence indicating population differentiation within *M. quadrifida* that may be the result of alternate backcrossing events (e.g., an F1 hybrid backcross with *M. quadrifida)*.

The remarkably similar genetic identity of individuals within *M. odorata* and *M. casturi* is consistent with hybrids that have formed only a few times and have thereafter been maintained clonally. In most cases of fruit trees, grafting is the primary means of clonal propagation of tree crops, allowing unique genetic individuals to be maintained in perpetuity (e.g., Warschefsky et al., 2016; Chapter III). Considering the small number of *M. casturi* individuals analyzed, it is possible that the four individuals we included were grafted clones. However, in the case of *M. odorata*, our 17 samples originated from different sources and the majority of individuals were not apparently grafted. So, if not by grafting, how could hybrid *M. odorata* be clonally maintained?

Nucellar polyembryony, a type of asexual reproduction or apomixis, produces multiple embryos within a single seed. One of the embryos in a polyembryonic seed is sexually derived, while the others develop from the maternal nucellar tissue (Aleza et al., 2010). While rare in angiosperms, nucellar polyembryony is well documented in many *Citrus* species (Wang et al., 2017) along with Southeast Asian cultivars of *M. indica* (Mukherjee and Litz, 2009), and has been reported at least three other species of *Mangifera,* including *M. odorata* and *M. casturi* (Kostermans and Bompard, 1993; Mukherjee and Litz, 2009; Lim, 2012a; b). Therefore, we propose that *M. odorata* may represent a cultivated hybrid lineage maintained by clonal reproduction via nucellar polyembryony. We can also speculate that *M. casturi* represents a similar cultivated, polyembryonic hybrid lineage, though additional individuals should be analyzed to confirm the genetic uniformity of the species.

Polyembryony is an important agricultural characteristic for tree crops like mango and citrus, allowing for the propagation of otherwise unattainable clonal rootstock

material. Recent genome sequencing of multiple *Citrus* species has helped to shed light on the evolution of polyembryony in the group (Wang et al., 2017). In the case of citrus, the majority of cultivated species are of interspecific hybrid origin, (e.g., sweet orange, grapefruit), and it has been suggested that these hybrid lineages have been maintained by polyembryony, with different citrus cultivars originating through somatic mutations in clonal lineages rather than through sexual reproduction (Wu et al., 2018). Evidence from whole genome sequencing of multiple *Citrus* species indicates that nucellar polyembryony is controlled by a single dominant allele that first arose in the mandarin, *C. reticulata* (Wang et al., 2017).

In mango, traditional analysis of phenotypic segregation from crosses between polyembryonic and monoembryonic cultivars also ascertained that a single dominant gene controls the trait (Aron et al., 1998; Kuhn et al., 2017). However, at this point many questions about the process of nucellar polyembryony in *Mangifera* remain unanswered. One issue central to the aims of our study is the fate of the sexually reproduced embryos, both in the case of original F1 hybrids and of further backcrosses. Since we found no evidence of recurrent backcrossing or introgression, it is possible that these individuals do not commonly survive because of some genetic incompatibility. However, the lack of F1 individuals could be explained by nomenclature, as it is possible that *M. odorata* and *M. casturi* are names applied to only very specific hybrid lineages, much like the use of cultivar names, while F1 hybrids may be given different common and/or scientific names.

Future Research

Here, we have provided preliminary evidence that indicates *M. casturi*, a species only known from cultivation and classified as extinct in the wild (IUCN, 2012), may be a hybrid of *M. indica* and *M. quadrifida.* More thorough sampling of *M. casturi* and *M. quadrifida* would provide greater insight into the origin of *M. casturi* and the population structure we observed within *M. quadrifida.* Although we analyzed more samples of *M. odorata* than *M. casturi*, the perplexing lack of diversity within *M. odorata* warrants additional sampling and investigation, particularly because the species was previously described as being polymorphic (Hou, 1978; Teo et al., 2002).

In the future, research should investigate whether bioinformatic parameters impact the ability to detect hybridization and introgression in ddRADseq datasets. As discussed in Chapter IV, bioinformatic parameters can alter the placement of hybrid taxa in phylogenetic analysis. Here, the amount of missing data permitted varied for the *M. odorata* and *M. casturi* datasets, which may have impacted the results, and verification of these findings using multiple different levels of missing data is an important forthcoming step.

One important avenue for research is to determine whether *M. odorata* and *M. casturi* can be re-created by controlled crosses between their respective parental taxa. However, hand pollination of *Mangifera* species is said to be very difficult and is inefficient because a large proportion of fruits are aborted prematurely (Iyer and Schnell, 2009). In addition, given that *M. odorata* and *M. indica* appear to be backcrosses, replicating these individuals would require at least two generations, or 6-20 years (Iyer and Schnell, 2009).

Overall, our research provides important insights into the consequence of hybridization within *Mangifera*. Coupled with our knowledge of hybrid citrus species, our findings reveal a pattern of perennial crop cultivation: maintenance of favorable hybrid perennial crop lineages through apomixis.

Literature Cited

Abbott, R., D. Albach, S. Ansell, J.W. Arntzen, S.J.E. Baird, N. Bierne, J. Boughman, et al. 2013. Hybridization and speciation. *Journal of Evolutionary Biology* 26: 229–246.

Adamack, A.T., and B. Gruber. 2014. PopGenReport: Simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution* 5: 384–387.

Agapow, P.M., and A. Burt. 2001. Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes* 1: 101–102.

Aleza, P., J. Juárez, P. Ollitrault, and L. Navarro. 2010. Polyembryony in non-apomictic citrus genotypes. *Annals of Botany* 106: 533–545.

Anderson, E. 1949. Introgressive hybridization. John Wiley & Sons, Inc., New York, NY.

Anderson, E., and G.L. Stebbins. 1954. Hybridization as an evolutionary stimulus. *Evolution* 8: 378–388.

Andrews. S. 2010. FastQC: a quality control tool for high throughput sequence data. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Arnold, M. 1992. Natural hybridization as an evolutionary process. *Annual Review of Ecology and Systematics* 23: 237–261.

Arnold, M.L., E.S. Ballerini, and A.N. Brothers. 2012. Hybrid fitness, adaptation and evolutionary diversification: lessons learned from Louisiana Irises. *Heredity* 108: 159–66.

Aron, Y., H. Czosnek, and S. Gazit. 1998. Polyembryony in mango (*Mangifera indica* L.) is controlled by a single dominant gene. *HortScience* 33: 1241–1242.

Barton, N.H., and H. G.M. 1985. Analysis of hybrid zones. *Annual Review of Ecology and Systematics* 16: 113–148.

Bompard, J.M. 2009. Taxonomy and Systematics. *In* R. E. Litz [ed.], The mango: Botany, production and uses, CAB International, Cambridge, MA.

Brown, A.H., M.W. Feldman, and E. Nevo. 1980. Multilocus Structure of Natural Populations of *Hordeum spontaneum*. *Genetics* 96: 523–536.

Buerkle, C.A. 2005. Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes* 5: 684–687.

Coart, E., S. Van Glabeke, M. De Loose, A.S. Larsen, and I. Roldán-Ruiz. 2006. Chloroplast diversity in the genus *Malus*: new insights into the relationship between the European wild apple (*Malus sylvestris* (L.) Mill.) and the domesticated apple (*Malus domestica* Borkh.). *Molecular ecology* 15: 2171–82.

Cornille, A., P. Gladieux, M.J.M. Smulders, I. Roldán-Ruiz, F. Laurens, B. Le Cam, A. Nersesyan, et al. 2012. New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genetics* 8: e1002703.

Coyne, J., and H.A. Orr. 2004. Speciation. Sinauer Associates Inc., Sunderland, MA.

Crane, J.H., and C.W. Campbell. 1994. The Mango. University of Florida Institute of Food and Agricultural Sciences, Gainesville, FL (pp. 1-7).

Dillon, N.L., I.S.E. Bally, C.L. Wright, L. Hucks, D.J. Innes, and R.G. Dietzgen. 2013. Genetic diversity of the Australian national mango genebank. *Scientia Horticulturae* 150: 213–226.

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

Earl, D.A., and B.M. vonHoldt. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361.

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

Ellstrand, N.C., R. Whitkus, and L.H. Rieseberg. 1996. Distribution of spontaneous plant hybrids. *Proceedings of the National Academy of Sciences of the United States of America* 93: 5090–5093.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* 14: 2611–2620.

Falush, D., M. Stephens, and J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567–87.

FAO (2003) Tropical Fruits. In *Medium-term prospects for agricultural commodities*. Food and Agriculture Organization of the United Nations, Rome, Italy. (pp. 1–6).

FAOSTAT (2013). Food and Agriculture Organization of the United Nations Statistics Division. Available from: http://www.fao.org/faostat/.

Gompert, Z., and C.A. Buerkle. 2010. Introgress: a software package for mapping components of isolation in hybrids. *Molecular ecology resources* 10: 378–84.

Griffith, W. 1854. Dicotyledonous Plants. *In* J. McLelland [ed.], Notulae ad Plantas Asiaticas, 414–420. Charles A. Serrao, Calcutta.

Harrison, R.G., and E.L. Larson. 2014. Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity* 105: 795–809.

Hou, D. 1978. Anacardiaceae. *In* Flora Malesiana Vol 8 Part 3, 395–548. Sijthoff & Noordhoff International Publishers, Alphen Aan Den Rijn, The Netherlands.

Hubisz, M.J., D. Falush, M. Stephens, and J.K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* 9: 1322–32.

Huson, D.H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution* 23: 254–67.

IUCN. 2012. World Conservation Monitoring Centre 1998. *IUCN Red List of Threatened Species. Version 2012.2.* Available at: www.iucnredlist.org.

Iyer, C.P.A., and R.J. Schnell. 2009. Breeding and genetics. *In* R.E. Litz [ed.], The mango: Botany, production and uses, 68–83. CAB International, Cambridge.

Jakobsson, M., and N.A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801–1806.

Jombart, T. 2008. Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403–1405.

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

Kamvar, Z.N., J.F. Tabima, and N.J. Grünwald. 2014. *Poppr* : an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2: e281.

Kiew, R. 2002. *Mangifera odorata* (Anacardiaceae) is a hybrid. 54: 205–206.

Kostermans, A.J.G.H., and J.M. Bompard. 1993. The mangoes: Their botany, nomenclature, horticulture, and utilization. Academic Press, San Diego, CA.

Kuhn, D.N., I.S.E. Bally, N.L. Dillon, D. Innes, A.M. Groh, J. Rahaman, R. Ophir, et al. 2017. Genetic map of mango: A tool for mango breeding. *Frontiers in Plant Science* 8: 1–11.

De Langhe, E., E. Hribová, S. Carpentier, J. Dolezel, and R. Swennen. 2010. Did backcrossing contribute to the origin of hybrid edible bananas? *Annals of Botany* 106: 849–57.

Lim, T.K. 2012a. *Mangifera laurina*. *In* Edible medicinal and non-medicinal plants, 124–126. Springer Netherlands, Dordrecht, New York.

Lim, T.K. 2012b. *Mangifera odorata*. *In* Edible medicinal and non-medicinal plants, 127–130. Springer Netherlands, Dordrecht, New York.

Mallet, J. 2007. Hybrid speciation. *Nature* 446: 279–83.

Miller AJ, Gross BL (2011) From forest to field: perennial fruit crop domestication. *American Journal of Botany* **98**(9), 1389–1414.

Mukherjee, S.K. 1949. The mango and its wild relatives. *Science and Culture* 26: 5–9.

Mukherjee, S.K., and R.E. Litz. 2009. Introduction: Botany and importance. *In* R. E. Litz [ed.], The mango: Botany, production and uses, 1–18. CAB International, Wallingford, UK.

Peterson, B.K., J.N. Weber, E.H. Kay, H.S. Fisher, and H.E. Hoekstra. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non model species. *PloS one* 7: 1–11.

Petit, R.J., and A. Hampe. 2006. Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics* 37: 187–214.

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–59.

Rieseberg, L.H. 1997. Hybrid origins of plant species. *Annual Review of Ecology and Systematics* 28: 359–389.

Rodríguez, F., M. Ghislain, A.M. Clausen, S.H. Jansky, and D.M. Spooner. 2010. Hybrid origins of cultivated potatoes. *Theoretical and Applied Genetics* 121: 1187–98.

Rosenberg, N.A. 2004. Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes* 4: 137–138.

Schnell, R.J., C.T. Olano, W.E. Quintanilla, and A. W. Meerow. 2005. Isolation and characterization of 15 microsatellite loci from mango (*Mangifera indica* L.) and cross-species amplification in closely related taxa. *Molecular Ecology Notes* 5: 625–627.

Servedio, M.R., and M. a. F. Noor. 2003. The role of reinforcement in speciation: Theory and data. *Annual Review of Ecology, Evolution, and Systematics* 34: 339–364.

Soltis, P.S., and D.E. Soltis. 2009. The role of hybridization in plant speciation. *Annual Review of Plant Biology* 60: 561–88.

Stebbins, G.L. 1950. Variation and Evolution in Plants. Columbia University Press, New York, NY.

