SYSTEMATICS AND EVOLUTION OF COMMIPHORA JACQ. (BURSERACEAE) IN MADAGASCAR

by

Morgan Robert Gostel

A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Environmental Science and Public Policy

Committee:

_____ Dr. Andrea Weeks, Dissertation Director

_____ Dr. Cody Edwards, Committee Member

_____ Dr. James Lawrey, Committee Member

_____ Dr. Donald Seto, Committee Member

_____ Dr. Albert Torzilli,
Graduate Program Director

_____ Dr. Robert Jonas, Department Chairperson

_____ Dr. Donna Fox, Associate Dean,
Student Affairs & Special Programs,
College of Science

_____ Dr. Peggy Agouris, Dean,
College of Science

Date: _____ Summer Semester 2015
George Mason University
Fairfax, VA

Systematics and Evolution of *Commiphora* Jacq. (Burseraceae) in Madagascar

A dissertation submitted in partial fulfillment of the requirements for the Doctor of Philosophy at George Mason University

By

Morgan Robert Gostel
Master of Science
Virginia Commonwealth University, 2010
Bachelor of Science
Virginia Commonwealth University, 2008

Director: Andrea Weeks, Associate Professor
Department of Biology

Summer Semester 2015
George Mason University
Fairfax, VA

Acknowledgements

I owe my deepest gratitude to so many for their support, both emotional and financial, and the ability to begin and complete this dissertation. The genealogy of this work extends quite far and begins of course, as all living things, with my family for their encouragement of curiosity, learning, and appreciation for the beauty in nature. I was first inspired to pursue a career in biological research by my friend and mentor, Dr. Peter deFur, whose levity of spirit, compassion, and earnestness is a reminder of what is most important in life. My interest in botany began as an undergraduate student at Virginia Commonwealth University in the laboratory of Dr. Gregory M. Plunkett, who first offered me the opportunity to pursue graduate research. My Masters research gave way to the opportunity to work with Dr. Andrea Weeks, who has been my lifeline, a rock, and the most supportive advisor I could ever imagine. My deepest thanks to Dr. Weeks for taking me on as a Ph.D. student and giving me the opportunity to study *Commiphora*, for the many late night and weekend responses to emails and edits, and for not giving up on me or my work and encouraging me to do the same.

At George Mason University (GMU) I would like to thank the rest of my committee, Drs. James Lawrey, Cody Edwards, and Don Seto for their support and guidance on my project and recommendations along the way. I am especially grateful to Dr. Edwards for the opportunity to work as a graduate research assistant in the Office of the Provost during the last year of my study, and thus expanding my interest in education. I owe much thanks to the Department of Environmental Science and Policy Executive Committee, especially Dr. Albert Torzilli, Sharon Bloomquist, Roslyn Cress, Joanne Zimmerman, and Susan Cheselka, for administrative support, financial support through teaching assistantships and two research fellowships that greatly improved my dissertation research, and ongoing assistance throughout the five years of my study at GMU. I also extend my thanks Dr. Reid Schwebach for offering guidance and assistance toward my pedagogical development and making me a better teacher.

I thank members of my wider professional network, who have supported me throughout the years, sometimes simply by 'being there' and other times for reminding me of my part in the 'bigger picture'. I especially thank the Botanical Society of America and Bill Dahl for inviting me to become involved in the BSA. Immense thanks to the many organizations that have helped to fund this research, including the National Science Foundation, the American Society of Plant Taxonomists, and the GMU Executive Committee. I am grateful for the help of colleagues throughout the course of this work, especially Pete Phillipson and Jeannie Raharimampionona, whose expertise and familiarity with the Malagasy *Commiphora* was of incomparable aide to this project. I hope to spend many more hours in the herbarium with you both! Many thanks to Dr. Pete

Lowry II, Richard Razakamalala, and Charles Rakotovao at the Missouri Botanical Garden and Jackie Andriantiana and Hanta Razafindraibe at PBZT in Madagascar, without whom my collections in Madagascar would not have been possible. I could not ask for better company in the field and I hope to see you all soon!

I give thanks to my friends, old and new, near and far, who have helped make this possible. This work is the culmination of a series of life events, good, bad, and everything in between. You've been there with me all along, but thanks especially to JD, Nate, Critter and Stephy. Thank you for being there and giving me support when I needed it, I love you! Thanks to my family, Gostel and Dal Forno, for their love and support over the years. John, Mary, Laura, and Robbie, you have each been rockstars! Thanks for being here for Manu and me, you have helped make Fairfax much more of a home for us. Obrigrado, Silvani, Luis, Anna, e Lara para todo seu amor e apoio, amo vocês! Most importantly I thank my wife, Manu, who brings happiness, love, and meaning to every day in my life.

All of you have played a role in one of the most important chapters in my life so far, thank you for reminding me that I will never stop learning and for giving me the chance to continue doing so.

Table of Contents

Page

List of Tables vi
List of Figures vii
List of Appendices ix
Abstract x
Chapter 1: Systematics of the myrrh genus, *Commiphora*, globally and in Madagascar, 1797–present 1
Introduction 1
References 5
Chapter 2: Phylogenetic reconstruction of the myrrh genus, *Commiphora* (Burseraceae), reveals multiple radiations in Madagascar and clarifies infrageneric relationships 7
Abstract 7
Introduction 8
Commiphora in Madagascar 11
Objectives and Hypotheses 12
Materials and Methods 12
Taxonomic Sampling 12
DNA Extraction and Molecular Sampling 14
Phylogenetic Inference 15
Divergence Dating Analysis 18
Results 19
Sequence Data 19
Discussion 22
Monophyly of Commiphora and its Taxonomic Sections 22
Species Evolution of Commiphora in Madagascar 23
Dating the Divergence of Commiphora 24
Well-supported Lineages within Commiphora 25
Conclusions 30
References 31
Chapter 3: Development of novel EPIC markers from EST databases and evaluating their phylogenetic utility in *Commiphora* (Burseraceae) 48
Abstract 48
Introduction 49
Methods and Results 49
Marker development 49
Taxonomic sampling and molecular methods 50

Phylogenetic analyses 51
Marker evaluation 52
Critical assessment of primer design criteria 53
Conclusions 53
References 54
Chapter 4: Nuclear phylogenomic analysis of angiosperms using multiplexed, PCR-based target enrichment: A case study using *Commiphora* (Burseraceae) 59
Abstract 59
Introduction 60
Materials and Methods 64
In-silico development: target selection and primer design 64
Primer design 66
Primer pooling & multiplexing 67
Taxon sampling, DNA extraction and standardization 68
Primer validation 69
Microfluidic PCR amplification, sequencing, and assembly of nuclear gene regions 70
Multiple sequence alignment and phylogenetic analysis 71
Results 72
Target selection, primer design, and validation 72
Multiplexed, microfluidic PCR enrichment and sequencing of nuclear gene regions 73
Phylogenetic analysis of individual loci and concatenated datasets 75
Discussion 78
Effective application of multiplexed microfluidic PCR for phylogenomics 79
Phylogenetic utility of targeted nuclear gene regions 81
References 83
Chapter 5: A partial taxonomic revision of the Rhynchocarpa clade of *Commiphora* (Burseraceae), endemic to Madagascar 104
Abstract 104
Introduction 105
Objectives 106
Materials and Methods 107
Key to species of *Commiphora* in the 'Rhynchocarpa' clade with stellate pubescence 109
Taxonomic treatment 110
1. *Commiphora andranovoryensis* Phillipson, Raharim., & Gostel, sp. nov. 110
2. *Commiphora aprevalii* Guillaumin Bull. Soc. Bot. France 56: 1909. 113
3. *Commiphora elliptica* Phillipson, Raharim., & Gostel, sp. nov. 119
4. *Commiphora falcata* Phillipson, Raharim., & Gostel, sp. nov. 124
5. *Commiphora morondavensis* Phillipson, Raharim., & Gostel, sp. nov. 128
6. *Commiphora razakamalalae* Gostel, Phillipson & A. Weeks 132
References 135

List of Tables

Table Page
Table 1. Taxonomic divisions of *Commiphora* according to treatments by Wild (1959), Vollesen (1985), and Gillett (1991). 36
Table 2. Statistics for each of the five markers employed in this study. 38
Table 3. Marker names, primer sequences and phylogenetic statistics for the novel nuclear EPIC markers and the benchmark nrDNA ETS and ITS regions. 56
Table 4. List of accessions sampled in this study with their DNA concentration ([DNA]), numbers of loci (No. loci) with at least 15 sequence reads mapped, raw reads, filtered and trimmed reads, and uniquely mapped reads, as well as the average and median number. 88
Table 5. Sequence characteristics for 49 nuclear loci used for phylogenetic analyses in this study. 95

List of Figures

Figure Page
Figure 1. Morphological features among species of *Commiphora* referred to throughout this manuscript. 40
Figure 2. Results of phylogenetic analysis of markers sampled for all 159 taxa included in this study. 43
Figure 3. Maximum likelihood tree resulting from ten-step, batch-run GARLI analyses. 45
Figure 4. 50% majority-rule consensus tree following 10 million generations of our fossil-calibrated divergence dating analysis in BEAST v1.6.1. 47
Figure 5. Phylogeny of 14 representative taxa in the Burseraceae sampled in this study. 57
Figure 6. Phylogeny of Malagasy *Commiphora* showing 'Arafy' and 'Rhynchocarpa' clades, resulting from RAxML analysis of 44 nuclear loci. 98
Figure 7. Phylogeny of Malagasy *Commiphora* showing 'Arafy' and 'Rhynchocarpa' clades, resulting from RAxML analysis of 41 nuclear loci containing <25% missing sequences. 99
Figure 8. Phylogeny of Malagasy *Commiphora* showing 'Arafy' and 'Rhynchocarpa' clades, resulting from RAxML analysis of 20 nuclear loci that contains ≤1% missing sequences. 100
Figure 9. Reduced, 65 taxon phylogeny of Malagasy *Commiphora* showing 'Arafy' and 'Rhynchocarpa' clades, resulting from RAxML analysis of 44 nuclear loci. 101
Figure 10. Reduced, 65 taxon phylogeny of Malagasy *Commiphora* showing 'Arafy' and 'Rhynchocarpa' clades, resulting from RAxML analysis of 41 nuclear loci. 102
Figure 11. Reduced, 65 taxon phylogeny of Malagasy *Commiphora* showing 'Arafy' and 'Rhynchocarpa' clades, resulting from RAxML analysis of 20 nuclear loci. 103
Figure 12. Reduced phylogeny of *Commiphora* from Gostel et al. (in press). Four Malagasy clades are highlighted in red. 137
Figure 13. Scanning electron microscopy (SEM) of indument features. 138
Figure 14. Illustration of *Commiphora andranovoryensis*. Drawn by Bobbi Angell. 139
Figure 15. Map of the known distribution of *Commiphora andranovoryensis*, based upon collections used for descriptions in this study. 140
Figure 16. Map of the known distribution of *Commiphora aprevalii*, based upon collections used for descriptions in this study. 141
Figure 17. Illustration of *Commiphora ellitpica*. Drawn by Bobbi Angell. 142
Figure 18. Map of the known distribution of *Commiphora ellitpica*, based upon collections used for descriptions in this study. 143
Figure 19. Map of the known distribution of *Commiphora falcata*, based upon collections used for descriptions in this study. 144
Figure 20. Illustration of *Commiphora morondavensis*. Drawn by Bobbi Angell. 145

Figure 21. Map of the known distribution of *Commiphora morondavensis*, based upon collections used for descriptions in this study. 146
Figure 22. Illustration of *Commiphora raxakamalalae*, drawn by Bobbi Angell. 147
Figure 23. Map of the known distribution of *Commiphora razakamalalae*, based upon collections used for descriptions in this study. 148

List of Appendices

Page
Appendix 1 159
Appendix 2 172
Appendix 3 177
Appendix 4 181
Appendix 5 189
Appendix 6 195
Appendix 7 199
Appendix 8 248

Abstract

SYSTEMATICS AND EVOLUTION OF *COMMIPHORA* JACQ. (BURSERACEAE) IN MADAGASCAR

Morgan Robert Gostel, Ph.D.

George Mason University, 2015

Dissertation Director: Dr. Andrea Weeks

The myrrh genus, *Commiphora*, comprises a clade of nearly 200 species of shrubs and trees that grow in warm tropical regions in Africa, Madagascar, the Arabian Peninsula, the Indian sub-continent, and South America. *Commiphora* is the most species-rich genus in the frankincense and myrrh family, Burseraceae, and species belonging to this genus are ecologically important throughout their range in eastern, sub-Saharan Africa and western Madagascar. Aromatic oleoresins extracted from several species have been used extensively as an olfactory aesthetic and medicine with demonstrated pharmacological benefits throughout both antiquity and in contemporary folk medicinal practices. Despite its diversity and ecological and economical significance, evolutionary relationships in the genus are poorly understood and few studies have sought to reconstruct the phylogenetic history of *Commiphora*. As a result, species boundaries and infrageneric relationships of this widespread group of plants are not well characterized. Species of *Commiphora* are

morphologically diverse and the genus provides an opportunity to study the evolutionary significance of traits such as the presence of thorns, production of oleoresins, and diverse types of indumentum. This dissertation research seeks to reconstruct the evolutionary history of *Commiphora*, with an emphasis on species in Madagascar, all of which are endemic. Toward this objective, we have applied molecular phylogenetic methods to resolve infrageneric relationships in the genus using molecular markers developed from three approaches. The first approach, detailed in chapter two, samples molecular markers that have been designed from previously published and widely sampled genetic loci for phylogenetic reconstruction in angiosperm genera, including two nuclear ribosomal markers (ETS and ITS) and three chloroplast spacers (*ndhF-rpl32*, *psbA-trnH*, and *trnD-trnT*). Our second approach, described in chapter three, uses molecular markers designed from a bioinformatics pipeline specifically targeting conserved nuclear loci predicted to be within close proximity to more informative, intronic regions of the nuclear genome. The third approach utilizes microfluidic PCR techniques and Illumina MiSeq to sample a set of putative shared, single-copy nuclear genomic loci. We screened 192 primer pairs for their phylogenomic utility in *Commiphora*. Ninety-one of these primer pairs amplified a single product and 49 sequenced loci were used for comparative phylogenetic analyses to reconstruct evolutionary relationships among species of Malagasy *Commiphora*. Our results suggest that previous attempts to circumscribe the diversity of *Commiphora* produce unnatural groups, *Commiphora* has experienced complex biogeographic radiations, diversity in the genus is characterized by strong geographic structure, and expanded taxonomic and genomic sampling improves our ability to discern infrageneric

groups. We have also begun a partial and ongoing revision of the genus in Madagascar, including a revision of six species, five of which are described as new species and all are categorized according to IUCN Red List criteria as either endangered or vulnerable. We outline priorities for future studies in this group, which include expanded taxonomic revision and molecular systematics research to improve species delimitation and better understand evolutionary trajectories. A key priority is to sample species from tropical east Africa.

Chapter 1: Systematics of the myrrh genus, *Commiphora*, globally and in Madagascar, 1797–present

Introduction

One of the most fundamental objectives of modern biological research is the unification of observations about life and biological diversity with the modern or evolutionary synthesis. This synthesis seeks to describe the origins of such diversity according to the theory of evolution. Throughout the 20th Century and now into the 21st Century, an accumulated knowledge base has provided important tools toward the study of evolutionary histories and phenomena. Indeed, cataloging the diversity of life and interpreting the evolutionary trajectory of biochemical, ecological, and morphological traits can shed light upon the ability of biological systems to adapt and survive. An estimated 400,000 species of angiosperms exist globally. As such, angiosperms represent one of the most diverse clades of organisms on Earth and having originated only ca. 180–140 mya (estimates vary, see Wikström et al. 2001, Magallón et al. 2015). Collectively, plant systematists have made important discoveries that reveal the complexity of organsimal evolution and improve our understanding of the diversification of flowering plant lineages. *Commiphora* is a diverse genus of shrubs and trees that belongs to the frankincense and myrrh family of angiosperms (order Sapindales). It is ecologically and

economically important and its evolutionary history is the principle subject of study in this dissertation.

The myrrh genus, *Commiphora*, is the most species-rich genus in Burseraceae, with ca. 200 species distributed throughout seasonally dry tropical forests in both the neo- and paleotropics. *Commiphora* is most diverse in tropical east Africa (especially Ethiopia and Somalia) and Madagascar, each home to ca. 100 and 50 species, respectively. Ecologically, *Commiphora* is a major constituent of the expansive *Acacia-Commiphora* woodland in tropical east Africa (Olson and Dinerstien 2002) and the western dry deciduous forest in Madagascar (Humbert 1964). Economically, several species of *Commiphora* have been regarded as highly valued commodities since antiquity (Van Beek 1960, Thulin and Claeson 1991, Kulhari et al. 2012). Despite its diversity and prevalence in human society and the environment, few studies have attempted to circumscribe species belonging to the genus. This most likely can be attributed to the large number of species in the genus, their remote location, and challenging state of collected material.

All species of *Commiphora* are deciduous, mostly dioecious, and usually precocious (producing their inflorescence prior to leaves). As a result, many herbarium specimens may only include some, but not all of the parts often necessary to make a proper species determination, including leaves, both bisexual and male flowers, fruit, and bark characteristics. The combination of these characters is difficult for botanists to collect and study, because it requires fieldwork that may span multiple growing seasons in remote tropical regions. Nevertheless, this challenge has not stopped some authors

from describing species based upon what might be considered inadequate material both long ago (noted in Gillet 1973) and more recently (e.g., Cheek and Rakotozafy 1991).

The nomenclatural history of *Commiphora* is also marked by some confusing details resulting from description of the type, *C. madagascariensis* Jacq., in 1797. Despite its epithet, which suggests the species was originally collected in either Madagascar or Mauritius, no collections of the species have been made from either country in the wild or from cultivation. The species closely resembles another species from tropical east Africa that may have been cultivated there and traded when this species was described (Gillett 1991). The type itself is quite distinct from most other species of *Commiphora* from Madagascar because it has true thorns – a feature shared with only one other species from the island (*C. simplicifolia* Engl.) – but *C. madagascariensis* and *C. simplicifolia* are clearly different species (leaflets have dentate margins and lateral leaflets are very rare and smaller in the latter than in the former). Regardless, progress toward the circumscription of this genus must be made carefully in the years ahead.

Five authors have attempted to circumscribe diversity in *Commiphora* since the mid-19th Century, including Berg (1862), Engler (1912), Wild (1959), Vollesen (1985), and Gillett (1991), summarized below in chapter 2. More recently, several studies have attempted to reconstruct the evolutionary history and test the monophyly of existing sectional classification of *Commiphora* and related genera in the Burseraceae using molecular methods (Clarkson et al. 2002, Weeks et al. 2005, Weeks and Simpson 2007, Becerra et al. 2012). Results from these recent molecular studies have suggested that existing sectional classification in *Commiphora* does not correspond to natural lineages, a

conclusion that merits further examination through expanded sampling. Chapter three includes a discussion on the monophyly of existing taxonomic sections.

With regard to the species of *Commiphora* in Madagascar, two authors, Perrier de la Bâthie and Capuron, are responsible for describing much of the currently recognized species diversity. Additional species have been described elsewhere, most notably by Engler (1883), but largely only one at a time and by non-specialists in the genus (Cheek and Rakotozafy 1991, Bardot-Vaucoulon 2002). Since this time, approximately 18 new species from Madagascar have been identified, but have not yet been published because they lack sufficient material for satisfactory description or are included in this dissertation (see Chapter 5). In total, we now recognize 50 species of *Commiphora* in Madagascar, including 28 described species, two new species that are currently recognized as varieties (but are distinct species), and an additional 20 undescribed species (five of which are described in this dissertation).

Traditional molecular markers, including proposed 'barcode' loci (Hollingsworth et al. 2011), have proven insufficiently variable to test hypotheses regarding interspecific relationships in *Commiphora* and instead, much of the phylogenetic resolution provided by these genetic loci has revealed clades within *Commiphora* that are large polytomies of multiple species. In order to better understand patterns of species evolution in *Commiphora*, particularly among species in Madagascar, we have developed two sets of novel molecular markers, which is described in chapters three and four of this dissertation.

Producing a fully resolved phylogeny for all species of *Commiphora* remains a challenge that will require ongoing research, however the work described in this dissertation represents important progress toward this objective. With improved phylogenetic resolution, *Commiphora* may serve as a model to test biogeographic hypotheses of species evolution in Madagascar and to infer the trajectory of important ecological and morphological innovations in the genus.

References

Bardot-Vaucoulon, M. 2002. Une nouvelle espèce de *Commiphora* (Burseraceae) du Nord de Madagascar. Adansonia 24: 43–47.

Becerra, J. X., K. Noge, S. Olivier, and D. L. Venable. 2012. The monophyly of *Bursera* and its impact for divergence times of Burseraceae. Taxon 61: 333–343.

Berg, O. 1862. Die Balsamodendron-Arten der Berliner Herbarien. Botanische Zeitung 20: 162.

Capuron, R. 1962. Contributions a l'étude de la flore forestière de Madagascar. VI. Note sure les Burseracees. Adansonia 2: 268–283.

Cheek, M. and A. Rakotozafy. 1991. The identity of Leroy's fifth subfamily of the Meliaceae, and a new combination in *Commiphora* (Burseraceae). Taxon 40: 231–237.

Clarkson, J. J., M. W. Chase, M. M. Harley. 2002. Phylogenetic relationships in Burseraceae based on plastid *rps* 16 intron sequences. Kew Bulletin 57: 183–193.

Engler, A. 1883. Burseraceae in De Candolle (Ed.), Monographiae Plantarum 4, pp. 3–60 and t. l.

Engler, A. 1912. Die Verbreitung der afrikanischen Burseraceae und die einteilung der Gattung *Commiphora*. Eng. Bot. Jahrb. 48: 449–490.

Gillett, J. B. 1973. *Commiphora* Jacq. (Burseraceae): Englerian species which "Disappear". Kew Bulletin 28: 25–28.

Gillett, J. B. 1980. *Commiphora* (Burseraceae) in South America and its relationship to *Bursera*. Kew Bulletin 34: 569–587.

Hollingsworth, P. M., S. W. Graham, and D. P. Little. 2011. Choosing and using a plant DNA barcode. PLoS One 6: e19254.

Kulhari, A., A. Sheorayan, S. Kalia, A. Chaudhury, R. K. Kalia. 2012. Problems, progress and future prospects for improvement of *Commiphora wightii* (Arn.) Bhandari, an endangered herbal magic, through modern biotechnological tools: a review. Genetic Resources and Crop Evolution 59: 1223–1254.

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

Perrier de la Bâthie, H. 1946. Burséracées. In: Humbert, H. (Ed.), Flore de Madagascar et des Comores. Imprimerie officielle, Tananarive, Madagascar, pp. 1–50.

Thulin, M. and P. Claeson. 1991. The botanical origin of scented myrrh (Bissabol or Habak Hadi). Economic Botany 45: 487–494.

Van Beek, G. W. 1960. Frankincense and myrrh. The Biblical Archaeologist 23: 69–95.

Vollesen, K. 1985. Studies in Burseraceae of Northeastern Africa. Kew Bulletin. 40: 39–76.

Weeks, A., D. C. Daly, and B. B. Simpson. 2005. The phylogenetic history and biogeography of the frankincense and myrrh family (Burseraceae) based on nuclear and chloroplast sequence data. Molecular Phylogenetics and Evolution 35: 85–101.

Weeks, A. and B. B. Simpson. 2007. Molecular phylogenetic analysis of *Commiphora* (Burseraceae) yields insight on the evolution and historical biogeography of an "impossible" genus. 42: 62–79.

Wikström, N., V. Savolainen, and M. W. Chase. 2001. Evolution of the angiosperms: calibrating the family tree. Proceedings of the Royal Society of London B 268: 2211–2220.

Chapter 2: Phylogenetic reconstruction of the myrrh genus, *Commiphora* (Burseraceae), reveals multiple radiations in Madagascar and clarifies infrageneric relationships

Gostel, M. R, Phillipson, P. B., and A. Weeks. Accepted with major revision to *Systematic Botany* 1 June 2015.

Abstract

The myrrh genus, *Commiphora* (Burseraceae), is the most species-rich in the frankincense and myrrh family, Burseraceae, and it presents several interesting distributions. The taxonomy at both species and infrageneric levels has been problematic and we evaluate existing classifications. Recent taxonomic investigations have revealed some 20 new species *Commiphora* in Madagascar in addition to those already described, for a total of approximately 25% of the entire genus. All of the Malagasy species are endemic to the island. Previous phylogenetic studies in *Commiphora* that included species from Madagascar have indicated that the genus immigrated to and radiated within the island twice. We seek to reconstruct species-level relationships in *Commiphora* using a more exhaustive taxon sampling to test this biogeographic hypothesis more thoroughly using a nearly comprehensive sampling of species from Madagascar. We employed phylogenetic reconstruction methods using five molecular markers (nrETS, nrITS, *psbA*–

trnH, *ndhF–rpl32*, and *trnD–trnT*). Results from this expanded sampling support a monophyletic *Commiphora* and indicate strong support for a global total of seven clades that we refer to as the 'Lasiodisca,' 'Granulifera,' 'Saxicola', 'Gariepensis,' 'Spinescens,' 'Arafy,' and 'Rhynchocarpa' clades. Furthermore, our results show that *Commiphora* is represented in Madagascar by four clades, two of which are species-rich. We describe the morphological and geographic affinities of each of these seven clades and identify priorities for future study in the group.

Introduction

Commiphora Jacq. is the most species-rich genus in the frankincense and myrrh family, Burseraceae, comprising approximately 190 tree and shrub species that are widely distributed in warm tropical regions including continental Africa, Madagascar and the other western Indian Ocean islands, the Arabian Peninsula, the Indian sub-continent, and South America. Species diversity is concentrated in the *Acacia-Commiphora* woodland of tropical east Africa (Olson and Dinerstein 2002) and the western dry deciduous forest of Madagascar (Humbert 1965). Species of *Commiphora* are typically dioecious, mostly frost-intolerant and restricted to seasonally dry tropical or sub-tropical forests and arid scrub/thickets, or rock outcrops. Most but not all species have exfoliating bark and produce aromatic oleoresins; xeric-adapted species are often pachycaulous and have branches armed with thorns. Their alternate leaves are trifoliolate or imparpinnate (rarely unifoliolate) and have brochidodromous venation (Fig. 1E). They can also be

recognized by their drupaceous fruits, which split at maturity into two (rarely four) valves to reveal a putamen that is partially surrounded by a fleshy red, orange, or yellow pseudaril. Their flowers are borne in paniculate or reduced (1-3 flowered) cymose inflorescences.

Many species descriptions have been based upon insufficient plant material, owed in part to the fact that species may be dioecious, deciduous in habit, and may flower before leafing out, which resulted in uncertainty in matching flowering, fruiting and vegetative material (Gillett 1973). Consequently, the interpretation of types and other original material can be challenging. Following early treatments, such as Engler's description of African Burseraceae (1912) and Wild's (1959) classification of the genus, the taxonomy of *Commiphora* has been driven by regional floristic treatments, such as those from northeastern and tropical east Africa (Vollesen 1985 and Gillett 1991, respectively) and southern Africa (Van der Walt 1973). For Madagascar, Perrier de la Bâthie's treatment in the *Flore de Madagascar et les Comores* (Perrier de la Bâthie 1946) recognized 20 species of *Commiphora* to which later authors have added another eight (Capuron 1962; Cheek and Rakotozafy 1991; Bardot-Vaucoulon 2002). However, recent taxonomic work revealed an additional 20 species from Madagascar (Phillipson, pers. comm.) that are currently in the process of being described. Several Malagasy species were described based on multiple syntypes (e.g., *C. arafy* H. Perrier, five syntypes; *C. guillauminii* H. Perrier, seven syntypes; *C. tetramera* Engl.. six syntypes) and many of these syntypes are clearly either distinct species or in some rare cases may not even represent species of *Commiphora*, a good example of the confusion caused by incorrect

association of material collected in different stages of growth. In these cases, the necessary lectotypification is ongoing (Phillipson et al. in mss.). Additional nomenclatural confusion exists with regard to the type species in the genus, *C. madagascariensis* Jacq., which was either collected originally in either Madagascar or Mauritius (as the epithet implies) when it was described in 1797. No collections of this species have since been made from either country, in the wild or otherwise, and Gillett (1991) suggested it most closely resembles a species found throughout various parts of India that was likely cultivated and traded widely at the time of its description. *C. madagascariensis* is armed with true thorns and the only species from Madagascar known to bear thorns, *C. simplicifolia* Engl., is clearly distinct.

Infrageneric classification within *Commiphora* is also unresolved. Engler (1912) recognized 43 sections that relied heavily on leaf characters, whereas Wild (1959) erected two subgenera, five sections, and 11 subsections that emphasized floral and fruit characters, such as the number and arrangement of stamens, disc lobing, and the shape and coverage of the pseudaril. Gillett (1991) recognized 14 sections in the genus, by following 12 of 13 outlined by Vollesen (1985), adding two (Sections *Abyssinicae* and *Hemprichia*) and including Vollesen's monotypic section, *Monoicae* within *Hemprichia*. A comparison of classifications is provided in Table 1. To date, only two studies have tested species-level phylogenetic relationships for *Commiphora* using molecular data (Becerra et al. 2012 and Weeks and Simpson 2007) and both suggest that at least one of the published infrageneric groups is artificial.

Commiphora in Madagascar

Madagascar harbors the second richest center of species diversity for *Commiphora* following tropical (North-) East Africa and likely includes the greatest number of endemic species by area. Madagascan *Commiphora* comprise 28 described and at least 16 undescribed species, all of which are endemic. The presence of this genus, along with species of the legume genera *Dalbergia* and *Hildegardia*, is regarded as an important ecological indicator for the western dry deciduous forest and southern arid spiny bush vegetation zones (Humbert 1965). Despite the concentration of unique *Commiphora* species in Madagascar, only a single author included them in an infrageneric taxonomic classification (185 species, Wild 1959). Previous molecular studies suggested that *Commiphora* is represented in Madagascar by two radiations (Becerra et al. 2012; Weeks and Simpson 2007), but each of these studies sampled only a minority of Malagasy species (13 and 11, respectively). Weeks and Simpson (2007) indicated that two Malagasy colonizations in *Commiphora* arrived on Madagascar during the early Miocene (ca. 18–19 mya) and radiated contemporaneously during the late Miocene (9–10 mya). It is unknown whether increased sampling of species of *Commiphora* from Madagascar may reveal additional independent, endemic lineages or change the inferred ages of their arrival and diversification throughout the island. For these reasons, we focused on Madagascar as a sampling priority for *Commiphora*, but when possible we have expanded our sampling to include all geographic localities from which the genus is known.

Objectives and Hypotheses

Our overall objective is to reconstruct phylogenetic relationships in *Commiphora* using a thorough taxonomic sampling and through the analysis of sequence data from five nuclear and chloroplast loci. We tested two hypotheses: (a) that the nine polytypic taxonomic sections of *Commiphora* are monophyletic; and (b) that the endemic Malagasy *Commiphora* originated via two independent dispersals from continental Africa. The results will allow us to evaluate the morphological and historical biogeographical evolution of the genus as a whole and with particular emphasis on the endemic lineages in Madagascar.

Materials and Methods

Taxonomic Sampling

Specimen sampling for this study included material from recent fieldwork in Namibia and Madagascar, as well as material from numerous herbaria (see Appendix 1). In total, our sampling included 159 ingroup and outgroup accessions. The majority (110/159, ca. 69%) of these accessions are represented in our molecular phylogeny by DNA from silica-preserved leaf tissue collected and dried in the field, however a subset of accessions (50/159, ca. 31%) is represented by material for DNA obtained from herbarium tissue. Ingroup sampling included silica- and herbarium-dried leaf tissue from 105 species of *Commiphora*, 22 of which were represented by multiple accessions. These 105 ingroup species represent a comprehensive sampling of subgenera, sections, and

subsections recognized by Wild (1959) and all but one (the monotypic section *Ugogenses*) of the fourteen described by Gillett (1991) (Appendix 1). In total, our species sampling included members from Gillett's (1991) nine polytypic sections: *Abyssinicae* (four of thirteen species), *Commiphora* (one of three species), *Campestres* (one of six species), *Africanae* (three of seven species), *Latifoliolatae* (five of seven species.), *Arillopsidium* (one of eight species), *Hildebrandtianae* (two of four species), *Hemprichia* (one of six species), and *Opobalsameae* (one of two species) and four monotypic sections: *Ciliatae*, *Coriaceae*, *Rostratae*, and *Pedunculatae*. Table 1 lists each taxonomic section and the species we have sampled from each. Our emphasis on Malagasy taxa resulted in nearly comprehensive species sampling (25 of 28 currently recognized species, 89%) as well as varieties (*C. aprevalii* var. *granulifera* and *C. orbicularis* var. *tulearensis*), subspecies (*C. brevicalyx* ssp. *vezorum*), plus sixteen of the species that are as yet undescribed (Phillipson et al. in mss.). Outgroup sampling included twelve species of *Bursera*: six species from Neotropical *B.* subg. *Elaphrium*, five species from Neotropical *B.* subg, *Bursera*, and one unplaced Vietnamese species, *B. tonkinensis*. These outgroups were selected on the basis of previous family-level phylogenetic studies that indicate that *Commiphora* is sister to or nested within *Bursera* (Clarkson et al. 2002, Weeks et al. 2005, Weeks et al. in mss.). Complete voucher information for each accession is included in Appendix 1.

DNA Extraction and Molecular Sampling

Whole genomic DNA was extracted from each accession using the FastPrep FastDNA® spin kit (Bio101 Systems, La Jolla, California). Five loci were sampled from the nuclear and chloroplast genomes. The two nuclear markers selected were the 3' terminus of nrDNA external transcribed spacer (ETS) and the complete nrDNA internal transcribed spacer (ITS), which included the ITS1 and ITS2 intergenic spacers flanking the 5.8S gene. The three chloroplast markers sampled included three intergenic spacer regions *psbA–trnH*, *ndhF–rpl32*, and *trnD–trnT*. Loci were selected based on their utility in other phylogenetic studies of *Commiphora* species (*psbA–trnH*; Weeks and Simpson 2007) or the utility in other angiosperm taxa (*ndhF–rpl32* and *trnD–trnT*; Shaw et al. 2005, 2007). The primers and thermocycling protocol used for PCR amplification of the loci are listed in Table 2, along with characteristics for each amplified region: aligned length, number of informative characters, and percent missing data, as calculated by SeqState (Müller 2005). For amplification of nrDNA ETS and cpDNA *psbA–trnH* regions, 15 μL PCR conditions included 0.75 μL of both forward and reverse primers (5 μM), 0.75 μL spermidine (4 mM), 3 μL H_2O, 2.25 μL total DNA, and 7.5 μL GoTaq green mastermix (Promega Corp., Madison, Wisconsin) *Taq* polymerase mix. 25 μL PCR conditions were used in the amplification of ITS, *ndhF–rpl32*, and *trnD–trnT* regions and included 1.25 μL of both forward and reverse primers (5 μM), 1.25 μL spermidine (4 mM), 5 μL H_2O, 3.75 μL total DNA, and 12.5 μL GoTaq green mastermix. All amplification products were purified prior to sequencing reactions using 1.5 μL exonuclease I and 3 μL shrimp alkaline phosphatase per 5 uL of PCR product (USB

Corp., Cleveland, Ohio) and a single thermocyler step, which included 30 minutes at 37°C followed by 15 minutes at 80°C. Purified PCR products were sequenced directly using a thermocycler program for 20 seconds at 94°C, 15 seconds at 55°C, and 1 minute at 60°C for 30 cycles. Sequencing reactions were carried out using Sanger dideoxy termination by Macrogen Inc. (Rockville, Maryland). Sequencing products were assembled and edited using Sequencher 4.7 (Gene Codes Corp., Ann Arbor, Michigan). The PCR primer sequence and statistics for each molecular marker used in this study are provided in Table 1. Sequence alignment was performed using MUSCLE version 3.7 (Edgar 2004) in the CIPRES web portal (Miller et al. 2010). Gap regions resulting from inferred insertion-deletion (indel) events among sequences in the multiple sequence alignment were treated as missing data. All sequences have been deposited in GenBank (Appendix 1) and multiple sequence alignment files have been uploaded to Dryad (DOI: http://dx.doi.org/10.5061/dryad.vb766).

Phylogenetic Inference

To determine phylogenetic congruence among the five loci sampled, we performed independent searches on each as well as on combined plastid datasets, combined ETS and plastid datasets, and combined five marker dataset from a reduced sample of sequences that included all five molecular markers. The reduced taxon sampling in the five marker dataset focused on two species-rich clades of Malagasy *Commiphora* species, the ‘Arafy’ and ‘Rhynchocarpa’ clades, and allowed us to include ITS sequences that could not be aligned with confidence across other ingroup species.

Maximum parsimony analyses were performed using PAUP* 4.0b10 (Swofford 2002) and a three-step heuristic search protocol modified from Plunkett et al. (2005). In the first step a heuristic search of 1,000 replicates was performed using random, stepwise addition and TBR branch swapping, with no more than 100 trees saved during each replicate. Trees saved during step one were loaded as starting trees for step two, which performed a search using one replicate, saving a maximum of 100,000 trees. The strict consensus from trees resulting from step two was loaded as a topological constraint for a third step, which repeated the same search protocol in step one. If no shorter trees were found during this third step, the strict consensus of step two was used as a conservative estimate of phylogenetic relationships. Branch support for maximum parsimony analyses was inferred by bootstrapping 1,000 replicates in PAUP* (Swofford 2002) using a heuristic search with TBR branch swapping and saving a maximum of 100 trees per replicate.

Model-based analyses (ML and BI) were performed on complete taxon datasets of ETS and the three chloroplast loci, as well as the reduced taxon dataset of ITS for the Malagasy 'Arafy' and 'Rhynchocarpa' clades. The correct model of nucleotide substitution was determined using JModelTest 0.1.1 (Posada 2008) for alignments of each marker and these models were implemented in Garli 1.0 (Zwickl 2006) and MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003) for maximum likelihood and Bayesian inference, respectively. Model-based analyses of the concatenated datasets were analyzed using partitioned analyses. Models for each partition corresponded to nucleotide substitution rate matrices identified by JModelTest 0.1.1 and additional parameters for site heterogeneity and the proportion of invariant sites, I. All model-based phylogenetic

reconstruction was performed using the CIPRES web portal (Miller et al. 2010). GARLI analyses consisted of a 10-series, partitioned analysis, with each series searching a maximum of 5,000,000 generations that was terminated if no new or significantly better topologies were found after 20,000 generations. Each GARLI series consisted of two replicates. The first series began with starting trees constructed by stepwise addition and the nine subsequent series used starting trees that represented the tree with the best likelihood score from the previous search series. The ML bootstrap analyses were performed with the same parameters as the 10-series searches, but with only one set of 100 replicates. Bayesian inference (BI) analyses were performed using MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003) with two simultaneous runs consisting of four chains each for 10 and 50 million generations, sampling every 1,000 and 2,000 generations, respectively. Stationarity resulting from BI analyses was determined by three criteria: a) runs must have achieved average standard deviation of split frequencies among chains less than 0.01; b) visual inspection of plotted likelihood scores showed convergence; and c) visual inspection of plotted model output parameters showed good mixing. All sampling prior to reaching stationarity in BI runs was discarded as burn-in and the remaining samples were used to determine branch support on topologies as posterior probability estimates. MP and ML analyses were not used for phylogenetic reconstruction of single marker analyses (except for reduced taxon datasets corresponding to the 'Arafy' and 'Rhynchocarpa' clades). Instead, congruence in phylogenetic signal between individual marker alignments was determined by analyses using BI only, using the same

search criteria as above, but with 50 million generations and sampling every 1,000 generations.