Teo, L.L., R. Kiew, O. Set, S.K. Lee, and Y.Y. Gan. 2002. Hybrid status of kuwini, *Mangifera odorata* Griff. (Anacardiaceae) verified by amplified fragment length polymorphism. *Molecular Ecology* 11: 1465–9.

Wang, X., Y. Xu, S. Zhang, L. Cao, Y. Huang, J. Cheng, G. Wu, et al. 2017. Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nature Genetics* 49: 765–772.

Warschefsky, E., R.V. Penmetsa, D.R. Cook, and E.J.B. von Wettberg. 2014. Back to the wilds: Tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. *American journal of botany* 101: 1791–800.

Warschefsky, E.J., L.L. Klein, M.H. Frank, D.H. Chitwood, J.P. Londo, E.J.B. von Wettberg, and A.J. Miller. 2016. Rootstocks: Diversity, domestication, and impacts on shoot phenotypes. *Trends in Plant Science* 21: 418–437.

Wu, G.A., J. Terol, V. Ibanez, A. López-García, E. Pérez-Román, C. Borredá, C. Domingo, et al. 2018. Genomics of the origin and evolution of *Citrus*. *Nature*.

Yakimowski, S.B., and L.H. Rieseberg. 2014. The role of homoploid hybridization in evolution: A century of studies synthesizing genetics and ecology. *American Journal of Botany* 101: 1247–1258.

Yamanaka, N., M. Hasran, D.H. Xu, H. Tsunematsu, S. Idris, And T. Ban. 2006. Genetic relationship and diversity of four *Mangifera* species revealed through AFLP analysis. *Genetic Resources and Crop Evolution* 53: 949–954.

Tables

Table 6.1. Samples in the *M. odorata* dataset. Sample ID consists of the individual ddRAD Sample ID, the study collection number and putative identification (cultivar or species name). The collection location (FTBG = Fairchild Tropical Botanic Garden, FTG = Fairchild Tropical Garden Herbarium, SBG = Singapore Botanic Garden, KRP = Purwodadi Botanic Garden, KRB = Bogor Botanic Garden, USDA = USDA Subtropical Horticulture Research Station at Chapman Field, FSP = Miami Dade Fruit and Spice Park, GBTB = Gardens by the Bay, FRIM = Forestry Research Institute of Malaysia, PA = Pasoh Forest Arboretum, PF = Pasoh Forest Research Station) and accession number within the respective collection are provided. The ddRAD library, sublibrary, and individual sample ID are given, as well as the number of raw reads for each individual.

Sample Name	Species ID	Collected From	Accession Number	Specimen Number	Provenance	Lane/ Sublibrary	Raw Reads
11C_MI_84_ Thai_Everbearing	*M. indica*	FTBG	2004-1177*A	–	–	1/11	993848
11G_MI_27_Pairi	*M. indica*	FTBG	2004-1081*A	–	–	1/11	1063491
12_EMI_100_M_pajang	*M. foetida*	FTBG	2012-2354*A	–	Brunei	1/12	244582
13F_MI_81_Mallika	*M. indica*	FTBG	2003-1722*A	–	–	3/13	1919662
14A_MI_92_ M_odorata_Row_E	*M. odorata*	FTBG	2008-1293	–	Malaya	3/14	1891440
14B_KRP_4_ M_foetida_cv_Pakel	*M. odorata*	KRP	IX.C.25	EW 117	Gedong Kuning	3/14	1260649
14C_KRP_31_M_indica_ cv_Gandik_luyung	*M. indica*	KRP	IX.B.24A	EW 141	–	3/14	612705
14H_KRP_9_M_indica	*M. indica*	KRP	XVI.E.21	EW 122	Sumba, NTT	3/14	1504680
15C_MI_16_Joe_Long	*M. odorata*	FTBG	2004-1197	–	–	3/15	1546871
15F_MI_103_M_foetida	*M. foetida*	FTBG	2014-0266*A	–	–	3/15	1420235
16C_MI_40_Cac	*M. indica*	FTBG	2006-1295*A	–	–	3/16	1555717
16F_FRIM_11_ M_cf_odorata	*M. odorata*	FRIM	Y02-1473, 33020193	EW 188	–	3/16	1189156

3A_MI_37_Golek	*M. indica*	FTBG	2006-1292*A	–	–	1/3	729910
3C_MI_20_Carabao	*M. indica*	FTBG	2005-1791*A	–	–	1/3	480442
3D_MI_35_Himsagar	*M. indica*	FTBG	2006-1278*A	–	–	1/3	586464
3E_MI_32_Gedong_Ginco	*M. indica*	FTBG	2004-1215*A	–	–	1/3	440891
3F_MI_69_Langra_Benarsi	*M. indica*	FTBG	2005-1945*A	–	–	1/3	798799
6A_MI_102_Chok_Anon	*M. indica*	FTBG	2006-1301*A	–	–	1/6	672634
6C_MI_11_Nam_Doc_Mai	*M. indica*	FTBG	2003-1724*A	–	–	1/6	603488
6F_MI_148_Gilas	*M. indica*	FTBG	2010-0366*A	–	–	1/6	737869
6G_MI_68_ Pohn_Sawadee	*M. indica*	FTBG	2004-1088*A	–	–	1/6	767434
9H_MI_34_Rumanii	*M. indica*	FTBG	2005-1742*A	–	–	1/9	1189430
AC_FRIM8_M_odorata	*M. odorata*	FRIM	33020171	EW 815	–	A/2	1287058
AD_FRIM10_M_odorata	*M. odorata*	FRIM	W05-1614, 33020185	EW 187	–	A/2	1282316
AF_K1_M_foetida	*M. foetida*	FRIM	NA	EW 223	Peninsular Malaysia	A/2	1104520
AG_K2_M_odorata	*M. odorata*	FRIM	NA	EW 224	Peninsular Malaysia	A/2	1483656
AH_KRB_1_M_sp	*M. foetida*	KRB	XIX.F.2.51	–	W. Java	A/2	1901836
BB_K3_M_indica	*M. indica*	FRIM	NA	EW 225	Peninsular Malaysia	B/2	814416
BC_K4_M_foetida	*M. foetida*	FRIM	NA	EW 226	Peninsular Malaysia	B/2	1107182
BG_KRP3_ M_foetida_pakel	*M. foetida*	KRP	IX.C.27	EW 116	Semarang (C. Java)	B/2	1091284
BH_KRP8_M_foetida	*M. foetida*	KRP	IX.C.11	EW 121	Blitar	B/2	1165602
CA_KRP_5_ M_odorata_cv_Kuweni	*M. odorata*	KRP	IX.C.37a	EW 118	Semarang (C. Java)	C/2	609735
CB_MI_91_Myatrynat	*M. indica*	FTBG	2006-1293*A	–	–	C/2	631319
CD_KRP_6_M_foetida_ cv_Pakel_lumut	*M. indica*	KRP	IX.C.9b	EW 119	Semarang (C. Java)	C/2	514920

CF_MI_42_Swethintha	*M. indica*	FTBG	2004-1211*A	–	–	C/2	537252
CG_MI_77_Amrapali	*M. indica*	FTBG	2012-2391*A	–	–	C/2	599107
CH_KRP_7_M_odorata	*M. odorata*	KRP	IX.C.10a	EW 120	Bantul	C/2	713827
DF_MI_80_Nam_Tam_Teem	*M. indica*	FTBG	2004-1228*A	–	–	D/2	702325
EB_MI_74_Pam_kai_mia	*M. indica*	FTBG	2012-2402*A	–	–	E/2	1315873
FA_MI_76_Chao_savoy	*M. indica*	FTBG	2003-1712*A	–	–	F/2	1760475
FF_MI51_Royal_Special	*M. indica*	FTBG	2004-1087*A	–	–	F/2	1311009
FG_MI_53_Sindhri	*M. indica*	FTBG	2004-1055*A	–	–	F/2	1876470
GH_MI_57_Jumbo_Kesar	*M. indica*	FTBG	2009-0822*A	–	–	G/2	1219797
HA_MI_88_Pancahdarakalasa	*M. indica*	FTBG	2004-1181*A	–	–	H/2	1561762
HC_MI_63_Dusheri	*M. indica*	FTBG	2005-1765*A	–	–	H/2	1599887
HE_MI_39_Pyu_Pyu_Kalay	*M. indica*	FTBG	2009-0816*A	–	–	H/2	1942535
HF_MI_79_Aslul_Mukararara	*M. indica*	FTBG	2008-1289*A	–	–	H/2	1222409
IA_MI_12_Tong_Dam	*M. indica*	FTBG	2003-1707*A	–	–	I/2	1032280
ID_MI_14_Alphonso	*M. indica*	FTBG	2004-1053*A	–	–	I/2	1518263
IF_MI_43_Saigon	*M. indica*	FTBG	2006-1276*A	–	–	I/2	832868
IG_MI_4_Phimsen_Mun	*M. indica*	FTBG	2003-1753*A	–	–	I/2	990893
IH_MI_36_Alampur_Baneshan	*M. indica*	FTBG	2010-0370*A	–	–	I/2	1092326
JC_MI_55_Frenz_odorata	*M. odorata*	FTBG	2010-0365*A	–	–	J/2	1317838
JD_MI_3_Cambodiana	*M. indica*	FTBG	2005-1753*A	–	–	J/2	1023296
JE_MI_61_Praya_Savoy	*M. indica*	FTBG	2006-1289*A	–	–	J/2	1142384
JG_MI_54_Sig_Siput	*M. indica*	FTBG	2005-1746*A	–	–	J/2	893846
JH_MI_70_M_rampagni	*M. odorata*	FTBG	2001-0889*A	–	Sarawak	J/2	1097061
KG_FRIM_17_M_foetida	*M. foetida*	FRIM	33020262	EW 194	–	K/2	1423614

LB_KRP_32_ M_indica_kepodang	*M. indica*	KRP	IX.B.18A	EW 142	–	L/2	972685
LC_PA_6_M_foetida	*M. foetida*	PA	NA	EW 200	Peninsular Malaysia	L/2	1062234
MD_KRP_13_M_sp	*M. odorata*	KRP	XVI.E.44	EW 126	S. Kalimantan	M/3	1808845
MH_FRIM_5_M_odorata	*M. odorata*	FRIM	33020097	EW 182	–	M/3	2014983
NA_MI_56_ Hindi_Besanara	*M. indica*	FTBG	2004-1233*A	–	–	N/3	1511170
OG_SBG_34_M_odorata	*M. odorata*	SBG	19970994*A	EW 176	–	O/3	1278585
PA_MI_89_Aeromanis	*M. indica*	FTBG	2004-1189*A	–	–	P/3	1549107
PD_KRB_31_M_sp	*M. foetida*	KRB	VII.E.179	–	N. Sulawesi	P/3	1543712
QD_MI_9_Totapuri	*M. indica*	FTBG	2004-1222*A	–	–	Q/3	1574796
RE_MI_31_Kaeo_Luemkon	*M. indica*	FTBG	2007-1056*A	–	–	R/3	1396821
RH_MI_15_Ivory	*M. indica*	FTBG	2003-1723*A	–	–	R/3	1658796
SF_SBG_25_M_foetida	*M. foetida*	SBG	NA	EW 169	–	S/3	1367395
TB_KRP_2_M_minor	*M. indica*	KRP	IX.B.32	EW 115	Sulawesi Tengah	T/1	1587805
UA_SBG_35_M_foetida	*M. foetida*	SBG	00/6994*A	EW 177	–	U/1	2400886
UB_SBG_30_M_foetida	*M. foetida*	SBG	20060155*F	EW 172	–	U/1	2399019
UD_SBG_22_M_foetida	*M. foetida*	SBG	NA	EW 166	–	U/1	2849945
UE_SBG_28_M_foetida	*M. foetida*	SBG	003135*A	EW 171	–	U/1	2537820
UH_SBG_32_M_foetida	*M. foetida*	SBG	20091887*A	EW 174	–	U/1	2343765
WF_MI_52_Maha_Chanok	*M. indica*	FTBG	2012-2399*A	–	–	W/1	2338767
XD_SBG_27_M_foetida	*M. foetida*	SBG	003135*A	EW 170	–	X/1	1443188
XE_SBG_23_M_odorata	*M. odorata*	SBG	NA	EW 167	–	X/1	1468589
XF_FRIM_3_M_indica	*M. indica*	FRIM	NA	EW 180	–	X/1	1676509
XG_FRIM_2_M_foetida	*M. foetida*	FRIM	NA	EW 179	–	X/1	1479415
XH_SBG_31_M_odorata	*M. indica*	SBG	NA	EW 173	–	X/1	1514572

YB_MI_48_Imam_Pasand	*M. indica*	FTBG	2004-1187*A	–	–	Y/1	1222060
YC_MI_65_Ratna	*M. indica*	FTBG	2012-2403*B	–	–	Y/1	1833551
YH_MI_73_Cowasji_patel	*M. indica*	FTBG	2004-1079*A	–	–	Y/1	3144639
ZB_FRIM_1_M_foetida	*M. foetida*	FRIM	NA	EW 178	–	Z/1	729331
ZC_SBG_33_M_odorata	*M. odorata*	SBG	20090008*A	EW 175	–	Z/1	1040003
ZD_SBG_2_M-foetida	*M. foetida*	SBG	200903556*C	EW 148	–	Z/1	576639
ZE_SBG_21_M_foetida	*M. foetida*	SBG	NA	EW 165	–	Z/1	996406
ZF_SBG_3_M-odorata	*M. odorata*	SBG	20093557*A	EW 149	–	Z/1	856966

Table 6.2. Samples in the *M. casturi* dataset. Sample ID consists of the individual ddRAD Sample ID, the study collection number and putative identification (cultivar or species name). The collection location (FTBG = Fairchild Tropical Botanic Garden, FTG = Fairchild Tropical Garden Herbarium, SBG = Singapore Botanic Garden, KRP = Purwodadi Botanic Garden, KRB = Bogor Botanic Garden, USDA = USDA Subtropical Horticulture Research Station at Chapman Field, FSP = Miami Dade Fruit and Spice Park, GBTB = Gardens by the Bay, FRIM = Forestry Research Institute of Malaysia, PA = Pasoh Forest Arboretum, PF = Pasoh Forest Research Station) and accession number within the respective collection are provided. The ddRAD library, sublibrary, and individual sample ID are given, as well as the number of raw reads for each individual.