Divergence Dating Analysis

Fossil-calibrated divergence dating was carried out using BEAST v1.6.1 (Drummond and Rambaut 2007). The .xml files for BEAST analyses were produced using the version of BEAUti which is packaged with this software and these configuration files are available at TreeBASE submission ID: 16632. Two nodes were designated as fossil calibration points for this analysis based on assignment of fossils in previous studies and the corresponding published divergence dates for the Burseraceae, which assigned minimum age constraints to stem *Bursera* + *Commiphora* (48.6 mya) and crown *Bursera* subg. *Elaphrium* (40.4 mya, De Nova et al. 2012). Two fossils were used to calibrate analyses that were the basis for these dates. The first fossil is *Protocommiphora europea* Reid and Chandler (Reid and Chandler 1933, Collinson 1983), which is assignable to stem *Commiphora* + *Bursera* on the basis of a pyrene surface that is assignable to either *Commiphora* or *Bursera* subgenus *Elaphrium*. The second fossil, *Bursera inaequalateris* (Lesquereux) MacGinitie (MacGinitie 1969) includes leaves that resemble *Bursera* subgenus *Elaphrium* (Weeks et al. 2005). Divergence dating analyses were carried out using a lognormal relaxed clock with normal priors on calibration points that captured the 95% HPD of dates published by De Nova et al. (2012). Our BEAST run utilized our combined ETS-plastid dataset, searching 10 million generations, saving every 2,000 generations. 30% of these generations were

discarded as burn-in following visual verification of stationarity as a function of likelihood scores.

Results

Sequence Data

579 new sequences were produced for analyses used in this study (see Table 2 and Appendix 1). Forty-three previously published sequences for ETS and 41 sequences for *psbA–trnH* were included in this study from Weeks and Simpson (2007), while all sequences for ITS, *ndhF–rpl3,* and *trnD–trnT* were newly generated. Statistics for each locus, including its aligned length, number of parsimony informative characters (PICs), and percent missing data, are summarized in Table 2. We were not able to amplify or sequence ITS sequences cleanly for 33 ingroup taxa. We attribute these difficulties to a long poly-C (15–20) region at the 5'-end of ITS2 and likely secondary structure in ITS1 that could not be surmounted using alternative reagents, polymerases or temperature cycling parameters. ITS chromatogram data from species that were sequenced successfully did not suggest the presence of multiple, divergent copies. ITS sequences across ingroup species were marked by numerous insertion/deletion mutations and regions of divergent nucleotides (data not shown), which resulted in highly speculative assumptions of homology in the alignments. Subsets of ITS sequences could be aligned with confidence for individual clades, as identified by preliminary analyses of other loci, so we elected to focus on those from the two largest clades of Malagasy species, in line with the objectives of study. The ITS phylogenies of the Malagasy 'Arafy' and

‘Rhynchocarpa’ clades are congruent with those produced using other nuclear and chloroplast loci (Fig. S1), thus we retained ITS sequence data as part of a reduced-taxon dataset.

The concatenated chloroplast dataset included an aligned length of 3,960 characters, 308 (7.8%) of which were parsimony informative. Analysis of the combined chloroplast data (Fig. 2A) clearly identified seven well-supported major clades that we informally named the ‘Lasiodisca’ (100% MPBS, 100% MLBS, and 1.0 PP, Figs. 2 and 3), ‘Arafy’ (94% MPBS, 95% MLBS, 1.0 PP, Figs. 2 and 3), ‘Spinescens’ (61% MPBS, 92% MLBS, and 1.0 PP, Figs. 2 and 3), ‘Granulifera’ (51% MPBS, 89% MLBS, 1.0 PP), ‘Gariepensis’ (<50% MPBS, 56% MLBS, 1.0 PP, Figs. 2 and 3), ‘Saxicola’ (100% MPBS, 100% MLBS, 1.0 PP, Figs. 2 and 3), and ‘Rhynchocarpa’ (97% MPBS, 99% MLBS, 1.0 PP, Figs. 2 and 3) clades for the purposes of further analysis and later discussion. Analyses of individual chloroplast marker datasets are provided for comparison in Figs. S2–S4.

Topologies resulting from analysis of ETS (Fig. S5) were congruent with those produced by chloroplast loci, with the exception of four species (*Commiphora kraeuseliana, C. wildii, C. anacardifolia,* and *C. glaucescens*). In the analysis of ETS, these four species are not resolved within the ‘Gariepensis’ clade, which indicates some incongruence between these branches of the nuclear and chloroplast gene trees (the ‘Granulifera’ and ‘Spinescens’ clades are not resolved, Fig. S5). Overall, the ETS tree is otherwise largely congruent with the chloroplast gene trees, but the ETS phylogeny did not resolve well-supported ‘Granulifera’ or ‘Spinescens’ clades (Fig. S5). In order to

maximize the amount of phylogenetic information from sequences produced in this study we concatenated ETS and chloroplast alignments (Fig 2B). The strict consensus of all trees presented in Fig. 2B does not resolve the larger, most inclusive ‘Granulifera’ and ‘Spinescens’ clades. However, parsimony bootstrap and Bayesian analyses do, with 61% MPBS and PP 1.0, and 51% MPBS and PP 1.0, respectively (data not shown). The smaller ‘Gariepensis’ clade is not supported by any parsimony-based analysis but does have a Bayesian posterior probability value of 1.0 (data not shown). Maximum likelihood analyses of the concatenated ETS and chloroplast data also supports the monophyly of ‘Granulifera’ (56% MLBS), ‘Spinescens’ (92% MLBS), and ‘Gariepensis’ (92% MLBS) clades (Fig. 3).

As described above, we performed phylogenetic analyses on reduced-taxon datasets for each of the 'Arafy’ and ‘Rhynchocarpa’ clades in order to retain the maximum amount of ITS data. These analyses included independent reduced-taxon datasets for the concatenated chloroplast markers, ETS, and ITS alignments (Fig. S1). Analyses of the ‘Arafy’ and ‘Rhynchocarpa’ clades were not capable of providing sufficient resolution of interspecific relationships; however, the analysis of both the chloroplast and ETS datasets independently recovered a well-supported clade of species with small leaflets (less than 30 mm wide) in the Arafy clade (1 and 0.97 PP, respectively, Fig. S1). Nearly half of all subclades resolved as a result of phylogenetic analysis of the reduced-taxon chloroplast and ETS datasets included only species-specific subclades (10/23 and 9/19 subclades were species-specific, respectively, Fig. 1). The majority of subclades resulting from the analysis of reduced-taxon datasets for ITS in

both the ‘Arafy’ and ‘Rhynchocarpa’ clades were unique to this dataset and indicate substantial incongruence between ITS sequences and the other marker sequences (Fig. S1). Furthermore, the results of phylogenetic analysis of the ‘Arafy’ and ‘Rhynchocarpa’ taxon sampling using ITS failed to recover several species-specific subclades that were resolved through analysis of other datasets (Fig. S1). Therefore, it appears the most incongruent of the three marker datasets for each of the ‘Arafy’ and ‘Rhynchocarpa’ clades was the ITS alignment.

Phylogenetic analyses of all 159 taxa revealed seven well-supported clades. The four most inclusive we informally refer to as the ‘Lasiodisca’, ‘Arafy’, ‘Spinescens’, and ‘Granulifera’ clades (Figs. 2 and 3). Nested within the ‘Granulifera’ clade are three additional well-supported subclades (Figs. 2 and 3), which we refer to the ‘Gariepensis’, ‘Saxicola’, and ‘Rhynchocarpa’ clades. Each of these clades is described in detail in the discussion, below.

Discussion

Monophyly of Commiphora and its Taxonomic Sections

Our analyses support a monophyletic *Commiphora*, as expected (Figs. 2, 3, and supplemental data). In this study we find limited support for our first hypothesis that Gillett’s (1991) polytypic infrageneric sections are individually monophyletic. All species included in this study that were also placed by Gillett (1991) into sections are included within the ‘Spinescens’ clade. Among those polytypic sections of Gillett represented by multiple taxa in our study, species circumscribed within three (*Abyssinicae*, *Africanae*,

and *Latifoliolatae*) are found within a weakly supported subclade within the 'Spinescens' clade (56% MLBS, Fig. 3, 0.95 PP, data not shown). The lack of resolution within this subclade means that we cannot rule out their reciprocal monophyly. This subclade also contains species circumscribed within two other polytypic sections for which we have sampled only a single taxon and one of Gillett's five monotypic sections (*Commiphora*, *Hemprichia*, and *Rostratae*, respectively). We sampled two species from section Hildebrandianae (*C. corrugata* and *C. alaticaulis*), which do not appear closely related in our phylogeny (Fig. 3). We sampled four of Gillett's (1991) five monotypic sections (*Rostratae*, *Coriaceae*, *Pedunculatae*, and *Ciliatae*, but not *Ugogenses*) but their resolution within the 'Spinescens' clade was insufficient to determine their relationship to the other polytypic sections. Five polytypic sections sampled in this study (*Arillopsidium*, *Campestres*, *Commiphora*, *Hemprichia*, and *Opobalsameae*) were represented by only a single species in this study and we were unable to test their monophyly. We recommend expanded sampling in each of the polytypic sections, which will require improved sampling from tropical east Africa, and we recommend broader infrageneric classification to include species from the additional six clades we have identified in this study.

Species Evolution of Commiphora in Madagascar

Results from our phylogenetic reconstruction do not support our second hypothesis, that *Commiphora* is represented in Madagascar by two endemic lineages. Instead, we find that four distinct lineages of endemic species inhabit Madagascar. Two

of these comprise the species-rich lineages previously identified in Weeks and Simpson (2007), the 'Arafy' and 'Rhynchocarpa' clades. The remaining two endemic lineages include species unsampled by Weeks and Simpson (2007) and have unexpected placements including *C. lasiodisca*, which is sister to all other *Commiphora* species, and *C. simplicifolia*, which is embedded within the 'Spinescens' clade and sister to widespread tropical East African species.

Dating the Divergence of Commiphora

Results of our divergence dating analyses support our expectation that present-day diversity in *Commiphora* underwent much of its radiation during the Miocene (23.03–5.33 mya). The crown ages for each of the seven well-supported clades identified in this study occur within the Miocene; however the crown age for the genus itself is slightly older (36.6±9.2 mya, Fig. 4) than previous estimates (27.8±4.5 mya, Weeks and Simpson 2007). The inclusion of *C. lasiodisca* and its position as sister to the rest of the species in *Commiphora* is likely responsible for this older crown age. Results from our divergence dating analysis date the split of *Commiphora* from *Bursera* at 48±5.3 mya (Fig. 4) within the estimates provided by previous studies (47.3±5.7 mya, Weeks and Simpson 2007). Below, we look at each of the subclades identified within *Commiphora* and discuss implications for morphological evolution, biogeography, and the relative timing of divergence.

Well-supported Lineages within Commiphora

Seven clades are recognized in *Commiphora* and each is described below. These clades do not correspond to any previously recognized taxonomic sections, but we have described general features that may help future classification.

COMMIPHORA LASIODISCA—The placement of this species as sister to all remaining species in the genus is surprising for a number of reasons. The branch that separates *Commiphora lasiodisca* from the rest of the species in the genus is quite long (Fig. 3); 9.5 million years elapse between its divergence and radiation of the remaining ingroup species. *C. lasiodisca* is readily distinguishable from all other species of *Commiphora* due to a combination of traits, including the absence of true thorns and an exceptionally dense, long, uniseriate pubescence that entirely covers the leaves and inflorescences at both juvenile and mature stages. The inflorescence of *C. lasiodisca* is pseudoracemose, with flowers born in dense glomerules on a long panicle. It is also the only Malagasy species that lacks a brightly colored, fleshy pseudaril covering the putamen. Naked putamens are present in a few other continental African species (*C. capensis*, *C. cervifolia*, and C. *kraeuseliana*) and in *Bursera*, the genus sister to *Commiphora*.

GRANULIFERA CLADE— The 'Granulifera' clade includes 38 species of *Commiphora* distributed throughout Namibia (14 species) and Madagascar (24 species and two infraspecific taxa). This clade is strongly supported in analyses of both the combined plastid (61% MPBS, 1.0 PP Fig. 2A) and combined ETS-plastid datasets (51% MPBS, 1.0 PP, and 89% MLBS, Figs. 2B and 3). In analyses of ETS data alone, the 'Granulifera' clade is not supported due to the unresolved position of four species (*Commiphora*

kraeuseliana, *C. anacardifolia*, *C. wildii*, and *C. glaucescens*). The name of this clade refers to the glandular indumentum present on many of its members. (Fig. 1D). Its presence and relative density on the reproductive and vegetative parts is variable but typically includes unicellular or uniseriate secretory trichomes. Other indument characteristics present on species belonging to this clade include long or short, uniseriate or stellate (stalked and unstalked) trichomes. Our divergence dating analysis indicated that this clade has crown age of 19.92 ma (11.933–27.76 ma, 95% HPD, Fig. 4).

This clade is characterized by a surprising disjunct distribution in southwest Africa and Madagascar. The continental African species form a sister group to the 'Rhynchocarpa' clade of Malagasy species in analyses of plastid datasets (Fig. 2A), but nuclear data suggest that the continental African clade is paraphyletic with the 'Saxicola' subclade sister to the 'Rhynchocarpa' clade (Figs. 2 and 3).

RHYNCHOCARPA CLADE— Species belonging to this clade are characterized by a predominance of ellipsoid fruits that taper to a narrow, apical beak (Fig. 1F) and all are endemic to Madgascar. The 'Rhynchocarpa' clade is the most-species rich of the Malagasy lineages (26 spp.) and is distributed throughout the latitudinal range of the western dry deciduous forests of Madagascar. Two species in the 'Rhynchocarpa' clade are widespread throughout the range of *Commiphora* in Madagascar, *C. grandifolia* and *C. marchandii*; however, the remaining 24 species are restricted more locally to dry, often calcareous or siliceous substrates including limestone karst or *tsingy* and low-elevation dry forest on sandy substrates. It is well supported in both combined ETS-plastid and plastid phylogenies (100% MPBS, 1.0 PP, and 97% MLBS, Figs. 2 and 3).

Divergence dating analyses (Fig. 4) suggest that the ‘Rhynchocarpa’ clade has radiated recently in Madagascar, diverging from its sister lineage approximately 8.4 ma (5.24–12.96 ma, 95% HPD, Fig. 4).

Within this clade, interspecific relationships are not well-resolved, perhaps due to the group’s relatively recent origin (Fig. 4A). 21 of the 26 species belonging to this clade share ellipsoid or beaked fruits, which are 1.5 – 5 times as long as they are wide. Improved phylogenetic resolution will be necessary to understand the morphological and geographic evolutionary trajectory the Rhynchocarpa clade in Madagascar.

SAXICOLA CLADE—The ‘Saxicola’ clade is named for the species *Commiphora saxicola*, which is common across the rocky, xeric shrubland of South Africa and Namibia. This clade, nested within the larger ‘Granulifera’ clade, includes three species (*C. saxicola*, *C. kuneneana*, and *C. crenato-serrata*) that form a sister group (MLBS: 81%, Fig. 3, PP: 1.0, data not shown) with the more species-rich ‘Rhynchocarpa’ clade. These three species, like many species placed within the larger ‘Granulifera’ clade, have a glandular indumentum and milky or opaque white oleoresins. Oleoresins of these Nambian species are fetid-smelling, whereas those of the Madagascan ‘Rhynchocarpa’ clade are pungently sweet-smelling or unscented. What is most striking is the presence of elongate, pedunculate inflorescences and infructescences, plus fruits that are that closely resemble those in the sister ‘Rhynchocarpa’ clade. The split between the ‘Saxicola’ and ‘Rhynchocarpa’ sister group is dated to 15.56 ma (8.55–23.27 ma, 95% HPD, Fig. 4) and the age of the crown group comprising the ‘Saxicola’ clade has been dated to 4.03 ma (1.42–7.64 ma, 95% HPD, Fig. 4).

GARIEPENSIS CLADE—This clade is comprised of eight species distributed throughout Angola, Namibia, and South Africa. The name ‘Gariepensis’ refers to the Gariep physiographic region in southern Namibia, as well as to a species in the clade, *Commiphora gariepensis*. Together with the ‘Rhynchocarpa’ and ‘Saxicola’ clades, the ‘Gariepensis’ clade comprises the remainder of the larger, ‘Granulifera’ clade. All of the species belonging to the ‘Gariepensis’ clade are characterized by sessile, glandular trichomes and the lack of true thorns. The pseudaril is incredibly variable in this clade. Most species belonging to this clade have a pseudaril with four shallow lobes that cover ≤ 1/3 of the lower putamen, however one species has four pseudaril arms that reach to approximately 3/4 of the putamen (*C. discolor*) and two species that lack a pseudaril altogether (*C. capensis* and *C. cervifolia*), Two species in this clade, *C. gracilifrondosa* and *C. oblanceolata*, are haplostemonous. The crown age for the ‘Gariepensis’ clade is dated to 16.95 ma (9.73–24.62 ma, 95% HPD, Fig. 4) in our divergence dating analyses.

ARAFY CLADE—The ‘Arafy’ clade is so named for the Malagasy vernacular term for several species belonging to this clade and for one of the species within the clade, *Commiphora arafy* (Fig. 1E). Species belonging to this clade are concentrated in the arid spiny bush vegetation zone of southern and southwestern Madagascar. Among the 16 species belonging to this clade, ten are restricted to this region, while the ranges for the remaining six species extend north into the western dry deciduous forest. The majority of species in this clade are characterized by a shrubby, often pachycaulous habit and small leaves, although northern species can reach over 20 m (Fig. 2). The ‘Arafy’ clade is not well resolved in our phylogenetic reconstruction; only one subclade has strong branch

support values (*C. arafy* - *C.* sp. nov. C clade). Divergence-dating analyses indicate that the 'Arafy' clade is the younger of the two large Malagasy radiations of *Commiphora*, with a mean crown age of 6.8 ma (3.48–11.79 ma, 95% HPD, Fig. 4).

SPINESCENS CLADE—This clade is so named because it contains all species of *Commiphora* armed with true thorns, as well as many species that are unarmed. The 'Spinescens' clade (61% MPBS, 1.0 PP, 92% MLBS) is composed of 48 species that encompass the greatest range of morphological and geographic diversity of any clade within the phylogeny. Among the most geographically disparate species belonging to this clade are *C. wightii* from India, *C. planifrons* and *C. foliaceae* from Socotra, and *C. leptophloeos*, the sole Neotropical species distributed in Bolivia and Brazil. All of these species are armed with thorns, The geographic extent of the 'Spinescens' clade suggests that this trait may have allowed species to undergo range expansion and to persist after immigration. It is surprising that while several other lineages in *Commiphora* (e.g., the 'Rhynchocarpa' and 'Arafy' clades) have undergone considerable speciation following their arrival in new areas, the geographically disparate species belonging to this clade have not. Results from divergence dating analyses suggest that this clade was established ca. 20.4 ma (14–28.8 ma, Fig. 4).

Two additional morphological traits appear to be restricted to this clade, although not universally shared among its species: a pericarp having four-valves (instead of two) and a pseudaril that has four arms ascending to the putamen apex. The four-valved pericarp has arisen twice, once in *C. ciliata* and again in a clade containing *C. chiovendana*, *C. gileadensis*, and *C. coronillifolia*. It is most parsimonious that the four-

armed pseudaril has arisen independently three times in this clade, once in a subclade containing 25 species (MLBS: 56%, Fig. 3, PP: 0.95, data not shown), once in *C. boranensis*, and once again in a subclade containing four species *(C. angolensis, C. neglecta, C. pedunculata,* and *C. tenuipetiolata*, MPBS: 95% and PP: 1.0, Fig. 2, MLBS: 97%, Fig. 3). In the latter subclade, the pseudaril arms never reach the apex of the putamen and a similar pseudaril shape is found in a species belonging to the 'Gariepensis' clade (*C. discolor*). Haplostemony is also present among some members of this clade (*C. gileadensis, C. coronillifolia*) although not restricted to it. This trait is rare within *Commiphora* generally but also occurs in the 'Rhynchocarpa' (*C. ankaranensis*), 'Arafy' (*C. monstruosa*) and 'Gariepiensis' (*C. oblanceolata*) clades.

Conclusions

By increasing the taxonomic sampling and the depth of molecular data over that of Weeks & Simpson (2007), we have demonstrated that existing taxonomic sections are poor predictors of evolutionary relationships in *Commiphora*. None of Gillett's (1991) nine polytypic sections were supported as monophyletic by our study, however we only sampled multiple species from four of these sections and additional taxonomic sampling. We could not reject the monophyly of three of Gillett's polytypic sections (*Africanae, Abyssinicae,* and *Latifoliolatae*), however their placement together within a single subclade suggests their classification may be superficial. Furthermore, our study has revealed that species evolution in *Commiphora* corresponds to strong geographic structure and with the exception of one large, geographically diverse clade (

'Spinescens'), major diversification events have likely resulted from geographic isolation in southwest continental Africa and Madagascar during the Miocene. The position of *C. lasiodisca* as sister to all other species of *Commiphora* suggests that the ancestral geographic area occupied by *Commiphora* included Madagascar. This is consistent with the findings of Weeks et al. (2005), who suggested that *Commiphora* had dispersed and radiated within continental Africa during the middle Eocene, but places the origin of Malagasy *Commiphora* much earlier than that study, during the late Eocene (Fig. 4). The phylogenetic structure of clades revealed by this study suggest several morphological synapomorphies that appear phylogenetically conserved (the presence of true thorns, fruit shape, and pubescence type), while other traits appear more evolutionarily labile and have a convergent history in multiple clades (haplostemony, leaflet shape, and pseudaril shape,). Future studies, including sectional classification, must emphasize expanded taxon sampling, particularly in tropical east Africa, as well as greater genomic sampling to improve phylogenetic resolution and allow for a more detailed inference of the evolutionary trajectories of morphological character states and biogeographic histories.

References

Agnarsson, I., and M. Kuntner. 2012. The generation of a biodiversity hotspot : Biogeography and phylogeography of the Western Indian Ocean islands. Pp. 33–82 in *Current topics in phylogenetics and phylogeography of terrestrial and aquatic systems*, ed. K. Anamthawat-Jonsson. Rijeka: In Tech Publishers.

Álvarez, I. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution.* 29: 417–434.

Bardot-Vaucoulon, M. 2002. Une nouvelle espèce de *Commiphora* (Burseraceae) du Nord de Madagascar. *Adansonia, sér. 3* 24: 43–47.

Buerki, S., D. S. Devey, M. W. Callmander, P. B. Phillipson, and F. Forest. 2013. Spatio-temporal history of the endemic genera of Madagascar. *Botanical Journal of the Linnean Society.* 171: 304–329.

Callmander, M. W., P. B. Phillipson, G. E. Schatz, S. Andriambololonera, M. Rabarimanarivo, N. Rakotonirina, J. Raharimampionona, C. Chatelain, L. Gautier, and P. P. Lowry II, 2011. The endemic and non-endemic vascular flora of Madagascar updated. *Plant Ecology and Evolution* 144: 121–125.

Capuron, R. 1962. Contributions a l'étude de la flore forestière de Madagascar. VI. Note sure les Burséracées. *Adansonia* 2: 268–283.

Cheek, M. and A. Rakotozafy. 1991. The identity of Leroy's fifth subfamily of the Meliaceae, and a new combination in *Commiphora* (Burseraceae). *Taxon* 40: 231–237.

Clarkson, J. J., M. W. Chase, and M. M. Harley. 2002. Phylogenetic relationships in Burseraceae based on plastid *rpsl6* intron sequences. *Kew Bulletin* 57: 183–193.

Collinson, M. E. 1983. Fossil plants of the London Clay. Oxford: Oxford University Press.

Daly, D. C., M. M. Harley, M-C. Martínez-Habibe, and A. Weeks. 2011. Burseraceae. Pp. 76–104 in *The families and genera of vascular plants* vol. 10, ed. K. Kubitzki. New York: Springer-Verlag.

Downie, S. R. and D. S. Katz-Downie. 1996. A molecular phylogeny of Apiaceae subfamily Apioideae: Evidence from nuclear ribosomal DNA internal transcribed spacer sequences. *American Journal of Botany* 83: 234–251.

Drummond, A. J. and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 214.

Edgar, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–7.

Engler, A. 1912. Die verbreitung der afrikanischen Burseraceen im Verhaltnis zu ihrer systematischen Gliederung und die Einteilung der Gattung *Commiphora*. *Botanische Jahrbücher für Systematik, Pflanzengeschichte, und Pflanzengeographie* 48: 443–490.

Gillett, J. B. 1973. *Commiphora* Jacq. (Burseraceae): Englerian species which "disappear". *Kew Bulletin* 28: 25–28.

Gillett, J. B. 1980. *Commiphora* (Burseraceae) in South America and its relationship to *Bursera*. *Kew Bulletin* 34: 569–587.

Gillett, J. B. 1991. Burseraceae. Pp. 1–95 in *Flora of tropical east Africa*, ed. R. M. Polhill. Rotterdam: A. A. Balkema.

Heath, T. A., D. J. Zwickl, J. Kim, and D. M. Hillis. 2008. Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. *Systematic Biology* 57: 160–166.

Humbert, H. 1965. Description des types de vegetation. Pp. 46–78 in *Notice de la carte Madagascar* vol. 6, eds. H. Humbert and G. Cours-Darne. Pondichéry: Travaux de la Section Scientifique et Technique de l'Institut Français de Pondichéry, hors série.

Katoh, K., K. Kuma, H. Toh, and T. Miyata. 2005. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* 33: 511–518.

Kress, W. J., K. J. Wurdack, E. A. Zimmer, L. A. Weigt, and D. H. Janzen. 2005. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences USA* 102: 8369–8374.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.

Luton, K., D. Walker, and D. Blair. 1992. Comparisons of ribosomal internal transcribed spacers from two congeneric species of flukes (Platyhelminthes: Trematoda: Digenea). *Molecular and Biochemical Parasitology* 56: 323–327.

MacGinitie, H. D. 1969. The Eocene Green River flora of northwestern Colorado and northeastern Utah. *University of California Publications in Geological Sciences* 83: 1–140.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Pp. 1-8 in Proceedings of the Gateway Computing Environments Workshop (GCE). New Orleans: Gateway Computing.

Müller, K. 2005. SeqState – primer design and sequence statistics for phylogenetic DNA data sets. *Applied Bioinformatics* 4:65–69.

Myers, N., R. Mittermeier, C.G. Mittermeier, G. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858.

Nieto Feliner, G. and J. A. Rosselló. 2007. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution* 44: 911–919.

Olson, D. M. and E. Dinerstein. 2002. The global 200: Priority ecoregions for global conservation. *Annals of the Missouri Botanical Garden* 89: 199–224.

Perrier de la Bâthie, H. 1946. Burséracées Pp. 1–50 in *Flore de Madagascar et des Comores*, ed. H. Humbert. Tananarive, Madagascar : Imprimerie officielle.

Posada, D. 2008. jModeltest: Phylogenetic model averaging. *Molecular Biology and Evolution* 25: 1253–1256.

Reid, E. M. and M. E. J. Chandler. 1933. The London Clay Flora. British Museum (Natural History), London, England.

Ronquist, F. and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 195–203.

Sang, T., D. J. Crawford, and T. F. Stuessy. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* 84: 1120–1136.

Schatz, G.E., C. E. Birkinshaw, P. P. Lowry II, F. Randriantafika, and F. Ratovoson. 2000. Endemism in the Malagasy tree flora. Pp. 1–8 in *Diversity and endemism in Madagascar*, eds. W. R. Lourenço and S.M. Goodman. Paris: Mémoires de la Société de Biogéographie.

Shaw, J., E. B. Lickey, J. T. Beck, S. B. Farmer, W. Liu, J. Miller, K. C. Siripun, C. T. Winder, E. E. Schilling, and R. L. Small. 2005. The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92: 142–166.

Shaw, J., E. B. Lickey, E. E. Schilling, and R. L. Small. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in Angiosperms: The tortoise and the hare III. *American Journal of Botany* 94:275–288.

Stanford, A. M., R. Harden, and C. R. Parks. 2000. Phylogeny and biogeography of *Juglans* (Juglandaceae) based on *matK* and ITS sequence data. *American Journal of Botany* 87: 872–882.

Swofford, D. L. 1998. PAUP* Phylogenetic analysis using parsimony (*and other methods), v. 4.0 beta 10. Sunderland: Sinauer Associates.

Van der Walt, J. J. A. 1973. The South African species of *Commiphora*. *Bothalia* 11: 53–102.

Vollesen, K. 1985. Studies in the Burseraceae of northeastern Africa. *Kew Bulletin* 40: 39–76.

Weeks, A. and B. B. Simpson. 2004. Molecular genetic evidence for interspecific hybridization among endemic Hispaniolan *Bursera* (Burseraceae). *American Journal of Botany* 91: 976–984.

Weeks, A., and B. B. Simpson. 2007. Molecular phylogenetic analysis of *Commiphora* (Burseraceae) yields insight on the evolution and historical biogeography of an "impossible" genus. *Molecular Phylogenetics and Evolution* 42: 62–79.

Weeks, A., D. C. Daly, and B. B. Simpson. 2005. The phylogenetic history and biogeography of the frankincense and myrrh family (Burseraceae) based on nuclear and chloroplast sequence data. *Molecular Phylogenetics and Evolution* 35: 85–101.

Weeks, A., F. Zapata, S. K. Pell, D. C. Daly, J. D. Mitchell, and P. V. A. Fine. 2014. To move or to evolve: Contrasting patterns of intercontinental connectivity and climatic niche evolution in "Terebinthaceae" (Anacardiaceae and Burseraceae). *Frontiers in Genetics* 5: 1–20.

Wells, N. A. 2003. Some hypotheses on the Mesozoic and Cenozoic paleoenvironmental history of Madagascar. Pp. 16–34 in *The natural history of Madagascar*, eds. S. M. Goodman and J. P. Benstead. Chicago: The University of Chicago Press.

Wild, H. 1959. A revised classification of the genus *Commiphora* Jacq. *Boletim da Sociedade Broteriana* 33: 67–95.

Zwickl, D. J. 2006. *GARLI. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion.* Ph.D. dissertation. Austin, Texas: University of Texas.

Zwickl, D. J. and D. M. Hillis. 2002. Increased taxon sampling greatly reduces phylogenetic error. *Systematic Biology* 51: 588–598.

Table 1. Taxonomic divisions of *Commiphora* according to treatments by Wild (1959), Vollesen (1985), and Gillett (1991).
Asterisks (*) indicate sections not sampled in this study. Sections or subsections in bold are either monotypic or only represented by a single species in our study sample. [a]Includes species from both sections *Commiphora* and *Campestres*, Vollesen (1985) and Gillett (1991). [b]Includes species from both sections *Abyssinicae* and *Campestres*, Vollesen (1985) and Gillett (1991). [c]Includes species from sections *Latifoliolata*, *Campestres*, *Opobalsameae*, *Hildebrandtianae*, and *Rostratae*, Vollesen (1985) and Gillett (1991). [d]Includes species from both sections *Campestres* and *Latifoliolata*, Vollesen (1985) and Gillett (1991). [e]Includes species from both sections *Hildebrandtianae* and *Ugogenses*, Vollesen (1985) and Gillett (1991). [f]Includes species from sections *Africanae*, *Arillopsidium*, and *Coriaceae*, Vollesen (1985) and Gillett (1991). [g]Includes species from both sections *Abyssinicae* and *Rostratae*, Vollesen (1985) and Gillett (1991). [h]Includes species from both sections *Abyssinicae* and *Coriaceae*, Vollesen (1985) and Gillett (1991). [i]Includes species from both sections *Arillopsidium* and *Latifoliolata*, Vollesen (1985) and Gillett (1991). [j]Includes species from section *Arillopsidium*, Vollesen (1985) and Gillett (1991). [k]Includes species from sections *Coriaceae*, *Hemprichia*, and *Pedunculata*. [l]Includes species from both sections *Commiphora* and *Abyssinicae*, Gillett (1991). [m]Includes species from both sections *Arillopsidium* and *Hemprichia*, Gillett (1991).

Wild (1959)	**Vollesen (1985)**	**Gillett (1991)**
Commiphora subgenus		
Commiphora		
Section 1: *Commiphora*	**Sect. *Commiphora*[l]**	**Sect. *Commiphora***
Subsection 1:		**Sect. *Abyssinicae***
Madagascarienses[a]		
Subsection 2:	**Sect. *Campestres***	**Sect. *Campestres***
Pyracanthoides[b]		
Subsection 3:		
Quadricinctae[c]		
Subsection 4:	Sect. *Latifoliolatae*	Sect. *Latifoliolatae*

Latifoliolatae[d]		
Subsection 5:	***Sect. *Ugogenses***	***Sect. *Ugogenses***
Ugogenses[e]		
Subsection 6:	**Sect. *Pedunculatae***	**Sect. *Pedunculatae***
Pedunculatae		
Section 2: *Africanae*[f]	Sect. *Africanae*	Sect. *Africanae*
Section 3: *Rostratae*[g]	**Sect. *Rostratae***	**Sect. *Rostratae***
Section 4: *Coriaceae*	**Sect. *Coriaceae***	**Sect. *Coriaceae***
Subsection 1:	**Sect. *Ciliatae***	**Sect. *Ciliatae***
Rangeanae[h]		
Subsection 2:	Sect. *Hildebrandtianae*	Sect. *Hildebrandtianae*
Teretifoliolatae		
Section 5: *Spondioideae*		
Subsection 1:	**Sect. *Arillopsidium*[m]**	**Sect. *Arillopsidium***
Cupulares[i]		
Subsection 2:		**Sect. *Hemprichia***
Pruinosae[j]		
Subsection 3:	***Sect. *Monoicae***	
Glaucidulae[k]		
Commiphora subgenus II:	**Sect. *Opobalsameae***	**Sect. *Opobalsameae***
Opobalsamum		

Table 2. Statistics for each of the five markers employed in this study.
[a]Weeks and Simpson (2004). [b]Stanford et al. (2000). [c]Luton et al. (1992). [d]Downie and Downie (1996). [e]Primer pairs not previously published. [f]Sang et al. (1997). [g]Shaw et al. (2005). [h]Marker statistics for the ITS region are reported from the 'Matambelo' and 'Arafy' clades of *Commiphora* only. [i]Thermocycler protocol values include temperature (°C):duration (sec) for denaturation / annealing / and extension; each is separated by a forward slash, respectively. All thermocycler protocols include 35 steps and an initial four minute denaturation at 94°C, a final seven minute extension at 72°C and were held at 15°C. [j]Parsimony informative characters.

DNA Region	Primer name (ref)	Primer Sequence (5'–3')	Thermocycler protocol[i]	Aligned length (bp)	Total taxa PICs[j] (# / %)	Ingroup PICs[j] (# / %)	Missing Data (%) / # seqs	Substitution Model
ETS	ETS1F[a]	TTCGGTATCCTG TGTTGCTTAC	94:60/53:60/72:45	418	217 / 52%	204 / 49%	4.4% / 150	TPM2uf+G
	18S2R[a]	GAGACAAGCAT ATGACTACTGGC AGGATCAACCA G						
ITS	ITS5A[b]	CCTTATCATTTA GAGGAAGGAG	94:60/53:60/72:45	(Arafy) 486 (Rhyncho) 533	(Arafy) [h]66 / 13.58% (Rhyncho) 199 / 37.336%	(Arafy) [h]30 / 6.173% (Rhyncho) 184 / 34.522%	25% / 40	(Arafy) TPM1uf+G (Rhynchocarpa) TIM+I+G
	ny43[c]	TATGCTTAAAYT CAGCGGCT						
	ny109Com[e]	GWGACACCCAG GCAGACG						
	ny45[d]	GCATCGATGAAG AACGTAGC						
psbA–trnH	psbAfor[f]	GTTATGCATGAA CGTAATGCTC	94:60/53:60/72:45	699	60 / 8.6%	45 / 6.4%	1.2% / 161	TIM2+G
	trnHrev[f]	CGCGCATGGTGG ATTCACAATC						
ndhF–rpl32	rpl32-F[e]	CCAATATCCCTT YYTTTTCCAA	94:60/53:60/72:45	1,458	124 / 8.5%	94 / 6.5%	2.5% / 158	TVMef+I+G
	5ndhF-F[e]	GAAAGGTATKAT CCAYGMATATT						
	ndhFrpl32-INTF[e]	CTCTGTCTATTTT CGGCATGC						
	ndhFrpl32-INTR[e]	GGCTGGCCCATA ACTTTAACC						

trnD–trnT	trnD-TFCom[e]	GGGAAATCAAATGTACAGC	94:30/54:35/70:60	1,803	126 / 7%	100 / 5.5%	4.7% / 156	TPM1uf+G
	trnDTR[g]	CTACCACTGAGTTAAAAGGG						
	trnD-TMF[e]	GTCGAATCCCCGCTGCCTCCTTG						
	trnD-TMR[e]	GTCCTTCCGATCTAGTCATAC						

Figure 1. Morphological features among species of *Commiphora* referred to throughout this manuscript.
A) Individual of *C. mafaidoha*, a tall, canopy tree ca. 18 m. B) Individual of *C. capuronii*, a small, saxicolous shrub. C) True thorn in *C. simplicifolia*. D) Granular pubescence, observed under 20X magnification in an individual of *C.* sp. nov. E (Gostel 140). E)

Large, globose fruit from individual of *C. arafy*. F) Elongated, lacriform fruit from individual of *C.* sp. nov. I. Scale bars shown.

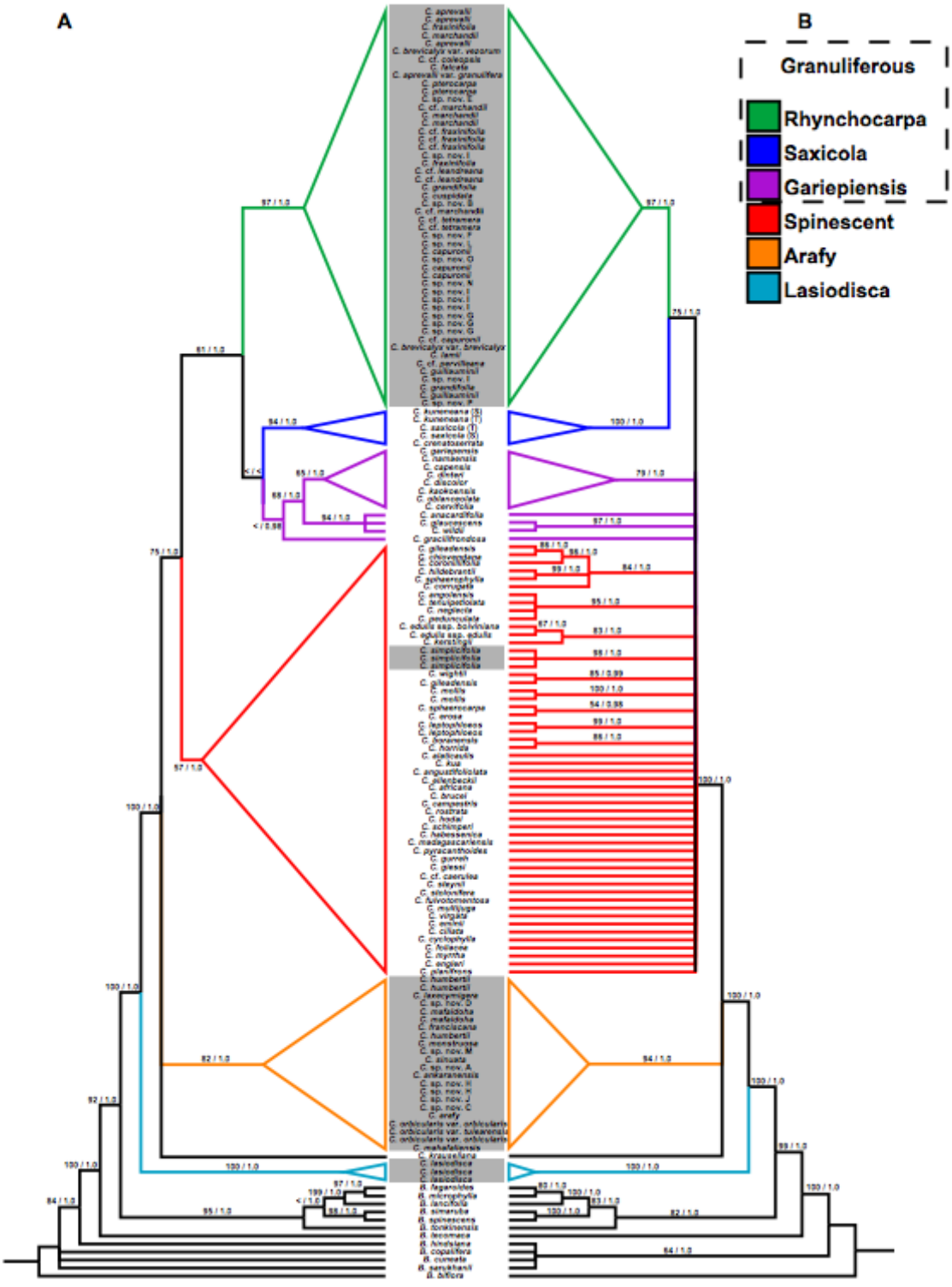

A
B
Granuliferous
Rhynchocarpa
Saxicola
Gariepiensis
Spinescent
Arafy
Lasiodisca

Figure 2. Results of phylogenetic analysis of markers sampled for all 159 taxa included in this study. A) Strict consensus of 100,000 trees resulting from maximum parsimony analysis of the concatenated plastid marker dataset. B) Strict consensus of 100,000 trees resulting from maximum parsimony analysis of the concatenated ETS and plastid marker dataset. MP bootstrap and Bayesian posterior probability support values are provided above branches and separated by a forward slash (e.g., # / #). Shaded taxon labels indicate Malagasy taxa.