Sample Name	Species ID	Collected From	Accession Number	Specimen Number	Provenance	Lane/ Sublibrary	Raw Reads
11C_MI_84_ Thai_Everbearing	*M. indica*	FTBG	2004-1177*A	–	–	1/11	993848
11G_MI_27_Pairi	*M. indica*	FTBG	2004-1081*A	–	–	1/11	1063491
11H_MI_86_M_casturi_ 982186A	*M. casturi*	FTBG	982186*A	–	Kalimantan	1/11	1335965
13F_MI_81_Mallika	*M. indica*	FTBG	2003-1722*A	–	–	3/13	1919662
14C_KRP_31_M_indica_ cv_Gandik_luyung	*M. indica*	KRP	IX.B.24A	EW 141	–	3/14	612705
14E_MI_10_ Tommy_Atkins	*M. indica*	FTBG	2003-1734*A	–	–	3/14	1588941
16C_MI_40_Cac	*M. indica*	FTBG	2006-1295*A	–	–	3/16	1555717
3A_MI_37_Golek	*M. indica*	FTBG	2006-1292*A	–	–	1/3	729910
3B_MI_117_Mabrouka	*M. indica*	FTBG	2003-1713*A	–	–	1/3	561681
3C_MI_20_Carabao	*M. indica*	FTBG	2005-1791*A	–	–	1/3	480442
3D_MI_35_Himsagar	*M. indica*	FTBG	2006-1278*A	–	–	1/3	586464
3E_MI_32_Gedong_Ginco	*M. indica*	FTBG	2004-1215*A	–	–	1/3	440891
6A_MI_102_Chok_Anon	*M. indica*	FTBG	2006-1301*A	–	–	1/6	672634
6B_MI_158_Baptiste	*M. indica*	FTBG	2006-1309*A	–	–	1/6	511517
6F_MI_148_Gilas	*M. indica*	FTBG	2010-0366*A	–	–	1/6	737869

6G_MI_68_ Pohn_Sawadee	*M. indica*	FTBG	2004-1088*A	–	–	1/6	767434
6H_MI_72_ M_quadrifida_RowA	*M. quadrifida*	FTBG	2012-2379A	–	–	1/6	1056687
9C_MI_130_ M_mempelam	*M. laurina*	FTBG	2012-2371*B	–	–	1/9	719950
AB_FRIM6_M_quadrifida	*M. quadrifida*	FRIM	T04-1467	EW 183	–	A/2	251955
AE_GBTB_3_M_indica	*M. indica*	GBTB	NA	–	–	A/2	1619145
CB_MI_91_Myatrynat	*M. indica*	FTBG	2006-1293*A	–	–	C/2	631319
CF_MI_42_Swethintha	*M. indica*	FTBG	2004-1211*A	–	–	C/2	537252
CG_MI_77_Amrapali	*M. indica*	FTBG	2012-2391*A	–	–	C/2	599107
DC_KRP_26_M_similis	*M. quadrifida*	KRP	XVI.D.II.16	–	E. Kalimantan	D/2	678691
DF_MI_80_ Nam_Tam_Teem	*M. indica*	FTBG	2004-1228*A	–	–	D/2	702325
EB_MI_74_Pam_kai_mia	*M. indica*	FTBG	2012-2402*A	–	–	E/2	1315873
ED_SBG2014_10_ M_magnifica	*M. quadrifida*	SBG	20110755*A	–	–	E/2	1323694
FA_MI_76_Chao_savoy	*M. indica*	FTBG	2003-1712*A	–	–	F/2	1760475
FD_KRP_28_M_casturi	*M. casturi*	KRP	XVI.D.II.14	EW 138	S. Kalimantan	F/2	1650119
FF_MI51_Royal_Special	*M. indica*	FTBG	2004-1087*A	–	–	F/2	1311009
FG_MI_53_Sindhri	*M. indica*	FTBG	2004-1055*A	–	–	F/2	1876470
GA_MI_47_Rawa	*M. quadrifida*	FTBG	2012-2356	–	–	G/2	1172739
GB_KRP_11_M_sp	*M. casturi*	KRP	XVI.E.22	EW 134	Maluku	G/2	1107713
GH_MI_57_Jumbo_Kesar	*M. indica*	FTBG	2009-0822*A	–	–	G/2	1219797
HA_MI_88_ Pancahdarakalasa	*M. indica*	FTBG	2004-1181*A	–	–	H/2	1561762
HC_MI_63_Dusheri	*M. indica*	FTBG	2005-1765*A	–	–	H/2	1599887

HE_MI_39_Pyu_Pyu_Kala y	*M. indica*	FTBG	2009-0816*A	–	–	H/2	1942535
HF_MI_79_ Aslul_Mukararara	*M. indica*	FTBG	2008-1289*A	–	–	H/2	1222409
IA_MI_12_Tong_Dam	*M. indica*	FTBG	2003-1707*A	–	–	I/2	1032280
ID_MI_14_Alphonso	*M. indica*	FTBG	2004-1053*A	–	–	I/2	1518263
IF_MI_43_Saigon	*M. indica*	FTBG	2006-1276*A	–	–	I/2	832868
IG_MI_4_Phimsen_Mun	*M. indica*	FTBG	2003-1753*A	–	–	I/2	990893
IH_MI_36_ Alampur_Baneshan	*M. indica*	FTBG	2010-0370*A	–	–	I/2	1092326
JD_MI_3_Cambodiana	*M. indica*	FTBG	2005-1753*A	–	–	J/2	1023296
JE_MI_61_Praya_Savoy	*M. indica*	FTBG	2006-1289*A	–	–	J/2	1142384
JG_MI_54_Sig_Siput	*M. indica*	FTBG	2005-1746*A	–	–	J/2	893846
LB_KRP_32_ M_indica_kepodang	*M. indica complex*	KRP	IX.B.18A	EW 142	–	L/2	972685
LF_SBG_19A_ M_quadrifida	*M. quadrifida*	SBG	20110756*A	EW 162	–	L/2	813211
MC_MI_101_ M_quadrifida	*M. quadrifida*	FTBG	2012-2356*A	–	–	M/3	2765794
NA_MI_56_ Hindi_Besanara	*M. indica*	FTBG	2004-1233*A	–	–	N/3	1511170
PA_MI_89_Aeromanis	*M. indica*	FTBG	2004-1189*A	–	–	P/3	1549107
QB_KRB_8_M_griffithii	*M. casturi*	KRB	VII.E.170a	–	S. Kalimantan	Q/3	1223203
QD_MI_9_Totapuri	*M. indica*	FTBG	2004-1222*A	–	–	Q/3	1574796
RD_PF_15_M_sp	*M. quadrifida*	PF	151873	EW 214	Peninsular Malaysia	R/3	1078866
RE_MI_31_Kaeo_Luemko n	*M. indica*	FTBG	2007-1056*A	–	–	R/3	1396821
RH_MI_15_Ivory	*M. indica*	FTBG	2003-1723*A	–	–	R/3	1658796
SB_MI_78_M_A_carle	*M. casturi*	FTBG	2007-1060*A	–	–	S/3	1612855

SC_KRP_27_M_indica	*M. indica*	KRP	XVI.D.II.8	EW 137	Sulawesi	S/3	1594169
TC_KRB_2_M_similis	*M. quadrifida*	KRB	VI.D.8a	–	Bangka I., S. Sumatra	T/1	1755500
WF_MI_52_Maha_Chanok	*M. indica*	FTBG	2012-2399*A	–	–	W/1	2338767
XB_SBG_6_M_quadrifida	*M. quadrifida*	SBG	NA	EW 152	–	X/1	1559839
XF_FRIM_3_M_indica	*M. indica*	FRIM	NA	EW 180	–	X/1	1676509
YB_MI_48_Imam_Pasand	*M. indica*	FTBG	2004-1187*A	–	–	Y/1	1222060
YC_MI_65_Ratna	*M. indica*	FTBG	2012-2403*B	–	–	Y/1	1833551
YH_MI_73_Cowasji_patel	*M. indica*	FTBG	2004-1079*A	–	–	Y/1	3144639

Table 6.3. Population genetic analysis of *M. odorata* and its parental taxa, *M. indica* and *M. foetida. Ho*: observed heterozygosity; *Hs*: heterozygosity within populations, aka 'gene diversity'; *Fis*: inbreeding coefficient; *π:* nucleotide diversity; rbarD: multilocus linkage disequilibrium; Ap%: percent private alleles.

Species	Ho	Hs	Fis	%poly	π	rbarD	Ap%
M. odorata	0.3005	0.1638	-0.6230	43.22%	0.0016	0.3984	0.26%
M. indica	0.1072	0.1186	0.0840	25.32%	0.0309	0.0270*	44.25%
M. foetida	0.0487	0.0644	0.3115	57.29%	0.0246	0.0407	11.76%

Table 6.4. Population genetic analysis of *M. casturi* and its putative parental taxa, *M. indica* and *M. quadrifida. Ho*: observed heterozygosity; *Hs*: heterozygosity within populations, aka 'gene diversity'; *Fis*: inbreeding coefficient; *π:* nucleotide diversity; rbarD: multilocus linkage disequilibrium; Ap%: percent private alleles.

Species	Ho	Hs	Fis	%poly	π	rbarD	Ap%
M. casturi	0.2869	0.1580	-0.7801	32.70%	0.0045	0.2522	1.72%
M. indica	0.0939	0.1112	0.1213	53.75%	0.0321	0.0455	40.26%
M. quadrifida	0.1473	0.1377	-0.0394	37.60%	0.0331	0.3011	25.96%

Figure Captions

Figure 6.1. Visualization of inferred population structure for the *M. odorata* dataset produced using the software distruct with *K* = 2 populations. Individuals within the dataset are represented by a vertical bar and colors represent assignment to each population. Individuals are labeled by species name *(M. foetida, M. odorata, M. indica)* and, for *M. indica,* region of origin (India, Southeast Asia [SEA]).

Figure 6.2. Principal component analysis of 88 samples of *M. foetida, M. odorata,* and *M. indica* from the *M. odorata* dataset, with individual samples colored according to a) species only (red = *M. foetida,* orange = *M. odorata,* green = *M. indica*) or b) species and region (red = *M. foetida*, orange = *M. odorata,* green = Indian *M. indica,* yellow = Southeast Asian *M. indica*).

Figure 6.3. Plot of hybrid index (0 = *M. foetida,* 1 = *M. indica*) and interspecific heterozygosity for 17 samples of *M. odorata* calculated in the program INTROGRESS.

Figure 6.4. Neighbor network tree for *M. odorata* (orange), *M. foetida* (red), and *M. indica* (green) inferred using SPLITSTREE.

Figure 6.5. Visualization of inferred population structure for the *M. casturi* dataset produced using the software distruct with *K* = 2 populations. Individuals within the dataset are represented by a vertical bar and colors represent assignment to each population. Individuals are labeled by species name *(M. quadrifida, M. casturi, M. indica).*

Figure 6.6. Principal component analysis of 65 samples of *M. quadrifida, M. casturi,* and *M. indica* from the *M. casturi* dataset, with individual samples colored according to a) species only (brown = *M. quadrifida,* purple = *M. casturi,* green = *M. indica*).

Figure 6.7. Plot of hybrid index (0 = *M. quadrifida,* 1 = *M. indica*) and interspecific heterozygosity for 4 samples of *M. casturi* calculated in the program INTROGRESS.

Figure 6.8. Neighbor network tree for *M. casturi* (purple), *M. quadrifida* (brown), and *M. indica* (green) inferred using SPLITSTREE.

Figures

Figure 6.1.