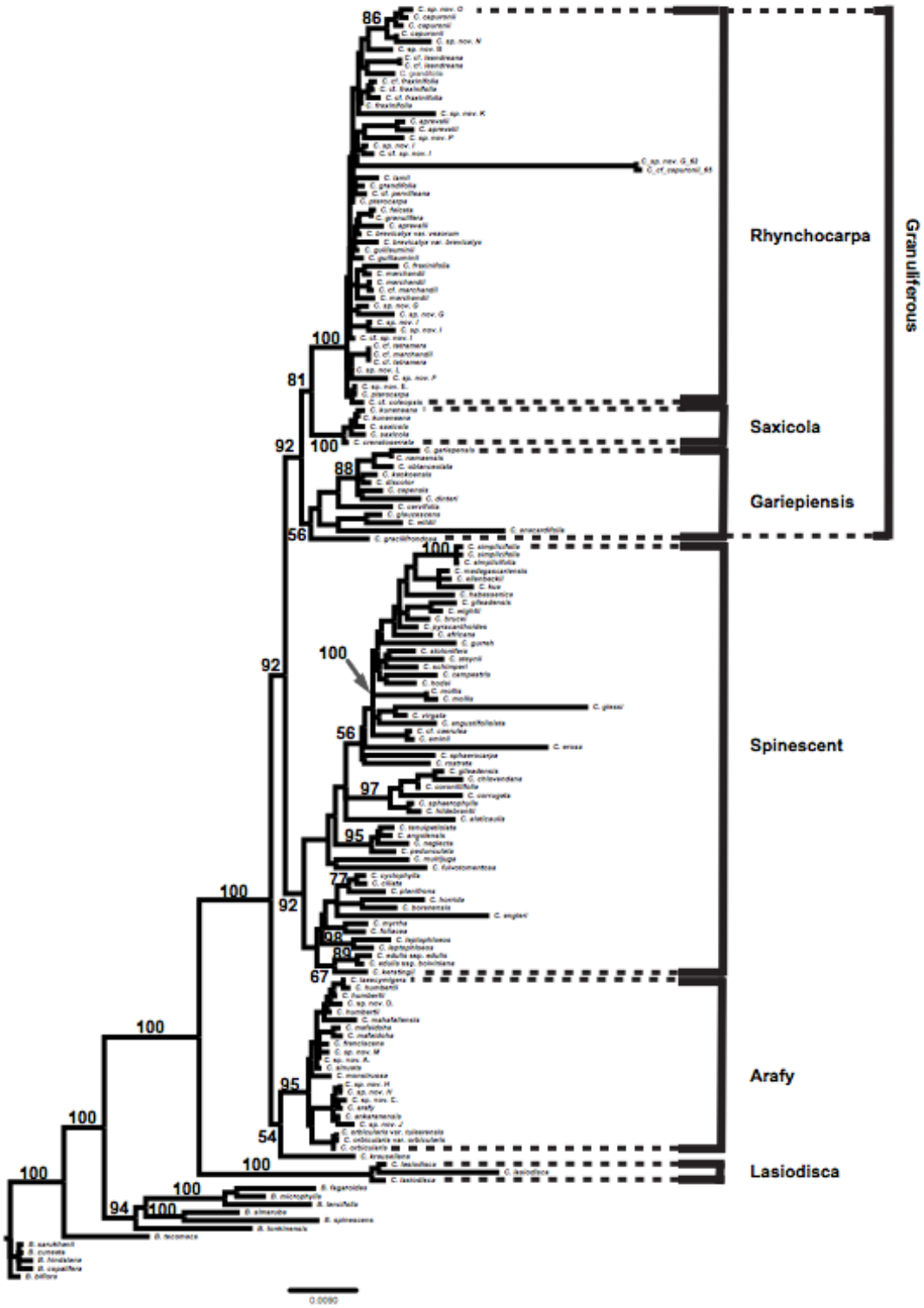

Rhynchocarpa
Saxicola
Gariepiensis
Granuliferous
Spinescent
Arafy
Lasiodisca

Figure 3. Maximum likelihood tree resulting from ten-step, batch-run GARLI analyses. Branch lengths were shortened after analysis for only the branches indicated in brackets in order to improve their visibility in this figure. ML bootstrap support values are provided above branches.

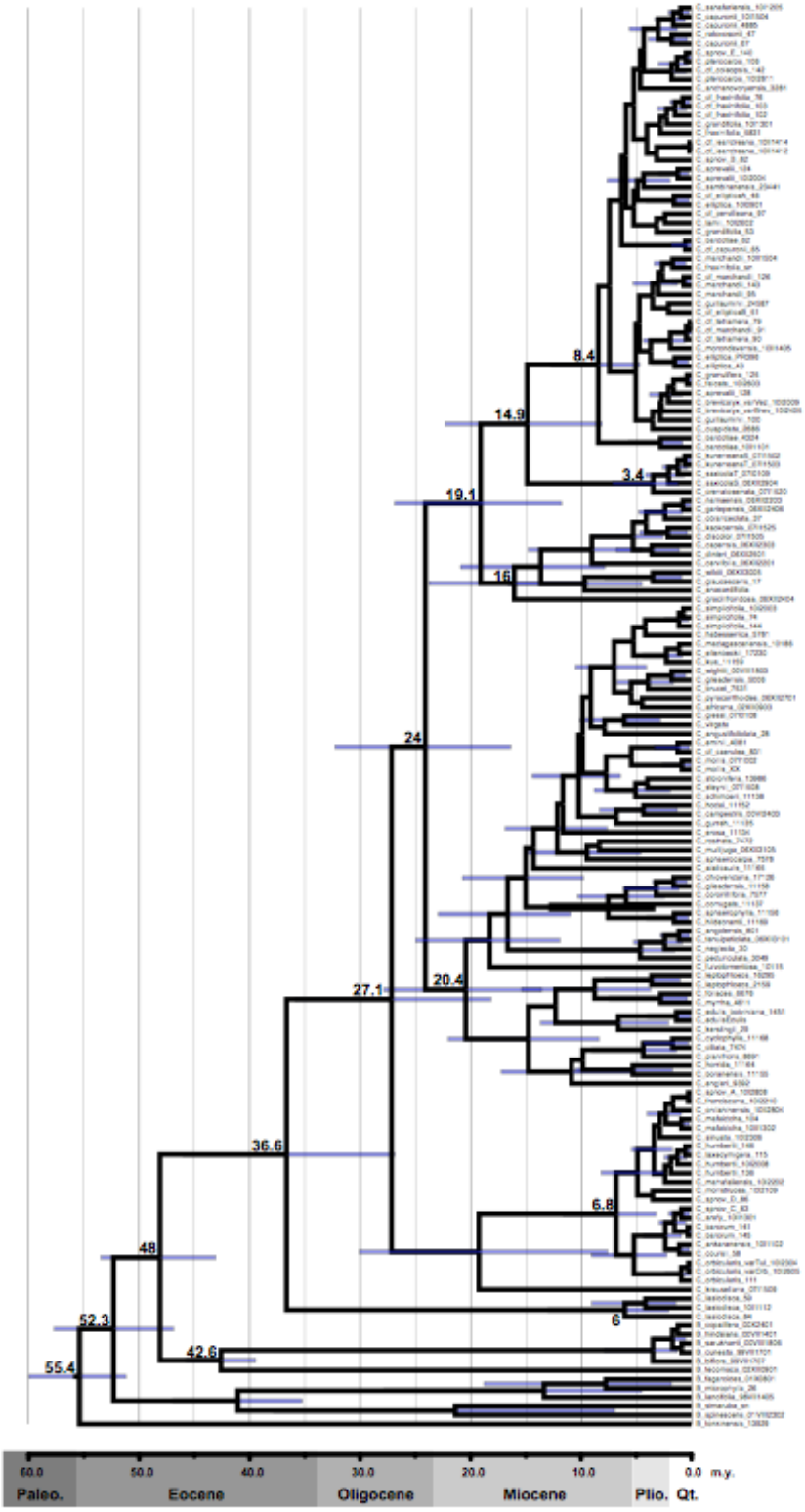

8.4
14.9
3.4
19.1
16
24
27.1
20.4
36.6
6.8
48
6
52.3
42.6
55.4
60.0
50.0
40.0
30.0
20.0
10.0
0.0 m.y.
Paleo.
Eocene
Oligocene
Miocene
Plio.
Qt.

Figure 4. 50% majority-rule consensus tree following 10 million generations of our fossil-calibrated divergence dating analysis in BEAST v1.6.1. Bars above nodes correspond to the 95% highest posterior density values. Estimated mean ages for nodes are provided within the 95% HPD bars.

Chapter 3: Development of novel EPIC markers from EST databases and evaluating their phylogenetic utility in *Commiphora* (Burseraceae)

Published in Applications in Plant Sciences 2(4): 2014

Abstract

Premise of the study: Novel nuclear exon-primed intron-crossing (EPIC) markers were developed to increase phylogenetic resolution among recently diverged lineages in the frankincense and myrrh family, Burseraceae, using *Citrus*, *Arabidopsis*, and *Oryza* genome resources.

Methods and Results: Primer pairs for 48 nuclear introns were developed using the genome resource IntrEST and were screened using species of *Commiphora* and other Burseraceae taxa. Four putative intron regions (*RPT6A*, *BXL2*, *mtATP Synthase D*, and *Rab6*) sequenced successfully for multiple taxa and recovered phylogenies consistent with those of existing studies. In some cases, these regions yielded informative sequence variation on par with that of the nrDNA internal transcribed spacer.

Conclusions: The combination of freely available genome resources and our design criteria have uncovered four, single-copy nuclear intron regions that are useful for phylogenetic reconstruction of Burseraceae taxa. Because our EPIC primers also amplify *Arabidopsis*, we recommend their trial in other rosid and eudicot lineages.

Introduction

Resolving phylogenetic relationships among closely related angiosperm species is frequently hindered due to limited variation in currently available markers (Li et al., 2008; Zimmer and Wen, 2012). This challenge is no less problematic in the myrrh genus, *Commiphora* Jacq. (Burseraceae), where complete, species-level resolution has not been achieved despite the use of multiple gene regions (Weeks and Simpson, 2007). We describe the development and evaluation of four novel, exon-primed, intron-crossing (EPIC) markers for Burseraceae (Sapindales) using a repository of putative, intron-flanking nuclear unigenes from 43 plant taxa and two complete reference genomes (IntrEST; Ilut and Doyle, 2012). Markers were evaluated for their phylogenetic utility at the species-level using a recently radiated lineage of *Commiphora* and a generic-level sampling in Burseraceae. Sequence variation from these novel markers was compared to existing nuclear markers and shows promise for resolving relationships at both shallow and deeper phylogenetic scales.

Methods and Results

Marker development

Development of EPIC markers for Burseraceae involved unigene datasets of *Citrus clementina* hort. ex Tanaka and *C. sinensis* (L.) Osbeck (Rutaceae; Sapindales) and two reference genomes available in IntrEST, *Oryza sativa* L. and *Arabidopsis*

thaliana (L.) Heynh. We developed twelve primer pairs for putative introns from each of four predicted amplicon size categories (200 bp increments between 400–1,200 bp), resulting in 48 total primer pairs. For each size category, six primer pairs were developed from a percent-identity criterion of either 80–89.9% or 90–100% between the unigene and the corresponding reference. We predicted that the lower percent-identity criterion (80–89.9%) might yield more informative variation among closely related species. Half of the primer pairs were generated using *C. clementina* and the other half from *C. sinensis* unigenes. Primer sequences were a consensus between unigene and the corresponding reference genome. Primers were preferentially designed using *A. thaliana*. Primers were designed between 18–30 bp, within 50 bp of putative intron splice regions in the reference genome, having a melting temperature (T_m) between 51–74 °C, without predicted dimers, and a 35–60% G-C content. Primer characteristics were evaluated using OligoEvaluator^{TM} (Sigma-Aldrich, Inc. St. Louis, Missouri, USA). Exceptions for T_m and %GC were made for 18 primers where it was not possible to meet all necessary criteria (Appendix 4). Each primer pair was tested by its ability to amplify a single PCR product from two species of Madagascan *Commiphora* (*C. lamii* H. Perrier and *C. ankaranensis* (J.-F. Leroy) Cheek & Rakot.) and a positive control (*A. thaliana*) and to sequence cleanly.

Taxonomic sampling and molecular methods

Markers that passed all above criteria were evaluated using 14 species of Burseraceae (Appendix 3), including eight *Commiphora* ingroup species and six

outgroup species from closely (*Bursera* Jacq. ex L., *Aucoumea* Pierre) and distantly (*Boswellia* Roxb. ex Colebr., *Protium* Burm. f., and *Beiselia* Forman) related genera, respectively (Weeks et al., 2005; Thulin et al., 2008). All ingroup taxa are Madagascan and seven correspond to one of two species-rich clades in Madagascar. We sampled densely from one clade to test phylogenetic utility at shallow-scales. Whole genomic DNA was extracted from specimens using the FastPrep FastDNA® Spin Kit (Bio101 Systems, La Jolla, California, USA). All markers were amplified in 25 µL PCR reactions including: 0.5 µL forward and reverse primers (5 µM), 0.5 µL spermidine (4 mM), 2 µL total DNA, and 5 µL GoTaq® Green Master Mix (Promega Corp., Madison, Wisconsin, USA). A ramp-up PCR thermocycler protocol followed a 4 min pre-soak at 94°C with 35 cycles of 30 s at 94°C (denaturation), 1 min at 48–56°C (annealing), and 50 s at 72°C (extension), followed by a 4 min post-soak at 72°C. PCR products were purified prior to sequencing reactions using ExoSAP (USB Corp., Cleveland, Ohio, USA) and sequenced by Macrogen, Inc. (Rockville, Maryland, USA) using an ABI 3730XL Analyzer with BigDye Terminator v3.1 (Applied Biosystems, Forster City, California, USA). Sequencing reactions (10 µL) for both directions included 40 ng/µL template. Products were assembled and edited using Sequencher 4.7 (Gene Codes Corp., Ann Arbor, Michigan, USA).

Phylogenetic analyses

Multiple sequence alignment (MSA) was performed using MUSCLE version 3.7 (Edgar, 2004). Gap regions in the MSA were treated as missing data. Markers were

evaluated using maximum parsimony (MP) and Bayesian inference (BI). MP analyses were conducted using PAUP* 4.0b10 (Swofford, 2002) with a two-step protocol modified from Plunkett et al. (2005). Branch support for internal nodes was inferred by bootstrapping 1,000 replicates in PAUP*. BI analyses were performed using MrBayes version 3.1.2 (Ronquist and Huelsenbeck, 2003). Two runs were performed for each dataset using the best-fitting model as determined by jModelTest (Posada, 2008) consisting of four chains each for 10 million generations sampled every 1,000 generations; 10% sampling was discarded as burn-in for each run. MSA and BI analyses were performed in the CIPRES Science Gateway (Miller et al., 2010).

Marker evaluation

Fifteen of the 48 EPIC primer pairs (31%) amplified at least one species and four pairs (8%; 10F–10R, 16F–16R, 39F–39R, and 43F–43R) produced amplicons that sequenced cleanly for multiple taxa. Provisional marker names are provided based upon gene ontology categories from reference taxa (Table 3). When searched in BLAST, sequences of putative intron regions for *RPT6A*, *BXL2*, and *Rab6* (Appendix 4) matched gene ontology categories predicted for the *Arabidopsis* and *Oryza* references in IntrEST. Sequence products for *mtATPSynthaseD* (Appendix 4) did not BLAST to predicted gene ontology categories. Sequences produced by this study have been deposited into GenBank (Appendix 3). Sequence alignment files are deposited in the Dryad Digital Repository (http://doi.org/10.5061/dryad.382p0; Gostel and Weeks, 2014). Phylogenetic statistics of new EPIC markers are presented in comparison (Table 3) with those from

nuclear markers developed for previous phylogenetic studies of Burseraceae (ETS: Weeks and Simpson, 2007; ITS: Gostel and Weeks, unpublished). Phylogenetic analysis of EPIC markers developed in this study recovered well-resolved phylogenies consistent with those from previous studies (Figure 5). The concatenated set of all four EPIC markers resulted in improved phylogenetic resolution compared to previously developed markers (Figure 5).

Critical assessment of primer design criteria

Each of the 15 primer pairs that amplified at least one species spanned the range of melting temperatures (51–74°C), differed from their pair by less than 10°C in T_m, and were developed from both *Citrus* unigene datasets and both reference genomes. Over half (9/15) of these markers were designed using 80–89.9% identity criteria, yet only two (16F-16R and 43F-43R) sequenced cleanly for multiple taxa. Two of the six 90–100% identity criteria primer pairs (10F-10R and 39F-39R) sequenced cleanly for multiple taxa and yielded the most informative variation among *Commiphora* species. These results do not support predictions that lower percent identity would provide better shallow-scale phylogenetic resolution, which suggests mutation rates between exon and intron regions are independent.

Conclusions

The EPIC markers developed in this study may also be useful for phylogenetic reconstruction in other angiosperm taxa. Most primer pairs amplified *Arabidopsis*

thaliana (Brassicales) and they may work in other rosid or eudicot taxa. Of the four markers, *RPT6A* is most promising for further evaluation. This ca. 400 bp region sequenced cleanly for all Burseraceae taxa and yielded a percentage of phylogenetically informative characters on par with ITS. Our study demonstrates how genomic resources from model organisms can be leveraged to advance the phylogenetic systematics of non-model organisms.

References

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughout. *Nucleic Acids Research* 32: 1792–1797.

Gostel, M. R. and A. Weeks. 2014. Data from: Development of novel EPIC markers from EST databases and evaluating their phylogenetic utility in *Commiphora* (Burseraceae). Dryad Digital Repository. http://doi.org/10.5061/dryad.382p0.

Ilut, D. C. and J. J. Doyle. 2012. Selecting nuclear sequences for fine detail molecular phylogenetic studies in plants: A computation approach and sequence repository. *Systematic Botany* 37: 7–14.

Li, M., J. Wunder, G. Bissoli, E. Scarponi, S. Gazzani, E. Barbaro, H. Saedler, and C. Varotto. 2008. Development of COS genes as universally amplifiable markers for phylogenetic reconstructions of closely related plant species. *Cladistics* 24: 727–745.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE)*. 14 November 2010, New Orleans, LA pp 1–8.

Plunkett, G. M., P. P. Lowry II, D. G. Frodin, and J. Wen. 2005. Phylogeny and geography of *Schefflera*: Pervasive polyphyly in the largest genus of Araliaceae. *Annals of the Missouri Botanical Garden* 92: 202–224.

Posada, D. 2008. jModeltest: Phylogenetic model averaging. *Molecular Biology and Evolution* 25: 1253–1256.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 195–203.

Swofford, D. L. 2002: PAUP* 4.0b10. Phylogenetic analysis using parsimony (*and other methods). Sunderland: Sinauer Associates.

Thulin, M., B.-A. Beier, S. G. Razafimandimbison, and H. I. Banks. 2008. *Ambilobea*, a new genus from Madagascar, the position of *Aucoumea*, and comments on the tribal classification of the frankincense and myrrh family (Burseraceae). *Nordic Journal of Botany* 26: 218–229.

Weeks, A., D. C. Daly, and B. B. Simpson. 2005. The phylogenetic history and biogeography of the frankincense and myrrh family (Burseraceae) based on nuclear and chloroplast sequence data. *Molecular Phylogenetics and Evolution* 25: 85–101.

Weeks, A., and B. B. Simpson. 2007. Molecular phylogenetic analysis of *Commiphora* (Burseraceae) yields insight on the evolution and historical biogeography of an "impossible" genus. *Molecular Phylogenetics and Evolution* 42: 62–79.

Zimmer, E. A. and J. Wen. 2012. Using nuclear gene data for plant phylogenetics: progress and prospects. *Molecular Phylogenetics and Evolution* 65: 774–785.

Table 3. Marker names, primer sequences and phylogenetic statistics for the novel nuclear EPIC markers and the benchmark nrDNA ETS and ITS regions.
Provisional marker names correspond to the predicted gene ontology category for the reference genome (*Arabidopsis thaliana* or *Oryza sativa*) that most closely matches unigene sequences indentified in IntrEST. Phylogenetic statistics are reported for the ingroup[1] (*Commiphora* spp.) and family-wide[2] (all Burseraceae spp.) sampling and correspond to aligned length, percent parsimony informative characters, consistency index, retention index, and corrected retention index, respectively. % Missing Data corresponds to the percent of missing sequence data in aligned data matrices.

Provisional marker name	Primer pair	Primer sequence	Ingroup[1] statistics	Family-wide[2] statistics	% Missing Data
RPT6A Intron	10F 10R	CTCCARCACATYCAYGARCTCCAGC AGCTGTAAYTCTTCTYTRAGCATCC	(454, 1.1, 0.95, 0.6, 0.569)	(454, 11.8, 0.91, 0.84, 0.764)	4.6
BXL2 Intron	16F 16R	CTTGTGGGAAKCCATCGGAC CGTTGTACATKGCYCTKGCYTCA	(1,049, 0.6, 0.96, 0.625, 0.601)	(1,049, 9.4, 0.916, 0.8, 0.761)	17.9
mtATPSynthase D Intron	39F 39R	TCCTYCCYTACRCMTCTGAGC GTTGATGCKGGAAYKATRACCA	(1,600, 0.2, 0.991, 0.8, 0.792)	(1,600, 5.4, 0.976, 0.864, 0.844)	46.7
Rab6 Intron	43F 43R	CCTTCAACAGATACAACAACATGCA TCCATGYCCCCACATATGCA	(984, 2.4, 0.974, 0.939, 0.915)	(984, 8.4, 0.979, 0.936, 0.916)	32.6
ETS	ETS1F 18S2R	TTCGGTATCCTGTGTTGCTTAC GAGACAAGCATATGACTACTGGCAGG ATCAACCAG	(389, 4.4, 0.85, 0.4, 0.34)	(389, 13.9, 0.723, 0.6, 0.435)	2.6
ITS	ITSny183 ITSny109Com	CCTTATCATTTAGAGGAAGGAG GWGACACCCAGGCAGACG	(850, 1.9, 0.878, 0.52, 0.456)	(850, 11, 0.809, 0.639, 0.517)	12.2

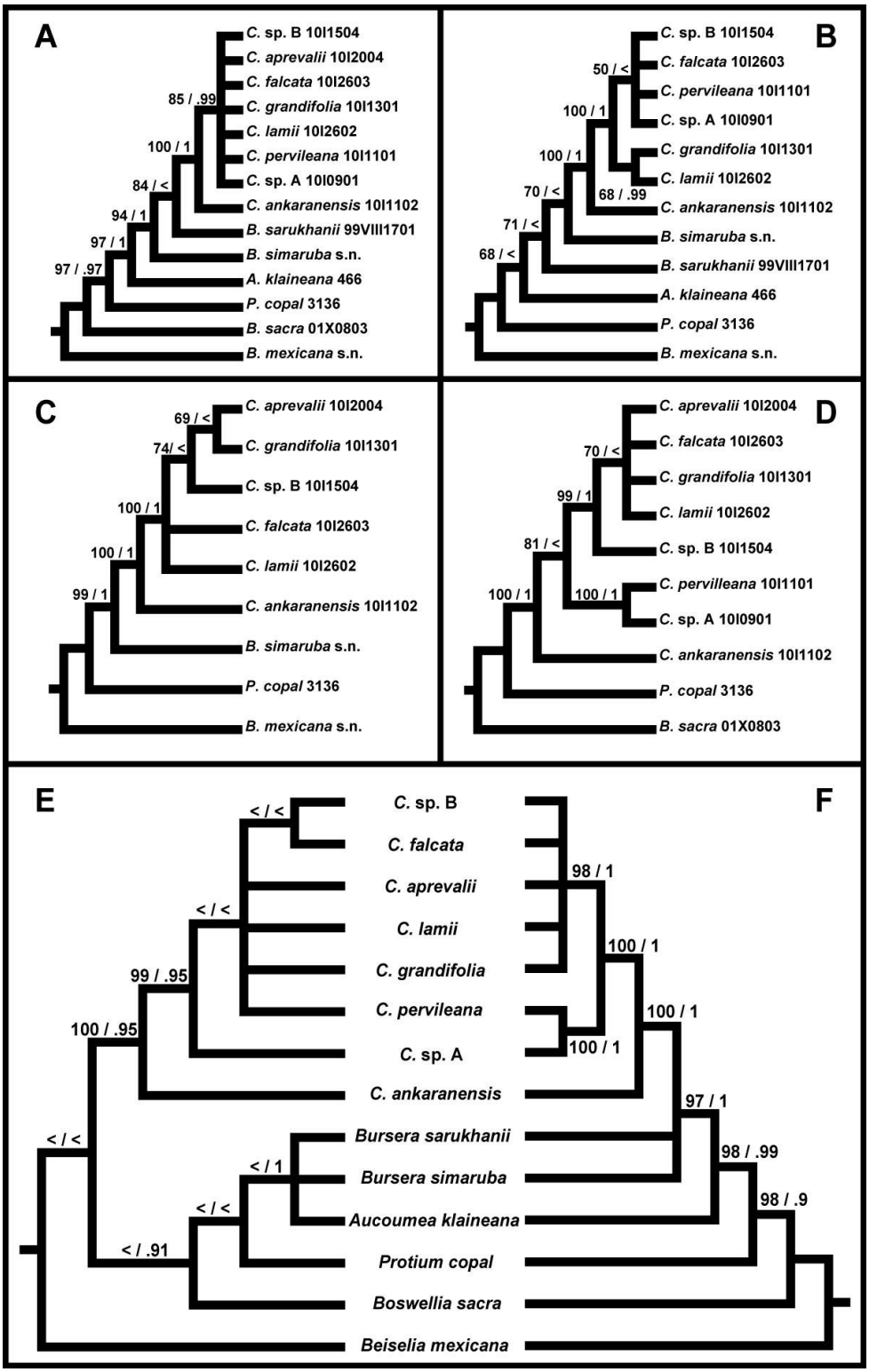


Figure 5. Phylogeny of 14 representative taxa in the Burseraceae sampled in this study.

Values above branches correspond to maximum parsimony bootstrap support values, followed by Bayesian posterior probabilities. “Concatenated new markers” refers to a concatenated data set of all four new markers.

Chapter 4: Nuclear phylogenomic analysis of angiosperms using multiplexed, PCR-based target enrichment: A case study using *Commiphora* (Burseraceae)

Accepted with revision to the Journal for Systematics and Evolution. 10 July 2015.

Abstract

Developing effective and cost-efficient nuclear phylogenomic datasets for angiosperm species is a continuing challenge to the systematics community. Here we describe the development and validation of a novel set of 91 nuclear markers for multiplexed, PCR-based enrichment by leveraging publicly available genome datasets. Using microfluidic PCR, we minimize the resources required to generate reduced representation nuclear genomic libraries for 96 angiosperm species simultaneously. We then evaluate the ability of the loci to resolve species level relationships within two recently radiated lineages of endemic Madagascan *Commiphora* (Burseraceae) species. Our results demonstrate that 1) effective nuclear phylogenomic markers can be designed for non-model angiosperm taxa from these publicly available datasets; 2) microfluidic PCR amplification followed by high throughput sequencing can produce highly complete taxon by locus sequence data matrices with minimal resource investment; and 3) these numerous nuclear phylogenomic markers can improve our understanding of phylogenetic relationships within *Commiphora*. We provide recommendations for future multiplexed, PCR-based target

enrichment studies to more effectively design primers, prepare multiplexed amplification reactions, and lower the cost of sequencing each locus.

Introduction

To date, completely sequencing nuclear genomes of angiosperm species remains out of reach for most comparative phylogenetics and systematics applications due to its high cost and analytical complexity. Consequently, systematists use reduced representations of the nuclear genome to pursue evolutionary questions and scale projects in a way that can maximize both taxonomic diversity and genome coverage, yet still remain financially and technologically feasible (reviewed in Cronn et al. 2012 and Lemmon and Lemmon 2013). Methods for gathering partitioned or reduced nuclear genome datasets for phylogenetic studies include array-based or in-solution target hybridization (see Mamanova et al. 2010), restriction-enzyme based methods (RAD-seq, sensu Baird et al. 2008), multi- and uniplex PCR-based enrichment (Bybee et al. 2011, Ho et al. 2014), and transcriptome mining (RNAseq). Each of these methods carries its own very different set of practical considerations and trade-offs for producing high-quality datasets suitable for the phylogenetic analysis of angiosperm species.

Angiosperms systematists have used target-hybridization, RAD-seq, and RNAseq to uncover the evolutionary relationships among multiple non-model taxa recently (i.e., Henriquez et al. 2014, Mandel et al. 2014, Weitemier et al. 2014), but fewer have used multiplexed, PCR-based enrichment for any group of organisms (Cronn et al. 2012,

Richardson et al. 2012, Lemmon and Lemmon 2013, Ho et al. 2014, Uribe-Convers pers. comm.). Multiplexed, PCR-based target enrichment generates sequencing libraries for individual samples through the simultaneous amplification of hundreds of targeted genomic regions. Samples are barcoded and then sequenced in parallel via high throughput sequencing. Each sequenced, targeted genomic region is then assembled into a matrix by barcode (i.e., multiplex identifier, MID) and analyzed. This method has been referred to as targeted amplicon sequencing (TAS, Bybee et al. 2011) and genome-tagged amplification (GTA, Ho et al. 2014), but we prefer the more descriptive term 'multiplexed, PCR-based target enrichment' because it emphasizes that multiple genome regions are targeted and sequenced in parallel. Microdroplet and microfluidic PCR have transformed the potential for multiplexed, PCR-based enrichment for phylogenomics research because they enable thousands of simultaneous, small volume reactions and can significantly reduce cost and labor.

With the exception of one population genetics study of *Artemesia* species (Richardson et al. 2012), no published studies have examined the utility these microdroplet or microfluidic techniques for multiplexed, PCR-based target enrichment in angiosperms. This is unfortunate because the combination of the two methods has advantages over other reduced representation methods for nuclear phylogenomic analysis, including high reproducibility, short sample processing time, and relatively lower costs. It does not require RNA, a previously sequenced genome scaffold, or intensive post-sequencing bioinformatics processing. One possible cause of this deficit of studies has been that few nuclear genomic markers have been well-characterized and

evaluated for species-level phylogenetic reconstruction across angiosperms (Zimmer and Wen 2012). However, recent advances in genomic resources are now eliminating the barrier to using multiplexed, PCR-based target enrichment for nuclear phylogenomic analysis of angiosperms.

Researchers now have access to a number of publicly-available genomic resources that can be used to discover phylogenetically informative, single copy orthologous loci within the nuclear genomes of their focal taxa. Although such loci (conserved ortholog sets; COSs), usually based upon expressed sequence tag (EST) or unigene datasets, have been available to plant systematists for more than a decade (Fulton et al. 2002, Chapman et al. 2007, Wu et al. 2006), recent studies (Yuan et al. 2009, Duarte et al. 2010, Zhang et al. 2012, and De Smet et al. 2013) have greatly improved their utility by comparing their copy number and presence in multiple angiosperm taxa. Duarte et al. (2010) compared gene loci in four annotated angiosperm genomes to produce a set of 959 putative, single-copy nuclear gene loci. This approach was expanded by De Smet et al. (2013) who recovered more than 2,800 putative orthologous, low-copy loci. To date, these studies remain largely an untapped resource; few studies have tested these loci for their phylogenetic utility in resolving species-level relationships (Naumann et al. 2011, Cacho and Strauss 2013). Moreover, the growing number of published transcriptome datasets, notably the 1,000 transcriptomes derived from species across the angiosperm tree of life and produced by the OneKP project (https://sites.google.com/a/ualberta.ca/onekp/home, Matasci et al. 2014) now make it possible to more closely tailor primers for these low- and single-copy nuclear gene loci to

one's taxon of interest. In the present study, we combine these resources to develop a suite of nuclear loci suitable for multiplexed, PCR-based target enrichment across species of a diploid, non-model angiosperm genus.

The objectives of our study are threefold: **1)** develop a set of phylogenomic targets from the nuclear genome of a diploid, non-model angiosperm genus using publicly available resources; **2)** test the efficacy of microfluidic methods for enriching these regions simultaneously in multiplex PCR; and **3)** evaluate the utility of these regions, individually and in combination, for resolving species-level phylogeny of *Commiphora* Jacq. (Burseraceae). Previous studies in the myrrh genus, *Commiphora*, have shown that two clades, 'Arafy' and 'Rhynchocarpa', have undergone rapid speciation in Madagascar (Weeks and Simpson, 2007; Gostel et al. in review). Phylogenetic reconstruction using multiple molecular markers commonly used for angiosperm species-level phylogenetics, including two proposed barcode loci (nrDNA ITS and cpDNA *psbA–trnH*), have not fully resolved interspecific relationships among the 41 species in these Madagascan clades. Gostel and Weeks (2014) subsequently developed novel nuclear, exon-primed, intron-crossing marker loci for *Commiphora* using *Citrus* spp. EST data and *Arabidopsis* genome resources from Ilut and Doyle (2012). However, this strategy produced only four new loci that were sufficiently variable to infer species relationships within the 'Arafy' and 'Rhynchocarpa,' clades and contained proportions of phylogenetically informative characters that were no greater than those of published nuclear loci. Gostel and Weeks (2014) demonstrated that a far greater number of loci would have to be sampled to produce well-resolved and well-

supported phylogenies of Malagasy *Commiphora* species. Towards this end, we outline here the development of nuclear phylogenomic targets for *Commiphora*, their application using microfluidic, multiplexed PCR-based target enrichment, and their ability to improve our understanding of the evolution of *Commiphora* in Madagascar.

Materials and Methods

In-silico development: target selection and primer design

Two publicly available genomic databases were used to develop nuclear phylogenomic targets for *Commiphora* (2N=24; Hanson et al. 2001 and Bennett and Leitch 2012). The first comprised 959 putatively single copy nuclear gene regions that are shared among four angiosperm species (*Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, and *Vitis vinifera*) (Duarte et al. 2010). Of these four taxa, *A. thaliana* diverged most recently from a common ancestor with *Commiphora*, approximately 104 Ma (Magallón et al. 2015). Given the large evolutionary distance between *Commiphora* and *Arabidopsis*, we refined primers for these nuclear gene regions using available genomic resources from a more closely related taxon, *Bursera* Jacq., which is sister to *Commiphora* (Weeks et al. 2014). The second comprised sequence data from a transcriptome of *Bursera simaruba* derived from leaf tissue (Johnson et al. 2012) that was provided by the OneKP project (Matasci et al. 2014). The transcriptome was assembled following the protocol outlined in Xie et al. (2014) and comprised 67,632 sequence reads, ranging from 100–5,324 bp (mean, 340.1 bp).

We imported the two genomic datasets into Geneious v 7.1.7 and mapped the reverse-translated RNA transcripts of *Bursera* to 950 of the Duarte et al. (2010) single copy nuclear gene regions of *Arabidopsis,* which were downloaded as "AGI genome locus sequences" using the "Sequence Bulk Download and Analysis" tool from the *Arabidopsis* Information Resource (http://www.arabidopsis.org; Accessed 9 April 2015). Nine of the single copy nuclear gene regions identified by Duarte et al. (2010) were not downloaded because they were no longer designated as shared single copy nuclear gene loci in the PlantTribes 2.0 database (Wall et al. 2008, http://fgp.bio.psu.edu/tribedb/10_genomes/index.pl?action=home). We used the "Map to Reference" function in Geneious to map transcript reads to the 950 nuclear gene regions.

We customized the sensitivity of mapping the *Bursera* transcript reads to the 950 *Arabidopsis* nuclear gene regions in order to achieve alignments that would span potential indel regions and allow for some ambiguity in the assembled contigs of *Bursera* transcripts. Under the "Custom Sensitivity" options, we selected random mapping in the case of multiple best matches, with gaps allowed (20% maximum per read and 500 maximum gap size). A minimum 25 bp overlap was enforced with a minimum overlap identity of 80%. We used an intermediate word length of 10 and selected not to ignore repeated words. The maximum number of mismatches per read was set to 50% and our maximum ambiguity was set quite high, to 60%. We selected to accurately map reads with errors to repeat regions and search more thoroughly for poorly matching reads.

Contigs were individually verified by eye for quality control purposes. In many cases, the assembled contigs included multiple reads, up to 27 *Bursera* transcript

sequences per *Arabidopsis* nuclear gene region. Many of these *Bursera* transcript reads appeared to have been spuriously assembled as a result of our low-stringency criteria. Consequently, we discarded most contigs with less than 35% pairwise identity. However, we reviewed each discarded contig (n=587) because Geneious v 7.1.7 artificially reduces the percent pairwise identity of contigs where gap regions are inferred. As a result, we retained ten initially discarded contigs that had very low pairwise identity scores (between 5–35%) but high sequence similarity.

Primer design

Primers were designed from the contigs of *Bursera* transcripts that matched the *Arabidopsis* nuclear gene regions in Geneious v 7.1.7 using Primer3 (Untergrasser et al. 2012). Primer3 search criteria included the following: product size between 95 and 600 bp, with an optimal product size of 450 bp; Santa Lucia 1998 T_m calculation; primer length between 18 and 27 bp with an optimal length of 24 bp; 59°C, 60°C, and 61°C minimum, optimal, and maximum T_m, respectively; 35%, 50, and 65% minimum, optimal, and maximum GC%, respectively; 50°C maximum dimer T_m; maximum 3' stability of 9; maximum poly-x 5; and maximum T_m difference of 100°C. In many cases, Primer3 selected multiple primers from each contig. The optimal primer pairs were selected from the range of possibilities based on a hierarchy of criteria that included target sequence length, primer melting point (T_m), and primer interactions (dimer formation and secondary structures such as hairpins).

Once primer pairs for the putatively single copy nuclear gene regions were designed, we modified them for amplification on the Fluidigm Access Array (Fluidigm Corporation, San Francisco, California) microfluidic instrument and sequencing on the Illumina platform. We first added Fluidigm Access Array custom sequence adapters CS1 and CS2 to the 5' ends of the forward and reverse primers of each target gene region (TS), respectively (Appendix 6). Fluidigm microfluidic PCR requires four primers per locus per taxon: the CS-TS primer pair and a pair of primers that includes the high-throughput sequencing adapter sequence (PE), a species-specific barcode sequence (BC) and the CS1 or CS2 sequence. PE-CS primer pair sequences and BC sequences were developed according to the Fluidigm Access Array User Guide (Fluidigm PN 100-3770, San Francisco, California) and prepared by Eurofins MWG Operon, LLC (Huntsville, Alabama). Species-specific barcode sequences were only added to the PE-CS reverse primers.

Primer pooling & multiplexing

The CS-TS primer pairs were organized into pools of two for co-amplification in the same well of the Integrated Flow Circuit (Fluidigm Corporation; San Francisco, California) during microfluidic PCR using criteria that reduced the chances of their interference. We determined the location of each nuclear gene region as inferred by *Arabidopsis thaliana* chromosome position and pooled primer pairs that had maximal separation within the genome, assuming that genetic distance is syntenic with *Commiphora*. Primers were paired if their target gene regions were found on different

Arabidopsis chromosomes. However, because the *A. thaliana* genome comprises only five chromosomes, if primers had to be paired on the same chromosome, we only paired primers separated by minimum of 10,000 bp. Our second pooling criterion was based on individual primer interactions. We used the MFE Primer Interaction tool (Qu et al. 2012) to test for interactions between primers in each pool and resorted the pairs if necessary.

Taxon sampling, DNA extraction and standardization

Whole genomic DNA was extracted using the FastPrep FastDNA® Spin Kit (Bio101 Systems, La Jolla, California) from 96 samples of silica-dried plant material. Our positive control was a sample of *Bursera simaruba* (Appendix 1), the species for which primers were designed. We sampled three other outgroup species, including two from *subgenus Bursera*, the subgenus in which *B. simaruba* is placed (*B. fagaroides* and *B. spinescens*) and one from subgenus *Elaphrium* (*B. tecomaca*). Our ingroup sampling included 92 accessions of *Commiphora* species endemic to Madagascar. Previous studies have demonstrated that *Commiphora* is represented in Madagascar by four lineages (Gostel et al. *in review*). We include here accessions only from the two most species-rich Malagasy lineages, the 'Arafy' and 'Rhynchocarpa' clades. For most species, multiple accessions were included (Appendix 5).

DNA was quantitated using a Qubit Fluorometer and dsDNA HS Assay Kit (Life Technologies, Grand Island, New York). When possible, DNA concentration (ng/µL) was normalized to between 5 and 10 ng/µL, as recommended (Fluidigm PN 100-3770, San Francisco, California). The C-value for *Commiphora* (Hanson et al. 2001, Bennett et

al. 2012, Kew C-value database) corresponds to a diploid genome size of ca. 611 Mbp, consequently a maximum of ca. 1520 nuclear genome copies are present in every ng of whole genomic DNA, excluding the contribution of chloroplast and mitochondrial DNA. Therefore, we would expect 7,600–15,200 genome copies in a 5 µL PCR containing 1 µL of 5–10 ng/µL DNA. The Fluidigm Access Array Integrated Flow Circuit (IFC) is divided into 48 x 0.033 µL reaction chambers and thus each reaction chamber should have ca. 50–100 genome copies. However, we could not obtain this concentration for some accessions despite multiple extractions (15 DNA extracts, <1.5 ng/µL). In these cases, we used DNA concentrations as low as 0.212 ng/µL (ca. 6–7 template genome copies per reaction chamber), which allowed us to test the lower threshold of IFC sensitivity to DNA concentration.

Primer validation

We verified that each primer pair could amplify a single product from *Bursera simaruba* following Fluidigm protocol (Fluidigm PN 100-3770, San Francisco, California). The thermocycler protocol included three alternating standard and C_0t (Mathieu-Daudé et al. 1996) cycles, beginning with 2 minutes at 50°C, 20 minutes at 70°C, and 10 minutes at 95°C. The first set of ten standard cycles included a denaturation step at 95°C for 15 seconds, annealing at 60°C for 30 seconds, and extension at 72°C for 1 minute. Two C_0t cycles followed, including four steps consisting of 95°C for 15 seconds, 80°C for 30 seconds, 60°C for 30 seconds, and 72°C for 1 minute. Standard and C_0t cycles alternated two more times with 8, 2, 8, and 5 cycles, respectively. After 35

cycles, samples were held at 4°C prior to being visually verified via agarose gel electrophoresis (2% agarose; 40 V for 90 min).

Microfluidic PCR amplification, sequencing, and assembly of nuclear gene regions

Microfluidic PCR amplification of the nuclear gene regions from the 96 taxon accessions used two Integrated Flow Circuits on the IFC Controller AX and FC1 cycler instruments following the "4-Primer Amplicon Tagging 48.48 Access Array IFC" protocol (Fluidigm PN 100-3770, San Francisco, California). The final product was a single volume containing an enriched library of the targeted gene regions barcoded by taxon accession. The library was cleaned prior to sequencing using the Agencourt AMPure XP kit (Beckman Coulter, Inc., Brea, California), quantitated using a Qubit fluorometer and diluted to 2 µM with DNA Suspension Buffer (TekNova T0221, Hollister, California), and sequenced on a MiSeq platform (Illumina) at the Johns Hopkins University (JHU) Genomics Resources Core Facility (GRCF) using a 600-cycle kit (Illumina). We utilized custom sequencing primers designed based upon Illumina TruSeq adapters and ordered as locked nucleic acids (LNA) from Exiqon, Inc (Woburn, MA).

Sequence quality filtering and end trimming were performed using CutAdapt 1.8.1 (https://cutadapt.readthedocs.org/en/stable/index.html#, Martin 2014). We filtered reads by removing ends with quality scores <Q20, removing reads shorter than 60 bp, trimming poly-N tails ≥ 6 bp, and trimming all Illumina adapter sequences from the ends of our reads. Demultiplexed, filtered, and trimmed .fastq sequence files were imported

into Geneious v 7.1.7 and mapped to reference sequences of the targeted nuclear gene regions using the "Map to Reference" feature. Reference sequences were either produced by Sanger sequences produced for two *Commiphora* samples or using the primer sequence as a reference for the "Map to Reference" feature. Target nuclear genomic reference sequences were sequenced using Sanger dideoxy termination at Macrogen Inc. (Rockville, Maryland) for one sample that did not produce microfluidic amplification product (*C. coleopsis* 142) and another sample that did produce DNA sequences (*C. laxecymigera* 115). We compared the sequences to those of conserved orthlogous sets (COSs) identified in other studies (i.e., Wu et al. 2006, Yuan et al. 2009, and De Smet et al. 2013) to check for overlap among COSs.

Multiple sequence alignment and phylogenetic analysis

Multiple sequence alignment was carried out for each locus using Muscle v.3.8.425 (Edgar 2004) in Geneious v.7.1.7 with default settings and a maximum of 8 iterations. All alignments were individually checked, manually adjusted if necessary, and deposited in Dryad (For reviewers only: we have deposited these files in a dropbox folder available at the following link: https://www.dropbox.com/sh/kednmcl8n5xtbco/AAC3jD-Rdz6UmI9XTkS96VvBa?dl=0). Fasta alignment files were uploaded to the CIPRES Science Gateway (Miller et al. 2010) and analyzed in RAxML v.8.1.11 (Stamatakis 2014) using the GTR+Γ model of substitution for all datasets and including 1,000 rapid bootstrap iterations. We analyzed each locus individually and in three concatenated alignments that included 1) all loci for all taxon accessions (highest level of missing

sequence data), 2) a subset of loci that were present in at least 75% of all taxon accessions (i.e., ≤ 25% missing sequences), and 3) a smaller subset of loci that were present in at least 99% of all taxon accessions (i.e., ≤ 1% missing sequences).

Results

Target selection, primer design, and validation

39,246 of 67,632 (58%) reverse-translated RNA sequence reads from the *Bursera simaruba* transcriptome mapped to our reference target loci and produced 950 contigs for each of the 950 nuclear gene regions from *Arabidopsis thaliana* in 147 hours and 8 minutes of computational time on a dual core, 1.7 Ghz AMD processor. From these 950 loci, 363 produced contigs that met our selection criteria and of these, 208 contained sequence characteristics suitable for primer design. A search of these 208 loci yielded 239 different primers (Appendix 6) that in some cases overlapped sections of the same locus. We reduced these primer pairs to 192 by discarding those that would have formed predicted dimers or would have amplified the shortest or longest fragments.

Of the 192 primer pairs, 85 produced single products during validation. Of these 85, 37 produced fragments of expected size based on the *Arabidopsis thaliana* reference regions and 48 were longer than expected. Each of the 48 was sequenced in order to design internal primers that would keep fragment size within the optimal range. Internal primers for eight of these longer loci were generated, six of which successfully amplified a single product within the predicted smaller size and were added to the final panel. In

total, 91 primer pairs were multiplexed for microfluidic PCR amplification. Twelve IFC primer pools included primers with predicted interactions and had to be resorted. The locations of the loci amplified by the primers were overrepresented on *A. thaliana* chromosomes 2 and 3 and, consequently, eight of our multiplexed IFC primer pools shared locations on either chromosome 2 or 3. Across 48 primer chambers of the IFC, 43 chambers contained two primer pairs and 5 chambers contained a single primer pair.

Multiplexed, microfluidic PCR enrichment and sequencing of nuclear gene regions

Microfluidic PCR reactions for each taxon were visualized using agarose gel electrophoresis to verify successful amplification. Eighty-eight of 96 taxa (ca. 93%) successfully amplified product after microfluidic PCR in the integrated flow cell prior to sequencing. Eight taxa (*Commiphora arafy* 10-II-12-01, *C. arafy* 10-II-13-01, *C. arafy* 10-II-14-02, *C.* cf. *coleopsis* 142, *C. fraxinifolia* 10-II-13-05, *C.* sp. nov. B 82, *C.* sp. nov. H 141, and *C.* sp. nov. N 47) did not produce visible product, likely due to the formation of embolisms within the IFC that blocked the movement of product from reaction chambers to sample wells. All three accessions for *C. arafy* did not amplify, which could indicate the presence of inhibitory secondary compounds in the DNA extracts.

Results from our MiSeq runs are presented in Table 4 and include species name, extracted DNA concentration, number of reads, and number of reads that mapped to unique targets. The 88 taxon accessions yielded varying numbers of sequenced loci. Five of these accessions either yielded no sequences or were omitted from our dataset because their initial alignments and phylogenetic placement were highly inconsistent with that of

conspecifics (*Commiphora ankaranensis* 10-I-12-11, *C. brevicalyx* var. *vezorum* 10-I-20-09, *C. monstruosa* 10-I-21-09, *C.* sp. nov. I 51, and *C.* sp. nov. J 58). Of these five excluded samples, four had original genomic DNA concentrations below the recommended level (<5 ng/µL). Of the 83 taxon accessions that did yield sequence data, each on average yielded 51 loci. These 83 taxon accessions included 16 of 18 samples with DNA concentration below 1.5 ng/µL and 45 of 53 (ca. 85%) samples with DNA concentration below 5 ng/µL. We removed an additional three taxa (*C. ankaranensis* 10-I-14-02, *C. mahafaliensis* 10-I-28-03, and *C. sp.* 54) from our analyses because both the overall quality and quantity of sequence data were poor. The median number of trimmed reads that mapped to reference locus targets for these three species was only 58, 35, and 56, respectively, compared to a median for all reads of 2,108 per locus per taxon (Table 4). However, the average number of reads per locus for these three species was quite high (8,625, 54,548, and 28,661, respectively). This indicates our sequencing run produced a disproportionately high number of reads from these accessions for some loci and not others.

Eighty-five of the 91 primer pairs produced sequence data from at least one taxon accession; only five primer pairs failed to work (Table 5 and Appendix 1). Of these 85 primer pairs, on average each locus is represented by sequences from 51 different taxon accessions. Sequencing success was greater for loci having shorter predicted lengths; loci < 1,000 bp succeeded on average for 57 taxa, whereas loci ≥1,000 bp succeeded on average for just 16 taxa. We verified the source of all sequence products by performing a

nucleotide BLAST search (http://blast.ncbi.nlm.nih.gov/Blast.cgi) for highly similar sequences and all search results matched other angiosperm taxa.

Phylogenetic analysis of individual loci and concatenated datasets

We selected 49 of the 85 sequenced loci (ca. 58%) for phylogenetic analysis (Table 5). The remaining 36 loci were excluded either because the number of sequences was too low per locus (<35) and unlikely to be reliable markers in the future studies, and/or because their alignments contained a large number of gaps (>50% characters) that lowered our confidence in the alignments' accuracy. The individual matrices included 35-78 taxa (mean: 72 and median: 78) ranging from 125-1,957 aligned bp (mean: 560 bp and median: 474 bp) and containing 10 (3.375%) - 155 (21.257%; mean: 57 and median: 49) phylogenetically informative characters. Of the 49 loci, 40 recovered well-supported (>70% maximum likelihood bootstrap) backbone topologies consistent with previous published studies that showed a monophyletic group of four *Bursera* outgroup species (Weeks et al. 2005, Weeks and Simpson 2007). Additionally, 26 of these 49 loci recovered at least one of the two clades of *Commiphora* species identified in previous studies as the 'Arafy' and 'Rhynchocarpa' lineages (Gostel et al. in review, see Appendix 7). Eight loci did not to recover a monophyletic *Bursera* (AT1G21840, AT1G65030, AT2G20790, AT2G36740, AT3G10400, AT3G29130, AT3G54460, and AT4G37510).

Of the 40 loci returning expected backbone topologies, species relationships within the major clades and at their tips often lack resolution and have very short branch lengths. Five loci individually recovered both the 'Arafy' and 'Rhynchocarpa' clades

(AT1G18060, AT1G77550A, AT1G77550B, AT2G05120, and AT2G21710F). Only two loci recovered a well-supported (ML bootstrap ≥60%) for the 'Arafy' clade, but not the 'Rhynchocarpa' clade (AT2G04740 and AT4G14605). Approximately half (19) of the analyses of individual loci recovered a well-supported (ML bootstrap ≥60%) 'Rhynchocarpa' clade (AT1G31780, AT1G65030, AT1G76450, AT1G77930A, AT2G03667, AT2G20330, AT2G31090A, AT2G44760, AT3G04650R, AT3G14910, AT3G26580F, AT3G46220, AT4G00560F, AT4G18810B, AT4G19900, AT4G21770, AT4G24590, AT4G31770, and AT5G15680). Among loci that recovered a well supported 'Rhynchocarpa' clade, but not a well supported 'Arafy' clade, species belonging to the latter clade often form grade at the base of the 'Rhynchocarpa' clade and additional sampling from species of *Commiphora* outside of these two Malagasy clades would likely result in resolution of a monophyletic 'Arafy' clade.

In the topologies of loci that have more resolution, species-specific subclades are formed from multiple accessions of single species. There are a few loci that contain important differences for a few sampled accessions, most notably the position of *Commiphora pervilleana* 93, which either forms a species-specific subclade with *C. pervilleana* 97 in the analysis of six loci (AT2G05120, AT2G04740, AT2G21710F, AT2G44760, AT4G29590, and AT5G52180) or forms a clade with *C.* cf. *coleopsis* 142 and *C.* cf. *leandreana* 10-II-14-12 in the analysis of twelve loci (AT1G10860, AT1G59990, AT1G65030, AT1G76540F, AT2G04620, AT2G20330, AT2G22370B, AT2G44660A, AT44660B, AT3G14910, AT3G18810B, and AT5G04910); *C. lamii* 10-I-26-04 which is resolved in the 'Arafy' clade in analyses of thirteen loci (AT1G10860,

AT1G31780, AT1G59990, AT1G76450R, AT1G77550A, AT2G20330, AT2G22370B, AT2G44660B, AT3G01380R, AT3G14910, AT3G46220, AT4G00560F, and AT4G31770), but more commonly resolved in the 'Rhynchocarpa' clade (all remaining loci).

The three concatenated sequence data matrices contained different numbers of loci. The first, a total evidence matrix, included all taxa and all loci and comprised 44 loci and 80 taxa, 26,138 aligned bp and 2,749 (10.5%) phylogenetically informative characters, and 19.4% missing data (Figure 6). The second matrix included all taxa and all loci having ≥75% taxon coverage, contained 41 loci and 80 taxa, 24,463 aligned bp and 2,645 (10.8%) phylogenetically informative characters, and 16.6% missing data (Figure 7). The third matrix included all taxa and all loci having ≥99% taxon coverage, contained 20 loci and 80 taxa, 9,678 aligned bp and 942 (9.7%) phylogenetically informative characters, and 1.8% missing data (Figure 8). Each of the concatenated datasets recovered overall topologies that were largely consistent with each other. Each recovered a monophyletic *Commiphora* with 100% bootstrap support and 96–100% bootstrap support for two clades that correspond to the 'Arafy' and 'Rhynchocarpa' clades (Gostel et al. in review). Most subclades contain all accessions of one species or include pairs of sister species. Species that form species-specific subclades in all three concatenated datasets are *C. ankaranensis*, *C. aprevalii*, *C. falcata*, *C. franciscana*, *C. fraxinifolia*, *C. grandifolia*, *C. granulifera*, *C. mafaidoha*, *C. orbicularis*, *C. pterocarpa*, *C.* sp. nov. I, *C.* sp. nov. L, and *C. tetramera*. The bootstrap support for these species clades ranges from 92–100%. *C. lamii* and *C. pervilleana* form species-specific subclades

in the 20 loci concatenated dataset (Figure 8), but not in the 44 or 41 loci datasets. A few species do not form species-specific subclades in any of the concatenated analyses and these include *C. brevicalyx, C.* cf. *capuronii*, *C. humbertii*, *C. marchandii*, *C. sinuata*, and C. sp. nov. M. All remaining species were not sampled sufficiently to test their monophyly.

An important difference among our three concatenated datasets is the identity of the most basal member of the 'Rhynchocarpa' clade. In the analyses of 44 and 41 loci datasets, *Commiphora lamii* 10-I-26-04 is sister to all remaining species in this clade; however, in the 20 loci dataset *C. guillauminii* 100 is sister to all remaining species and both accessions of *C. lamii* form a species-specific subclade. Our analyses revealed strong branch support (>70% bootstrap) for sister species relationships among several groups of taxa. In the 'Arafy' clade, for example there are two subclades that contain 5 and 11 species, respectively and are present in all concatenated analyses with 93–100% bootstrap support.

Discussion

Our marker development strategy resulted in a set of 49 nuclear loci, which corresponds to a nearly 7-fold increase in the amount of analyzed sequence data from a recent study in *Commiphora* (26,138 compared to 3,960 aligned characters, Gostel et al. in review). These markers recover a backbone topology that is consistent that topology recovered from markers used in previous studies of *Commiphora* and with greater

phylogenetic resolution at shallower nodes. Our nuclear marker development strategy therefore appears to be successful in recovering loci for species-level phylogenetics and we recommend molecular systematists consider this methodology among the suite of available options. Publicly-available resources, including the transcriptome database of OneKP project (Matasci et al. 2014), the interactive marker design pipeline, MarkerMiner (Chamala et al. 2015), as well as the growing set of COS for angiosperms (e.g. De Smet et al. 2013), could be used to develop cost-effective multiplex PCR markers for most, if not all, major angiosperm clades.

Effective application of multiplexed microfluidic PCR for phylogenomics

Microfluidic PCR has been used infrequently (see Richardson et al. 2012, Uribe-Convers pers. comm.) for target enrichment in phylogenetics studies, although two recent reviews have cited its potential for phylogenetics (Cronn et al. 2012 and Lemmon and Lemmon 2013). While the latter study estimated that 96 samples with 96 loci would cost upwards of $18,000, we found that were able to amplify 91 loci in 96 samples for less than $2,000 and discovered opportunities to reduce these costs even further in the future. This cost efficiency is even greater when one considers that the method does not require additional sequence library preparation or extensive bioinformatics post-sequencing data processing.

The efficiency of the method can be illustrated in its savings of time and reagent costs as compared to uniplex PCR and Sanger sequencing. Using two IFC devices, we performed 8,736 amplification reactions using 30 total units of *Taq* polymerase in a

single day. This is equivalent to 91 96-well PCR plates that would have required a considerably greater quantity of extracted DNA (ca. 48–480x) as well as reagents costing an order of magnitude greater than required by the microfluidics method. By multiplexing our barcoded loci on the Illumina platform, we sequenced each locus for each taxon for ca. $0.18 each for a total of $1,400, as compared to $3 per sequence ($55,000, total) for the Sanger method. Given the high median number of Illumina sequence reads per locus, per species (2,108), we could theoretically add as many as 20 additional multiplexed integrated flow circuit (IFC) libraries per Illumina lane, achieve ca. 100X coverage per locus per species, and reduce the cost of sequencing to less than $0.01 per locus per species.

Our study has also uncovered the upper and lower bounds of target amplicon length and DNA concentration, respectively, that will refine future applications of this method. We found that the IFC microfluidic device is more effective in returning complete datasets for loci smaller than 1,000 bp, so primer design criteria should be modified to stay below this threshold. Moreover, we have found that small quantities of DNA are not necessarily a barrier to using the method effectively. We were able to recover high quality sequences from low concentration DNA (as low as 0.268 ng/μL, Table 4), despite recommendations by the developers of this technology to use an equivalent of 5–10 ng/μL. A majority of our samples were from field-collected, silica-dried leaf and bark tissue collected since 2009, however our four outgroups were from leaf tissue collected as long as fourteen years ago. We predict that microfluidic PCR could be useful for much older, degraded tissues, including preserved museum

specimens, as well as very small amounts of tissues, such as those gathered from rare or infrequently collected species.

Phylogenetic utility of targeted nuclear gene regions

Compared to the existing molecular markers used for comparative phylogenetics in *Commiphora* (Gostel et al. in review), the 49 nuclear loci in this study performed well in preliminary phylogenetic analysis. Among existing markers sequenced in *Commiphora*, the external transcribed spacer of the nuclear ribosome (nrETS) has been most phylogenetically informative, and resolves three and two well-supported (ML bootstrap > 70% and Bayesian poster probability PP > 0.9) species-specific subclades in the 'Arafy' (*C. orbicularis*, *C. mafaidoha*, and *C.* sp. nov. H) and 'Rhynchocarpa' (*C. aprevalii*, and *C. grandifolia*) clades, respectively (Gostel et al. in review). Beyond this resolution, prior markers resolved very few subclades or sister-species relationships within either the 'Arafy' or 'Rhynchocarpa' clades. One of these, *C. arafy* and *C.* sp. nov. C, is not testable in this study, because we did not recover sequence data for specimens of *C. arafy*. The nuclear gene regions explored in this study have yielded individual gene trees that have well-supported species-specific subclades, which suggest single orthologous loci are indeed being targeted by the primers.

The nuclear gene regions explored in this study have also yielded individual gene trees containing groups of expected sister species. These clades, which highlight the utility of these loci to elucidate patterns of speciation, include five and two clades containing sister species pairs in the 'Arafy' and 'Rhynochocarpa' clades, respectively

(Figures 6–8). These clades corroborate our previous expectations, such as the clade containing *Commiphora* sp. nov. I and *C.* sp. nov. L. Although previous molecular phylogenetic studies have not tested their placement, they are thought to be closely related on the basis of several morphological synapomorphies, including pubescence and leaf characters. The phylogenies also reveal unanticipated relationships. Within the 'Arafy' clade, species consistently belong to one of two well-supported subclades (Figures 6–8) that are informally named the 'small-leaved' and 'large-leaved' clades and contain species with leaflets < 10 cm long and < 3 cm wide (with the exception of *C. laxecymigera*) and ≥ 10 cm long and ≥3 cm wide, respectively.

Instances of phylogenetic incongruence among the individual loci and concatenated datasets merit further study beyond the preliminary analyses presented here. The most immediate priority is to determine the sensitivity of phylogenetic signal of these loci to alternative model parameters (Brown and Lemmon 2007). Incorporating multi-species coalescent and other coalescent methods (Yang and Rannala 2014, Liu et al. 2015) and partitioning sequence data using signal- and model-based criteria (Lanfear et al. 2012, Frandsen et al. 2015) will be important steps in their complete analysis and the development of a complete species tree that can be used to test evolutionary hypotheses in the genus. Future research will also focus on assessing the utility of the 91 primer pairs for nuclear phylogenomic studies of other taxa. Cross-amplification studies will explore the upper and lower taxonomic bounds of the phylogenetic signal from these loci to delimit the application of this novel suite of markers to resolving the angiosperm tree of life.

References

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, and Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS One 3: e3376.

Bennett MD and Leitch IJ. 2012. Plant DNA C-values database (release 6.0, Dec. 2012) [online]. Available from http://www.kew.org/cvalues/. [accessed 12 October 2014].

Brown JM and Lemmon AR. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Systematic Biology 56: 643–655.

Bybee SM, Bracken-Grissom H, Haynes BD, Hermansen RA, Byers RL, Clement MJ, Udall JA, Wilcox ER, and Crandall KA. 2011. Targeted amplicon sequencing (TAS): A scalable next-gen approach to multilocus, multitaxa phylogenetics. Genome Biology and Evolution 3: 1312–1323.

Cacho NI and Strauss SY. 2013. Single-copy nuclear gene primers for *Streptanthus* and other Brassicaceae from genomic scans, published data, and ESTs. Applications in Plant Sciences 1: 12000002.

Chamala S, García N, Godden GT, Krishnakumar V, Jordon-Thaden IE, De Smet R, Barbazuk WB, Soltis DE, and Soltis PS. 2015. MarkerMiner 1.0: A new application for phylogenetic marker development using angiosperm transcriptomes. Applications in Plant Sciences 3: 1400115.

Chapman MA, Chang J, Weisman D, Kesseli RV, and Burke JM. 2007. Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae). Theoretical and Applied Genetics 115: 747–755.

Cronn R, Knaus BJ, Liston A, Maughan PJ, Parks M, Syring JV, and Udall J. 2012. Targeted enrichment strategies for next-generation plant biology. American Journal of Botany 99: 291–311.

De Smet R, Adams KL, Vandepoele K, Van Montagu MCE, Maere S, and Van de Peer Y. 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proceedings of the National Academy of Sciences 110: 2898–2903.

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

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792–1797.

Frandsen PB, Calcott B, Mayer C, and Lanfear R. 2015. Automatic selection of partitioning schemes for phylogenetic analyses using iterative *k*-means clustering of site rates. BMC Evolutionary Biology 15: 13.

Fulton TM, Van der Hoeven R, Eannetta NT, and Tanksley SD. 2002. Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. The Plant Cell 14: 1457–1467.

Gostel MR and Weeks A. 2014. Development of novel, exon-primed intron-crossing (EPIC) markers from EST databases and evaluation of the phylogenetic utility in *Commiphora* (Burseraceae). Applications in Plant Sciences 2: 1300098.

Gostel MR, Phillipson PB, and Weeks A. In Review. Phylogenetic reconstruction of the myrrh genus, *Commiphora* (Burseraceae), reveals multiple radiations in Madagascar and clarifies infrageneric relationships. Systematic Botany.

Hanson L, McMahon KA, Johnson MAT, and Bennett MD. 2001. First nuclear DNA C-values for another 25 angiosperm families. Annals of Botany 91: 1–8.

Henriquez CL, Arias T, Pires JC, Croat TB, and Schaal BA. 2014. Phylogenomics of the plant family Araceae. Molecular Phylogenetics and Evolution 75: 91–102.

Ho T, Cardle L, Xu X, Bayer M, Prince KSJ, Mutava RN, Marshall DF, and Syed N. 2014. Genome-tagged amplification (GTA): A PCR-based method to prepare sample-tagged amplicons from hundreds of individuals for next generation sequencing. Molecular Breeding 34: 977–988.

Ilut DC and Doyle JJ. 2012. Selecting nuclear sequences for fine detail molecular phylogenetic studies in plants: A computation approach and sequence repository. Systematic Botany 37:7–14.

Johnson MT, Carpenter EJ, Tian Z, Bruskiewich R, Burris JN, Carrigan CT, Chase MW, Clarke ND, Covshoff S, dePamphilis CW, Edger PP, Goh F, Graham S, Greiner S, Hibberd JM, Jordon-Thaden I, Kutchan TM, Leebens-Mack J, Melkonian M, Miles N, Myburg H, Patterson J, Pires JC, Ralph P, Rolf M, Sage RF, Soltis D, Soltis P, Stevenson D, Stewart CN Jr, et al. 2012. Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. PLoS One 7: e50226.

Lanfear R, Calcott B, Ho SYW, and Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution 29: 1695–1701.

Lemmon EM and Lemmon AR. 2013. High-throughput genomic data in systematics and phylogenetics. Annual Reviews in Ecology, Evolution, and Systematics 44: 99–121.

Liu L, Xi Z, and Davis CC. 2015. Coalescent methods are robust to the simultaneous effects of long branches and incomplete lineage sorting. Molecular Biology and Evolution 32: 791–805.

Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, and Hernández-Hernández T. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytologist doi: 10.1111/nph.13264.

Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, and Turner DJ. 2010. Target-enrichment strategies for next-generation sequencing. Nature Methods 7: 111–118.

Mandel JR, Dikow RB, Funk VA, Masalia RR, Staton SE, Kozik A, Michelmore RW, Rieseberg LH, and Burke JM. 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. Applications in Plant Sciences 2: 1300085.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal 17: 10–12.

Matasci N, Hung L-H, Yan Z, Carpenter EJ, Wickett NJ, Mirarab S, Nguyen N, Warnow T, Ayyampalayam S, Barker M, Burleigh JG, Gitzendanner MA, Wafula E, Der JP, dePamphilis CW, Roure B, Philippe H, Ruhfel BR, Miles NW, Graham SW, Mathews S, Surek B, Melonian M, Soltis DE, Soltis, PS, Rothfels C, Pokorny L, Shaw JA, DeGironimo L, Stevenson DW, Villarreal JC, Chen T, Kutchan TM, Rolf M, Baucom RS, Deyholos MK, Samudrala R, Tian Z, Wu X, Sun X, Zhang Y, Wang J, Leebens-Mack J, and Wong GK-S. 2014. Data access for the 1,000 plants (1KP) project. GigaScience 3: 17.

Mathieu-Daudé F, Welsh J, Vogt T, and McClelland M. 1996. DNA rehybridization during PCR: the 'C_ot effect' and its consequences. Nucleic Acids Research 24: 2080–2086.

Miller MA, Pfeiffer W, and Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE) 14 November 2010, New Orleans, Louisiana pp 1–8.

Naumann J, Symmank L, Samaín M-S, Müller KF, Neinhuis C, dePamphilis CW, and Wanke S. 2011. Chasing the hare – evaluating the phylogenetic utility of a nuclear single copy gene region at and below species level within the species rich group *Peperomia* (Piperaceae). BMC Evolutionary Biology 11: 357.

Qu W, Zhou Y, Zhang Y, Lu Y, Wang X, Zhao D, Yang Y, Zhang C. 2012. MFEprimer-2.0: A fast thermodynamics-based program for checking PCR primer specificity. Nucleic Acids Research 40: W205-8.

Richardson BA, Page JT, Bajgain P, Sanderson SC, and Udall JA. 2012. Deep sequencing of amplicons reveals widespread intraspecific hybridization and multiple origins of polyploidy in big sagebrush (*Artemisia tridentata*; Asteraceae). American Journal of Botany 99: 1962–1975.

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

Untergrasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, and Rozen SG. 2012. Primer3 – new capabilities and interfaces. Nucleic Acids Research 40: e115.

Wall PK, Leebens-Mack J, Müller KF, Field D, Altman NS, and dePamphilis CW. 2008. PlantTribes: A gene and gene family resource for comparative genomics in plants. Nucleic Acids Research 36: D970-6.

Weeks A, Daly DC, and Simpson BB. 2005. The phylogenetic history and biogeography of the frankincense and myrrh family (Burseraceae) based on nuclear and chloroplast sequence data. Molecular Phylogenetics and Evolution 35: 85–101.

Weeks A and Simpson BB. 2007. Molecular phylogenetic analysis of *Commiphora* (Burseraceae) yields insight on the evolution and historical biogeography of an "impossible" genus. Molecular Phylogenetics and Evolution 42: 62–79.

Weeks A, Zapata F, Pell, SK, Daly DC, Mitchell JD, and Fine PVA. 2014. To move or to evolve: Contrasting patterns of intercontinental connectivity and climatic niche evolution in "Terebinthaceae" (Anacardiaceae and Burseraceae). Frontiers in Genetics 5: 409.

Weitemier K, Straub SCK, Cronn RC, Fishbein M, Schmickl R, McDonnell A, and Liston A. 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. Applications in Plant Sciences 2: 1400042.

Wu F, Mueller LA, Crouzillat D, Petiard V, and Tanksley SD. 2006. Combining bioinformatics and phylogenetics to identify large sets of single copy, orthologous genes (COSII) for comparative, evolutionary and systematics studies: A test case in the euasterid plant clade. Genetics 174: 1407–1420.

Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, Zhou X, Lam TW, Li Y, Xu X, Wong GK, Wang J. 2014. SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30: 1660–1666.

Yang Z and Rannala B. 2014. Unguided species delimitation using DNA sequence data from multiple loci. Molecular Biology and Evolution 31: 3125–3135.

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

Zhang N, Zeng L, Shan H, and Ma H. 2012. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. New Phytologist 195: 923–937.

Zimmer EA and Wen J. 2012. Using nuclear gene data for plant phylogenetics: Progress and prospects. Molecular Phylogenetics and Evolution 65: 774–785.

Table 4. List of accessions sampled in this study with their DNA concentration ([DNA]), numbers of loci (No. loci) with at least 15 sequence reads mapped, raw reads, filtered and trimmed reads, and uniquely mapped reads, as well as the average and median number.
[†]Designates accessions that we excluded from phylogenetic analyses and alignments because the median number of mapped reads was erratically low, suggesting skewed read sequencing.
[‡]Designates accessions that were excluded because they did not produce any sequence data, sequence data that was produced was very low, or sequenced reads did not map to targets.
[§]Designates accessions that were excluded because sequence reads suggested cross contamination.
[¶]Designates accessions that failed during microfluidic PCR.

Specimen	[DNA] (ng/µL)	No. loci	No. reads	No. filtered & trimmed reads	No. unique mapped reads	Average reads per locus	Median reads per locus
Bursera fagaroides Weeks 01-X-08-01	15.6	71	1001606	982688	975646 (99.3%)	11214	2179
B. simaruba Pell s.n.	21.6	68	643458	609446	603939 (99.1%)	7276	2729
B. spinescens Weeks	11.3	65	874760	860980	854794 (99.3)	11247	3252
B. tecomaca Weeks	9.8	69	1037074	1009156	999314 (99%)	11619	3167
Commiphora ankaranensis Weeks 10-I-11-02	8.24	56	364506	395730	392384 (99.2%)	6131	1445
C. ankaranensis Weeks 10-I-12-03	13.2	51	347776	342328	339840 (99.3%)	5859	1737
C. ankaranensis Weeks 10-I-14-02[†]	0.248	37	1852618	327128	319151 (97.6%)	8625	58
C. aprevalii Gostel 124	1.64	53	502556	466758	462645 (99.1%)	7119	2833
C. aprevalii Gostel 128	6.28	59	958996	912566	904397 (99.1%)	12058	1810

C. aprevalii Weeks 10-I-20-04	3.02	62	1234450	1175170	1163320 (99%)	14914	2389
C. brevicalyx var. *brevicalyx* Weeks 10-I-24-05	7.24	53	441534	372008	356072 (95.7%)	6035	1896
C. brevicalyx var. *vezorum* Weeks 10-I-23-08	2.42	53	737486	355524	353189 (99.3%)	6094	2296
C. capuronii Weeks 10-I-15-04	0.42	39	1419724	1268966	1260415 (99.3%)	30741	2156
C. capuronii Gostel 64	0.684	50	788548	757394	753130 (99.4%)	12764	2659
C. capuronii Gostel 67	4.92	41	1180892	606480	599419 (98.8%)	11753	279
C. coleopsis Gostel 142¶	9.4	35	NA	NA	NA	NA	NA
C. sp. Weeks 10-II-14-13	6.36	58	870972	858496	850867 (99.1%)	12512	3529
C. sp. Weeks 10-II-15-02	2.9	43	723286	717340	712947 (99.4%)	15843	11886
C. falcata Weeks 10-I-26-03	1.42	51	260132	249192	247564 (99.3%)	3994	1366
C. falcata Weeks 10-I-27-04	1.49	54	1363152	1197910	1184630 (98.9%)	19743	3908
C. franciscana Weeks 10-I-22-10	4.44	54	468628	455964	451077 (98.9%)	6732	2327
C. franciscana Weeks 10-I-23-02	5.72	51	429422	421128	417746 (99.2%)	6962	2694
C. fraxinifolia Gostel 76	3.7	42	1030322	979908	973984 (99.4%)	21173	9638
C. fraxinifolia Gostel 102	10.8	53	497418	458744	455270 (99.2%)	7463	2079
C. grandifolia Weeks 10-I-13-01	13.4	56	435606	413610	410117 (99.2%)	5943	1113

C. grandifolia Weeks 10-II-14-15	23.2	56	478292	423502	419963 (99.2%)	6175	1833
C. grandifolia Gostel 53	13.6	54	681934	672092	667248 (99.3%)	10591	1345
C. grandifolia Gostel 121	11.9	58	800278	755816	749434 (99.2%)	10706	2471
C. granulifera Gostel 120	6.8	50	353180	343562	341377 (99.4%)	5786	2312
C. granulifera Gostel 125	0.596	49	892422	668106	663653 (99.3%)	11850	3578
C. granulifera Gostel 127	21	57	972990	827444	820229 (99.1%)	11552	2223
C. guillauminii Gostel 100	23	60	1625864	1349706	1337474 (99.1%)	18073	2131
C. humbertii Weeks 10-I-20-08	3.73	55	974682	962818	957300 (99.4%)	15195	3745
C. humbertii Gostel 136	4.4	55	475424	473634	469778 (99.2%)	7577	1765
C. humbertii Gostel 146	5.58	50	457394	418918	415981 (99.3%)	7050	1765
C. lamii Weeks 10-I-26-02	0.912	55	1295252	1223652	1211604 (99%)	20193	3934
C. lamii Weeks 10-I-26-04	0.828	65	1845768	1537726	1523070 (99%)	19526	2568
C. laxecymigera Gostel 115	2.16	55	474136	429740	427219 (99.4%)	6781	1985
C. cf. *leandreana* Weeks 10-II-14-12	0.268	55	749438	527266	522306 (99.1%)	8162	2166
C. cf. *leandreana* Weeks 10-II-14-14	2.26	56	1579098	1249420	1235551 (98.9%)	19008	3659
C. mafaidoha Gostel 104	1.6	53	504302	438468	432775	6762	1726

					(98.9%)		
C. mafaidoha Weeks 10-II-13-02	1.66	43	286998	232624	230247 (99%)	4604	1800
C. mahafaliensis Weeks 10-I-21-02	13.6	41	328568	319218	317086 (99.3%)	6893	3122
C. mahafaliensis Weeks 10-I-22-02	12.8	54	415666	411402	408597 (99.3%)	6286	1793
C. mahafaliensis Weeks 10-I-28-03†	8.32	24	1503284	1372594	1363710 (99.4%)	54548	35
C. marchandii Weeks 10-II-15-04	5.12	56	1377574	659814	656201 (99.5%)	10095	3278
C. marchandii Gostel 95	7.48	59	1499164	1317502	1304775 (99%)	16727	2397
C. marchandii Gostel 143	2.89	60	1055546	893304	886074 (99.2%)	12306	1718
C. monstruosa Weeks 10-I-21-07	14.4	53	383266	374494	371954 (99.3%)	5999	1405
C. orbicularis Weeks 10-I-23-04	6.28	53	652140	608382	603844 (99.3%)	9739	2764
C. orbicularis Weeks 10-I-26-05	10	37	1004066	717004	711864 (99.3%)	20338	12862
C. orbicularis Gostel 111	3.18	54	604080	593424	587535 (99%)	9180	3593
C. ankaranensis Gostel 48	2.89	41	270982	267256	264894 (99.1%)	5636	2722
C. sp. Gostel 69	4.2	31	1680016	1456184	1446453 (99.3%)	45201	2105
C. cf. *arafy* Gostel 87	8.52	53	363412	315484	312758 (99.1%)	5044	2023
C. pervilleana Gostel 93	0.7	54	409880	334292	332003 (99.3%)	5354	2374

C. pervilleana Gostel 97	3.19	50	535284	340538	337113 (99%)	6360	1973
C. pterocarpa Weeks 10-I-28-11	0.64	46	226298	215518	213030 (98.8%)	3873	1476
C. pterocarpa Gostel 106	9.04	60	577484	478292	473971 (99.1%)	6582	1394
C. sinuata Weeks 10-I-23-01	2.76	53	310402	295722	292063 (98.8%)	4563	2135
C. sinuata Weeks 10-I-23-06	7.88	44	484448	461540	457976 (99.2%)	8979	2770
C. sp. Gostel 54[†]	1.65	28	2363426	844588	831171 (98.2%)	28661	56
C. sp. nov. A Weeks 10-I-28-08	0.792	48	361512	284810	281746 (98.9%)	4857	1382
C. sp. nov. C Gostel 83	3.38	48	383538	378434	375823 (99.3%)	6593	1536
C. sp. nov. D Gostel 86	3.96	55	374088	338970	336195 (99.2%)	5336	1791
C. sp. nov. E Gostel 140	3.35	53	570122	258396	256197 (99.1%)	4201	1752
C. sp. nov. G Weeks 10-I-11-01	3.22	54	740600	688338	683230 (99.3%)	11019	2856
C. sp. nov. O Weeks 10-I-12-05	3.46	43	252496	238232	236485 (99.3%)	4826	2593
C. sp. nov. G Weeks 10-I-14-04	5.04	52	605854	558368	554666 (99.3%)	9563	3845
C. sp. nov. G Weeks 10-I-14-10	17.9	54	192346	182674	181185 (99.2%)	2832	1073
C. sp. Weeks 10-I-12-01	11.8	49	492640	482134	478122 (99.2%)	8854	3955
C. sp. nov. G Gostel 62	1.32	53	344394	287904	285256	4754	1729

					(99.1%)		
C. sp. nov. H Gostel 145	0.699	54	712758	631118	625377 (99.1%)	9333	2524
C. sp. nov. I Weeks 10-I-09-01	2.36	38	327434	313646	311099 (99.2%)	7234	3564
C. sp. nov. I Gostel 43	5.02	55	402746	379390	376132 (99.1%)	5786	1930
C. sp. nov. I Gostel 46	4.4	57	450504	429334	425723 (99.2%)	6354	1241
C. sp. nov. J Gostel 44	4.46	38	454764	447588	444492 (99.3%)	10841	5477
C. sp. nov. L Weeks 10-II-14-05	4.44	51	368118	306102	303528 (99.2%)	5144	1627
C. sp. nov. L Weeks 10-II-14-11	5.16	40	577572	543476	539736 (99.3%)	11994	7214
C. sp. nov. M Weeks 10-I-28-04	0.464	52	511554	481502	477933 (99.3%)	7708	1774
C. sp. nov. M Weeks 10-I-28-10	0.536	52	1727870	1463088	1449375 (99.1%)	25427	1300
C. tetramera Gostel 79	3.82	50	817976	800538	793631 (99.1%)	13683	4222
C. tetramera Gostel 90	4.32	54	158362	155078	153711 (99.1%)	2606	1069
C. ankaranensis Weeks 10-I-12-11[§]	3.06	37	231232	225664	223839 (99.2%)	2833	966
C. arafy Weeks 10-II-12-01[‡]	10	NA	130	0	0	NA	NA
C. arafy Weeks 10-II-13-01[‡]	2.74	NA	224	0	0	NA	NA
C. arafy Weeks 10-II-14-07[‡]	5.24	NA	194	70	23 (32.9%)	NA	NA
C. brevicalyx var. *vezorum* Weeks 10-I-20-09[‡]	7.4	NA	0	0	0	NA	NA
C. fraxinifolia Weeks 10-II-13-05[‡]	11.1	NA	0	0	0	NA	NA

C. monstruosa Weeks 10-I-21-09[§]	6.12	68	351776	340418	337325 (99.1%)	4113	1734
C. sp. nov. B Gostel 82[¶]	9.68	NA	NA	NA	NA	NA	NA
C. sp. nov. H Gostel 141[¶]	2.12	NA	NA	NA	NA	NA	NA
C. sp. nov. I Gostel 51[‡]	1.63	NA	346320	161276	NA	NA	NA
C. sp. nov. J Gostel 58[§]	2.48	65	681342	668872	662909 (99.1%)	8391	1586
C. sp. nov. L Weeks 10-II-14-06[§]	10.5	49	392182	365784	363550 (99.4%)	6732	2707
C. sp. nov. N Gostel 47[¶]	0.215	NA	NA	NA	NA	NA	NA
Average, all taxa	NA	51.4	NA	52416864	51797426 (98.8%)	10396	2108

Table 5. Sequence characteristics for 49 nuclear loci used for phylogenetic analyses in this study.
[†]Gel length refers to the observed length of amplified PCR product estimated following gel electrophoresis and visual inspection of primer validation step products.
[‡]Designates target nuclear gene regions that were too long for sequence reads to cover the entire locus length and for which we only have forward and reverse read alignment matrices. Note the difference between aligned length and predicted/gel lengths.
[§]Designates loci that were excluded from phylogenetic analyses because they did not recover a monophyletic *Commiphora*.