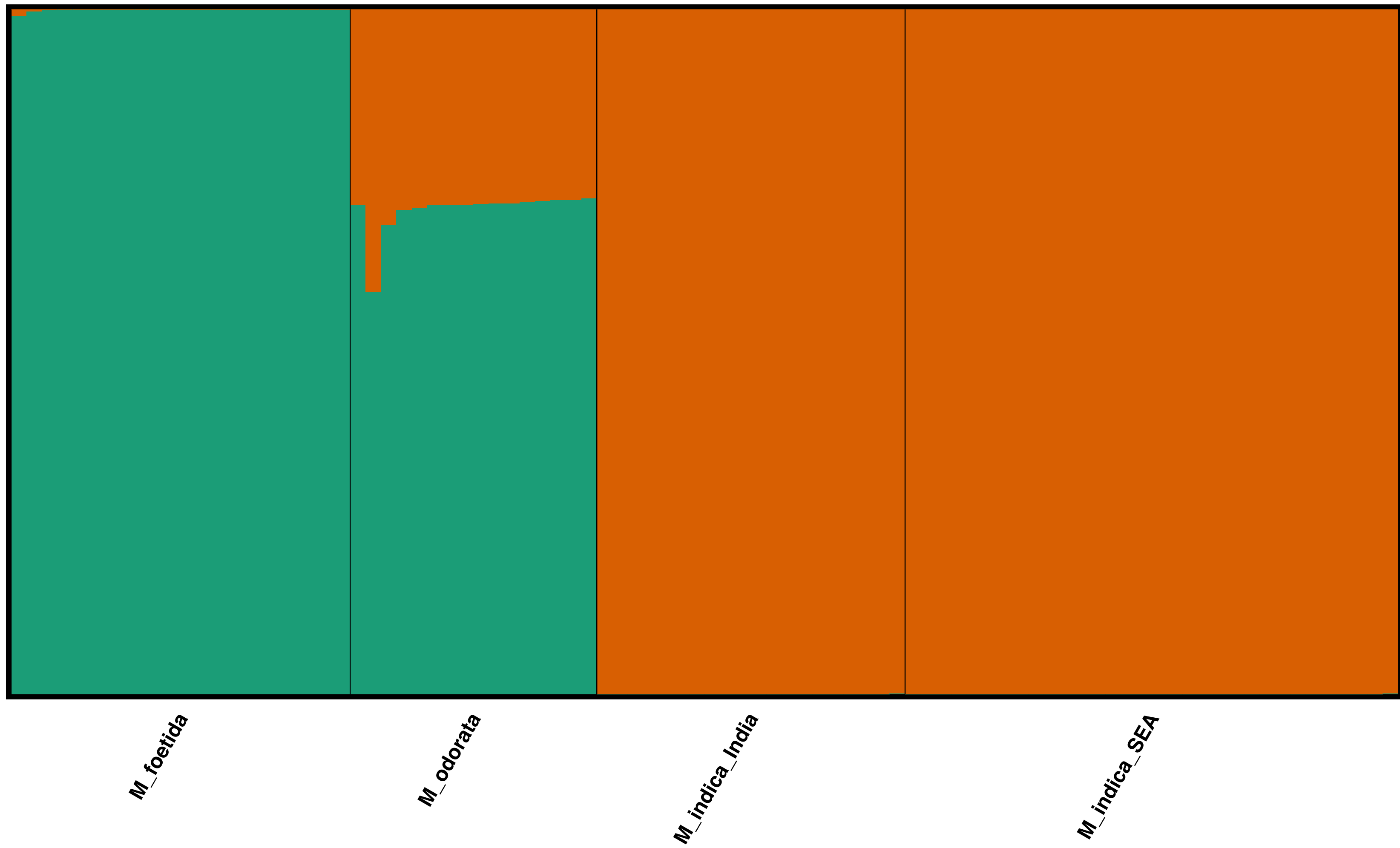


Figure 6.2.

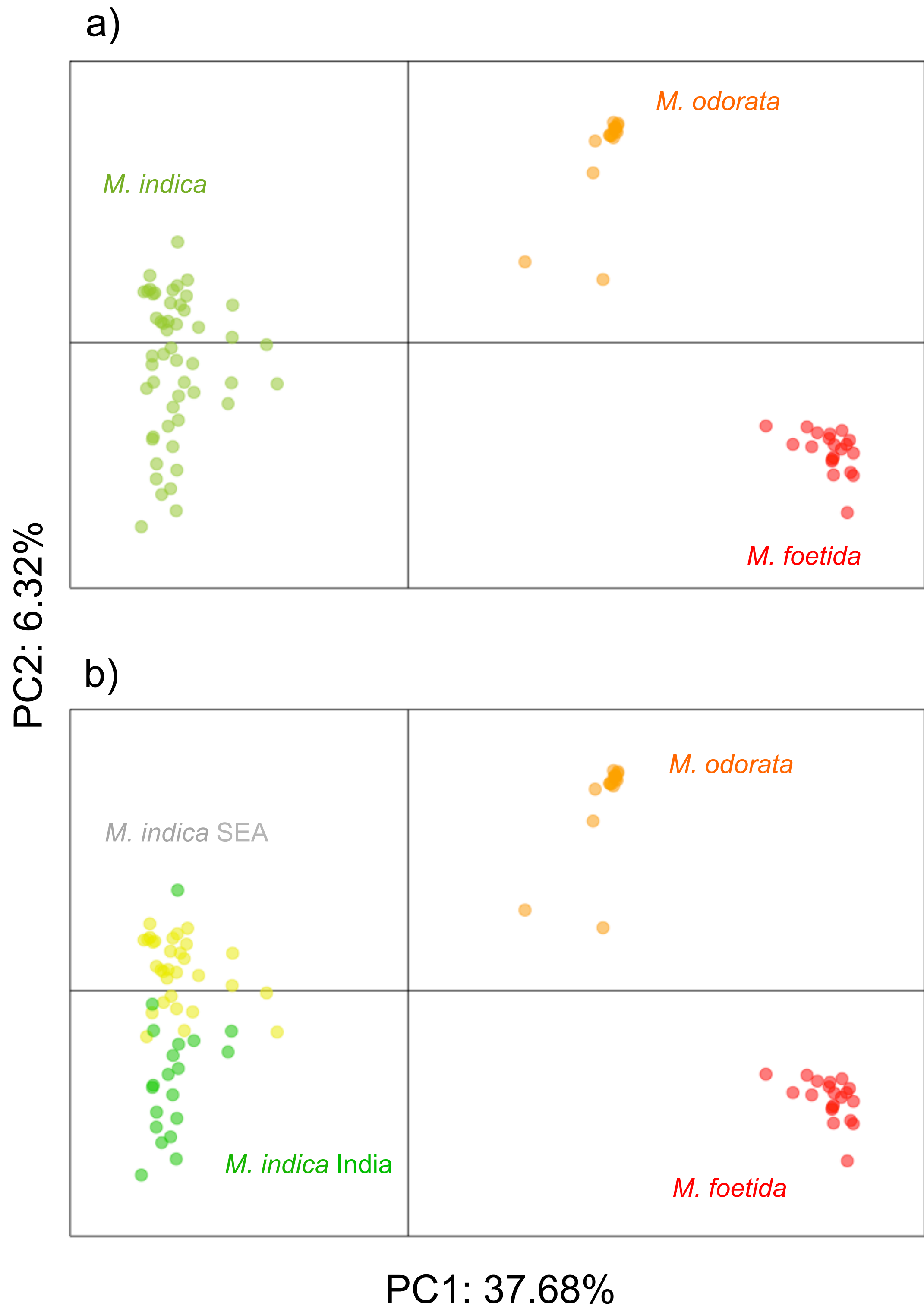


Figure 6.3.

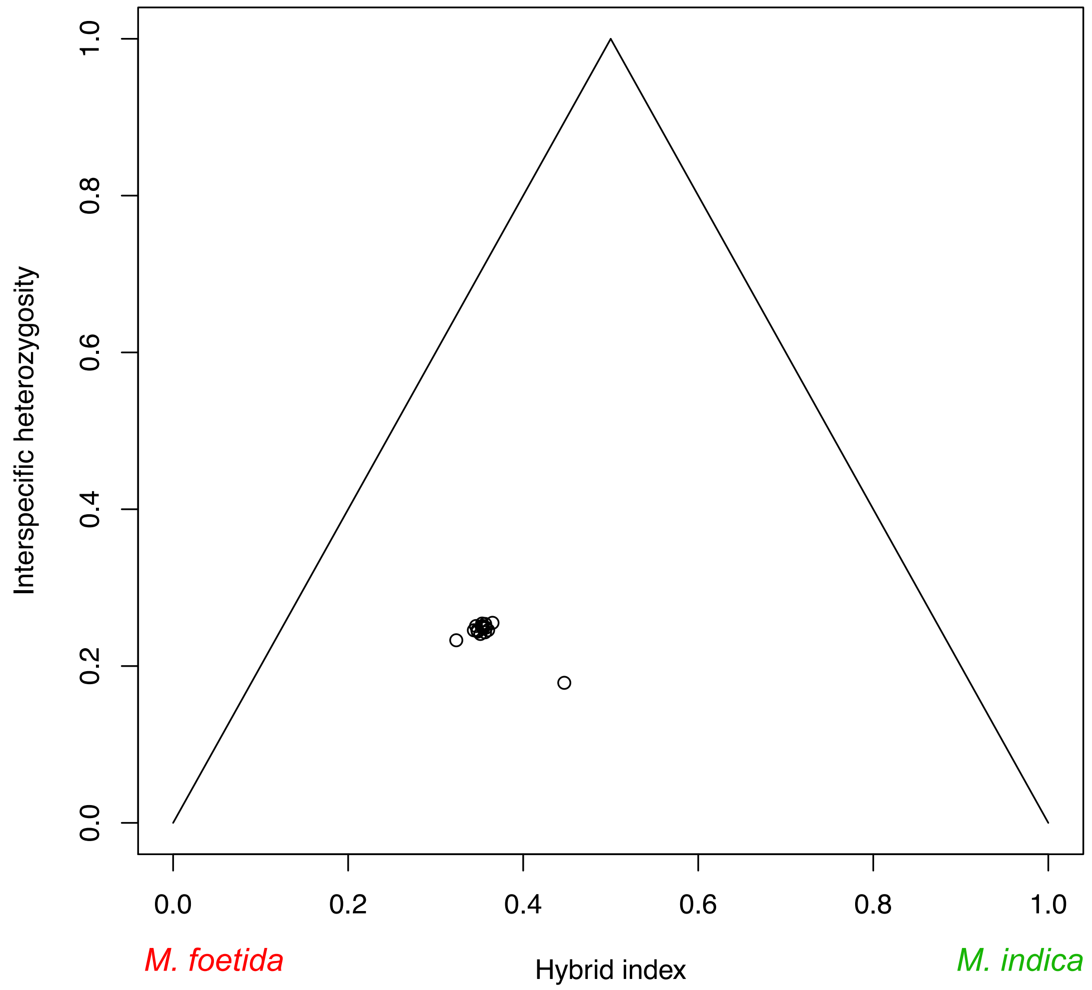


Figure 6.4.

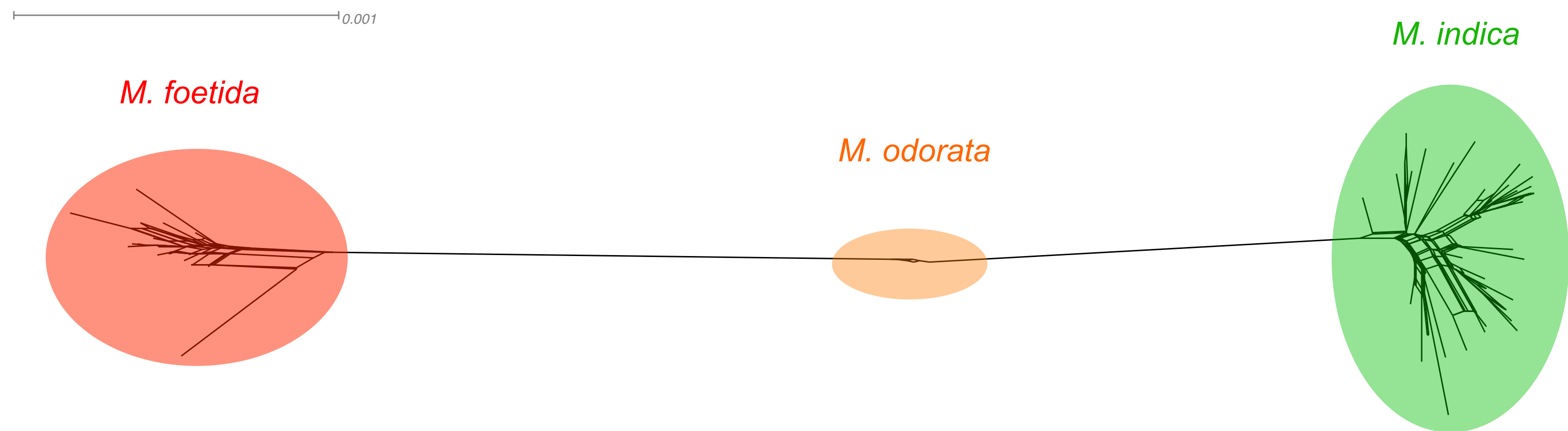


Figure 6.5.