Locus ID	No. taxa	Predicted length	Gel[†] length	Aligned length	No. PICs (%)	Min. length	Max. length	% GC	Missing Data (%)
AT1G18060	63	613	950	1044	90 (8.6%)	600	1017	36.8	37.2
AT1G31780	79	900	900	540	63 (11.7%)	533	540	41.1	1.3
AT1G59990	66	318	600	521	73 (14%)	510	520	39.7	1.9
AT1G65030	79	365	400	404	44 (10.9%)	386	401	51.4	3.4
AT1G65070	80	450	600	539	50 (9.3%)	538	539	39.9	0.2
AT1G76450F[‡]	60	450	1200	304	38 (12.5%)	265	289	38.5	2.8
AT1G76450R[‡]	60	450	1200	334	71 (21.3%)	267	332	39.5	13.6
AT1G77550A	65	450	1100	1957	121 (6.2%)	603	1884	36.2	34.8
AT1G77550B	49	450	1100	807	65 (8.1%)	594	807	37.7	23.1
AT1G77930A	80	450	800	735	122 (16.6%)	574	729	38.4	6.8
AT2G03667	79	230	600	582	61 (10.5%)	549	575	34.6	4
AT2G04620	80	450	500	473	30 (6.3%)	457	473	39.6	0
AT2G04740	80	523	700	515	53 (10.3%)	515	515	54.4	0
AT2G05120	80	478	800	493	33 (6.7%)	493	493	41.2	0.1
AT2G05320	78	450	600	474	23 (4.9%)	474	474	43.1	0
AT2G17265	80	497	400	771	60 (7.8%)	596	761	53.4	16.9
AT2G20330	66	450	600	649	65 (10%)	567	647	41.2	4.5
AT2G21710F[‡]	77	450	600	312	38 (12.2%)	200	299	32.5	7.1
AT2G22370B	68	450	900	1700	138 (8.1%)	755	1684	38.7	49.8
AT2G31890	80	445	600	445	50 (11.2%)	445	445	42.6	0
AT2G31890A	80	450	1000	538	56 (10.4%)	538	538	42.6	0
AT2G40760	78	262	250	339	30 (8.8%)	338	339	47.6	1.3

AT2G44660A	80	450	600	454	40 (8.8%)	454	454	44.6	0.1
AT2G44660B	72	450	800	1135	145 (12.8%)	207	1072	38.2	33.1
AT2G46100	80	164	200	371	60 (16.2%)	361	371	35.9	1.3
AT3G01380F‡	67	304	800	349	49 (14%)	298	304	39.1	15.2
AT3G01380R‡	67	304	800	312	43 (13.8%)	301	311	34.7	3.4
AT3G04650F‡	75	303	1000	306	30 (9.8%)	301	305	39	1
AT3G04650R‡	76	303	1000	271	39 (14.4%)	216	254	34.7	14.4
AT3G10400	80	268	400	325	32 (9.8%)	320	325	55.4	1.7
AT3G14910	64	233	800	945	120 (12.7%)	499	882	32.9	35.2
AT3G21540F‡	64	450	800	125	10 (8%)	114	124	33.1	7.6
AT3G21540R‡	64	450	800	306	28 (9.2%)	299	306	40.5	1.9
AT3G22660	80	241	400	437	24 (5.5%)	437	437	45.1	0
AT3G26580F‡	78	449	600	316	25 (7.9%)	300	316	37.7	4.8
AT3G26580R‡	78	449	600	307	31 (10.1%)	304	307	45.4	2.9
AT3G46220	19	450	1000	563	19 (3.4%)	562	563	42.3	0.2
AT3G54460	80	192	250	236	13 (5.5%)	233	236	42	0.6
AT4G00560	41	450	1200	332	25 (7.5%)	300	306	35.3	10.4
AT4G14605	79	545	700	540	37 (6.9%)	493	537	43.3	1
AT4G18810B	72	450	900	732	85 (11.6%)	602	726	40.1	16.9
AT4G19900	79	200	400	360	47 (13.1%)	309	327	36.9	10.2
AT4G21770	60	449	900	905	84 (9.3%)	534	895	35.7	37.1
AT4G29590	78	450	600	527	76 (14.4%)	500	526	40.4	7.2
AT4G31770	76	450	600	639	71 (11.1%)	608	636	39	5.4
AT5G04910	77	209	400	273	47 (17.2%)	273	273	45.1	1
AT5G15680	79	450	900	549	49 (8.9%)	453	547	40.7	1.5
AT5G52180	80	450	550	459	31 (6.8%)	450	459	42	1.5
AT2G44760	77	450	700	883	155 (17.6%)	468	742	38	20.3
AT1G21840R‡§	34	622	800	303	36 (11.9%)	302	303	43.6	1.8
AT2G20790§	82	271	400	328	12 (3.7%)	328	328	43.6	0.8
AT2G36740§	35	273	750	821	64 (7.8%)	588	794	36.2	25.1
AT3G29130F‡§	31	159	1200	306	23 (7.5%)	302	305	29.5	3.4

AT3G29130R[‡§]	32	159	1200	336	27 (8%)	299	323	35.8	14.3
AT4G37510[§]	58	226	250	280	21 (7.5%)	268	269	48.9	3.9

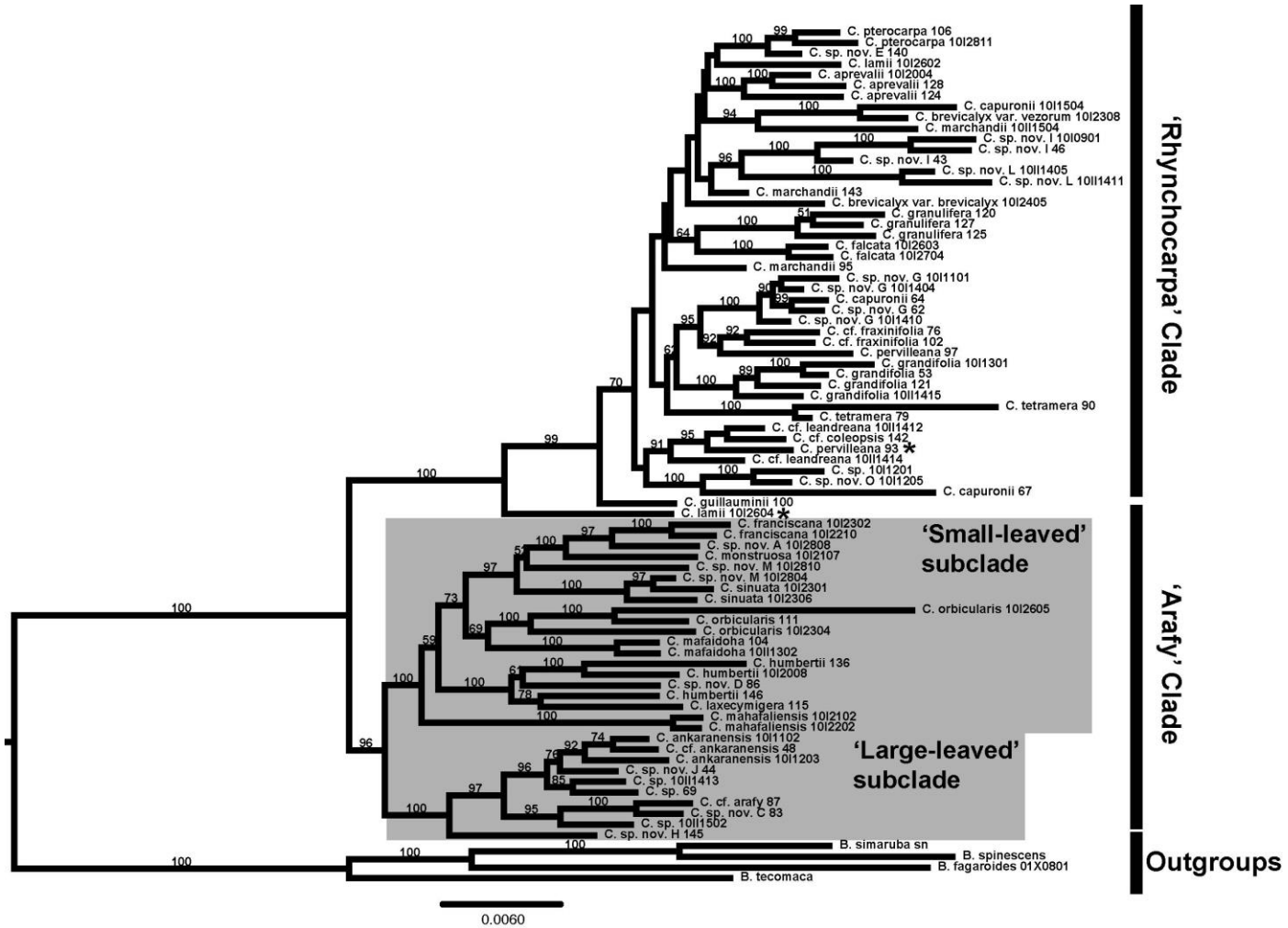


Figure 6. Phylogeny of Malagasy *Commiphora* showing 'Arafy' and 'Rhynchocarpa' clades, resulting from RAxML analysis of 44 nuclear loci. Asterisks are used to identify taxa that change position between analyses. Shaded boxes correspond to 'large' and 'small-leaved' subclades.

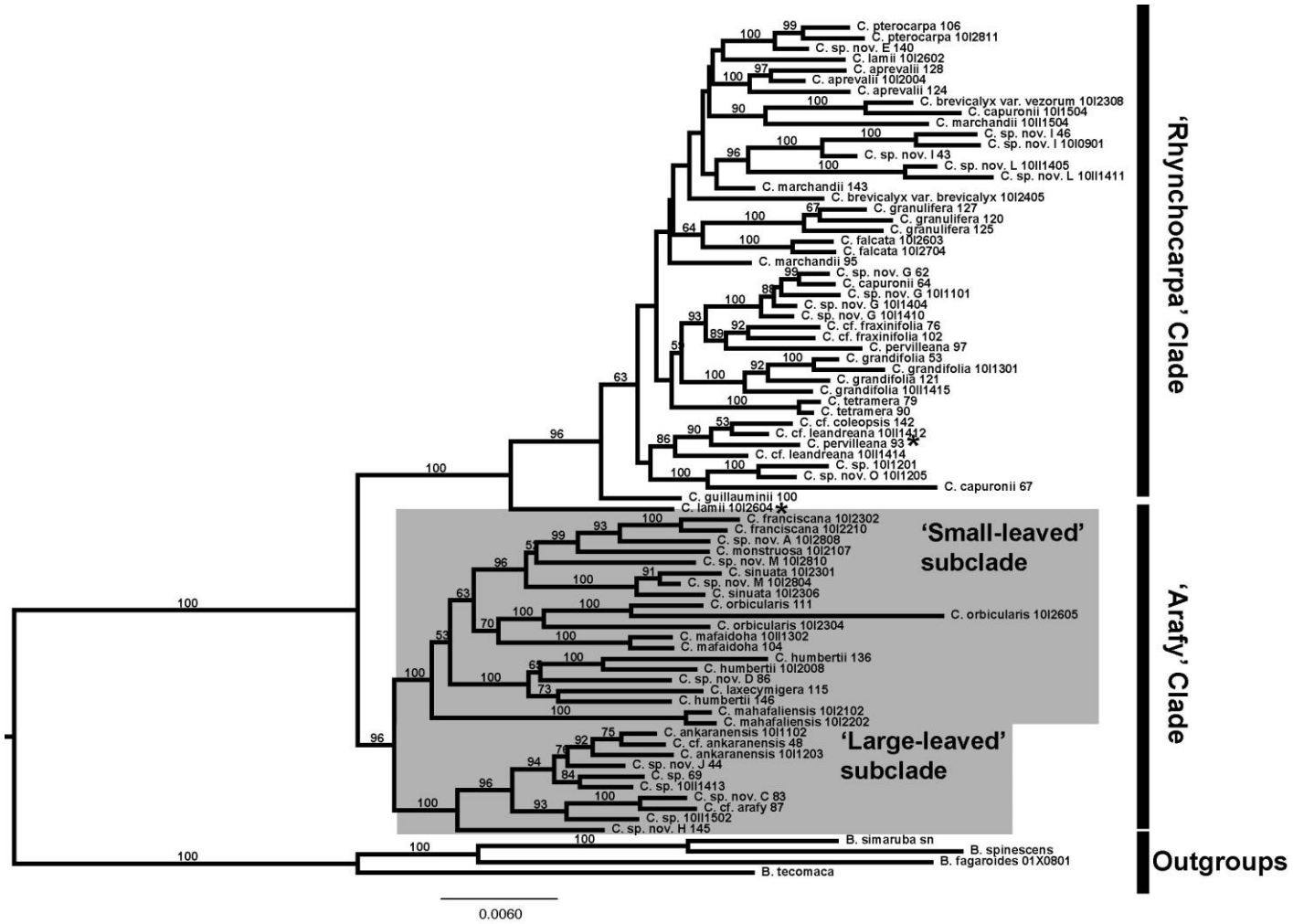


Figure 7. Phylogeny of Malagasy *Commiphora* showing 'Arafy' and 'Rhynchocarpa' clades, resulting from RAxML analysis of 41 nuclear loci containing <25% missing sequences. Asterisks are used to identify taxa that change position between analyses. Shaded boxes correspond to 'large' and 'small-leaved' subclades.

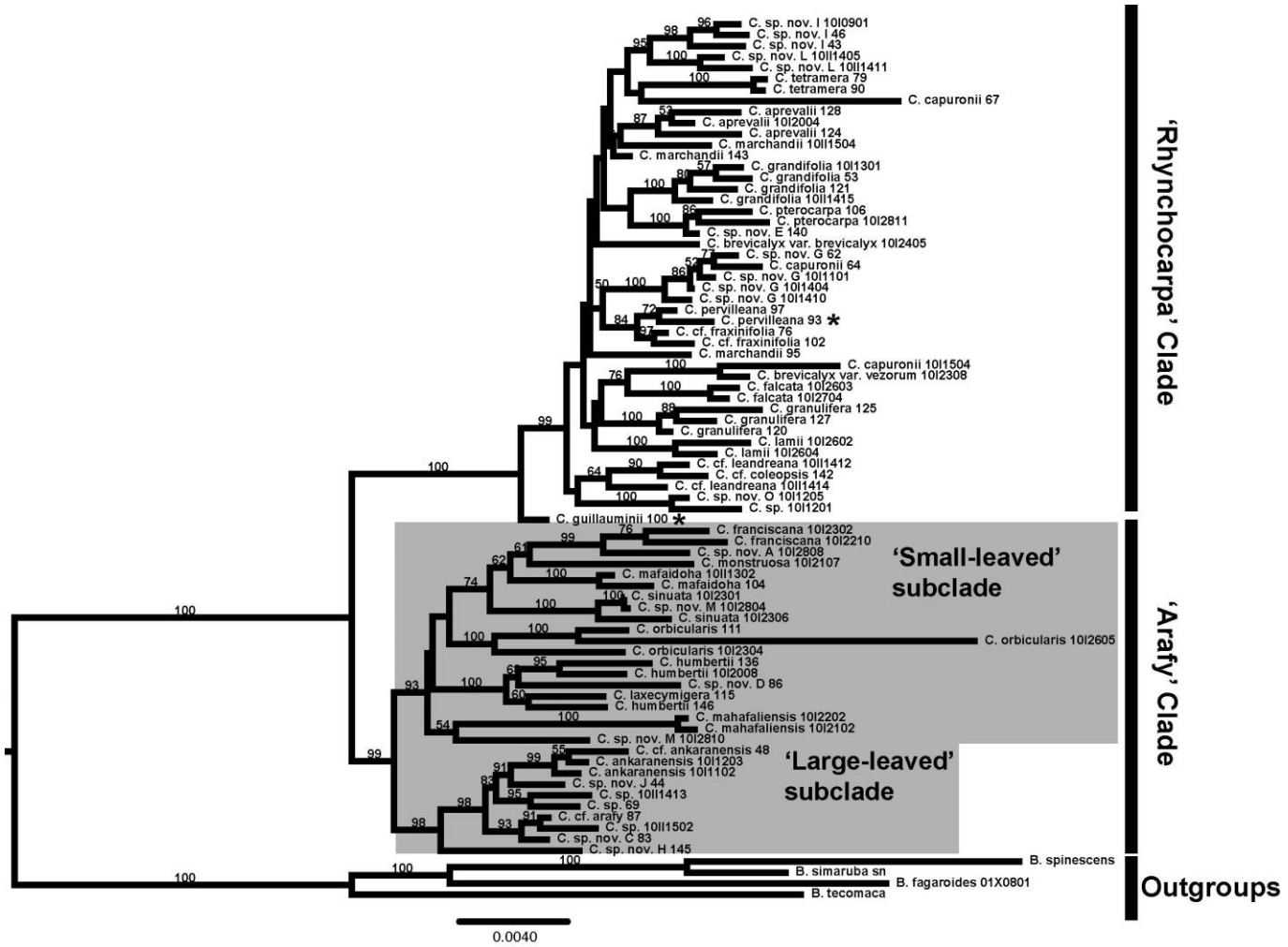


Figure 8. Phylogeny of Malagasy *Commiphora* showing 'Arafy' and 'Rhynchocarpa' clades, resulting from RAxML analysis of 20 nuclear loci that contains ≤1% missing sequences. Asterisks are used to identify taxa that change position between analyses. Shaded boxes correspond to 'large' and 'small-leaved' subclades.

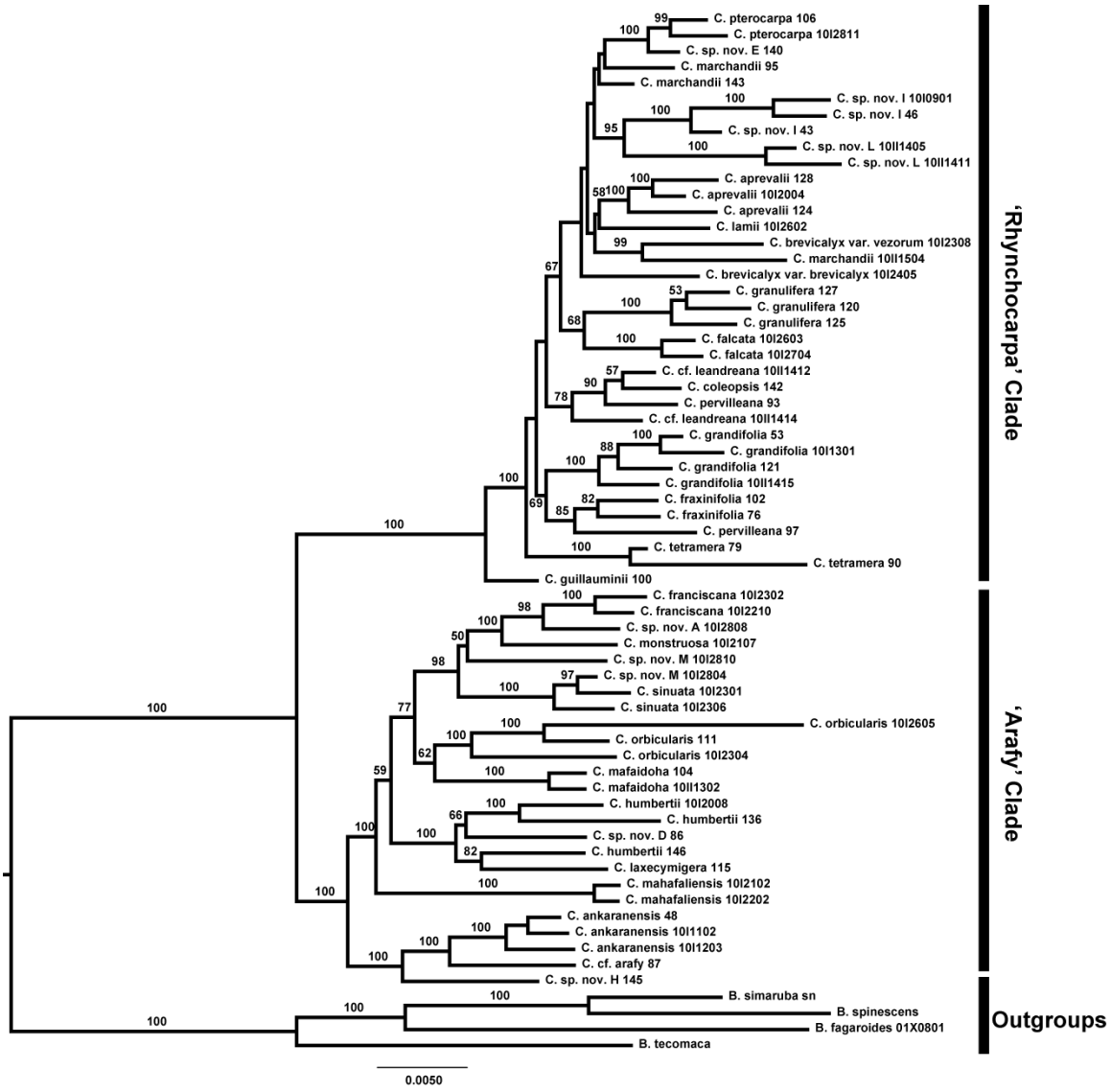


Figure 9. Reduced, 65 taxon phylogeny of Malagasy *Commiphora* showing 'Arafy' and 'Rhynchocarpa' clades, resulting from RAxML analysis of 44 nuclear loci.

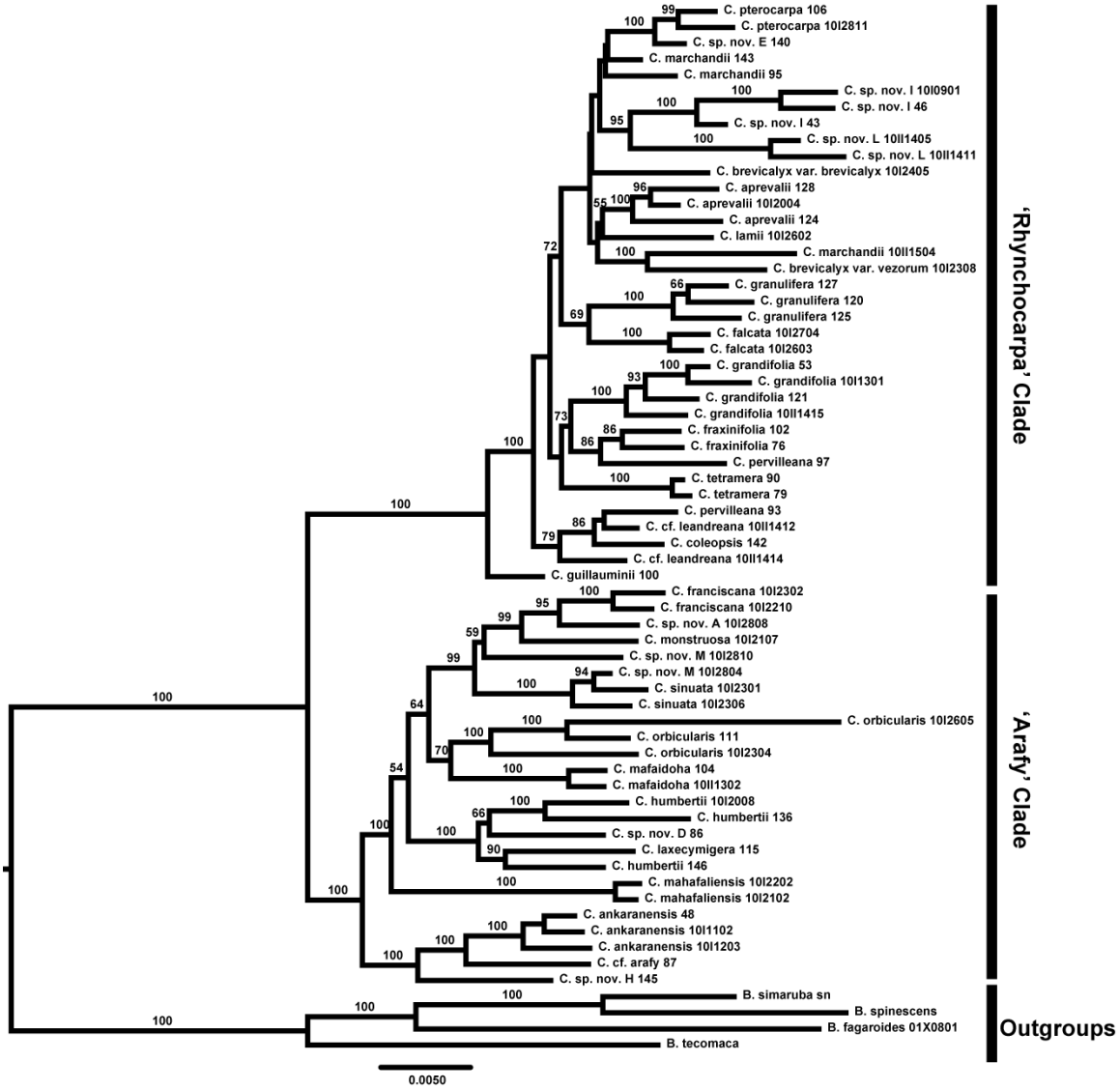


Figure 10. Reduced, 65 taxon phylogeny of Malagasy *Commiphora* showing 'Arafy' and 'Rhynchocarpa' clades, resulting from RAxML analysis of 41 nuclear loci.

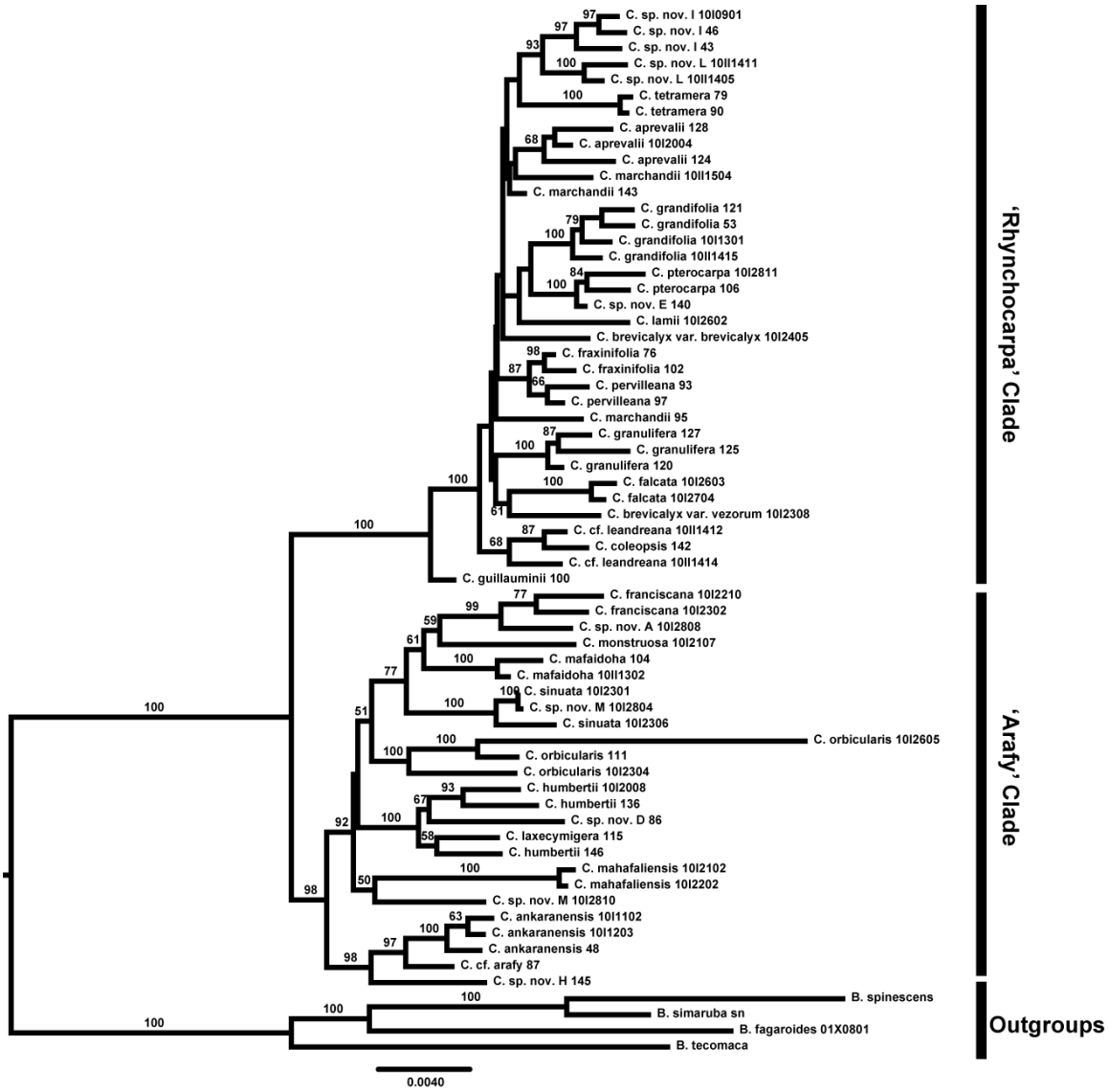


Figure 11. Reduced, 65 taxon phylogeny of Malagasy *Commiphora* showing 'Arafy' and 'Rhynchocarpa' clades, resulting from RAxML analysis of 20 nuclear loci.

Chapter 5: A partial taxonomic revision of the Rhynchocarpa clade of *Commiphora* (Burseraceae), endemic to Madagascar

To be submitted to Systematic Botany.

Abstract

Recent studies in *Commiphora* Jacq. (Burseraceae) have revealed a complex history of species evolution in this genus that has resulted in four colonization and radiation events in Madagascar. Two of these radiations are, morphologically, nearly indistinguishable by traditional morphological characters historically used for species circumscription in *Commiphora* and more broadly in Burseraceae. Apparently, convergent evolution has played a role in the diversification of *Commiphora* throughout Madagascar. In this study we present a partial taxonomic revision for the most species-rich clade of *Commiphora* in Madagascar that is referred to as the 'Rhynchocarpa' clade. Previous work suggests that this clade contains 26 species and because of the species richness belonging to this clade, we have decided to revise only a subset of its species. One character state shared by seven species in the 'Rhychocarpa' clade is the presence of stellate pubescence on vegetative and reproductive parts of the plant. This trait is shared only by one other species in *Commiphora* from continental Africa. Due to the rarity of stellate pubescence

in the genus, we feel our approach is an important step toward the necessary treatment of an interesting group.

Introduction

Commiphora Jacq. is the most species-rich genus in the frankincense and myrrh family, Burseraceae, with ca. 190 species of shrubs and trees distributed predominantly throughout dry and seasonally dry tropical forests in continental Africa, Madagascar, the Arabian Peninsula, and south Asia. Recent molecular phylogenetic work (Gostel et al. in press) has indicated that the Malagasy species of *Commiphora* comprise belong to four distinct clades (Fig. 12), two of which are species-rich ('Arafy' and 'Rhynchocarpa' clades), both endemic to the island. Twenty-eight species of *Commiphora* have been described from Madagascar, but recent studies suggest that diversity is much greater there, and at least 16 new Malagasy species are awaiting description (Gostel et al. in press; Phillipson and Raharimampionona, unpublished data). Species of *Commiphora* are deciduous, dioecious, and flowers and fruit are often produced only after plants lose their leaves. As a result of this combination of features, accurate species descriptions depend upon multiple collections. Taxonomic revision is sorely needed for *Commiphora*, but must be carefully targeted to ensure descriptions are made from suitable material. In this manuscript we have focused on the Malagasy endemic 'Rhynchocarpa' clade and, in particular, only the seven species in this clade that have stellate pubescence.

The 'Rhynchocarpa' clade is the most diverse of four clades in Madagascar with 15 described and currently 11 undescribed species are known (Gostel et al. in press).

Among these undescribed species, four have a stellate pubescence. Due to limited phylogenetic resolution, it is currently unknown whether the stellate species form a discreet sub-clade, however due to the rarity of this trait in the genus and more broadly in the family, this group of species is a logical target for taxonomic revision, and they very likely do represent a natural group. Only three currently described species of *Commiphora* are known to have stellate hairs, including two from Madagascar (*C. aprevalii* Guillaumin and *C. stellulata* H. Perr., both from the 'Rhynchocarpa' clade) and one species from continental Africa (*C. stellatopubescens* Gillett ex Thulin) that has not been placed in any infrageneric group because it has not been sampled in any molecular phylogenetic study. One variety, *C. aprevalii* var. *granulifera* Capuron, and one species, *C. stellulata* H. Perr., are excluded from this manuscript. The variety is excluded because it lacks stellate pubescence and we regard it as a distinct species. This variety is very distinct from *C. aprevalii* in that is has smooth bark, much shorter inflorescence. We have made this designation in another article (Phillipson et al. in prep.). The excluded species, *C. stellulata*, is excluded because we lack sufficient material to adequately describe it, we could only find one specimen for this species from 1962 (*Service Forestier* 21,677 (TEF).

Objectives

A taxonomic treatment of seven species of *Commiphora* is provided, including careful description of each species (including five novelties) in order to provide a better

understanding of diversity and the distribution and conservation status of species in this genus. Descriptions of five new species and one previously recognized species (*C. aprevalii* (Baill.) Guillaumin) are included. We also provide a conservation assessment (when possible), distribution maps, and a key to the six species of this easily distinguished selection of taxa. Illustrations are provided with descriptions of new species and also images of important morphological features that distinguish these species. Representative species of the 'Rhynchocarpa' clade included in this treatment are unique among Malagasy *Commiphora* because they produce stellate pubescence on vegetative and/or reproductive parts during some or all stages of growth. Our selection of these species in this revision is a firm approach toward necessary treatment of an interesting group of Malagasy *Commiphora*.

Materials and Methods

All specimens listed in this revision were examined in their respective herbaria (GMUF, MO, P, TAN, TEF) or through loans from P. Distribution maps were generated using QGIS v.2.9 based upon coordinates from collections, or from post-facto geo-referencing as far as possible (presented in square brackets in specimen lists below). Leaf architecture characteristics were measured using terminology and features described in (Ellis et al. 1999). Additional morphological features were examined from material available with freshly collected specimens in the field, herbarium preserved material, and also on the basis of characters historically used in taxonomic treatments of *Commiphora*.

Floral characters are not used extensively in our taxonomic treatment because 1) variation among flower characters is limited in this group of species and 2) material for three species is limited such that we did not have specimens with either male or bisexual flowers. The lack of good flowering material therefore limited our ability to make confident taxonomic comparisons of flower characters for key purposes. This is a challenge in *Commiphora* that results from the presence of precocious, androdioecious flowers and a deciduous habit – collection of material for species description therefore often requires collections from multiple seasons to ensure sampling an adequate suite of characters (Gillett 1991). Measurements were made using a digital caliper and (when necessary) a stereoscope.

Scanning electron microscopy (SEM) was carried out using a Zeiss SUPRA55-VP Electron Beam Lithography NPGS at the Georgetown University GNULab (http://gnulab.georgetown.edu/facilities). For designation of IUCN Red List categories in our conservation assessments, we utilized criterion B and estimated the extent of occurrence (EOO) and area of occupancy (AOO) for each species using GeoCAT (Bachman et al. 2011, http://www.geocat.kew.org). For AOO calculations, we used the IUCN Red List recommended cell size of 4 km^2. Estimates of continuing decline in extent of occurrence, area of occupancy, and quality of habitat were inferred by predicted future decline (PFD), which is a calculation of the percent AOO outside of protected areas (Callmander et al. 2007).