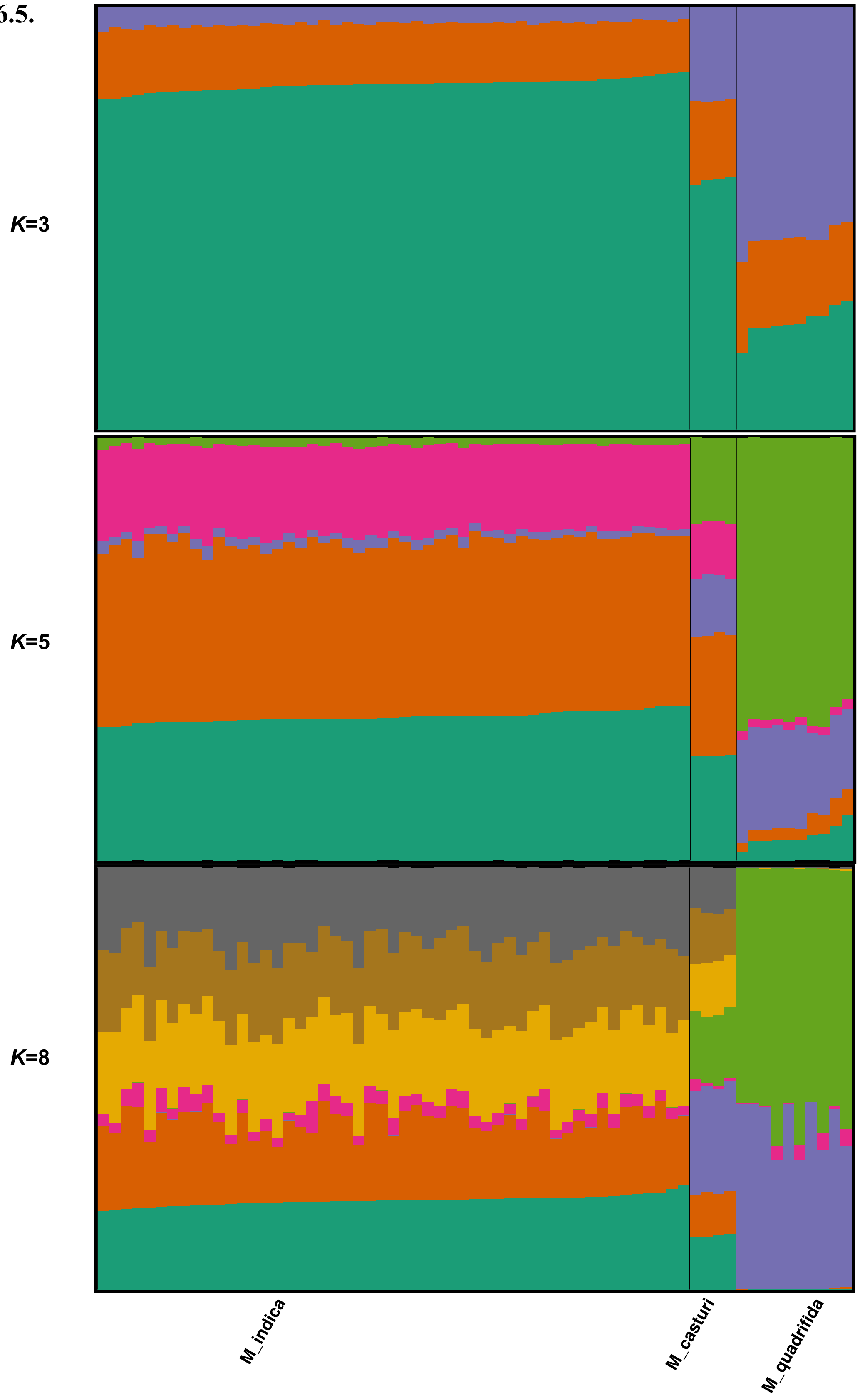


Figure 6.6.

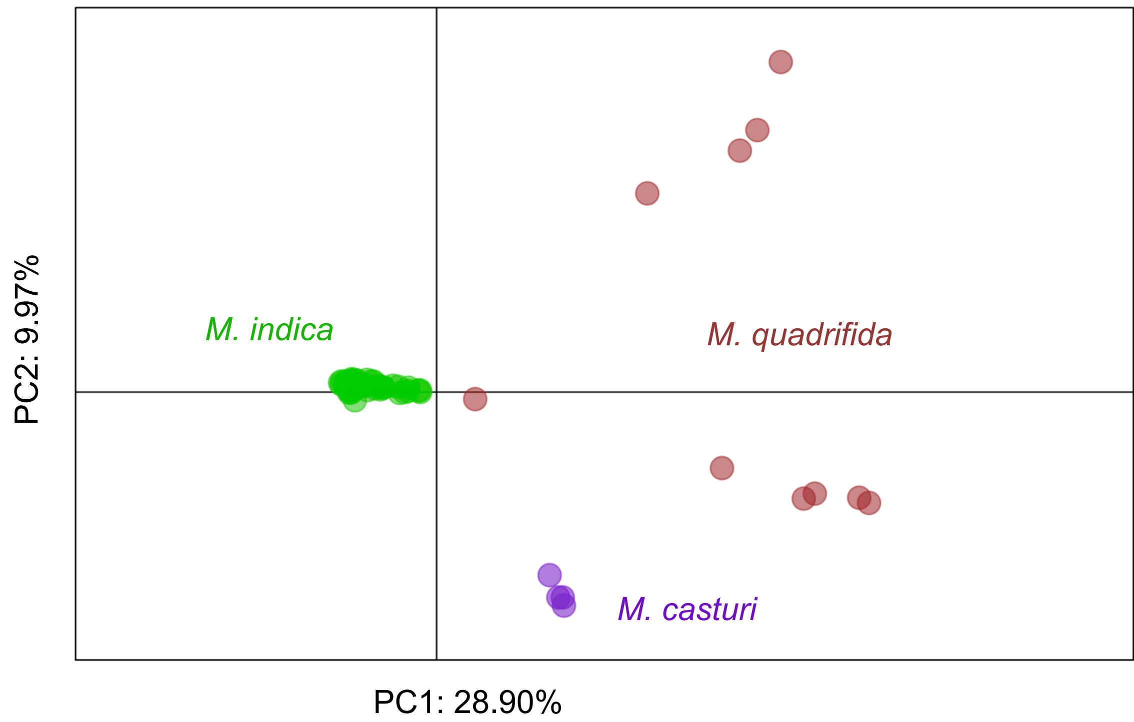


Figure 6.7.

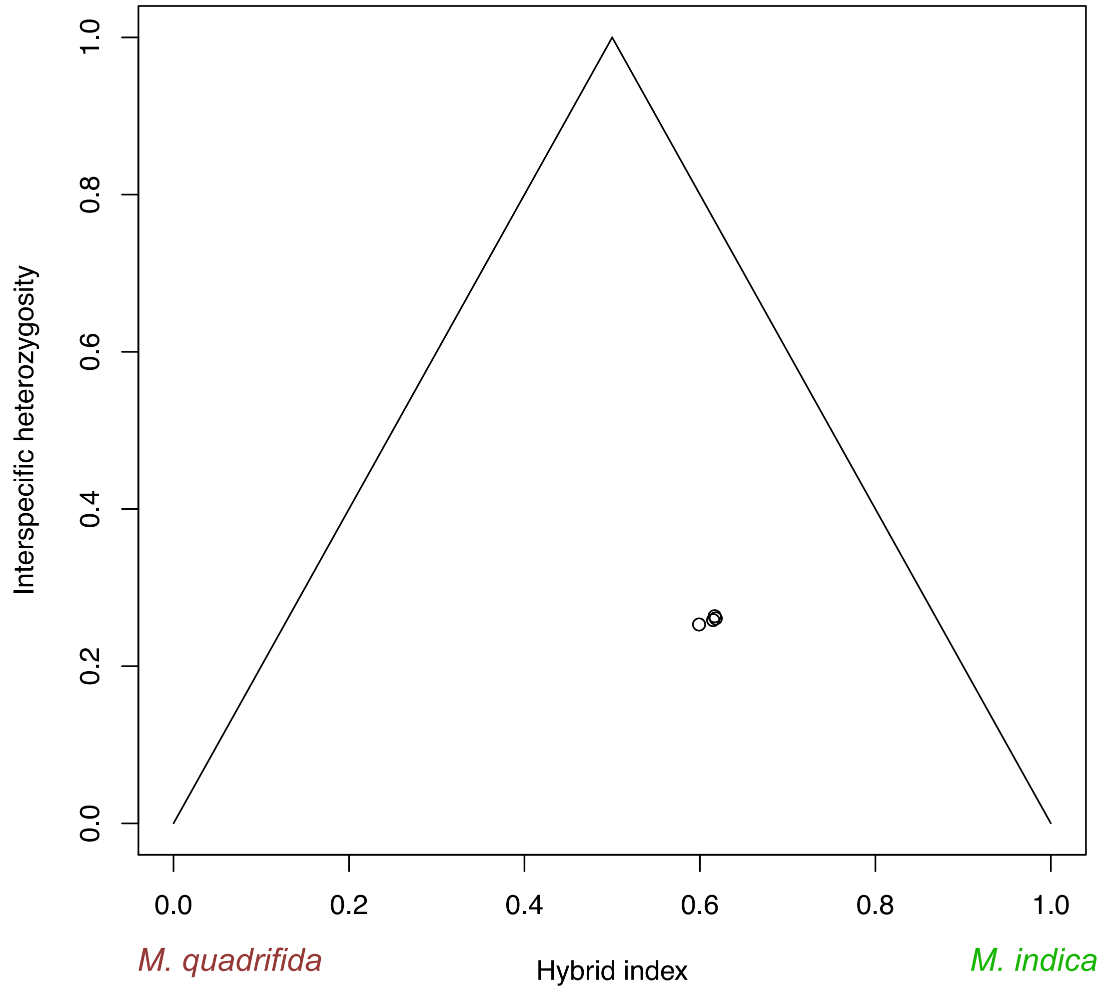


Figure 6.8.

Appendices Captions

Appendix 6.1. Plot of ΔK metric as described by Evanno et al. (2005) for the *M. odorata* dataset. The optimal number of ancestral populations is deemed to be that which has the highest value of ΔK.

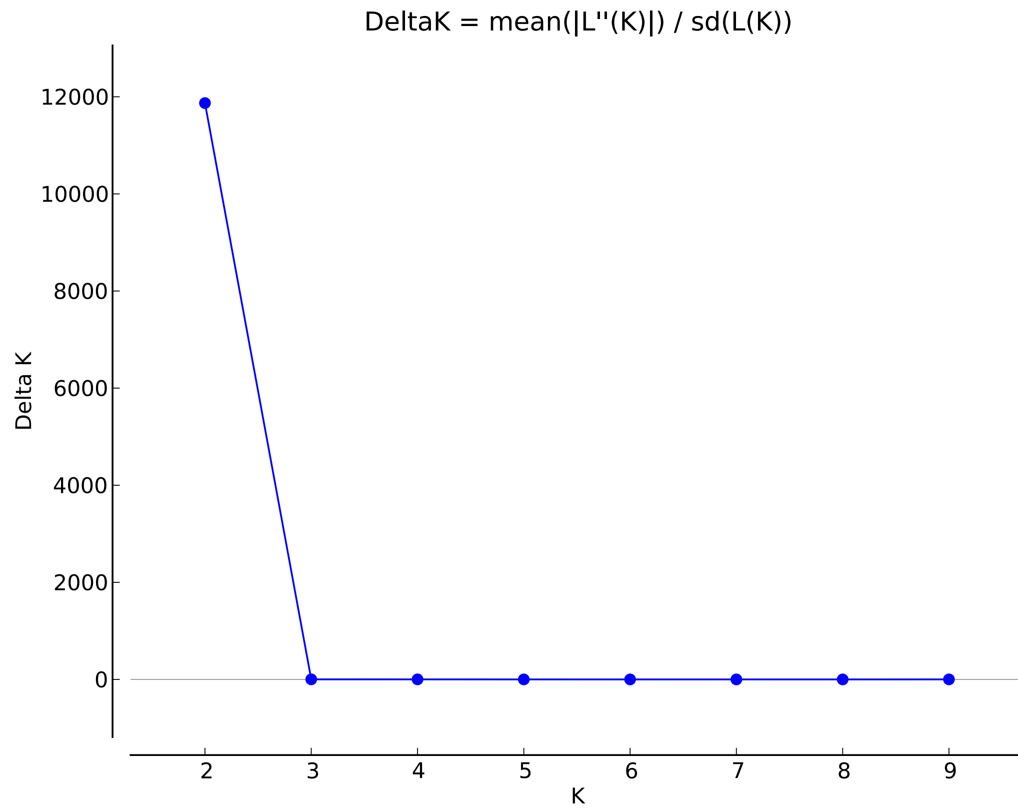
Appendix 6.2. Plot of expected distribution of rbarD for unlinked loci using 999 permutations (grey bars) and actual rbarD for *M. odorata* samples.

Appendix 6.3. Plot of ΔK metric as described by Evanno et al. (2005) for the *M. casturi* dataset. The optimal number of ancestral populations is deemed to be that which has the highest value of ΔK.

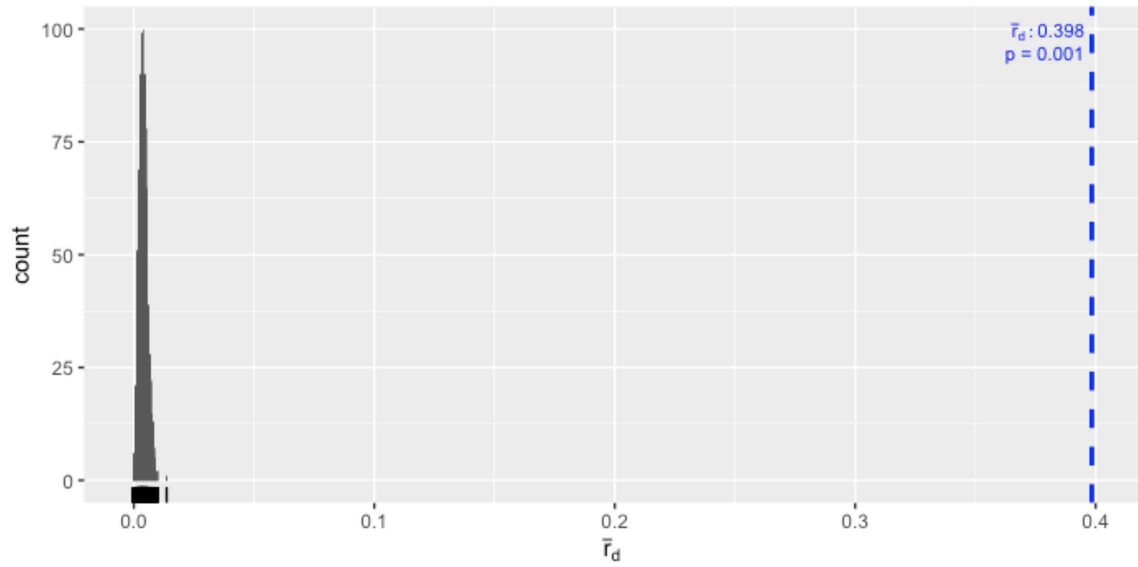
Appendix 6.4 Plot of expected distribution of rbarD for unlinked loci using 999 permutations (grey bars) and actual rbarD for **(A)** *M. casturi* and **(B)** *M. quadrifida* samples.

Appendices

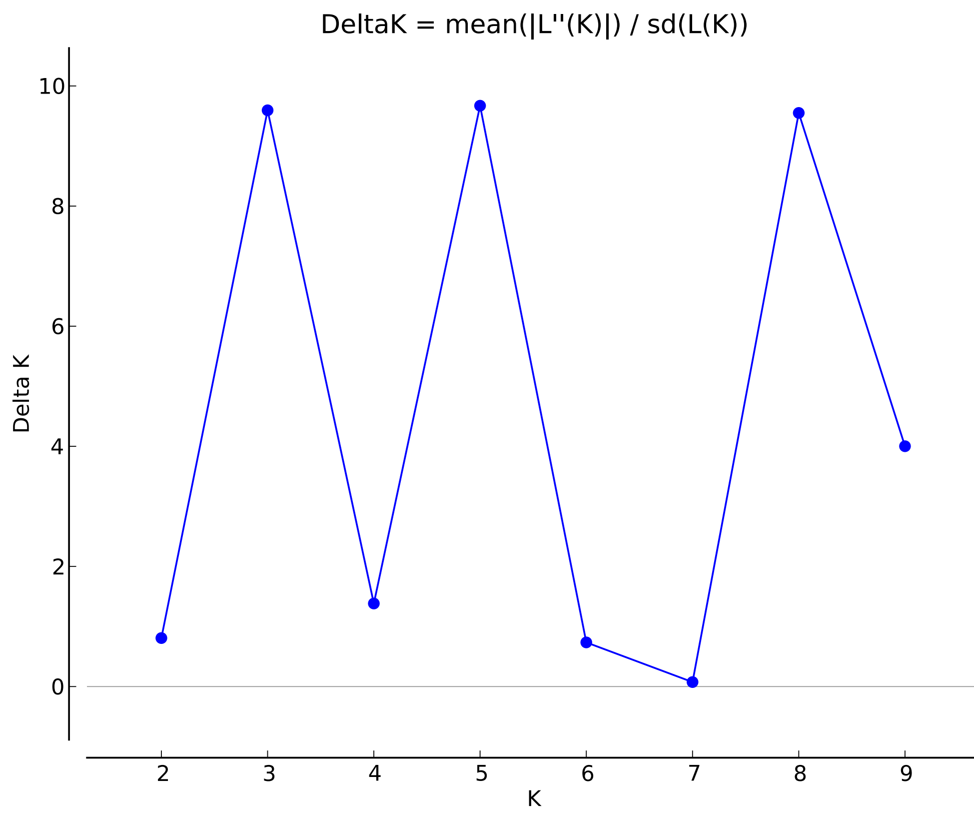
Appendix 6.1.



Appendix 6.2

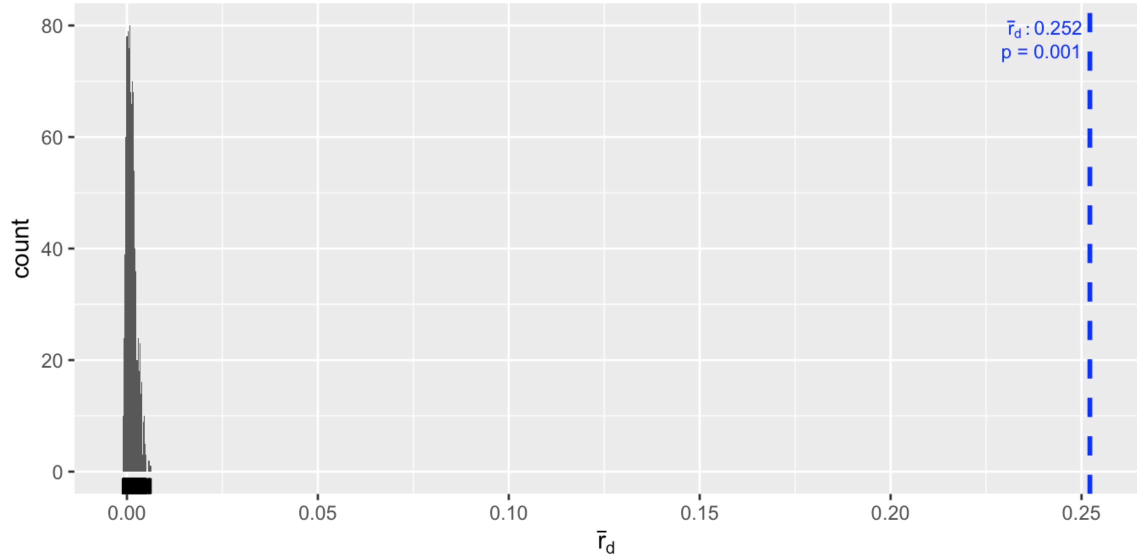


Appendix 6.3.

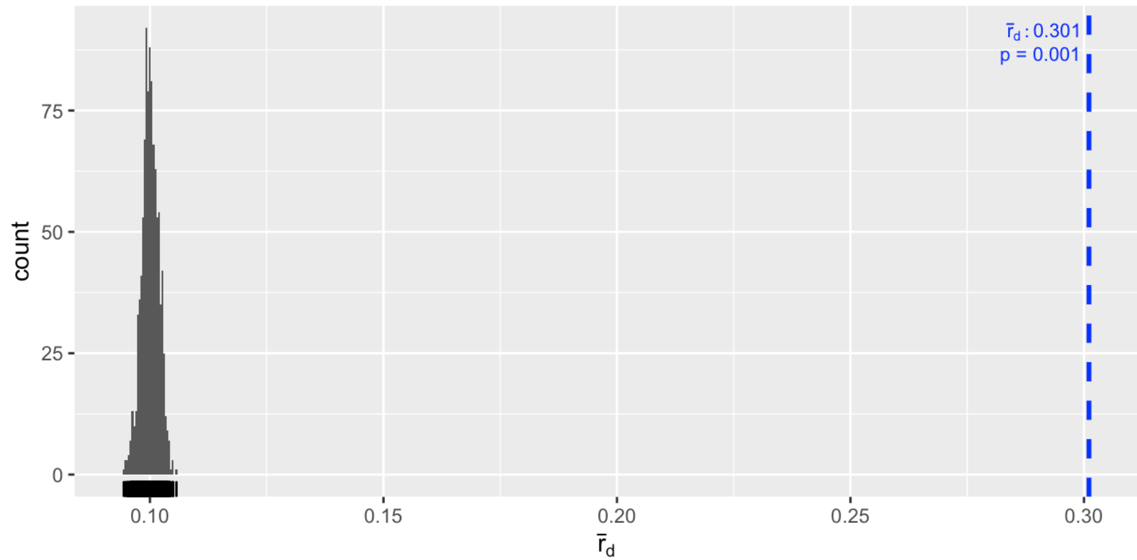


Appendix 6.4

A)



B)



CHAPTER VII:

CONCLUSIONS AND FUTURE DIRECTIONS

Introduction

Domesticated species served an important role in the formulation of Darwin's theory of evolution by natural selection and continue to inform our understanding of evolutionary processes today. In my dissertation, I have provided a comprehensive framework for understanding the evolution and domestication of one of the world's most important fruit crops, *Mangifera indica.* The work presented here is a novel and integrative approach to the study of crop domestication, applying advanced sequencing techniques to phylogeny, population genomics, and hybridization genomics in a non-model system of domestication. The results of my research fill existing gaps in our knowledge about domestication by expanding our perspective to include a broader taxonomic scope, species at various levels of domestication, and a system with a perennial life history.

As climate change alters the landscape, agriculturists must strive to produce crops adapted to new environmental conditions in order to meet the needs of a rapidly expanding human population (The Hague Conference 2010; Beddington et al. 2012; Hatfield and Takle 2014). In Chapter II, I proposed a multistep framework to improve crop performance by using naturally occurring genetic variation in crop wild relatives. While crop wild relatives often have higher levels of standing genetic variation and exhibit tolerance to agriculturally relevant stressors (e.g., flooding, drought, heat, cold) (Hajjar and Hodgkin 2007), their use in breeding programs has been limited (Ford-Lloyd et al. 2011). The workflow I proposed begins by building comprehensive collections of crop wild relatives that encompass the full range of the species' phenotypic, geographic, and environmental diversity. After genotyping and phenotyping, predictive association

networks can be developed with the goal of deploying wild relatives into prebreeding programs. Ultimately, the work presented in Chapter II provides an important perspective to the discussion of how to simultaneously improve plant health and production while reducing reliance on unsustainable agricultural inputs like irrigation, pesticides, and fertilizers.

Although the process of domestication has long been of interest to agriculturists and evolutionary biologists alike, the majority of formative crop research focused on a relatively small set of highly domesticated annual species like cereals and legumes (e.g., Singh et al. 1991; Wang et al. 1999; Matsuoka et al. 2002; Li et al. 2006; Londo et al. 2006; Huang et al. 2012; Hufford et al. 2013; Saintenac et al. 2013). In Chapter III, I contributed to the growing body of research examining domestication in woody perennial species by reviewing the biology of rootstocks. Surveying the literature, I found more than 70 woody perennial fruit crops are grown on rootstocks and 20 of the 25 most-produced fruit and nut crops are grafted in certain circumstances. Notably, species used as rootstocks are often closely related to but genetically distinct from the scion species they support, yet for any given crop, relatively few rootstock genotypes are typically employed in commercial settings. Grafting allows for independent selection of traits in the root and shoot system, in part mitigating the difficulties of domesticating long-lived outcrossing species. I found that rootstocks are selected for traits inherent to the root system itself (e.g., ease of vegetative propagation, flooding tolerance, resistance to root pests/pathogens), but also for traits imparted to the scion (e.g., precocity, dwarfing, productivity, fruit quality). Rootstocks may also have important effects on the plant microbiome, which is increasingly recognized as a critical factor in plant health,

resilience, and productivity. As a whole, the review suggests that while grafting is nearly ubiquitous in perennial crop cultivation, diverse rootstocks such as those from crop wild relatives remain an underutilized resource for perennial crop improvement.

The Mango

The core of my dissertation focused on a tropical perennial fruit tree, the mango *(M. indica)*, which is an important food source and export in many developing countries. I used an emerging molecular technique, restriction site associated DNA sequencing (RADseq) to examine the evolution of the mango and its wild relatives from phylogenomic and population genomic perspectives.