Key to species of *Commiphora* in the 'Rhynchocarpa' clade with stellate pubescence

1. Leaflets markedly falcate, sometimes with finely dentate margin. Stellate pubescence very sparse on leaflet blade and petiole, but always present at least on younger growth ..... *C. falcata*
1. Leaflets not markedly falcate, margin entire. Stellate pubescence persistent, and often dense, at least on the young growth. 2

    2. Fruit sub-globose or obovate, leaflets concolorous 3

        3. Pubescence stellate only, dense on young growth, becoming sparse at maturity. Bark gray or gray-brown, flaking in small, thin peels to reveal green underbark. ..... *C. aprevalii*

        3. Pubescence polymorphic with glandular, unstalked trichomes interspersed with stellate trichomes, bark brown/dark brown, exfoliating in plaques, not flaking in peels 4

            4. Bark quercicorticate, shaggy, and exfoliating in rough, thick, dark-brown plaques. Fruit 14–14.5 mm. long. Leaflets 7 .....*C. razakamalalae*

            4. Bark not quericorticate, exfoliating in papery sheets. Fruit large, 18–26 mm. long. Leaflets (5) 11–17 ..... *C. morondavensis*

    2. Fruit beaked, lacriform, leaflets discolorous 5

            5. Pubescence polymorphic with both glandular and stellate trichomes. Stellate trichomes born on a stalk, 6.5 – 110 µm long. Petiolules long, 4–7 mm. .....*C. andranovoryensis*

5. Pubescence not polymorphic (stellate only). Stellate trichomes not stalked. Petiolules short, 1.75–4 mm..... *C. elliptica*

Taxonomic treatment

1. ***Commiphora andranovoryensis*** Phillipson, Raharim., & Gostel, sp. nov.
TYPE: MADAGASCAR. Toliara: RN7, between Tulear–Sakaraha near PK 60 W of Andranovory, [23°08'S, 44°08'E] Dec 1961, *Service Forestier 29099* (holotype: TEF!, isotypes: P!).

Tree 3–8 m. tall. Growing on calcareous, limestone outcrops. Leaves imparpinnately compound with (3)–5–7–(9) leaflets, 85–215 mm. in length. Petiole 25–60 mm. long and 1–1.5 mm wide. Both petiole and leaflets covered with a dense, ferrugineous-brown, stalked stellate pubescense with sessile glandular trichomes more sparse, interspersed throughout. Leaflets subcordate, ovate, or narrowly ovate 40–85 mm in length with apiculate apex and aequilateral, rounded, or inaequilateral base. Terminal petiolules 8.5–30 mm. in length and 0.5–0.8 mm. in width. Lateral petiolules 4–7.5 mm. long and 0.5–1 mm wide. Female inflorescence and flowers unknown; male inflorescence is a paniculate cyme, 160–190 mm. long with three inflorescence orders, entirely covered in a dense stellate pubescence with glandular trichomes interspersed throughout. Stellate pubescence is stalked, with 5–15 arms (Fig. 13B). Arm length on stellate pubescence 40–260 µm. Each arm is flattened with a central line of compression/depression; margins of

arms are slightly revolute and arms arms spiral or corkscrew along their path. Peduncle is 65–85 mm. long and 0.7–0.85 mm. wide. Second inflorescence order is 1–11 mm. long, 0.4–0.5 mm. wide, and pedicels (articulated) 2–2.5 mm. long, 0.4–2 mm. wide. Lanceolate bracts are caducous at maturity, 1.4 mm. long and 0.4 mm wide. Perianth and male parts of flowers are 4-merous. Ovary is two-locular, one is abortive. The calyx is cup-like during anthesis, approximately 2.5 mm. in length and 2 mm. in width. Sepals are triangular; 1.1 mm. long and 0.75 mm. wide. Petals spatulate, 2 mm. long and 0.8 mm. wide. 8 stamen arranged diplostemonously. Antesepalous stamen with filament and anther 1 mm. long and 0.75 mm. in length, respectively. Antepetalous stamen with filament and anther 0.6 mm. and 0.75 mm. in length, respectively. Fruits are drupaceous, born on a paniculate cyme 43–190 mm. long. Secondary infructescence order 30–51 mm. long and 0.5–1.2 mm. wide, with pedicels 2.4–4.5 mm. long and 0.9–2 mm. wide. Fruits with persistent stamens are lacriform, covered in a densely red-brown stellate/glandular pubescence; 13.5–17 mm. long and 8.3–10 mm. wide. Pericarp splits at maturity into two valves, pseudaril unknown.

Leaf architecture—Leaflets with 6–8 major secondary veins, regularly spaced and markedly visibile on both ab- and adaxial surfaces. Major secondary vein angle proximally decurrent, becoming distally perpendicular. 2–(3) parallel intersecondaries of mixed length. Epimedial and intercostals tertiary fabric mixed. Exterior tertiaries and marginal ultimate veins (MUVs) looped. Areolation moderate with irregularly branched FEVs.

Illustration—Fig. 14

Phenology—Flowering November; fruiting December–January.

Distribution and Habitat—Elevation 0–50 m. *Commiphora andranovoryensis* appears to be a very rare species, known only from three collections from limestone outcrops in southwest Madagascar, extending from east of Toliara and west of Sakaraha village, north to Antsalova on karst in Tsingy de Bemaraha National Park. Distribution map provided in Fig. 15.

Conservation assessment—*Commiphora andranovoryensis* is endemic to Madagascar and is only known from three collections. The first two collections are from localities west of Andranovory village (including the type). This species has not been collected from Andranovory since and forest habitats in this area have been severely degraded in the time since these collections were made. It is very likely that this species is extirpated from Andranovory and only remains in small relict patches of habitat perhaps between Andranovory and Tsingy de Bemaraha. Nonetheless, we have included these two likely extirpated collections in our calculation of EOO (1,379 km^2) and AOO (12 km^2), which both lead us to designate this species as endangered. The calculated PFD for *C. andranovoryensis* is 66.6, which is quite high. The only known surviving populations remain in Tsingy Bemaraha National Park.

Etymology—We first identified *Commiphora andranovoryensis* from material collected by the Service Forestier near RN7 between the village of Andranovory and Tulear in southwest Madagascar.

Discussion— The combination of lacriform fruits on paniculate cymes, with the entire infructescence covered with a dense, ferrugineous stellate pubescence is shared only with

another species known from northwest Madagascar (*C. elliptica*). This species is distinct from the latter in the number and shape of leaflets, size of fruits, length of pedicel, and the number of inflorescence branching orders (always three versus two or sometimes three).

Representative Specimens Examined—MADAGASCAR. Mahajanga: Tsingy de Bemaraha, north of the Manambolo river, 29 Nov 1996, *Jongkind 3281* (P), Toliara: Along RN7 between Toliara and Sakaraha, near PK 55–65 west of Andranovory, [23°10'S, 44°04'E], Dec 1961, *Service Forestier 20,718* (TEF).

2. ***Commiphora aprevalii*** Guillaumin Bull. Soc. Bot. France 56: 1909.

TYPE: MADAGASCAR. Toliara: Morondava (Mouroundava), xxxxxx, *Grevé, 83*. (Holotype: P [P00048469]; Isotype: S 08-12595).

Synonym: *Balsamea aprevalii* Baill. Hist. Phys. Madagascar pl. 226Dbis. 1893. *nomen nudum*.

Tree 2–12 m. tall. Growing on diverse substrates including limestone karst outcrops, sand, or highly degraded remnant forest patches. Often planted in villages along roadsides as natural fencing. Bark gray-brown or light gray, peeling copiously in sheets or flaking over green underbark. Resin white–opaque white. Leaves imparipinnately compound with (3)–5–11–(13) leaflets, 55–200 mm. in length. Petiole 25–55 mm. long and 0.5–1.5 mm wide. Leaflets covered with a stellate (often sparse) stellate pubescence, sometimes interspersed with more sparse glandular indumentum. Petiole also covered

(usually more densely) in the same indumentum as the leaflets. Pubescence is more dense on juvenile and young leaves. Leaflets (narrowly) lanceolate–narrowly elliptical or round–ovate 10–90 mm in length with apiculate or subacuminate apex and aequilateral, rounded, or subinaequilateral base. Terminal petiolules 4.5–15 mm. in length and 0.4–1 mm. in width. Lateral petiolules 2–5 mm. long and 0.5–1 mm wide. Female inflorescence is a paniculate cyme 30–45 mm. long with two (rarely three) inflorescence orders, entirely covered in a somewhat dense, stellate pubescence with glandular indumentum interspersed throughout. Peduncle is 25–40 mm. long and 0.5–0.7 mm. wide. When present, the secondary inflorescence order is short, 5–8 mm. long. Pedicels (articulated) 1.5–1.8 mm. long and 0.25–0.35 mm. wide. Flower and inflorescence bracts appear to be absent. Perianth and female parts of flowers are 4-merous. Ovary is two locular, one is abortive. The calyx in female flowers is cup-like during anthesis, ca. 1.6 mm. long and 1.5 mm. wide. Sepals on female flowers are triangular, 1.4 mm. long and 0.9 mm. wide. 8 stamen arranged diplostemonously. Antesepalous stamen with anthers 1 mm. and filament 2 mm. in length, respectively. Antepetalous stamen with anthers 0.7 mm. and filament 0.8 mm. in length, respectively. Male inflorescence is a paniculate cyme, 40–260 mm. long with two or three inflorescence orders, entirely covered in a dense stellate pubescence with glandular trichomes interspersed throughout. Stellate pubescence not stalked with 10–12 arms (Fig. 13A). Arm length between 50–175 µm. Each arm is flattened laterally with occasional central compressed/depression. Arms are not revolute at margins and spiral or corkscrew outward distally. Peduncle is 27–45 mm. long and 0.5–0.8 mm. wide. Secondary inflorescence axis (if present) 0.1–2 mm. long and 0.1–0. 5

mm. wide. Pedicels (articulated) subsessile–1.75 mm. long and up to 0.25 mm. wide. Floral bracts absent; triangular inflorescence bracts are caducous at maturity, 0.2–0.8 mm. long and 0.2–0.6 mm. wide. Perianth and male parts of flowers are 4-merous. The calyx is cup-like during anthesis, approximately 2.75–2.8 mm. in length and 2 mm. in width. Sepals are triangular; 2–2.3 mm. long and 0.6–0.9 mm. wide. Petals spathulate or narrowly triangular and sharply recurved at apex, 1.6–1.8 mm. long and 0.6 mm. wide. 8 stamen are arranged diplostemonously. Antesepalous stamen with filament and anther 0.3–1 mm. long and 0.25–1 mm. in length, respectively. Antepetalous stamen with filament and anther 0.2–0.35 mm. and 0.25–0.75 mm. in length, respectively. Fruits are drupaceous, born on a paniculate cyme 30–125 mm. long. Fruiting peduncles and pedicels 17–85 mm. and 2–4 mm. long and 0.5–1.5 and 0.75–2 mm. wide, respectively. Fruits with persistent stamens are sub-globose or obovate, covered in a sparse stellate pubescence with even more sparse glandular indumentum interspersed; 9.5–13.5 mm. long and 7.5–11 mm. wide. Pericarp splits at maturity into two valves. The pseudaril is cup-like, covering lower 1/3 of the putamen and slightly lobed (to 1.5 mm.) on the lateral and sutural faces of putamen.

***Leaf architecture*—**Leaflets with (8)10–16 major secondary veins, regularly spaced. Major secondary vein angle occasionally proximally decurrent, becoming excurrent distally. 1–(2) parallel intersecondaries, ≥ 50% of major secondary length. Epimedial tertiary vein fabric mixed/irregular and intercostal tertiary veins opposite percurrent. Exterior tertiaries and MUVs looped. FEVs irregularly branched (mostly dichotomous) with moderate areolation.

Vernacular name—Daro, Daro fotsy

Phenology—Flowering November–January (very rarely May), fruiting November–June.

Distribution and Habitat—Elevation 45–715 m. A very common and widespread species throughout southern Madagascar, often growing in disturbed patches of forest along roadsides and also common in villages, where cuttings are planted for live fencing (Fig. 16). This species is known from several protected areas, including Andohahela National Park, Beza Mahafaly Special Reserve, and Kirindy Mitea National Park.

Conservation assessment—*C. aprevalii* is a very common species with a broad range throughout western and southern Madagascar. This species is designated as "Endangered", with an AOO = 140 km^2; however, the calculated EOO for this species 288,418 km^2, which suggests the species is "least concern". Our calculation of PFD for this species is high, however (52.6), which indicates that much of the occurrence for this species is outside of protected areas. *C. aprevalii* is commonly cultivated and grown as live, natural fencing in villages, which likely has inflated our EOO calculation. We designate this species as "Vulnerable".

Discussion—This is a geographically widespread species, readily distinguishable by its (sub-) lanceolate and subcoriaceous leaflets, sparse stellate indument, and comparatively long flowering and fruiting inflorescences. Fruits are also quite distinctive by their large size and shape (sub-globose or obovate). Morphologically, this species is most similar to *C. falcata*, which shares a similar number of leaflets, distribution of indument, and inflorescence length, but has distinctively falcate leaflets that are often toothed along margins.

Representative Specimens Examined—MADAGASCAR. Toliara: *Bosser 13842* (P). Toliara: along route 10 between Egeda and Ampanihy; between PK 200 and 250, scrub forest, 24°31'S 044°36'E, 200–250 m., 15 Feb 1975, *Croat 31,351* (MO). Toliara: road from Betioky to Tongobory, 15 km north of Betioky on calcareous plateau, 23°35'S 044°19'E, 100 m., 21 Mar 1985, *Dorr 4,108* (MO). Toliara: Andohahela National Park, 24°45'S 046°45'E, 200–300 m., 10–20 May 1994, *Eboroke 809* (MO). Toliara: growing alongside RN 7 in Ihosy village, 22°24'05"S 46°07'54"E, 637 m., 28 May 2013, *Gostel 105* (GMUF). Toliara: Along RN10, south of Andranovory, 22°19’11.2”S, 44°17’45.9”E, 375 m., 1 June 2013, *Gostel 130* (GMUF). Toliara: In Sakaraha village, growing in gardens at the ANGAP office, 22°54'32.4"S 44°31'13.4"E, 470 m., 1 Jun 2013, *Gostel 133* (GMUF). Toliara: growing alongside RN 7 near the junction for road leading to Sakalalina village, 22°17'44.87"S 46°17'23.2"E, 715 m., 3 Jun 2013, *Gostel 147* (GMUF). Toliara: Mandrare Basin, between 700–1,200 m., 20–22 Nov 1928, *Humbert 672I* (P), Toliara: Anadabolava, [24°12’S 46°19’E], 200–250 m., Dec 1933, *Humbert 12545* (P), Toliara: Manambolo valley, Isomono, [24°31’S 46°37’E], 400–900 m., Dec 1933, *Humbert 12937* (P), Toliara: Malio basin (affluent of the Mangoky) near Ambalabe in Ambatosola village, [21°58’S 45°15’E], 400–450 m., 23–27 Nov 1946, *Humbert 19404* (P), Toliara: Fiherenana, between Beantsy and Anjamala, [23°11’S 43°57’E], 30–300 m., 16–19 Jan 1947, *Humbert 19905bis* (P). Toliara: Fiherenana between Beantsy and Anjamala, [23°12'S 043°56'E], 30–300 m., 16 Jan 1947, *Humbert 19,888* (P). Toliara: plateau south of Fiherenana between Andranohinaly and Andranovory, [23°12'S 044°03'E], 300–400 m., 3 Feb 1947, *Humbert 20,117* (P). Toliara: Onilahy valley, near

Tongobory, [23°34'S 044°20'E], 80–200 m., 6 Feb 1947, *Humbert 20,193* (P). Toliara: Antanimoro (Androy) 30–35 km to the north, [24°49'S 045°40'E], 200–500 m., 6–9 Feb 1955, *Humbert 28,862* (P). Toliara: Fort Dauphin, west of town on northwest boundary of Reserve Integral #1, Andohahela, ca. 17 km north of Amboasary Sud, 24°55'S 046°25'E, 80–120 m., 7 Jan 1990, *McPherson 14,899* (MO). Toliara: 20 km southeast of Toliara on La Table, 23°23'S 043°45'E, 100 m., 5 Feb 1998, *McPherson 17,441* (MO). Toliara: along RN 7, 25 km southeast of Tulear, 23°20'S 043°51'E, 60 m., 25 Mar 1991, *Miller 6160* (MO). Toliara: Beahitsy, 21 Nov 1959, *Peltier 1427* (P). Toliara: Beza Mahafaly Reserve near Betioky. Parcelle 2, 23°40'S 044°35'E, 160 m., 17 May 1987, *Phillipson 1814* (MO). Toliara: Berenty Reserve, northwest of Amboasary, 25°00'S 046°16'E, 50 m., 22 Dec 1987, *Phillipson 2708* (MO). Toliara: RN 7, 27 km east of Tulear, 23°21'S 043°51'E, 100 m., 30 Dec 1987, *Phillipson 2765* (MO). Toliara: Andohahela Reserve, Parcelle 2, southeast of Hazofotsy, 24°51'S 046°33'E, 200 m., 21 Dec 1988, *Phillipson 2953* (MO). Toliara: RNI Andohahela #11, Parcelle 2, 24°49'49"S 046°32'15"E, 30–50 m., Mar 1994, *Rakotomalaza 166* (MO). Toliara: RNI Andohahela, Parcelle 2, south of Ambatoambo peak, 24°49'49"S 046°32'15"E, 30–50 m., 17 Mar 1994, *Randriambololona 97* (MO). Toliara: Miary in Bemia village, ca. 20 km northeast of Tulear, 23°18'S 043°48'E, 80–145 m., 26 Apr 1998, *Randrianaivo 206* (MO). Toliara: Berenty Private Reserve, ca. 9 km northwest of Amboasary Sud, near Mandrare river, 25°00'09"S 046°18'02"E, 30 m., 21 Jan 2006, *Rogers 919* (MO). Toliara: Bekily village, [24°13'S 045°19'E], 8 Feb 1954, *Service Forestier 8,409* (TEF). Toliara: Maintirano, [18°04'S 044°01'E], 20 Mar 1954, *Service Forestier 10,019* (TEF). Toliara: Menarahaka,

Ihosy, [22°32'S 046°29'E], 29 May 1954, *Service Forestier 10245* (TEF). Toliara: Antanimora, near route between Antanimora and Ambovombe, [24°50'30"S 045°48'E], 24–25 Jan 1955, *Service Forestier 11,727* (TEF), Toliara: Adramy forest, Canton Betanatanana, [17°58'S 044°47'E], 180 m., 15 Nov 1955, *Service Forestier 15,725* (TEF). Toliara: [23°20'S 043°51'E], *Service Forestier* (MO). Toliara: Antsangabe-Antoibe, Analalava, [15°04'S 047°14'E], 0–50 m., 29 Oct – 3 Nov 1958, *Service Forestier 18868* (TEF). Toliara: Mangoboka forest near Belaingo forest, Ankirihitra, [16°47'S 046°35'30"E], 15 May 1958, *Service Forestier 19354* (TEF). Toliara: [15°25'S 049°52'E], *Service Forestier 21677* (TEF), Toliara: Near Andranovory, south side of RN 7, 22°57.27'S, 44°20.66'E, *Weeks 10-I-20-04* (GMUF). Toliara: near Ankiliberengy, 23°19.82'S, 43°55.36'E, *Weeks 10-I-20-07* (GMUF). Toliara: Fiherenana river valley, on top of limestone plateau, 23°18.18'S, 43°44.87'E, *Weeks 10-I-24-03* (GMUF). Toliara: Ranobe, on unconsolidated sands, 23°1.48'S, 43°36.99'E, *Weeks 10-I-27-03* (GMUF). Toliara: Kirindy Research Station, east of Kirindy river, 20°4.36'S, 44°40.57'E, *Weeks 10-II-13-06* (GMUF). Toliara: Morondava, Kirindy Research Station, walking west on Conoco road from buildings, 20°4.36'S, 44°40.57'E, *Weeks 10-II-14-10* (GMUF).

3. ***Commiphora elliptica*** Phillipson, Raharim., & Gostel, sp. nov.

TYPE: MADAGASCAR. Antsiranana: Ankarana plateau, surrounding Ampandriampanihy cave, north of Mahamarina village, [12°57'S, 49°08'E], 17 Feb 1962, *Service Forestier 22,046* (TEF)

Large shrubs and trees 2–15 m. tall. Growing on diverse substrates including limestone karst outcrops, sand, or highly degraded remnant forest patches. Often planted in villages along roadsides as natural fencing. Bark brown, dark brown, or light gray-white, peeling in thin, papery sheets or flakes or occasionally in large, smooth plaques to reveal green underbark. Resin white–opaque white. Leaves imparpinnately compound with 5–9 leaflets, 110–240 mm. in length. Petiole 25–58 mm. long and 0.75–2.25 mm wide. Both petiole and leaflets covered with a dense, ferrugineous-brown, stellate pubescense. Stellate pubescence not stalked with 5–12 arms (Fig. 13C). Each arm between 105–320 μm in length. Arms laterally flattened or occasionally slightly compressed centrally. Arms spiraling distally. Leaflets elliptical 25–105 mm. in length with acuminate–subapiculate apex and subaequilateral or rounded base. Terminal petiolules 6–30 mm. in length and 0.25–0.8 mm. in width. Lateral petiolules 1.5–4 mm. long and 0.5–1.25 mm. wide. Female inflorescences are paniculate cymes 35–75 mm. with peduncle 15–25 mm. in length and 0.5–0.65 mm. in width. Female flowers subsessile with calyx 4 mm. long and 2 mm. wide. Sepals triangular, 3 mm. long, 1 mm. wide, petals lanceolate 2.7 mm. long, 0.7 mm. wide. Antesepalous anthers on female (bisexual) flowers 0.6 mm. long with filament 0.8 mm. long, antepetalous anthers unknown. Male inflorescence is a paniculate cyme, 70–115 mm. long with two inflorescence orders. Both female and male inflorescence entirely covered in a dense stellate pubescence. Peduncle is 6–95 mm. long and 0.5–2.75 mm. wide. Pedicels 1–4 mm. long, 1.15–3.75 mm. wide. Lanceolate or narrowly-triangular inflorescence bracts are caducous at maturity, 1–3 mm. long and 0.4–0.8 mm wide. Perianth and male parts of flowers are 4-merous. Ovary is two-locular, one

is abortive. The calyx in female flowers is obovate, but becoming elongated during anthesis. Calyx in male flowers 1.95–2.25 mm. in length and 1.75 mm. in width. Sepals on male flowers triangular (or narrowly-triangular); 2 mm. long and 1 mm. wide. Petals spathulate, 1–2.5 mm. long and 0.7–0.8 mm. wide. 8 stamen arranged diplostemonously on both female and male flowers. Antesepalous stamen on male flowers with filament and anther 0.7 mm. long and 0.6 mm. in length, respectively. Antepetalous stamen with filament and anther 0.4 mm. and 0.5 mm. in length, respectively. Fruits are drupaceous, born on a paniculate cyme 35–115 mm long. Fruits without persistent stamens are lacriform, covered in a densely red-brown stellate/glandular pubescence; 11.5–22 mm. long and 6.5–16 mm. wide. Pericarp splits at maturity into two valves, pseudaril is cup-like, covering lower 1/3 of putamen.

***Leaf architecture*—**Leaflets with 6–8 major secondary veins, regularly spaced and markedly conspicuous on both ab- and adaxial surfaces of blade. Major secondary vein angle proximally subdecurrent, becoming excurrent distally. 2–3 parallel intersecondary veins ≤ 50% of major secondary vein lengths. Epimedial tertiary vein fabric opposite percurrent. Intercostal tertiary vein fabric proximally alternate percurrent, becoming opposite percurrent distally. Exterior tertiary veins and MUVs looped. FEVs unbranched with good areolation.

Vernacular name—Matambelo, Matambelona, Mahafay.

Illustration—Fig. 17.

Phenology—Flowering December – January; fruiting January – August.

***Distribution and Habitat*—**Elevation 10 – 380 m. North west Madagascar from Ankarana north to Mahavango and east to Daraina (Fig. 18). Growing on variable substrate, often calcareous, but also in mixed siliceous, sandy dry forest.

Conservation assessment— This species is rather widespread in northern Madagascar, throughout the Antsiranana Province, with an EOO of 5,512 km^2 (vulnerable) and an AOO of 92 km^2 (endangered). Like *C. aprevalii*, *C. elliptica* is commonly cultivated as live, natural fencing in villages, which may inflate the calculated EOO and is reflected by a high PFD (56). Several subpopulations are known to exist throughout the Antsiranana province and this species is known from three protected areas, including Ankarana and Montagne d'Ambre National Parks.

Etymology—The specific epithet, *elliptica* has been given in reference to the elliptical shape of leaves.

Discussion—This species is most similar to *C. stellulata*, but differs in several important ways. *C. stellulata* was initially described on the basis of (5?) syntypes, which we feel did not represent a single species. Like *C. stellulata*, C. *elliptica* has discolorous leaflets, but the adaxial blade surface is much darker than the abaxial blade surface. The color of the pubescence is also very different between the two species, whereas *C. elliptica* has (almost always) a very dark red pubescence and *C. stellulata* is pale gray-white. Most importantly, the leaflet shape of *C. stellulata* is much more elongated than *C. elliptica*.

Representative Specimens Examined—MADAGASCAR. Antsiranana: Ankarana Reserve, 19 Jan 1991, *Bardot-Vaucoulon 389* (P). Antsiranana: Ankarongana (Karongana), Irodo, Analafondro forest, 12°37'46"S, 49°31'37"E, 68 m., 24 Feb 2006,

Birkinshaw 1576 (MO). Antsiranana: Ankarana Reserve, road towards Lac Vert, 12°50’47”S, 49°6’18”E, 82 m., 26 May 1999, *De Block 1,026* (P). Antsiranana: near Vohemar, rural community of Daraina, Bekaraoka forest, Andranotsimaty, 13°11.07’S, 49°42.35’E, 200 m., 16 Mar 2003, *Gautier 4,399* (G). Antsiranana: Vohemar, rural community of Daraina, Ambohitsitondroina forest, 13°07.9’S, 49°28.15’E, 130 m., 17 Mar 2004, *Gautier 4,619* (G). Antsiranana: Along roadside on RN6, planted as natural fencing near house in Anivorano Nord, 12°45’18”S, 49°14’15”E, 376 m., *Gostel 43* (GMUF). Antsiranana: near Windsor Castle in Baie de Courrier, 12°12’51”S, 49°10’03”E, 318 m., *Gostel 46* (GMUF). Antsiranana: near Windsor Castle in Baie de Courrier, 12°15’58”S, 49°11’44”E, 10 m., *Gostel 51* (GMUF). Antsiranana: Ankarana Special Reserve, 12°35’01.8”S, 49°26’56.3”E, 110 m., *Gostel 56* (GMUF). Antsiranana: along RN 6 roadside, south of Ankarana Special Reserve, 13°01’S, 49°08’E, 110 m., *Gostel 72* (GMUF). Antsiranana: Ramena, Andavakoera, 8 km. south of Andavakoera, 12°20’53”S, 49°21’27”E, 50 m., 11 Aug 2004, *Guittou 80* (MO). Antsiranana: Vohemar, Daraina, southern Bekaraoka, 13°10’17”S, 49°42’12”E, 149 m., 31 May 2005, *Hong-Wa 275* (MO). Antsiranana: Ankarana Reserve, 30–350 m., 24 Jan–29 Feb 1960, *Humbert 43,707* (P). Antsiranana: Ankarana Reserve near Campement des Anglais, 12°54’42”S, 49°06’43”E, 240–260 m., 22 May 1993, *Jongkind 976* (MO). Antsiranana: Ankarana Reserve, near Campement des Anglais, 12°54’S, 49°08’E, 150 m., 29 Jan 1994, *Leeuwenberg 14,346* (WAG). Antsiranana: Ramena, Sakalava Bay, 3.5 km. east of fokontany Ankorikihely, 12°16’26”S, 49°23’20”E, 10 m., 16 Aug 2004, *Leopold 8* (TAN). Antsiranana: Andranovondronina, Antsisikala, west of Ampasikely, 12°11’35”S,

49°12’45”E, 197 m., 13 May 2005, *Ramananjanahary 278* (MO). Antsiranana: Andramaimbo, Baie de Courrier, Mangrove, 12°12’33”S, 49°09’21”E, 28 m., 14 May 2005, *Ramananjanahary 302* (MO). Antsiranana: Andranovondronina, Anjiabe, north of Antsaravy, Bobaomby, 12°08’46”S, 49°19’49”E, 60 m., 16 Mar 2006, *Callmander 477* (MO). Antsiranana: Ramena, Nosy Longo, 12°18’10”S, 49°19’06”E, 50 m., 4 Apr 2006, *Callmander 502* (MO). Antsiranana: near Vohemar, rural community of Daraina, Bekaraoka forest, 340 m. from side point 96, 13°06.39’S, 49°42.65’E, 150 m. 10 Feb 2004, *Ranirison PR 396* (G). Antsiranana: Andrafiabe, between Ambolobozobe and Ambolobozokely, 12°29’39”S, 49°34’04”E, 10 m., 9 Feb 2005, *Ratovoson 936* (MO). Antsiranana: Andrafiabe, Ambolobozokely, Rigny Ray, Nosy Voanio, 12°26’11”S, 49°33’10”E, 10 m., 23 Mar 2006, *Ratovoson 1,114* (MO). Antsiranana: Ankarana Reserve, [12°49’S, 49°01’E], 11 Mar 1954, *Service Forestier 9379* (TEF). Antsiranana: Ankarana Reserve, northeast of Ambondromifehy, [12°53’30”S, 49°12’30”E], 13 Nov 1958, *Service Forestier 18,999* (TEF). Antsiranana: Oronjia Peninsula, on unconsolidated sands, 12°14.18’S, 49°21.29’E, 9 Jan 2010, *Weeks 10-I-09-01* (GMUF). Antsiranana: Oronjia Peninsula, on unconsolidated sands. 12°16.39’S, 49°23.31’E, 10 Jan 2010, *Weeks 10-I-10-04* (GMUF).

4. ***Commiphora falcata*** Phillipson, Raharim., & Gostel, sp. nov.

TYPE: MADAGASCAR. Toliara: Edge of calcareous plateau between La Table and Sarodrano, 8 Feb 1962, *Service Forestier 20,825* (holotype: TEF!, isotypes: MO!, P!).

Tree 5–8 m. tall. Growing on sand or calcareous, rocky outcrops. Bark gray-brown or gray, flaking to reveal green underbark. Resin opaque white. Leaves imparpinnately compound with 5–11 leaflets, 140–210 mm. in length. Petiole 40–60 mm. long and 0.5–1.25 mm wide. Petiole and leaflets covered in an extremely sparse (rarely absent) stellate and glandular pubescence. Stellate pubescence not stalked with 9 arms (Fig 13A). Each arm between 40–110 µm in length. Arms are strongly revolute, such that the edges fold into an elongated, spiraling, and compacted tube or spike. Leaflets distinctly falcate, 35–85 mm in length with apiculate apex and inaequilateral base. Terminal petiolules 4.5–25 mm. in length and 0.25–0.6 mm. in width. Lateral petiolules 2.75–10 mm. long and 0.25–0.75 mm wide. Female inflorescence is unknown. Male inflorescence is a paniculate cyme, 95–200 mm. long with two or three inflorescence orders, covered in a very sparse (rarely absent) stellate and glandular pubescence. Peduncle is 25–85 mm. long and 0.5–1 mm. wide. Secondary inflorescence axis (when present) 1.5–11.5 mm. long and 0.25–1.25 mm. wide. Pedicels (articulated) 2.5–6.5 mm. long and 0.5–1 mm. wide. Floral bracts absent; triangular inflorescence bracts are mostly caducous at maturity, 0.45 mm. long and 0.25–0.3 mm. wide. Perianth and male parts of flowers are 4-merous. The calyx is cup-like during anthesis, approximately 2–2.5 mm. long and 0.75–1 mm. wide. Sepals are narrowly triangular; 1.4–1.5 mm. long and 0.3–0.4 mm. wide. Petals spathulate and sharply recurved at apex, 2.75–3 mm. long and 0.9–1 mm. wide. 8 stamen are arranged diplostemonously. Antesepalous stamen with filament and anther 1.25–1.5 mm. and 0.7–0.8 mm. long, respectively. Antepetalous stamen with filament and anther 0.5–0.6 mm. and 0.6–0.7 mm. in length, respectively. Fruits are drupaceous, born on a paniculate

cyme 70–115 mm. long. Fruiting peduncles and pedicels 30–60 mm. and 2.5–6.5 mm. long and 0.5–0.6 and 0.7–1 mm. wide, respectively. Fruits with long (to 1.5 mm.) persistent stamens are sub-lacriform, sub-obovate, or sub-globose, covered (usually) in a very sparse stellate and glandular pubescence; 9–11.5 mm. long and 4.5–7.5 mm. wide. Pericarp splits at maturity into two valves. The pseudaril is cup-like, covering lower 1/3 of the putamen and very slightly lobed (to 1 mm.) on the lateral and sutural faces of putamen.

***Leaf architecture*—**Leaflets with 13–24 irregularly spaced major secondary veins. Major secondary vein angle proximally subdecurrent, becoming excurrent and perpendicular distally. One parallel intersecondary, > 50% length of major secondaries. No epimedial tertiary veins. Intercostal tertiary vein fabric alternate percurrent. Exterior tertiary veins looped with mixed MUVs. FEVs dichotomously branching with moderate areolation. Some (but not all) leaflets with numerous (30–45), small teeth, retroflexed convex with no apical features.

Vernacular name—Unknown

Phenology—Flowering in January, Fruiting in December–February.

***Distribution and Habitat*—**Found in elevations between 0–100 m. Rare trees growing in southern and southwest Madagascar in the province of Toliara on calcareous, rocky, or sandy substrates (Fig. 19).

Conservation assessment—The calculated EOO and AOO values for this species are 4,089 km^2 (endangered) and 12 km^2 (endangered), respectively. This species is quite rare and is only known from seven collections (four with georeferenced coordinates). Due to

the rarity of collections, low number of subpopulations, and EOO and AOO values, we designate this species as endangered.

***Etymology*—**The specific epithet is derived from Capuron's original description of the species, which was not validly published, but we retain the name because of its apt description of the most distinctive structure belonging to this species.

Discussion—Capuron's original description of the species (Capuron 1965) was not validly published since he designated more than one type specimen, which is a violation of ICBN Art. 40. It is validly published here for the first time, and although our circumscription of the species essentially follows that of Capuron (1965), the broader specimen base available to us now has enabled us to provide some additional information.

Representative Specimens Examined—MADAGASCAR. Toliara: Road to Sarodrano, Feb 1962, *Keraudren 1382* (P). Toliara: Antanimora route to Angavo, [24°51'S 45°45'E], 200–250 m., 6 Dec 1961, *Service Forestier 20424* (TEF). Toliara: north of Amboasary, on the route to Behara, [25°02'30"S 46°23'E], 50–100 m., 7 Dec 1961, *Service Forestier 20449* (TEF). Toliara: at the base of plateau between La Table and Sarodrano, Jan 1962, *Service Forestier 20816* (TEF). Toliara: Ranobe, on unconsolidated sands, 23°1.48'S, 43°36.99'E, 106 m., 26 Jan 2010, *Weeks 10-I-26-03* (GMUF). Toliara: Ranobe, on unconsolidated sands, 23°1.48'S, 43°36.99'E, 106 m., 27 Jan 2010, *Weeks 10-I-27-04* (GMUF).

5. ***Commiphora morondavensis*** Phillipson, Raharim., & Gostel, sp. nov.

TYPE: MADAGASCAR. Morondava, Andranomena forest, between Andranomena and Marofandilia, [20°07’ S 44°25’ E – 20°12’ S 44°35’ E], 0–50 m., 19 Jan 1962, *Service Forestier 20,887* (holotype: TEF!, isotypes: P!).

Tree to 20 m. Bark ferrugineous, exfoliating in papery sheets. Resin transparent or more often opaque-white. Leaves imparpinnately compound with (5)–11–17 leaflets, (35)–70–400 mm. in length. Petiole (7)–25–120 mm. long and 1–3 mm wide. Both petiole and leaflets covered with a variously dense stellate pubescense with sessile glandular trichomes interspersed throughout; pubescense very dense on young growth. Stellate pubescence not stalked with 6–13 arms (Fig. 13D). Each arm between (55) 140–750 µm in length. Each arm is exceptionally elongated (rarely stunted) strongly revolute such that the sides fold into an elongated, spiraling, and compacted tube or spike. Leaflets narrowly elliptical, elliptical, or rarely widely elliptical 30–105 mm. in length with acuminate (rarely apiculate) apex and rounded (rarely oblique) base. Terminal petiolules 3–40 mm. in length and 0.25–1.25 mm. in width. Lateral petiolules 0.25–4 mm. long and 0.5–1.75 mm wide. Functionally female inflorescences are paniculate cymes 80–225 mm. with peduncle 20–60 mm. in length and 1–1.1 mm. in width. Functionally female flowers subsessile (pedicel 0.1–0.2 mm. in length and width) with calyx 2–3 mm. long and 1–1.5 mm. wide. Sepals triangular, 2.5 mm. long, 1 mm. wide, petals spatulate, 3.2 mm. long, 1.1 mm. wide. Antesepalous filaments and anthers on female (functionally female) flowers 1.6 and 0.4 mm. long, respectively; antepetalous filaments and anthers 0.8 and

0.3 mm., respectively. Male inflorescence is a paniculate cyme, 65–210 mm. long with two inflorescence orders, entirely covered in a dense stellate pubescence with glandular trichomes interspersed throughout. Peduncle is 18–60 mm. long and 0.5–1.75 mm. wide. Pedicels 1.75–6.5 mm. long, 1–2 mm. wide. Lanceolate–triangular inflorescence bracts are caducous at maturity, 1.2–4 mm. long and 0.75–1.5 mm wide. Perianth and male parts of flowers are 4-merous. Ovary is two-locular, one locule is abortive. The calyx is globose-obovate in bud and cup-like during anthesis, approximately 1.5–3.5 mm. in length and 1–3 mm. in width. Sepals are triangular; 2–3.75 mm. long and 0.9–2.2mm. wide. Petals spathulate, 2–3 mm. long and 0.75–1.1 mm. wide. 8 stamen arranged diplostemonously. Antesepalous stamen with filament and anther 1.8–2 mm. long and 0.5–0.9 mm. in length, respectively. Antepetalous stamen with filament and anther 1.1 mm. and 0.4–0.9 mm. in length, respectively. Fruits are drupaceous, born on a paniculate cyme 70–105 mm. long. Fruits without persistent stamens are lacriform-subglobose, covered in a dense stellate/glandular pubescence when young (more sparse at maturity); 18–26 mm. long and 13–20 mm. wide. Pericarp splits at maturity into two valves, pseudaril is cup-like, with two wide ‘arms’ on the lateral, non-suture side, covering the lower third of the putamen.

***Leaf architecture*—**Leaflets with 5–8 (9–10–12) regularly spaced major secondary veins. Major secondary vein angle proximally decurrent, becoming excurrent distally. 1 subparallel intersecondary < 50% length of major secondary veins. Epimedial tertiary vein fabric mixed. Intercostal tertiary fabric mixed percurrent proximally, becoming

opposite or convex percurrent distally. Exterior tertiary veins and MUVs looped. FEVs irregularly branched with moderate areolation.

Vernacular name—Arafy mena, Daro mavo, Daromena,

Illustration—Fig. 20.

Phenology—Flowering: October–December, Fruiting: December–March.

Distribution and Habitat— Found in elevations between 0–450 m. A relatively common species in western and southwestern Madagascar (Fig. 21). Its narrow geographic distribution is recognized in the specific epithet, *morondavensis*.

Conservation assessment—This species is known from 16 collections in western Madagascar, including three protected areas (Kirindy Mitea National Park, Reniala Private Reserve, and Tsingy de Bemaraha National Park). The calculated EOO and AOO for *C. morondavensis* are 23,431 km^2 (not threatened) and 36 km^2 (endangered), respectively. A large proportion of subpopulations exist outside of protected areas (PFD = 40) and we therefore designate this species as vulnerable.

Etymology—The specific epithet, *morondavensis*, has been attributed because the range of this species is restricted to habitat nearby the city of Morondava in western Madagascar, although the range extends from ca. 50 km. north of Morondava to Morombe, further south.

Discussion—This species is distinctive as having a large number of leaflets, stellate pubescence, and sub-globose fruits. Prior to our description, specimens of *C. morondavensis* have been collected and initially recognized as *C. arafy*, *C. guillauminii*, or *C. stellulata*. In each of these characters, the specimens differ from their initial

determination substantially for several reasons. Unlike *C. arafy* and *C. guillauminii*, *C. morondavensis* has a stellate pubescence. Unlike *C. stellulata* (which is clarified in this revision, below), *C. morondavensis* does not have discolorous leaves and also has many more leaflets per leaf than *C. stellulata*.

Representative Specimens Examined—MADAGASCAR**.** Toliara, 50 km. north of Morondava, Marosalaza forest, 0 m., 9 May 1974, *Abraham 180A* (P). Toliara, Morondava, Malio basin (affluent of the Mangoky river), near Ambalabe, 400–450 m., 23–27 Nov 1946, *Humbert 19,356* (P!). Toliara, Faritra Atsimo Andrefana, Morombe, Basibasy, Andohasakoa, 21°58'35"S 43°37'29"E, 64 m., 6 Mar 2008, *Manjakahery 351* (MO). Toliara, Forest south of Berenty (Ankazoabo), Feb 1967, *Morat 2,532* (TAN). Toliara, North of Toliara, near Mangoky river, 21°45'S, 43°49'E, 50 m., 1 Jan 1989, *Phillipson* 3070 (MO). Toliara, Ankililoka, Ankatepoka, Betaimboraky, Anjahampolo, 22°44'13.3"S, 43°31'19.4"E, 120 m., 8 Nov 1998, *Rakotomalaza 1,803* (G). Toliara, Mangalakandoha, Sakaraha, [22°56'S, 44°53'E], 17 Nov 1951, *Service Forestier 4,120* (TEF). Toliara, Morombe, [22°16'S, 43°39'E], 13 Mar 1955, *Service Forestier 13,310* (TEF). Toliara: Andranomena forest, between Andranomena and Marofandilia, [20°07'S, 44°25'E], 19 Jan 1962, *Service Forestier 20,887* (TEF). Toliara, Morondava, Belo Andranomena, [20°10'30"S, 44°25'30"E], 10 m., 18 Oct 1962, *Service Forestier 21,079* (TEF). Toliara, Morondava, Andranomena-Marofandilia forest, 28–29 Nov 1969, *Service Forestier 28,909* (TEF). Toliara, between Tanandava and Morombe, along the bank of the Mangoky river, 1 Dec 1969, *Service Forestier 28,956* (TEF). Toliara, between Tanandava and Morombe, along the bank of the Mangoky river, 1 Dec 1969, *Service*

Forestier 28,957 (TEF). Toliara: Morondava, ca. 23 km. north of Kirindy Research Station on the road to Belo Tsiribina, 19°51’54”S, 44°36’47”E, 14 Feb 2010, *Weeks 10-II-14-05* (GMUF). Toliara: Morondava, ca. 23 jm. North of Kirindy Research Station on the road to Belo Tsiribina, 19°51’54”S, 44°36’47”E, 14 Feb 2010, *Weeks 10-II-14-06* (GMUF). Toliara: Morondava, Kirindy Research Station, walking west on Conoco road from buildings, 20°4.36’S, 44°40.56’E, 14 Feb 2010, *Weeks 10-II-14-11* (GMUF).

6. ***Commiphora razakamalalae*** Gostel, Phillipson & A. Weeks

TYPE: MADAGASCAR. Toliara: Atsimo-Andrefana Region, Sakaraha, Andranolava, Ampondrabe II, Lika, western limit of Zombitsy Special Reserve, 22°46’21”S, 44°40’25”E, 540 m., 8 Apr 2006, *Andriamihajarivo 885* (P).

Tall trees, 9–14 m. tall. Growing on sandy, siliceous substrates and apparently highly locally endemic. Bark distinctive, dark brown, peeling in large, thick plaques. Resin white–opaque white. Leaves imparpinnately compound with 5–9 leaflets, 115–250 mm. in length. Petiole 20–65 mm. long and 0.75–1.5 mm wide. Both petiole and leaflets covered with a stellate pubescense with sessile glandular trichomes, interspersed throughout. Stellate pubescence not stalked with 6–13 arms. Each arm between 50–340 µm in length. Each arm is strongly revolute (sensu aprevalii type, Fig. 13A), such that the sides fold into an elongated, spiraling, and compacted tube or spike. Leaflets discolorous, elliptical or sometimes obovate–round, 30–115 mm. long and 15–40 mm. wide with apiculate or acuminate apex and oblique base. Terminal petiolules 10–22 mm.

long and 0.5–0.75 mm. wide. Lateral petiolules 2.25–5.5 mm. long and 0.5–1 mm wide. Female inflorescence and flowers unknown; male inflorescence is a paniculate cyme, 70–95 mm. long with two inflorescence orders, entirely covered in a mostly sessile, glandular pubescence interspersed throughout. Peduncle is 10–35 mm. long and 1.4 mm. wide. Pedicels 3.5 mm. long, 1.5 mm. wide. Narrowly triangular bracts are caducous at maturity, 0.4 mm. long and 0.25 mm wide. Perianth and male parts of flowers are 4-merous. Ovary is two-locular, one is abortive. The calyx is obovate in bud and cup-like during anthesis, approximately 0.8 mm. in length and 0.5–1 mm. in width. Sepals are narrowly-triangular; 1.3 mm. long and 0.9 mm. wide. Petals narrowly-spatulate, 1.3 mm. long and 0.75 mm. wide. 8 stamen arranged diplostemonously. Antesepalous stamen with filament and anther 1.7 mm. long and 0.9 mm. in length, respectively. Antepetalous stamen with filament and anther 0.9 mm. and 0.8 mm. in length, respectively. Fruits are drupaceous, born on a paniculate cyme 15–85 mm. long. Fruiting peduncle is 10–35 mm long and 0.75–1.5 mm. wide. Fruiting pedicels 1.75–4.5 mm. long and 1.25–2.25 mm. wide. Fruits without persistent stamens are obovate–globose, covered in stellate and sessile glandular pubescence; 9–15 mm. long and 6.5–15 mm. wide. Pericarp splits at maturity into two valves, pseudaril with very shallow lobes on facial and sutural sides of putamen, covering lower quarter or fifth of putamen.

***Leaf architecture*—**Leaflets with 8–10 major secondary veins. Major secondaries proximally subdecurrent, becoming excurrent distally. Subparallel intersecondaries >50% length of major secondaries. Epimedial and intercostals tertiary vein fabric mixed.

Exterior tertiary veins and MUVs looped. FEVs irregularly branched with moderate areolation.

Vernacular name—Daro, Daro fotsy

Illustration—Fig. 22.

Phenology—Flowering June; fruiting February–April.

Distribution and Habitat—Found in elevations of 80–800 m. This species is known only from the type and four other collections and grows on rocky, calcareous, or sandy substrates. Four collections are from the Parc National de Zombitsy in Toliara in southwest Madagascar and one is from the Onilahy valley (Fig. 23).

Conservation assessment—*C. razakamalalae* has a calculated EOO of 7,353 km^2 (vulnerable) and AOO of 20 km^2 (endangered). Although two collections exist from the Onilahy river basin, near Toliara, this species has only been collected outside of Zombitsy National Park prior to 1967. Very few known subpopulations therefore exist and the PFD (20) suggests this species is threatened. We designate this species as endangered.

Etymology—The specific epithet *razakamalalae* is in reference to our friend and colleague, Richard Razakamalala, who was also the first identify this species in the field during a collecting trip in June 2013. Upon seeing the specimen (Gostel 140) it was immediately obvious that it was a previously undescribed species, due to the distinctively rough bark.

Discussion—This species is only represented by one other collection, from the same locality in 2006. The leaves of this species are strikingly discolorous and most closely

resemble those of *C. elliptica*. The bark of this species (Fig. 22) is extraordinarily unique for species of *Commiphora* in Madagascar. Further collection is recommended from Zombitsy and surrounding habitat to determine if the range of this species is more broad than suggested by existing collections.

Representative Specimens Examined—MADAGASCAR. Toliara: Parcel of Zombitsy Special Reserve north of RN7, 22°52'43.6"S, 44°41'35.3"E, 800 m., *Gostel, M. R. 140* (GMUF!). Toliara: Onilahy valley, Andranomay near Tongobory, [23°34'S 44°20'E], 80–200 m., 6 Feb 1947, *Humbert 20193* (P). Toliara: Zombitsy, [22°46'S 44°42'E], Feb 1967, *Morat 2478* (P). Fianarantsoa: Menarahaka, Ihosy, [22°32'S 46°29'E], 29 May 1954, *Service Forestier 10245* (TEF).

References

Bachman, S. J. Moat, A. W. Hill, J. de la Torre, and B. Scott. 2011. Supporting Red List threat assessments with GeoCAT : Geospatial conservation assessment tool. Pp. 117–126 in *e-Infrastructures for data publishing in biodiversity science*. Eds. V. Smith and L. Penev, *ZooKeys* 150.

Callmander, M. W., G. E. Schatz, P. P. Lowry II, M. O. Laivao, J. Raharimampionona, S. Andriambololonera, T. Raminosoa, and T. K. Consiglio. 2007. Identification of priority areas for plant conservation in Madagascar using Red List criteria : Rare and threatened Pandanaceae indicate sites in need of protection. *Oryx* 41: 168–176.

Capuron, R. 1962. Contributions a l'étude de la flore forestière de Madagascar. VI. Note sur les Burséracées. *Adansonia* 2: 268–283.

Daly, D. C., M. M. Harley, M-C. Martínez-Habibe, & A. Weeks. 2011. Burseraceae. Pp. 76–104 in *The families and genera of vascular plants* vol. 10, ed. K. Kubitzki. New York: Springer-Verlag.

Ellis, B., D. C. Daly, L. J. Hickey, K. R. Johnson, J. D. Mitchell, P. Wilf, and S. L. Wing. 2009. *Manual of Leaf Architecture*. New York: Comstock Publishing Associates.

Engler, A. 1912. Die verbreitung der afrikanischen Burseraceen im Verhaltnis zu ihrer systematischen Gliederung und die Einteilung der Gattung *Commiphora*. Botanische Jahrbücher für Systematik, Pflanzengeschichte, und Pflanzengeographie 48: 443–490.

Gillett, J. B. 1980. *Commiphora* (Burseraceae) in South America and its relationship to *Bursera*. *Kew Bulletin* 34: 569–587.

Gillett, J.B. 1991. Burseraceae. Pp. 1–95 in *Flora of tropical east Africa*, ed. R. M. Polhill. Rotterdam: A. A. Balkema.

Gostel, M. R., P. B. Phillipson, A. Weeks. *In Press*. Phylogenetic reconstruction of the myrrh genus, *Commiphora* (Burseraceae), reveals multiple radiations in Madagascar and clarifies infrageneric relationships. *Systematic Botany*.

Humbert, H. 1965. Description des types de vegetation. Pp. 46–78 in *Notice de la carte Madagascar* vol. 6, eds. H. Humbert and G. Cours-Darne. Pondichéry: Travaux de la Section Scientifique et Technique de l'Institut Français de Pondichéry, hors série.

IUCN. 2001. *2001 categories and criteria* (version 3.1). IUCN. Gland, Switzerland.

Perrier de la Bâthie, H. 1946. Burséracées Pp. 1–50 in *Flore de Madagascar et des Comores*, ed. H. Humbert. Tananarive, Madagascar : Imprimerie officielle.

Weeks, A., D. C. Daly, and B. B. Simpson. 2005. The phylogenetic history and biogeography of the frankincense and myrrh family (Burseraceae) based on nuclear and chloroplast sequence data. *Molecular Phylogenetics and Evolution* 35: 85–101.

Weeks, A., and B. B. Simpson. 2007. Molecular phylogenetic analysis of *Commiphora* (Burseraceae) yields insight on the evolution and historical biogeography of an "impossible" genus. *Molecular Phylogenetics and Evolution* 42: 62–79.

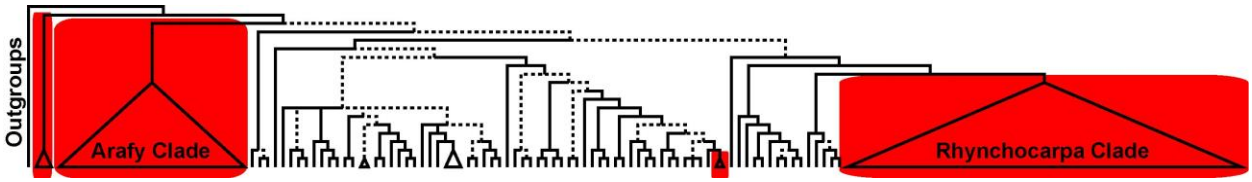


Figure 12. Reduced phylogeny of *Commiphora* from Gostel et al. (in mss.). Four Malagasy clades are highlighted in red.
Two are monotypic (*C. lasiodisca* and *C. simplicifolia*), while two others are species rich and named "Arafy clade" and "Rhynchocarpa", respectively.

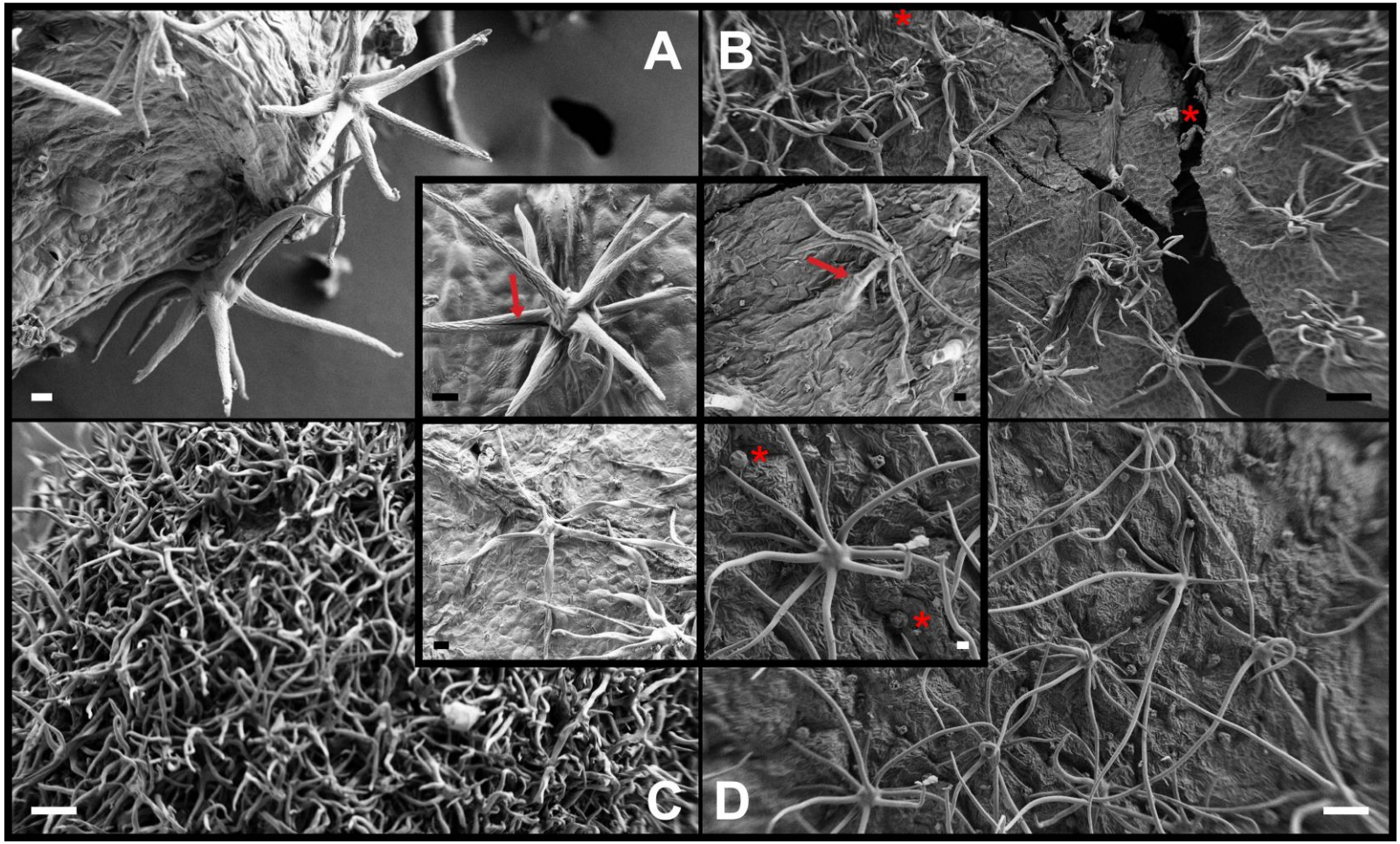


Figure 13. Scanning electron microscopy (SEM) of indument features.
A: *Commiphora aprevalii* type pubescence, scale bars (both) = 10 µm.; B: *C. andranovoryensis* type pubescence, scale bar = 100 µm., inset = 20 µm.; C: *C. elliptica* type pubescence, scale bar = 100 µm., inset = 20 µm.; and D: *C. morondavensis* type pubescence, scale bar = 100 µm., inset = 20 µm.

Figure 14. Illustration of *Commiphora andranovoryensis*. Drawn by Bobbi Angell.

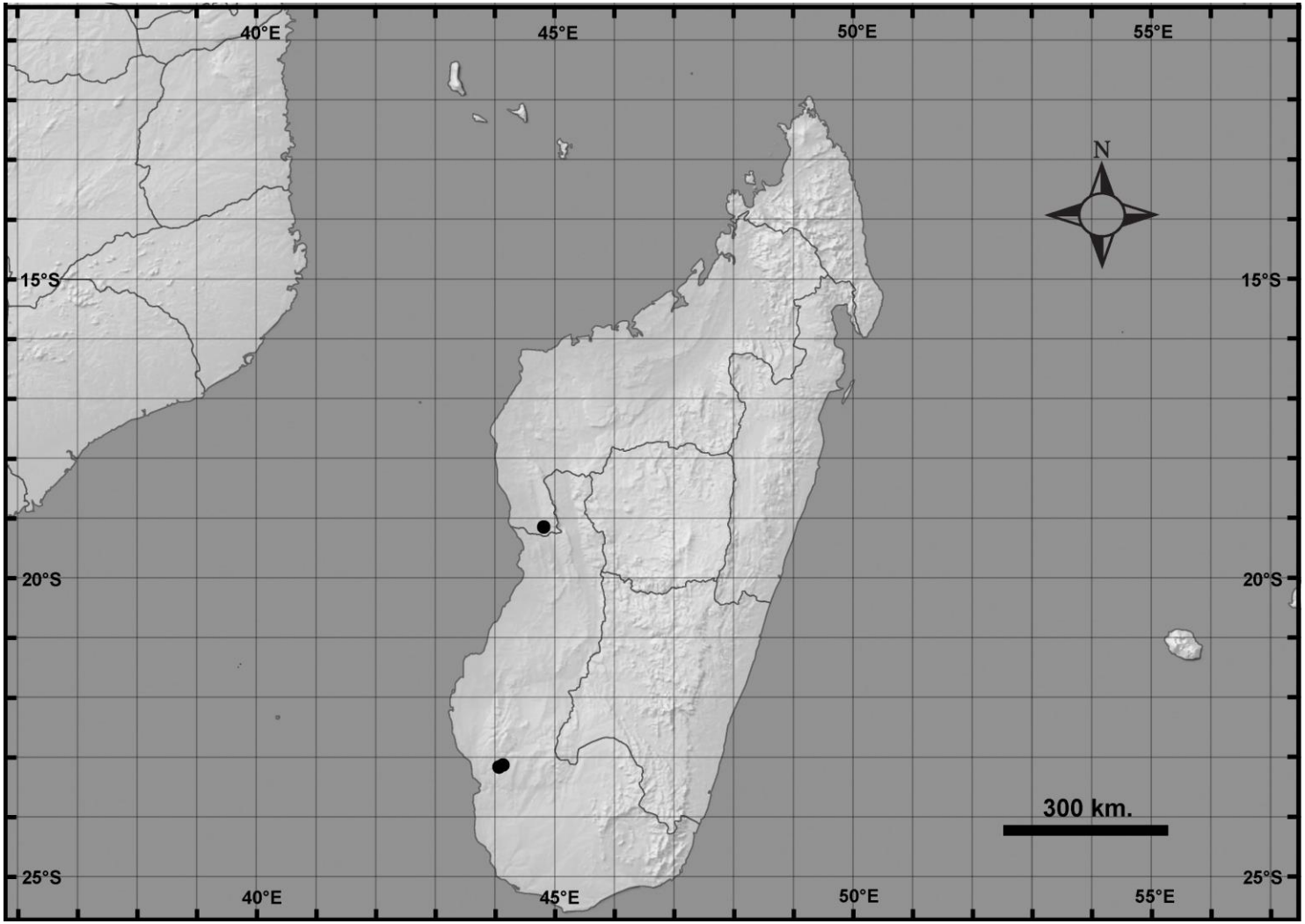


Figure 15. Map of the known distribution of *Commiphora andranovoryensis* in Madagascar, based upon collections used for descriptions in this study.

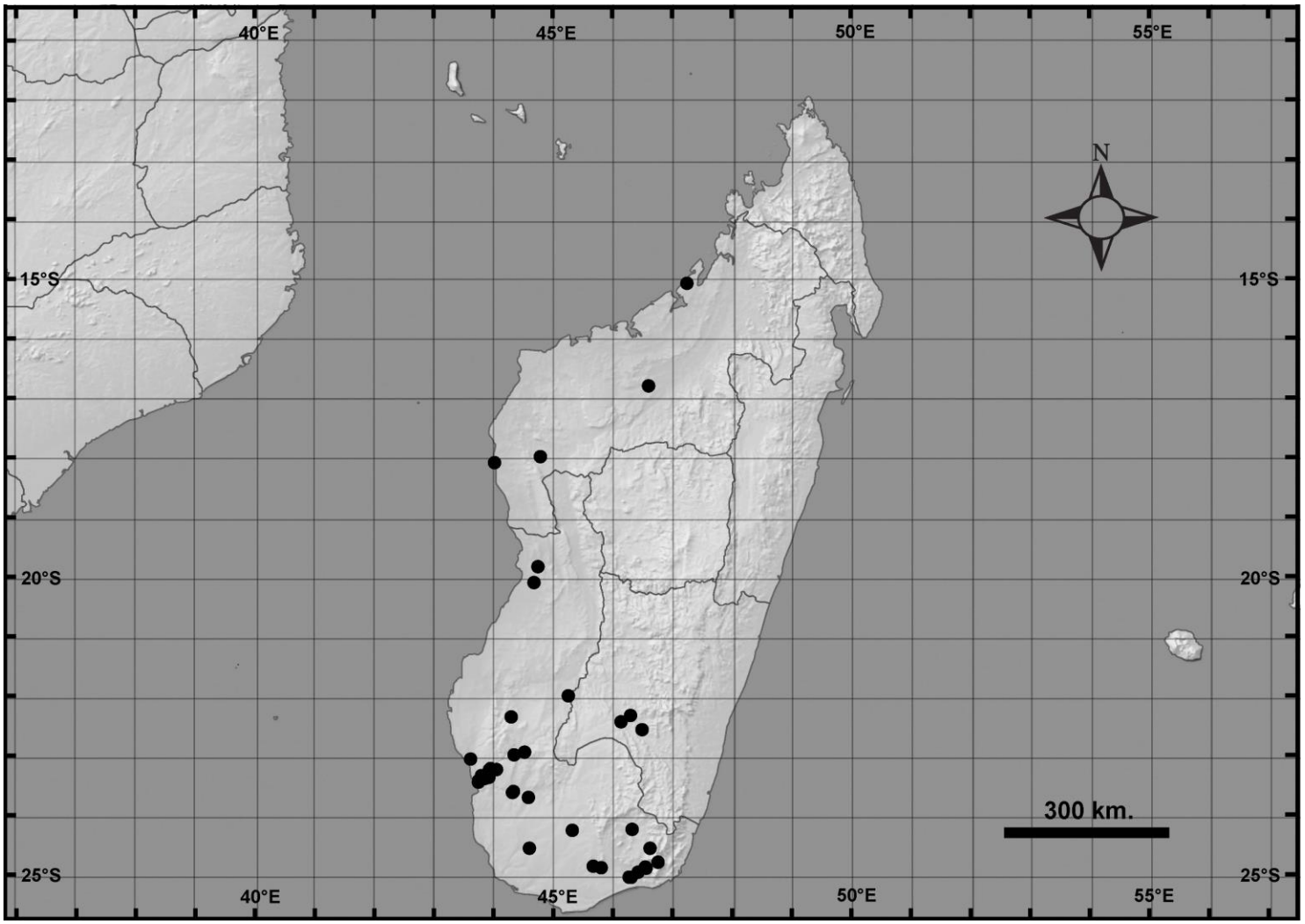


Figure 16. Map of the known distribution of *Commiphora aprevalii* in Madagascar, based upon collections used for descriptions in this study.

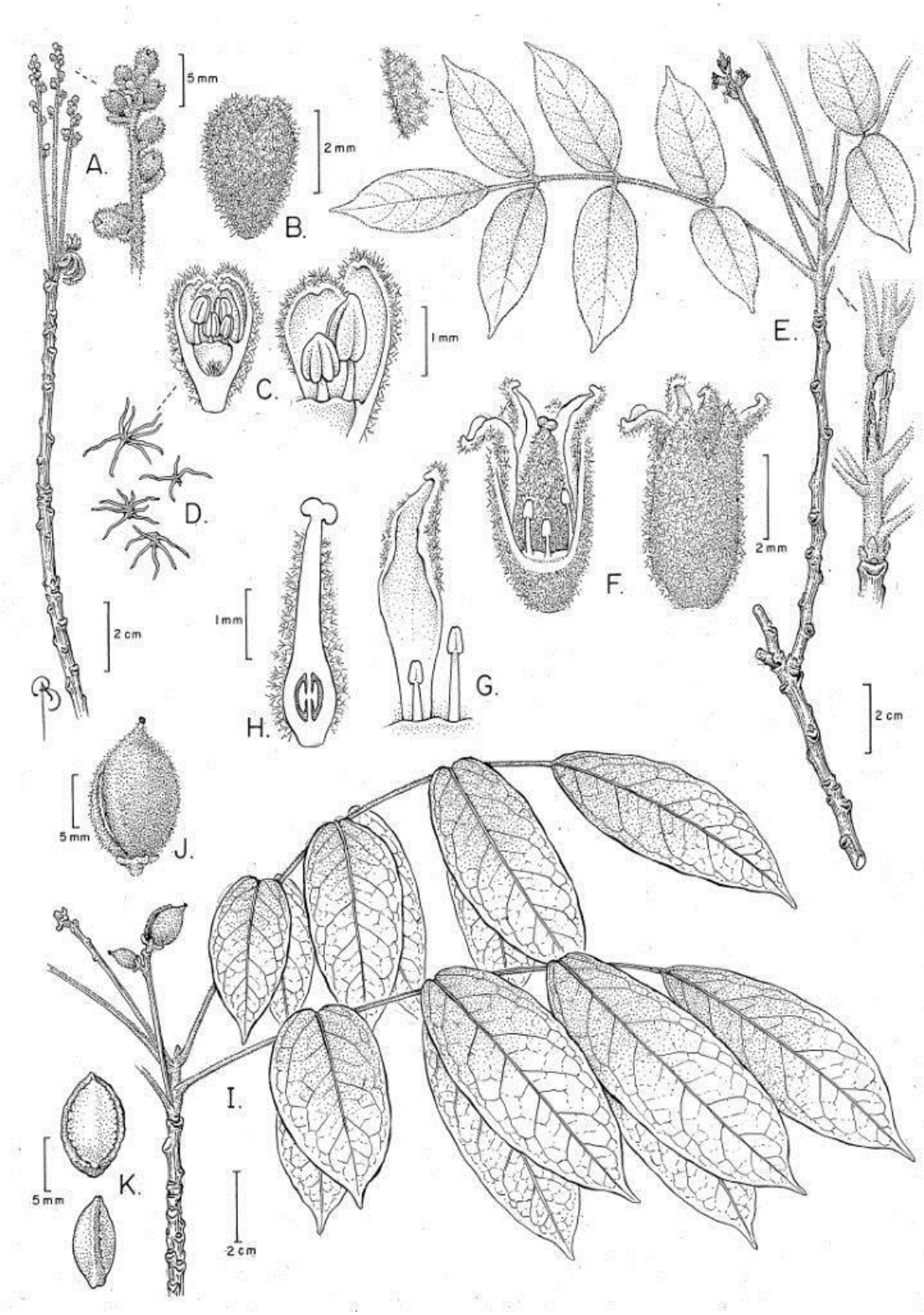


Figure 17. Illustration of *Commiphora elliptica*. Drawn by Bobbi Angell.

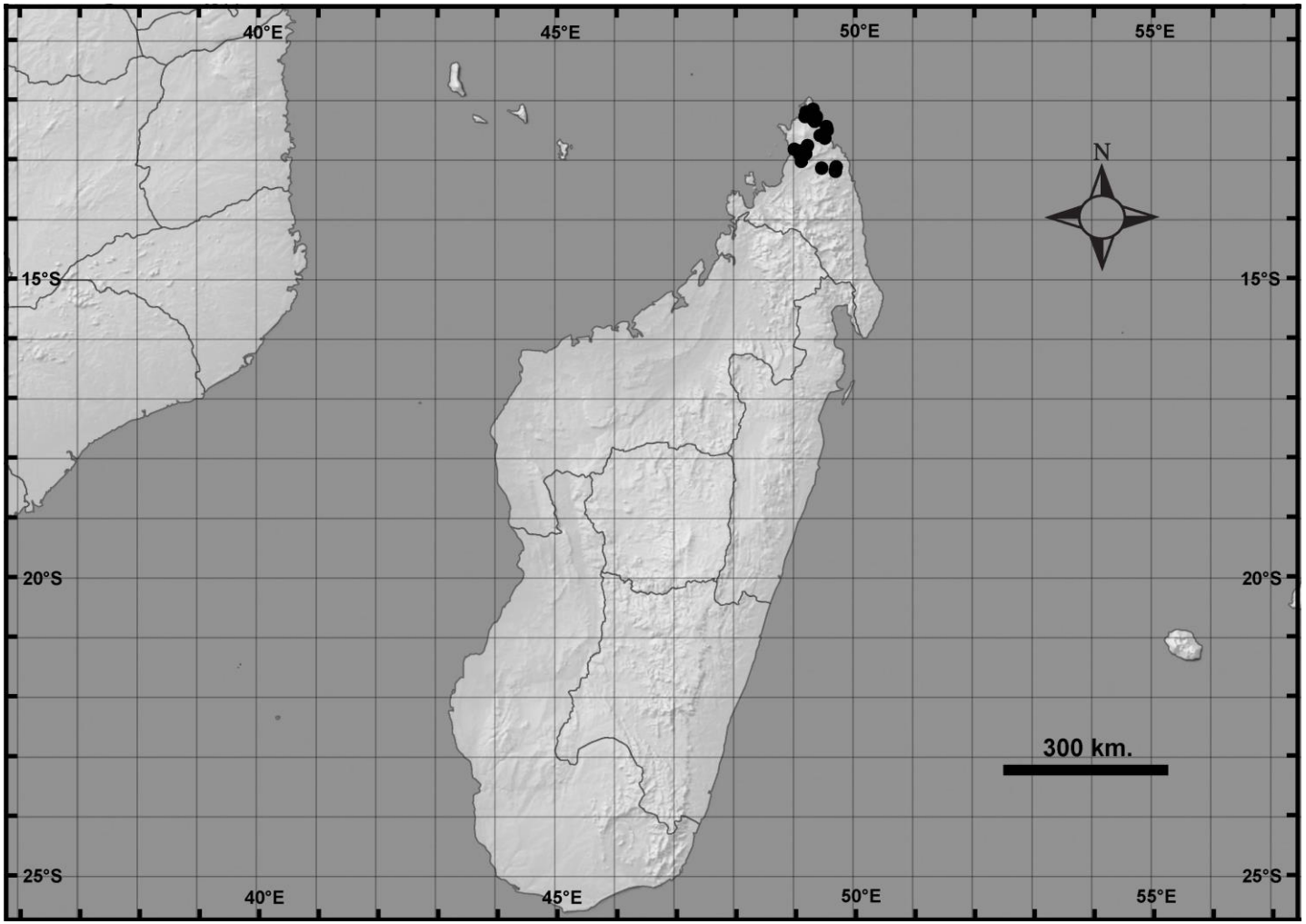


Figure 18. Map of the known distribution of *Commiphora elliptica* in Madagascar, based upon collections used for descriptions in this study.

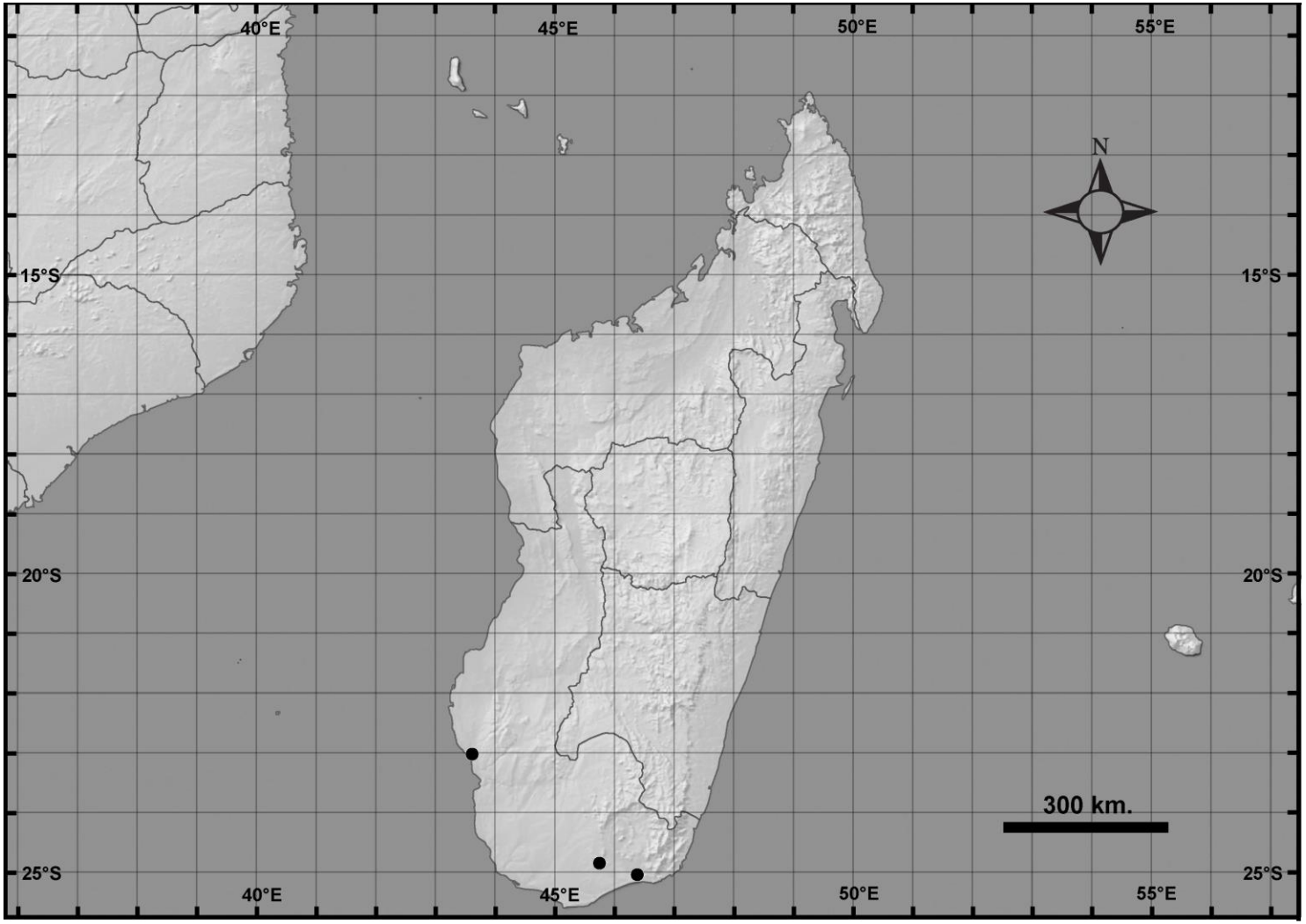


Figure 19. Map of the known distribution of *Commiphora falcate* in Madagascar, based upon collections used for descriptions in this study.

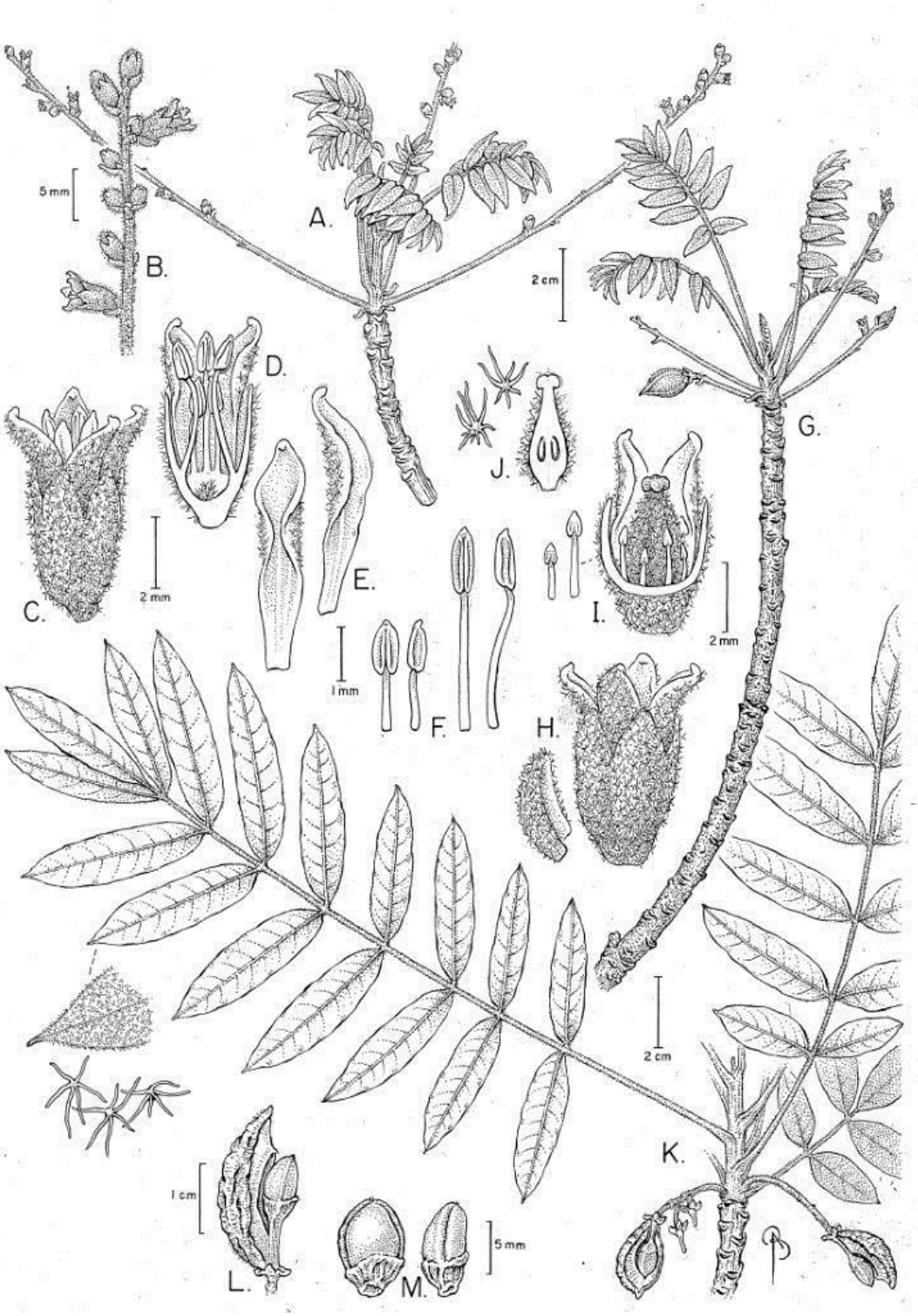


Figure 20. Illustration of *Commiphora morondavensis*. Drawn by Bobbi Angell.