Phylogenetics

In Chapter IV, I produced the most comprehensive phylogenetic hypothesis for the genus *Mangifera* to date, revealing the evolutionary relationships between mango and its wild and semi-domesticated relatives. The economically important species *M. indica* was found to be a member of a closely related clade that includes *M. gedebe* along with cultivated species like *M. zeylanica*, *M. pentandra, M. lalijiwa, M. laurina,* and *M. casturi.* The *Mangifera* phylogeny also clarified infrageneric relationships within *Mangifera*, which had previously been hypothesized based on morphology alone, revealing that the genus consists of three primary clades. Unexpectedly, the phylogenetic inference showed that the genus *Mangifera,* as traditionally circumscribed, is not monophyletic: two species included in the analysis, *M. superba* and *M. caesia,* form a separate lineage that is sister to the clade of *Mangifera* and *Bouea*. On the basis of

morphological descriptions, I will propose that a total of 5 species currently classified as *Mangifera* represent a new genus of Anacardiaceae (Warschefsky and Pell, in prep.). Some of the species in the new genus are well known, such as *M. caesia,* which is widely cultivated in Malaysia, Indonesia, and parts of Thailand (Kostermans and Bompard 1993).

Phylogenetic analysis of RADseq data is still an emerging technique and the relative importance of different bioinformatic parameters has been debated in recent literature (e.g., Leaché et al. 2015; Huang and Knowles 2016; Eaton et al. 2017; Leaché and Oaks 2017; Tripp et al. 2017). To explore how bioinformatic parameter settings impact downstream results including topology and branch support values, I interrogated the data obtained in Chapter IV. My findings demonstrated that RADseq datasets with high levels of missing data (80% missing) are able to produce well-supported topologies. While the amount of missing data permitted in a dataset had a strong impact on resulting phylogenies, the clustering parameter, which is associated with ortholog identification, showed a relatively small effect on downstream topology. The study presented in Chapter IV was one of the first explicitly examine the impact of intraspecific sampling on topology using RADseq datasets. Results showed that intraspecific sampling can affect resulting topologies, particularly in the case of hybrid taxa. Overall, the examination of bioinformatic parameters in Chapter IV advances our understanding of how RADseq datasets should be analyzed for phylogenetic inference.

Future Directions

While the phylogeny of *Mangifera* presented in Chapter IV clarifies the systematics of the mango genus, it also underscores the need for additional work within the group. In the present study, I was unable to obtain samples from a number of *Mangifera* species, including many that are thought to be close allies of *M. indica,* such as *M. sylvatica* and *M. caloneura*. Future efforts to include additional taxa, including endemic species from Indochina, the Philippines, and the Andaman and Nicobar Islands, would allow for a clearer picture of the biogeography of *Mangifera.* Taxonomic revision of the genus in its entirety should emphasize species delimitation in certain complex taxa, such as *M. gedebe* and *M. quadrifida,* along with the closely related clade of species that includes *M. indica.* The suggested future work would improve our understanding of the evolutionary history of *Mangifera* species and their possible genetic contributions to cultivated mango.

The phylogeny of *Mangifera* lays the groundwork for future studies investigating the evolution of important traits in the genus, such as fruit morphology. Because each major clade recovered in the phylogeny contains both cultivated and wild species (Chapter IV, Fig. 4.4), *Mangifera* represents a novel system in which to examine the evolution and domestication of closely related tree taxa. In the case of *Mangifera,* it would be particularly interesting to explore whether traits such as fruit size, fibrousness, and exudate toxicity show any clear associations with domestication. Fruit characters like size and color may also be examined through the lens of their association with seed dispersers, which, in the case of *Mangifera,* include imperiled megafauna like orangutans, rhinos, and elephants (Phillipps and Phillipps 2016).

Population Genetics

Crop improvement depends on a foundational understanding of how domestication affects crop genetic diversity as well as the standing genetic variation within cultivated germplasm (e.g., Maxted and Guarino 2000; Iqbal et al. 2001; Burke et al. 2002; Mohammadi and Prasanna 2003; Esquinas-Alcázar 2005; Doebley et al. 2006; Ferreira 2006; Pickersgill 2007; Gross and Olsen 2010; Miller and Gross 2011; Kassa et al. 2012). In Chapter V, I examined the composition and geographic distribution of genetic diversity in the mango by analyzing RADseq data from 108 mango cultivars representing eight geographic regions. I found no evidence of a genetic bottleneck associated with the introduction of *M. indica* into Africa and the Americas, and calculated similar levels of diversity among all eight geographic regions. However, I found that mango cultivars from Southeast Asia contain novel genetic diversity that was not observed in cultivars from any other part of the world. Combined with their distinctive morphology, the unique genetic diversity of Southeast Asian cultivars suggests mango has a more complex history of domestication than previously assumed. My results are consistent with a multiple scenarios, some of which I outline below.

1. *One species, one domestication event, divergent selection: M. indica* may have been domesticated a single time from a single wild species, with divergent post-domestication selection driving differentiation between Indian and Southeast Asian cultivar types.

2. *One species, one domestication event, hybridization with a wild relative: M. indica* may have been domesticated a single time from a single species. When introduced into Southeast Asia, a region of high congeneric diversity, the cultivated mango may have undergone introgression with one or more wild relatives.

3. *One species, two domestication events: M. indica* may have had a broad native range and Indian and Southeast Asian populations may have been domesticated independently from divergent populations of wild *M. indica.*

4. *Two species, two domestication events: M. indica* may have been independently domesticated from two different interfertile species (with subsequent hybridization) and the differentiation observed in cultivated mango reflects these two ancestral species.

Future Directions

In the future, the above-mentioned scenarios should be tested using population demographic inference. However, one of the most important components of understanding the domestication history of the mango is the identification and analysis of wild *M. indica* populations, which are poorly known (see below).

As is the case for many species, the genomic resources available for mango are rapidly improving. No less than three genome sequences of different mango cultivars are nearing completion, and mango transcriptomics studies are also becoming increasingly common. The mango genome sequences will enable cost-effective methods of obtaining

whole genome sequence data from hundreds of mango cultivars, bypassing the limitations associated with RADseq datasets (e.g., missing data). Notably, the three mango genomes in progress are all Indian-type cultivars, so resequencing of Southeast Asian cultivars is an important early goal that will help to answer more questions about the history of the mango as well as identify candidate genes for traits that distinguish the two cultivar types, including polyembryony.

Hybridization

Hybridization is known to occur frequently in perennial species, and many perennial crop species are of confirmed or putative hybrid ancestry (Warschefsky et al. 2014; Chapter II, Appendix 2.1). In Chapter VI, I examined the occurrence and consequences of interspecific hybridization between native Southeast Asian *Mangifera* species and *M. indica,* which was introduced to the region. In the case of *M. odorata,* I found support for previous research identifying the species as a hybrid of *M. indica* and *M. foetida*. Notably, I did not observe any evidence of genetic introgression between *M. foetida* and Southeast Asian cultivars of *M. indica* via the hybrid intermediate *M. odorata.* Preliminary evidence reported in Chapter VI also suggests *M. casturi,* a species considered extinct in the wild (IUCN 2012), is in fact a garden hybrid of *M. indica* and *M. quadrifida*. Surprisingly, both hybrid lineages exhibited low genetic species-level diversity and high genetic identity between samples, suggesting that *M. odorata* and *M. casturi* may be predominantly clonal. While observed clonality may be the result of grafting, given that both *M. odorata* and *M. casturi* are reported to produce

polyembryonic seeds, I suggest that the lineages may be maintained by polyembryonic regeneration.

Future directions

According to the phylogeny produced in Chapter IV, the parental species *M. indica, M. foetida,* and *M. quadrifida,* are not closely related, being from clades I, II, and III, respectively (Chapter IV, Fig. 4.4). Therefore, the results of Chapter VI show that hybridization can occur across the 3 major clades of *Mangifera*, though perhaps infrequently. Putative hybrids between other *Mangifera* species have been reported in the literature (e.g., *M. laurina* x. *M. gedebe; M. foetida* x *M. pajang;* Kostermans and Bompard 1993) but because the genus is poorly studied, no molecular analysis has confirmed their hybrid status. Future efforts should aim to better characterize the prevalence and phylogenetic limits of hybridization in *Mangifera*, including whether any species are interfertile with *Bouea* species.

The results of Chapter V demonstrate that Southeast Asian mango cultivars contain novel diversity that could have come from hybridization and subsequent introgression with a congeneric species. While Chapter VI showed that the novel diversity of Southeast Asian mango cultivars is not the result of introgression with *M. foetida*, a few other candidate parental species exist. *Mangifera sylvatica* is native to Myanmar and northern Thailand and is said to have fruits with a very strongly curved or beaked shape reminiscent of the characteristic shape of Southeast Asian mango cultivars. Additionally, *M. laurina,* an Indonesian species, is closely related to *M. indica;* like Southeast Asian mango cultivars, *M. laurina* has polyembryonic seeds. Since I was

unable to obtain samples of *M. sylvatica* and limited sampling precluded a robust analysis of *M. laurina,* investigating these species should be a priority for future research.

Albeit a logical connection, the association of polyembryony and hybridization suggested by Chapter VI has never been explicitly investigated. The trait of polyembryony in *Mangifera* is both evolutionarily interesting and agriculturally important; polyembryonic cultivars are often used as clonal rootstock sources. However, the origin of polyembryony in the genus and in *M. indica* – when it first arose, whether it has independently evolved in multiple species, whether wild populations of *M. indica* showed variability in this trait, and why it is only present in one of the mango cultivar types – has never been explored. An important first step to understanding the evolution of polyembryony in *Mangifera* is phenotyping the species. For the majority of *Mangifera* species, Kostermans and Bompard (1993) do not mention either mono- or polyembryony in their descriptions, leaving questions about the prevalence of polyembryony within the genus.

Applications For Genebank Management

The maintenance of crop genetic variation in genebanks is essential to pre-breeding efforts that will ensure the future success of many of today's most important crops (Ferreira 2006). Therefore, it is critical that genetic repositories are managed in a way that maximizes the amount of genetic diversity preserved in each collection and accurate identification of accessions is of extreme importance (Schoen and Brown 1993; FAO 2010). Efficient genebank management is even more imperative for tree species, which require large investments of time, space, care, and money to sustain. Results from

Chapters IV, V and VI have confirmed the identities of many samples that were unlabeled or mislabeled in living collections. In total, 20 misidentified/mislabeled samples were identified, along with 12 unlabeled/unidentified samples. An additional 7 cases of possibly misidentified/mislabeled samples were also found, but could not be determined at this time. The new identifications will be disseminated to each respective living collection in the hopes that the information will be used to inform management practices.

Collectively, wild and semidomesticated *Mangifera* species represent a source of novel genetic variation that could be introduced into cultivated *M. indica* through pre-breeding programs. Of particular significance, some *Mangifera* species demonstrate resistance to mango anthracnose *(Colletotrichum gloeosporoides)* (Bompard 2009), which is one of the most problematic diseases in mango (Prusky et al. 2009). Additionally, a few *Mangifera* species native to high altitudes and subtropical forests may exhibit cold tolerance that could be introduced into *M. indica*, allowing it to be cultivated in subtropical and Mediterranean climates (Bompard 2009). Along with being used in traditional breeding programs some *Mangifera* species have been successful in rootstock trials (Campbell 2007). However, relatively few species have been tested, and few, if any non-*indica* rootstocks are used in commercial mango production.

Both interspecific grafting as well as traditional and modern breeding techniques require some degree of genetic compatibility between the species involved (Mudge et al. 2009). The phylogenetic hypothesis for *Mangifera* presented in Chapter IV (Fig. 4.4) provides insight into the genetic diversity and relatedness among *Mangifera* species, which in turn informs breeding and rootstock selection. Species within the same clade as

M. indica are the most likely to be compatible rootstocks and to be able to hybridize. However, to thoroughly explore the phylogenetic limits of interspecific hybridization and rootstock compatibility in *Mangifera,* future efforts should test compatibility between individuals from different clades.

Concluding Remarks

Crop wild relatives are invaluable genomic reserves that can readily be used in crop breeding programs. However, many crop wild relatives are in dire need of conservation; in the case of *Mangifera,* the need for conservation is particularly acute. Human population growth in South and Southeast Asia is driving rapid deforestation in the regions, threatening *Mangifera* trees themselves as well as many of the megafaunal frugivores they depend on for seed dispersal. Of the 35 *Mangifera* species for which the IUCN has sufficient data, two are considered extinct in the wild, one is critically endangered, 9 are endangered, 12 are vulnerable, and 3 are near threatened (IUCN 2012). In part because of their recalcitrant seeds, which do not survive drying and freezing (Mukherjee and Litz 2009), *Mangifera* species are also severely underrepresented in genebank collections (Castañeda-Álvarez et al. 2016). In order to understand and preserve the diversity of *Mangifera* species, the global community of mango breeders, researchers, and germplasm repositories must lead concerted efforts to document, collect, and conserve *Mangifera* both *in situ* and *ex situ*.

Wild ancestors of perennial species are often difficult to locate and identify, as is the case in mango. Wild *M. indica* has historically been reported from Northeastern India (particularly Assam and Sikkim provinces) and is thought to have a range that extends

into parts of Nepal, Bhutan, Bangladesh, and Myanmar (Kostermans and Bompard 1993). However, no recent studies have sought to survey wild populations of *M. indica,* and these individuals have never been subjected to molecular analysis. While germplasm collections for many important perennial crops like apple and citrus contain accessions of wild ancestors and relatives, it appears no accessions of wild *M. indica* exist in global germplasm repositories. Therefore, the location, identification, and preservation of wild *M. indica* is of utmost importance to understanding the evolution and domestication history of the mango and to conserving the priceless diversity of the 'The King of Fruits'.

Literature Cited

Beddington, J., M. Asaduzzaman, M. Clark, A. Fernández, M. Guillou, M. Jahn, L. Erda, T. Mamo, N. Van Bo, C. A. Nobre, R. Scholes, R. Sharma, and J. Wakhungu. 2012. Achieving food security in the face of climate change: Final report from the Commission on Sustainable Agriculture and Climate Change. CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS). Copenhagen, Denmark. Available online at: www.ccafs.cgiar.org/commission.

Bompard, J. M. 2009. Taxonomy and Systematics. Pp. 19–41 *in* R. E. Litz, ed. The Mango: Botany, Production and Uses. CAB International, Cambridge, MA.

Burke, J. M., S. Tang, S. J. Knapp, and L. H. Rieseberg. 2002. Genetic Analysis of Sunflower Domestication. Genetics 161:1257–1267.

Campbell, R. J. 2007. The Potential of New *Mangifera* Species in Florida. Proc. Florida State Hortic. Soc. 120:11–12.

Castañeda-Álvarez, N. P., C. K. Khoury, H. A. Achicanoy, V. Bernau, H. Dempewolf, R. J. Eastwood, L. Guarino, R. H. Harker, A. Jarvis, N. Maxted, J. V. Müller, J. Ramirez-Villegas, C. C. Sosa, P. C. Struik, H. Vincent, and J. Toll. 2016. Global conservation priorities for crop wild relatives. Nat. Plants 2:16022. Nature Publishing Group.

Doebley, J. F., B. S. Gaut, and B. D. Smith. 2006. The Molecular Genetics of Crop Domestication. Cell 127:1309–1321.

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

Esquinas-Alcázar, J. 2005. Science and society: protecting crop genetic diversity for food security: political, ethical and technical challenges. Nat. Rev. Genet. 6:946–53.

FAO. 2010. The Second Report on The State of the World's Plant Genetic Resources for Food and Agriculture. Rome.

Ferreira, M. E. 2006. Molecular analysis of gene banks for sustainable conservation and increased use of crop genetic resources. Pp. 1–5 *in* J. Ruane and A. Sonnino, eds. The Role of Biotechnology in Exploring and Protecting Agricultural Genetic Resources.

Ford-Lloyd, B. V., M. Schmidt, S. J. Armstrong, O. Barazani, J. Engels, R. Hadas, K. Hammer, S. P. Kell, D. Kang, K. Khoshbakht, Y. Li, C. Long, B.-R. Lu, K. Ma, V. T. Nguyen, L. Qiu, S. Ge, W. Wei, Z. Zhang, and N. Maxted. 2011. Crop Wild Relatives—Undervalued, Underutilized and under Threat? Bioscience 61:559–565.

Gross, B. L., and K. M. Olsen. 2010. Genetic perspectives on crop domestication. Trends Plant Sci. 15:529–37. Elsevier Ltd.

Hajjar, R., and T. Hodgkin. 2007. The use of wild relatives in crop improvement: a survey of developments over the last 20 years. Euphytica 156:1–13.

Hatfield, J., and G. Takle. 2014. NCA 2014: Chapter 6 Agriculture. Clim. Chang. Impacts United States Third Natl. Clim. Assess. 150–174.

Huang, H., and L. L. Knowles. 2016. Unforeseen consequences of excluding missing data from next-generation sequences: Simulation study of rad sequences. Syst. Biol. 65:357–365.

Huang, P., J. Molina, J. M. Flowers, S. Rubinstein, S. A. Jackson, M. D. Purugganan, and B. A. Schaal. 2012. Phylogeography of Asian wild rice, *Oryza rufipogon*: a genome-wide view. Mol. Ecol. 21:4593–604.

Hufford, M. B., P. Lubinksy, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand, and J. Ross-Ibarra. 2013. The genomic signature of crop-wild introgression in maize. PLoS Genet. 9:e1003477.

Iqbal, M. J., O. U. K. Reddy, K. M. El-Zik, and A. E. Pepper. 2001. A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. TAG Theor. Appl. Genet. 103:547–554.

IUCN. 2012. World Conservation Monitoring Centre 1998. *IUCN Red List of Threatened Species. Version 2012.2.* Available at: www.iucnredlist.org.

Kassa, M. T., R. V. Penmetsa, N. Carrasquilla-Garcia, B. K. Sarma, S. Datta, H. D. Upadhyaya, R. K. Varshney, E. J. B. von Wettberg, and D. R. Cook. 2012. Genetic patterns of domestication in pigeonpea (*Cajanus cajan* (L.) Millsp.) and wild *Cajanus* relatives. PLoS One 7:e39563.

Kostermans, A. J. G. H., and J. M. Bompard. 1993. The Mangoes: Their Botany, Nomenclature, Horticulture, and Utilization. Academic Press, San Diego, CA.

Leaché, A. D., B. L. Banbury, J. Felsenstein, A. N. M. De Oca, and A. Stamatakis. 2015. Short tree, long tree, right tree, wrong tree: New acquisition bias corrections for inferring SNP phylogenies. Syst. Biol. 64:1032–1047.

Leaché, A. D., and J. R. Oaks. 2017. The utility of single nucleotide polymorphism (SNP) data in phylogenetics. Annu. Rev. Ecol. Evol. Syst. 48:69–84.

Li, C., A. Zhou, and T. Sang. 2006. Rice domestication by reducing shattering. Science. 311:1936–9.

Londo, J. P., Y. Chiang, K. Hung, T. Chiang, and B. A. Schaal. 2006. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. Proc. Natl. Acad. Sci. 103:9578–9583.

Matsuoka, Y., Y. Vigouroux, M. M. Goodman, J. Sanchez G, E. Buckler, and J. Doebley. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. Proc. Natl. Acad. Sci. U. S. A. 99:6080–4.

Maxted, N., and L. Guarino. 2000. Chapter 3: Planning Plant Genetic Conservation. Pp. 37–78 *in* R. D. Smith, J. B. Dickie, S. H. Linington, H. W. Pritchard, and R. J. Probert, eds. Seed Conservation: Turning Science Into Practise. Kew Publishing, London.

Miller, A. J., and B. L. Gross. 2011. From forest to field: perennial fruit crop domestication. Am. J. Bot. 98:1389–414.

Mohammadi, S. A. A., and B. M. Prasanna. 2003. Analysis of Genetic Diversity in Crop Plants — Salient Statistical Tools. Crop Sci. 43:1235–1248.

Mudge, K., J. Janick, S. Scofield, and E. E. Goldschmidt. 2009. A History of Grafting. Pp. 437–494 *in* J. Janick, ed. Horticultural Reviews. John Wiley & Sons, Inc.

Mukherjee, S. K., and R. E. Litz. 2009. Introduction: Botany and Importance. Pp. 1–18 *in* R. E. Litz, ed. The Mango: Botany, Production and Uses. CAB International, Wallingford, UK.

Phillipps, Q., and K. Phillipps. 2016. Mammals of Borneo. Princeton University Press, Princeton, NJ.

Pickersgill, B. 2007. Domestication of plants in the Americas: insights from Mendelian and molecular genetics. Ann. Bot. 100:925–40.

Prusky, D., I. Kobiler, I. Miyara, and N. Alkan. 2009. Fruit Diseases. Pp. 210–230 *in* R. E. Litz, ed. The Mango: Botany, Production and Uses. CAB International, Cambridge, MA.

Saintenac, C., D. Jiang, S. Wang, and E. Akhunov. 2013. Sequence-based mapping of the polyploid wheat genome. G3 Genes, Genomes, Genet. 3:1105–14.

Schoen, D. J., and A. H. Brown. 1993. Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. Proc. Natl. Acad. Sci. U. S. A. 90:10623–7.

Singh, S. P., P. Gepts, and D. G. Debouck. 1991. Races of Common Bean. Econ. Bot. 45:379–396.

The Hague Conference. 2010. The Hague Conference on Agriculture, Food Security and Climate Change. Rome, Italy.

Tripp, E. A., Y. H. E. Tsai, Y. Zhuang, and K. G. Dexter. 2017. RADseq dataset with 90% missing data fully resolves recent radiation of *Petalidium* (Acanthaceae) in the ultra-arid deserts of Namibia. Ecol. Evol. 7:7920–7936.

Wang, R., A. Stec, J. Hey, L. Lukens, and J. Doebley. 1999. The limits of selection during maize domestication. Nature 398:236–239.

Warschefsky, E., R. V. Penmetsa, D. R. Cook, and E. J. B. von Wettberg. 2014. Back to the wilds: Tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. Am. J. Bot. 101:1791–800.

VITA

EMILY WARSCHEFSKY

2009	B.A., Biology Reed College Portland, Oregon
2009-2011	Plant Conservation Research Intern Fairchild Tropical Botanic Garden & Gemini Botanic Gardens Miami, Florida
2011-2018	PhD Student Florida International University Miami, Florida

PUBLICATIONS AND PRESENTATIONS

2018 von Wettberg, E.J.B. and 40 others (E. Warschefsky, author number 17). Ecology and community genomics of an important crop wild relative expands access to adaptations as a prelude to agricultural innovation. Nature Communications 9(649): 1-13. doi:10.1038/s41467-018-02867-z

2018 Warschefsky, E. and E.J.B. von Wettberg. Mango genebank diversity. Annual Plant and Animal Genome Conference. San Diego, CA. Oral presentation.

2018 Warschefsky, E. and E.J.B. von Wettberg. A fruitful phylogeny: Revealing the evolutionary relationships of *Mangifera* species. Annual Plant and Animal Genome Conference. San Diego, CA. Oral presentation.

2017 Warschefsky, E. and E.J.B. von Wettberg. A fruitful phylogeny: Using RADseq to reveal the evolutionary relationships of *Mangifera* species. XIX International Botanical Congress. Shenzhen, China. Oral presentation.

2017 Warschefsky, E. and E.J.B. von Wettberg. A fruitful phylogeny: Using RADseq to reveal the evolutionary relationships of *Mangifera* species. Annual Meeting of the Society for the Study of Evolution. Portland, OR. Oral presentation.

2017 Warschefsky, E. and E.J.B. von Wettberg. Domestication genetics of the mango. Annual Plant and Animal Genome Conference. San Diego, CA. Oral presentation.

2016 Warschefsky, E.J, L.L. Klein, M.H. Frank, D.H. Chitwood, J.P. Londo, E.J.B. von Wettberg, and A.J. Miller. COVER, FEATURE REVIEW: Rootstocks: diversity,

domestication and impacts on shoot phenotypes. Trends in Plant Science 21(5): 418-437. doi:10.1016/j.tplants.2015.11.008.

2016 Warschefsky, E. and E.J.B. von Wettberg. A fruitful phylogeny: Using RADseq to reveal the evolutionary relationships of *Mangifera* species. Annual Meeting of the Botanical Society of America. Savannah, GA. Oral presentation.

2016 Warschefsky, E. A Phylogeny for the genus *Mangifera*. The 10th International Flora Malesiana Symposium. Edinburgh, Scotland, UK. Oral presentation.

2016 Warschefsky, E. and E.J.B. von Wettberg. Tree Crop Diversity. Gene Conservation of Tree Species Workshop. Chicago, IL. Oral presentation.

2014 Roorkiwal, M., E.J.B. von Wettberg, H. Upadhyay, E. Warschefsky, A. Rathore, and R. Varshney. Exploring germplasm diversity to understand the domestication process in *Cicer* spp. using SNP and DArT markers. PLoS ONE 9(7): e102016. doi:10.1371/journal.pone.0102016

2014 Warschefsky, E., R. Penmetsa, D. Cook, and E.J.B. von Wettberg. Back to the wilds: Tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. American Journal of Botany 101(10): 1791-1800. doi:10.3732/ajb.1400116

2014 von Wettberg, E.J.B., E. Warschefsky, P. Chang, J. Berger, A. Greenspan, B. Bukun, A. Aydogan, A. Kahraman, S. Nuzhdin, R.V. Penmetsa, D. Cook. Population genetics of wild chickpea. Annual Meeting of the Botanical Society of America. Boise, ID. Oral presentation.

2014 Warschefsky, E., R. Campbell, and E.J.B. von Wettberg. Insight into the genetic composition of the mango *(Mangifera indica* L.). Annual Meeting of the Botanical Society of America. Boise, ID. Poster presentation.

2013 Warschefsky, E., R. Campbell, T. Hidayat, and E.J.B. von Wettberg. A preliminary phylogeny for *Mangifera.* Annual Meeting of the Botanical Society of America. New Orleans, LA. Poster presentation.

2011 Warschefsky, E.*, B. Jestrow, and E.J.B. von Wettberg. Patterns of genetic variation in *Chamaecrista* species of South Florida. Annual Meeting of the Botanical Society of America. St. Louis, MO. Poster presentation.