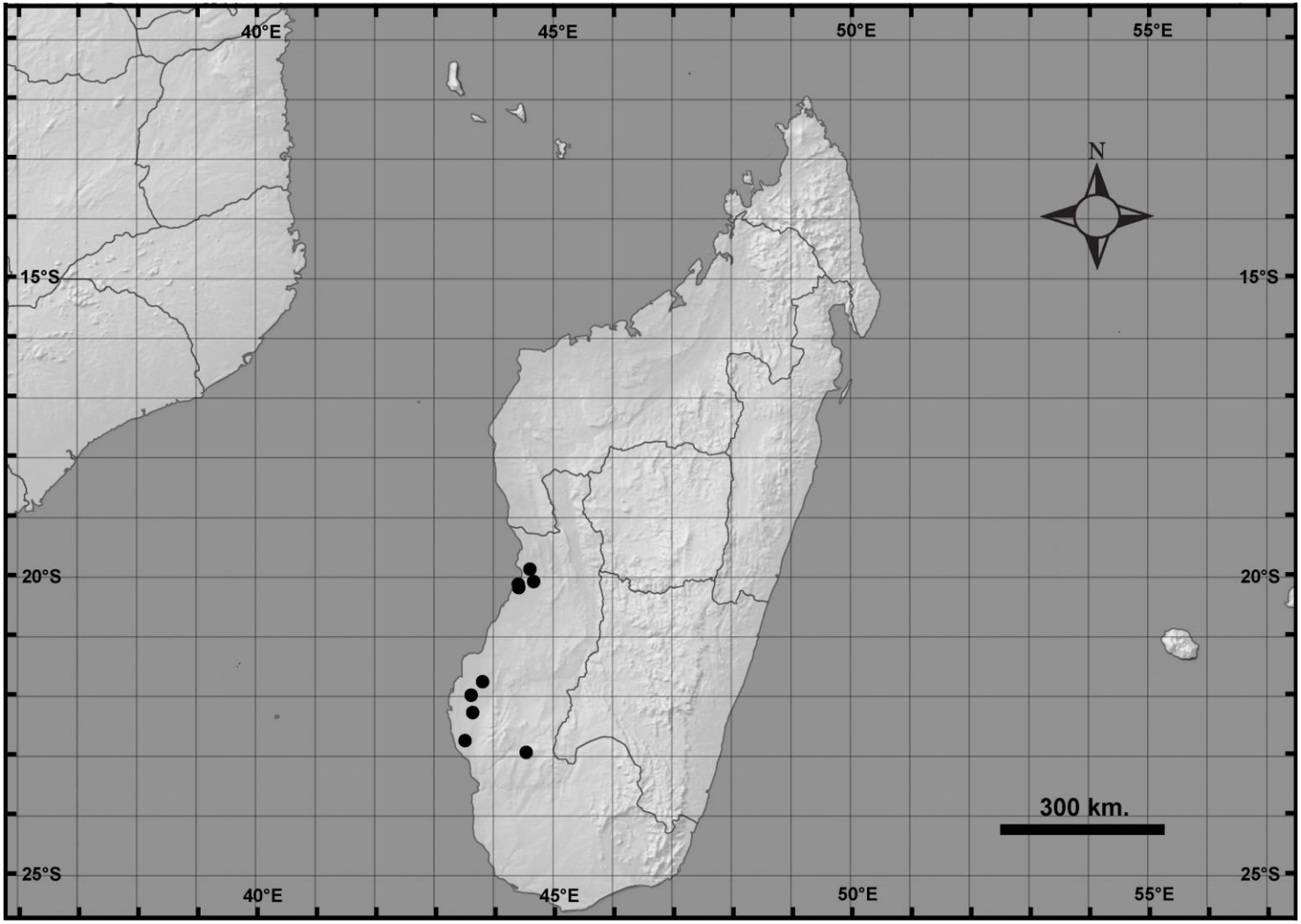


Figure 21. Map of the known distribution of *Commiphora morondavensis* in Madagascar, based upon collections used for descriptions in this study.

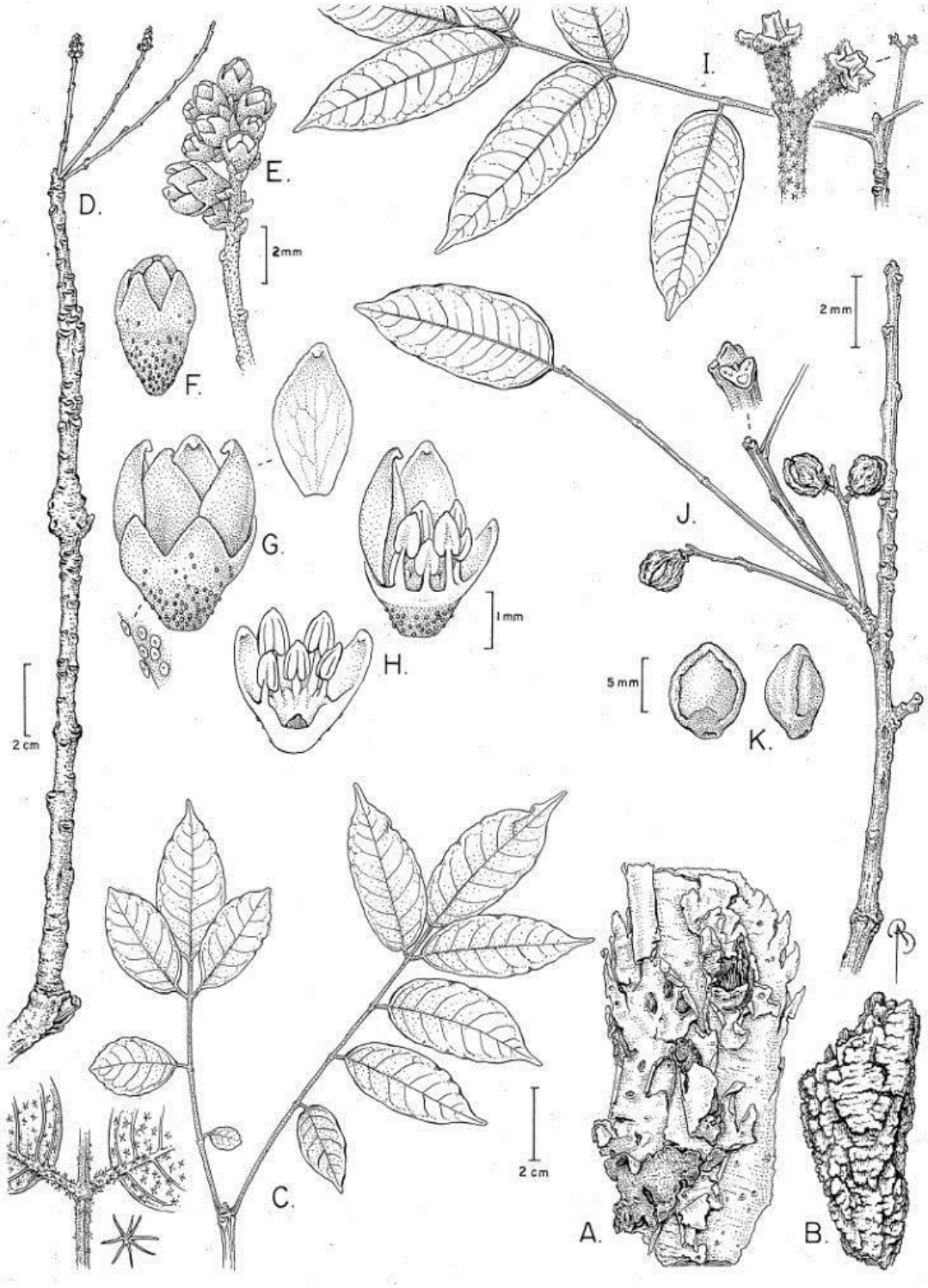


Figure 22. Illustration of *Commiphora razakamalalae*, drawn by Bobbi Angell.

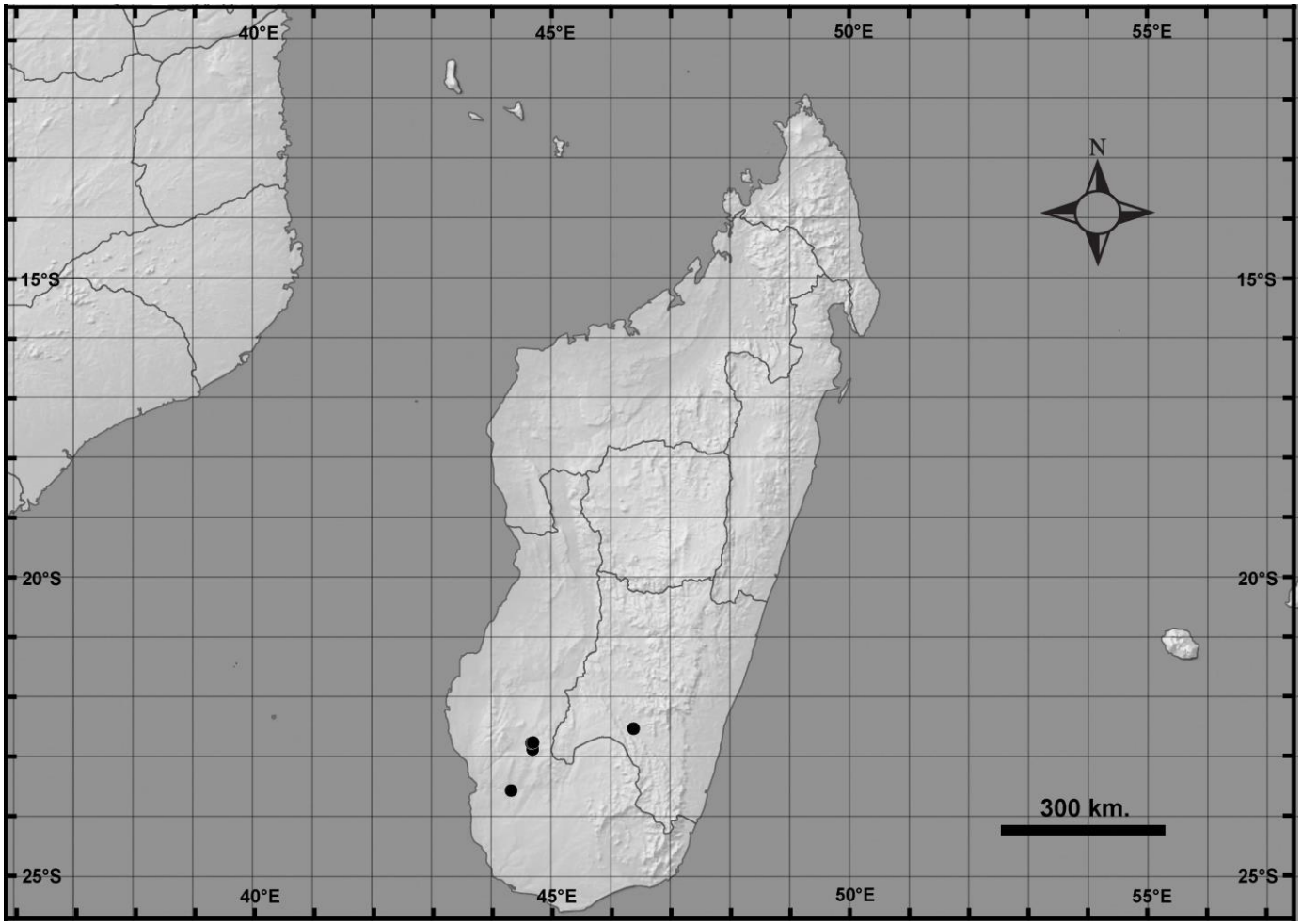


Figure 23. Map of the known distribution of *Commiphora razakamalalae* in Madagascar, based upon collections used for descriptions in this study.

Appendix 1

List of species sampled, with collector, voucher, geographic locality, and Genbank accession number for sequences used in this study.

Outgroup: *Bursera biflora* (Rose) Standl., Weeks 99-VII-17-07 (TEX), Mexico, AY315039/AY315041, AF445807, AY831896, KP061484, KP061332, *B. copallifera* (DC.) Bullock, Weeks 00-X-24-01 (TEX), Mexico, AY315042/AY315044, AF445833, AY831897, KP061485, KP061333, *B. cuneata* (Schltdl.) Engl., Weeks 99-VII-17-01 (TEX), Mexico, AY315045/AY315047, AY315045/AY315047, AY831898, KP061486, KP061334, *B. fagaroides* (Kunth) Engl., Weeks 01-X-08-01 (TEX), Baja California Sur, Mexico, AY309308/AY309310, AF445843, AY309391, KP061487, KP061335, *B. hindsiana* (Benth.) Engl., Weeks 00-VII-14-01 (TEX), Baja California Sur, Mexico, AY315048/AY315050, GQ378136, AY831899, KP061488, KP061336, *B. lancifolia* (Schltdl.) Engl., Weeks 98-VII-14-05 (TEX), Mexico, AY309317/AY309319, AF445857, AY309394, KP061489, KP061337, *B. microphylla* A. Gray, Weeks 01-X-08-02 (TEX), Arizona, USA, AY309326/AY309328, AF445855, AY309396, KP061490, KP061338, *B. sarukhanii* Guevara & Rzed., 00-VIII-18-06 (TEX), Mexico, AY315051/AY315053, AF445820, AY831900, KP061491, KP061339, *B. simaruba* (L.) Sarg., Pell s.n., Florida, USA, GQ378038, GQ378104, GQ377902, KP061492, KP061340, *B. spinescens* Urb. & Ekman, Weeks 01-VIII-23-02 (TEX), Mexico,

AY309356/AY309358, NA, AY309403, KP061493, KP061341, *B. tecomaca* (DC.) Standl., Weeks 02-XII-09-01 (TEX), Mexico, AY309359/AY309361, AF445838, AY309409, KP061494, KP061342, *B. tonkinensis*, Daly 13929 (NY), Vietnam, KP061640, NA, KP061215, KP061495, KP061343 ***Abyssinicae***: *C. brucei* Chiov., Gilbert 7531 (MO), Sidamo, Ethiopia. 4°14'N 42°03'E, KP061650, NA, KP061228, KP061511, KP061359, *C. ellenbeckii* Engl., Kuchar 17230 (MO), Hiiraan, Somalia. 3°42'N 44°54'E, KP061679, NA, KP061258, KP061544, KP061394, *C. habessenica* (O. Berg) Engl., Gereau 5781 (MO), Kigoma, Tanzania. 04°40'04"S 29°37'29"E, KP061696, NA, KP061278, KP061566, KP061415, *C. kua* Vollesen, Thulin 11159 (UPSV), Ethiopia: Harerge Region, KP061703, NA, KP061286, KP061576, KP061423 ***Africanae*:** *Commiphora africana* (A. Rich.) Engl., Weeks 02-XII-09-03 (TEX), South Africa, AY831869, NA, AY831901, KP061496, KP061344, *C. schimperi* (O. Berg) Engl., Thulin 11138 (UPSV), South Africa, AY315075NA7**, NA, AY831930, KP061615, KP061462, *C. stolonifera* Burtt, Greenway 13986 (MO), Iringa, Tanzania. No coordinates., AY831892, NA, AY831932, KP061634, KP061479, **Arillopsidium:** C. *edulis* subsp. *boiviniana* (Engl.) J.B. Gillett, Mwangoka 1451 (MO), Tanga, Tanzania. 04°58'44"S 38°58'02"E, NA, NA, KP061256, KP061542, KP061392, *C. edulis* subsp. *edulis* (Klotzsch) Engl., Weeks 00-VI-14-03 (TEX), Zimbabwe, KP061678, NA, KP061257, KP061543, KP061393, ***Campestres*:** *C. campestris* Engl., Weeks 00-VI-24-03 (TEX), Tropical East Africa, AY831873, NA, AY831906, KP061512, KP061360, ***Ciliatae***: *C. ciliata* Vollesen, Gilbert 7474 (MO), Sidamo, Ethiopia. 4°52'N 40°55'E, AY831874, NA, AY831907, KP061536, KP061385, ***Coriaceae***: *C. myrrha* (T. Nees)

Engl., Wieland 4611 (MO), Hobyo, Somalia. 06°32’12”N 047°42’24”E, NA, NA, KP061303, KP061597, KP061443, ***Hemprichia*****:** *C.* cf. *caerulea* Burtt, Kibure 831 (MO), Iringa, Tanzania. 08°54’30”S 34°24’08”E, KP061656, NA, KP061233, KP061521, KP061369, ***Hildebrandtianae*****:** *C. alaticaulis* J.B. Gillett & Vollesen, Thulin 11165 (UPSV), Ethiopia: Harerge Region, KP061641, NA, KP061216, KP061497, KP061345, *C. corrugata* J.B. Gillett & Vollesen, Thulin 11137 (UPSV), Ethiopia: Harerge Region, KP061672, NA, KP061249, KP061537, KP061387, ***Latifoliolata*****:** *C. eminii* Engl., Kayambo 4081 (MO), Arusha, Tanzania. 3°23’20”S 36°47’26”E, AY315060NA2**, NA, NA, KP061547, KP061396, *C. erosa* Vollesen, Thulin 11134 (UPSV), Ethiopia: Harerge Region, KP061682, NA, KP061262, KP061549, KP061398, *C. mollis* (Oliv.) Engl., Weeks 07-I-10-02 (GMUF), Namibia, NA, NA, KP061299, KP061592, KP061438, *C. mollis* (Oliv.) Engl., Weeks 14-IX-05-01 (GMUF), South Africa, NA, NA, KP061300, KP061593, KP061439, *C. sphaerocarpa* Chiov., Gilbert 7578 (MO), Sidamo, Ethiopia. 4°04’N 41°33’E, AY831891, NA, AY831931, KP061628, KP061474, ***Opobalsameae*****:** *C. boranensis* Vollesen, Thulin 11155 (UPSV), Ethiopia: Harerge Region, KP061649, NA, KP061226, KP061508, KP061356, ***Pedunculatae*****:** *C. pedunculata* (Kotchy & Peyr.) Engl., Gereau 3049 (MO), Iringa, Tanzania. 07°36’S 36°19’E, AY831889, NA, AY831927, KP061605, NA, **Incert. sed.** *C. anacardifolia* Dinter & Engl., Weeks 07-I-01-11 (GMUF), Namibia, KP061642, NA, KP061217, KP061498, KP061346, *C. angolensis* Engl., Raal 801 (TEX), Transvaal. Soutvlei-Grootfontein, Namibia", AY315054NA6**/KP061644, NA, AY831902, KP061500, KP061348, *C. angustifoliolata* Mendes, Swanpoel 28, Angola, KP061645, NA,

KP061219, KP061501, KP061349, *C. ankaranensis* (J.-F. Leroy) Cheek & Rakot., Weeks 10-I-11-02 (GMUF), Antsiranana, Madagascar. 12°18.886'S 49°20.306'E, KF035010, KP061782, KF035052, KF035032/KP061502, KP061350, *C. aprevalii* (Baill.) Guillaumin, Weeks 10-I-20-04 (GMUF), Toliara, Madagascar. 22°57.268'S 44°20.658'E, KF034992, KP061749, KF035034, KF035013/KP061503, KP061351, *C. aprevalii* (Baill.) Guillaumin, Gostel 124 (GMUF), Toliara, Madagascar. Beza Mahafaly Reserve. 23°39'18"S 44°37'45"E, KP061654, NA, KP061221, KP061518, KP061366, *C. aprevalii* (Baill.) Guillaumin, Gostel 128 (GMUF), Toliara, Madagascar. 23°40'59"S 44°38'03.7"E, KP061646, KP061750, KP061220, KP061504, KP061352, *C. aprevalii* var. *granulifera* H. Perrier, Gostel 125 (GMUF), Toliara, Madagascar. Beza Mahafaly Reserve. 23°39'18"S 44°37'45"E, KP061693, KP061767, KP061274, KP061564, KP061412, *C. arafy* H. Perrier, Weeks 10-II-13-01 (GMUF), Morondava, Madagascar. 20°9.914'S 44°26.736'E, KF034993, KP061783, KF035035, KF035014/KP061505, KP061353, *C. brevicalyx* subsp. *brevicalyx* H. Perrier, Weeks 10-I-24-05 (GMUF), Toliara, Madagascar. 23°18.177'S 43°44.868'E, AY831872, KP061753, KP061227, KP061509, KP061357, *C. brevicalyx* subsp. *vezorum* Capuron, Weeks 10-I-20-09 (GMUF), Toliara, Madagascar. 23°19.82'S 43°55.355'E, KF035011, KP061754, KF035053, KF035032/KP061510, KP061358, *C. capensis* Engl., Weeks 06-XII-23-03 (GMUF), Namibia, KP061651, NA, KP061229, KP061513, KP061361, *C. capuronii* Bardot-Vaucoulon, Weeks 10-I-15-04 (GMUF), Antsiranana, Madagascar. 12°34.821'S 49°27.515'E, KF034991, KP061755, KF035033, KP061514, KP061362, *C. capuronii* Bardot-Vaucoulon, Gautier 4885 (MO), Antsiranana, Madagascar. 13°06'12"S

49°34’48”E, KP061652, KP061756, KP061230, KP061515, KP061363, *C.* cf. *capuronii* Bardot-Vaucoulon, Gostel 65 (GMUF), Antsiranana, Madagascar. Ankarana NP. 12°56’28.5”S 49°06’55.8”E, NA, KP061758, KP061234, KP061516, KP061364, *C. capuronii* Bardot-Vaucoulon, Gostel 67 (GMUF), Antsiranana, Madagascar. Ankarana NP. 12°56’28.5”S 49°06’55.8”E, KP061736, KP061757, KP061231, KP061624, KP061470, *C. cervifolia* Van Der Walt, Weeks 06-XII-22-01, Namibia, KP061653, NA, KP061232, KP061517, KP061365, *C. chiovendana* J.B. Gillett ex Thulin, Kuchar 17126 (MO), Hiiraan, Somalia. 4°11’N 45°22’E, KP061670, NA, KP061247, NA, KP061384, *C.* cf. *coleopsis* H. Perrier, Gostel 142 (GMUF), Toliara, Madagascar. Zombitse NP. 22°53’10.84”S 44°41’31.04”E, KP061732, NA, KP061235, KP061620, KP061466, *C. coronillifolia* Chiov., Gilbert 7577 (MO), Dolo, Somalia. 4°04’N 41°53’E, KP061671, NA, KP061248, NA, KP061386, *C. crenatoserrata* Engl., Weeks 07-I-15-20 (GMUF), Namibia, KP061673, NA, KP061251, KP061538, KP061388, *C. cyclophylla* Chiov., Thulin 11168 (UPSV), Ethiopia: Harerge Region, KP061675, NA, KP061253, KP061539, NA, *C. dinteri* Engl., Weeks 06-XII-25-01 (GMUF), Namibia, KP061676, NA, KP061254, KP061540, KP061390, *C. dinteri* Engl., Weeks 07-I-15-05 (GMUF), Namibia, KP061677, NA, KP061255, KP061541, KP061391, *C. engleri* Guillaumin, Luke WRQ 9392 (MO), Udzungwa Mountain National Park, Tanzania", KP061681, NA, KP061261, KP061548, KP061397, *C. falcata* Capuron, Weeks 10-I-26-03 (GMUF), Toliara, Madagascar. 23°1.479’S 43°36.999’E, KF034994, KP061763, KF035036, KF035015/KP061550, KP061399, *C. foliacea* Sprague, Hein 6676 (E), Al Mahra, Yemen. No coordinates.", KP061683, NA, KP061263, KP061551, KP061400, *C.*

franciscana Capuron, Weeks 10-I-22-10 (GMUF), Toliara, Madagascar. 23°24.619'S 23°24.619'E, KF034995, KP061785, KF035037, KF035016/KP061552, KP061401, *C. fraxinifolia* Baker, Phillipson 5831 (MO), Antananarivo, Madagascar. 19°05'54"S 46°46'28"E, KP061684, KP061764, KP061264, KP061553, KP061402, *C. fraxinifolia* Baker, Service Forestiere s.n., Fianarantsoa, Madagascar. Ambinda Forest on route to Ivohibe, KP061685, NA, KP061265, KP061554, KP061403, *C.* cf. *fraxinifolia* Baker, Gostel 76 (GMUF), Mahajanga, Madagascar. Katsepy Village. 15°45'56"S 46°14'30"E, KP061734, NA, KP061241, KP061622, KP061468, *C. cf. fraxinifolia* Baker, Gostel 102 (GMUF), Mahajanga, Madagascar. Kily Village. 16°55'34"S 46°57'40"E, KP061659, NA, KP061239, KP061523, KP061372, *C.* cf. *fraxinifolia* Baker, Gostel 103 (GMUF), Antananarivo, Madagascar. 16°56'15"S 46°56'54"E, KP061665, NA, KP061240, KP061530, KP061379, *C. fulvotomentosa* Engl., Luke WRQ 10115 (MO), Mueda Plateau, Mozambique. 1123S, 3922E, KP061686, NA, KP061266, KP061555, KP061404, *C. gariepensis* Swanepoel, Weeks 06-XII-24-06 (GMUF), Namibia, NA, NA, KP061267, KP061556, KP061405, *C. giessi* Van Der Walt, Weeks 07-I-01-08 (GMUF), Namibia, KP061687, NA, KP061268, KP061557, KP061406, *C. gileadensis* (L.) C. Chr., Thulin 11158 (UPSV), Ethiopia: Harerge Region, KP061688, NA, KP061269, KP061558, NA, *C. gileadensis* (L.) C. Chr., Hein 5006 (E), Al Mahra, Yemen. No coordinates., KP061689, NA, KP061270, KP061559, KP061407, *C. glaucescens* Engl., Swanpoel 17, Namibia, KP061690, NA, KP061271, KP061560, KP061408, *C. gracilifrondosa* Dinter ex Van Der Walt, Weeks 06-XII-24-04 (GMUF), Namibia, KP061691, NA, KP061272, KP061561, KP061409, *C. grandifolia* Engl., Weeks 10-I-13-

01 (GMUF), Antsiranana, Madagascar. 12°34.821’S 49°27.515’E, KF034996, KP061765, KF035038, KF035017/KP061562, KP061410, *C. grandifolia* Engl., Gostel 53 (GMUF), Antsiranana, Madagascar. Baie de Courrier. 12°16’51”S 49°11’06”E, KP061692, KP061766, KP061273, KP061563, KP061411, *C. guillauminii* H. Perrier, Gostel 100 (GMUF), Mahajanga, Madagascar. Andranofantsika Village. 16°20’13”S 46°50’40”E, KP061660, KP061768, KP061275, KP061524, KP061373, *C. guillauminii* H. Perrier, S.F. 24587 (TEF), Mahajanga, Madagascar. 18°49’30”S 44°22’30”E, KP061694, NA, KP061276, NA, KP061413, *C. gurreh* Engl., Thulin 11135 (UPSV), Ethiopia: Harerge Region, KP061695, NA, KP061277, KP061565, KP061414, *C. hildebrantii* Engl., Thulin 11169 (UPSV), Ethiopia: Harerge Region, KP061697, NA, KP061279, KP061567, KP061416, *C. hodai* Sprague, Thulin 11152 (UPSV), Ethiopia: Harerge Region, KP061698, NA, KP061280, KP061568, KP061417, *C. horrida* Chiov., Thulin 11164 (UPSV), Ethiopia: Harerge Region, KP061699, NA, KP061281, KP061569, KP061418, *C. humbertii* H. Perrier, Weeks 10-I-20-08 (GMUF), Toliara, Madagascar. 23°19.82’S 43°55.255’E, KF034997, KP061786, KF035039, KF035018, KP061419, *C. humbertii* H. Perrier, Gostel 136 (GMUF), Toliara, Madagascar. Sakamena. 22°54’32.4”S 44°31’13.4”E, KP061658, KP061787, KP061282, KP061571, KP061371, *C.* cf. *humbertii* H. Perrier, Gostel 146 (GMUF), Toliara, Madagascar. 22°17’44.87”S 22°18’46.44”E, KP061700, NA, KP061283, KP061572, KP061420, *C. kaokoensis* Swanepoel, Weeks 07-I-15-25 (GMUF), Namibia, KP061701, NA, KP061284, KP061573, KP061421, *C. kerstingii* Engl., Etkin 29 (MO), Kano, Nigeria. 11°42’05”N 08°52’31”E, AY831879, NA, AY831915, KP061574, NA, *C. krauseliana*

Heine, Weeks 07-I-15-09 (GMUF), Namibia, KP061702, NA, KP061285, KP061575, KP061422, *C. kuneneana* Swanepoel, Weeks 07-I-15-02 (GMUF), Namibia, KP061704, NA, KP061287, KP061577, NA, *C. kuneneana* Swanepoel, Weeks 07-I-15-03 (GMUF), Namibia, KP061705, NA, KP061288, KP061578, KP061424, *C. lamii* H. Perrier, Weeks 10-I-26-02 (GMUF), Toliara, Madagascar. 23°1.479'S 43°36.999'E, KF034998, KP061769, KF035040, KF035019, KP061425, *C. lasiodisca* H. Perrier, Weeks 10-I-11-12 (GMUF), Antsiranana, Madagascar. 12°19.233'S 49°20.292'E, KP061706, KP061746/KP061779, KP061289, KF035020/KP061580, KP061426, *C. lasiodisca* H. Perrier, Gostel 59 (GMUF), Antsiranana, Madagascar. Ankarana NP. 12°35'01.8"S 49°26'56.3"E, KP061707, KP061747/KP061780, KP061290, KP061581, KP061427, *C. lasiodisca* H. Perrier, Gostel 84 (GMUF), Mahajanga, Madagascar. Tsingy de Namoroka NP. 16°29'49"S 45°25'20"E, KP061709, KP061748/KP061781, KP061291, KP061582, KP061428, *C. laxecymigera* H. Perrier, Gostel 115 (GMUF), Toliara, Madagascar. Beza Mahafaly Reserve. 23°41'17.4"S 44°34'34.9"E, KP061710, KP061788, KP061292, KP061583, KP061429, *C.* cf. *leandriana* H. Perrier, Weeks 10-II-14-12 (GMUF), Morondava, Madagascar. 20°4.361'S 44°40.565'E, KP061661, NA, KP061242, KP061525, KP061374, *C.* cf. *leandriana* H. Perrier, Weeks 10-II-14-14 (GMUF), Morondava, Madagascar. 20°4.361'S 44°40.565'E, NA, NA, NA, KP061526, KP061375, *C. leptophloeos* (Mart.) J.B. Gillett, Abbott 16295 (TEX), Santa Cruz, Bolivia. 16°35'S 61°52'W", KP061711, NA, KP061293, KP061584, KP061430, *C. leptophloeos* (Mart.) J.B. Gillett, Dal Forno 2159 (GMUF), Nossa Senhora da Glória, Sergipe, Brazil. 10°06'13.5"S 37°25'12.9"W, KP061662, NA, KP061294, KP061527, KP061376, *C.*

madagascariensis Jacq., Luke WRQ 10186 (MO), Mtwara, Tanzania. Mnazi Bay. 10°17’S 04°22’E, NA, NA, KP061295, KP061585, KP061431, *C. mafaidoha* H. Perrier, Weeks 10-II-13-02 (GMUF), Morondava, Madagascar. 20°9.914’S 44°26.736’E, KF035000, KP061789, KF035042, KF035021/KP061587, KP061433, *C. mafaidoha* H. Perrier, Gostel 104 (GMUF), Toliara, Madagascar. 22°24’05”S 46°07’54”E, KP061712, NA, KP061296, KP061586, KP061432, *C. mahafaliensis* Capuron, Weeks 10-I-22-02 (GMUF), Toliara, Madagascar. 23°24.619’S 43°46.8’E, KF035001, KP061790, KF035043, KF035022/KP061588, KP061434, *C. marchandii* Engl., Weeks 10-II-15-04 (GMUF), Morondava, Madagascar. 20°25.843’S 45°22.872’E, KF035002, KP061771, KF035044, KF035023/KP061589, KP061435, *C. marchandii* Engl., Gostel 95 (GMUF), Mahajanga, Madagascar. Ankarafantsika NP. 16°19’42”S 46°47’20”E, KP061714, NA, KP061298, KP061591, KP061437, *C. marchandii* Engl., Gostel 143 (GMUF), Toliara, Madagascar. 22°20’10.4”S 46°17’23.2”E, KP061713, KP061772, KP061297, KP061590, KP061436, *C.* cf. *marchandii* Engl., Gostel 91 (GMUF), Mahajanga, Madagascar. Katsepy Village. 15°45’18”S 46°14’39”E, KP061664, NA, KP061244, KP061529, KP061378, *C.* cf. *marchandii* Engl., Gostel 126 (GMUF), Toliara, Madagascar. Beza Mahafaly Reserve. 23°39’18”S 44°37’45”E, KP061663, NA, KP061243, KP061528, KP061377, *C.* cf. *marchandii* Engl., Gostel 117 (GMUF), Toliara, Madagascar. 23°39'02""S 44°37'46""E, NA, KP061770, NA, NA, NA, *C. monstruosa* (H. Perrier) Capuron, Weeks 10-I-21-09 (GMUF), Toliara, Madagascar. 23°24.686’S 43°46.816’E, KF035003, KP061791, KF035045, KF035024/KP061594, KP061440, *C. multijuga* K. Schum., Weeks 06-XII-22-03 (GMUF), Namibia, KP061716, NA, KP061302,

KP061596, KP061442, *C. namaensis* Schinz, Weeks 06-XII-31-05 (GMUF), Namibia, KP061717, NA, KP061304, KP061598, KP061444, *C. neglecta* Verdc., Weeks 00-VII-14-04 (TEX), South Africa, AY831886, NA, AY831924, KP061599, KP061445, *C. oblanceolata* Schinz, Swanpoel 37, Namibia, KP061718, NA, KP061305, KP061600, KP061446, *C. orbicularis* var. *orbicularis* Engl., Weeks 10-I-26-05 (GMUF), Toliara, Madagascar. 23°1.479'S 43°36.999'E, KF035004, NA, KF035046, KF035025/KP061603, KP061449, *C. orbicularis* var. *orbicularis* Engl., Gostel 111 (GMUF), Toliara, Madagascar. 23°32'05"S 44°19'09"E, KP061720, KP061793, KP061307, KP061602, KP061448, *C. orbicularis* var. *tulearensis* Capuron, Weeks 10-I-23-04 (GMUF), Toliara, Madagacar. 23°29.576'S 43°45.709'E, KP061721, KP061794, KP061308, KP061604, KP061450, *C.* cf. *pervilleana* Engl., Gostel 97 (GMUF), Mahajanga, Madagascar. Ankarafantsika NP. 16°20'13"S 46°47'11"E, KP061667, NA, KP061246, KP061533, KP061380, *C. planifrons* Engl., Miller 8691 (E), S of Shihali, Socotra. No coordinates., KP061722, NA, KP061309, KP061606, KP061452, *C. pterocarpa* H. Perrier, Weeks 10-I-28-11 (GMUF), Toliara, Madagascar. 23°26.849'S 43°57.484'E, KP061724, KP061775, KP061311, KP061608, KP061454, *C. pterocarpa* H. Perrier, Gostel 106 (GMUF), Toliara, Madagascar. 22°24'50"S 46°01'13"E, KP061723, KP061774, KP061310, KP061607, KP061453, *C. pyracanthoides* Engl., Weeks 06-XII-27-01 (GMUF), Namibia, KP061725, NA, KP061312, KP061609, KP061455, *C. rostrata* Engl., Gilbert 7472 (MO), Sidamo, Somalia. 4°52'N 40°55'E, KP061726, NA, KP061314, KP061611, KP061457, *C. saxicola* Engl., Weeks 06-XII-29-04 (GMUF), Namibia, KP061729, NA, KP061317, KP061613, KP061460, *C. saxicola*

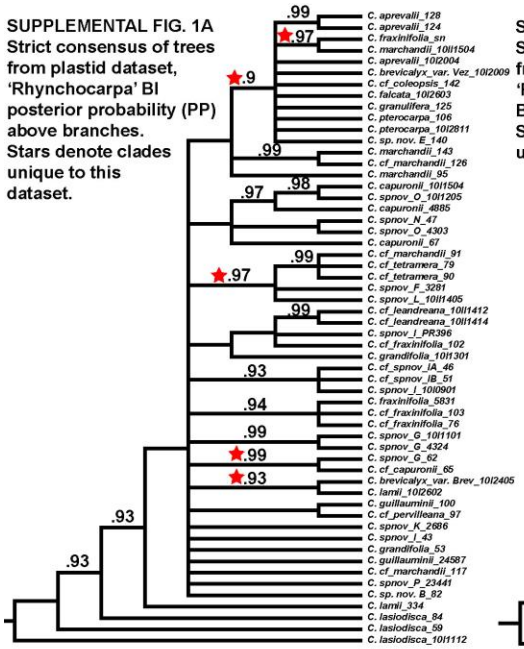
Engl., Weeks 07-I-01-09 (GMUF), Namibia, KP061730, NA, KP061318, KP061614, KP061461, *C. simplicifolia* H. Perrier, Weeks 10-I-20-03 (GMUF), Toliara, Madagascar. 22°54.614'S 44°20.401'E, KF035006, NA, KF035048, KF035027/KP061616, KP061463, *C. simplicifolia* H. Perrier, Gostel 74 (GMUF), Mahajanga, Madagascar. Katsepy Village. 15°45'18"S 46°14'39"E, KP061731, NA, KP061320, KP061618, KP061464, *C. simplicifolia* H. Perrier, Gostel 144 (GMUF), Toliara, Madagascar. 22°20'10.4"S 46°17'23.2"E, KP061669, NA, KP061319, KP061617, KP061382, *C. sinuata* H. Perrier, Weeks 10-I-23-06 (GMUF), Toliara, Madagascar. 23°31.554'S 43°45.44'E, KF035007, KP061795, KF035049, KF035028/KP061619, KP061465, *C. sphaerophylla* Chiov., Thulin 11156 (UPSV), Ethiopia: Harerge Region, KP061739, NA, KP061323, KP061629, NA, *C.* sp. nov. A, Weeks 10-I-28-08 (GMUF), Toliara, Madagascar. 23°26.849'S 43°57.484'E, KF035008, NA, KF035050, KF035029/KP061630, KP061475, *C.* sp. nov. B, Gostel 82 (GMUF), Mahajanga, Madagascar. Tsingy de Namoroka NP. 16°29'49"S 45°25'20"E, KP061737, KP061778, KP061324, KP061626, KP061472, *C.* sp. nov. C, Gostel 83 (GMUF), Mahajanga, Madagascar. Tsingy de Namoroka NP. 16°29'49"S 45°25'20"E, KP061738, NA, KP061325, KP061627, KP061473, *C.* sp. nov. D, Gostel 86 (GMUF), Mahajanga, Madagascar. Tsingy de Namoroka NP. 16°29'49"S 45°25'20"E, KP061741, NA, KP061326, KP061632, KP061477, *C.* sp. nov. E, Gostel 140 (GMUF), Toliara, Madagascar. Zombitse NP. 22°52'43.6"S 44°41'35.3"E, KP061740, NA, KP061327, KP061631, KP061476, *C.* sp. nov. F, Jongkind 3281 (P), Mahajanga, Madagascar. Tsingy de Bemaraha NP. 19°09'S 44°49'E, KP061643, NA, KP061218, KP061499,

KP061347, *C.* sp. nov. G, Weeks 10-I-11-01 (GMUF), Antsiranana, Madagascar. 12°18.886'S 49°20.306'E, KP061647, NA, KP061222, KF035026/KP061506, KP061354, *C.* sp. nov. G, Gostel 62 (GMUF), Antsiranana, Madagascar. Ankarana NP. 12°56'28.5"S 49°06'55.8"E, KP061708, NA, KP061223, KP061519, KP061367, *C.* sp. nov. G, Schatz 4324, Antsiranana, Madagascar. 12°19'10"S 49°20'16"E, KP061666, NA, KP061245, KP061531, KP061451, *C.* sp. nov. H, Gostel 141 (GMUF), Toliara, Madagascar. Zombitse NP. 22°52'59.3"S 44°41'56.5"E, KP061648, KP061784, KP061224, KP061507, KP061355, *C.* sp. nov. H, Gostel 145 (GMUF), Toliara, Madagascar. Zombitse NP. 22°22'10.09"S 46°26'18.15"E, KP061655, NA, KP061225, KP061520, KP061368, *C.* sp. nov. I, Weeks 10-I-09-01 (GMUF), Antsiranana, Madagascar. 12°14.177' 49°21.293'E, KF035009, KP061761, KF035051, KF035030/KP061545, KP061395, *C.* sp. nov. I, Gostel 43 (GMUF), Antsiranana, Madagascar. Anivorano Nord. 12°45'18"S 49°14'15"E, KP061735, KP061762, KP061259, KP061623, KP061469, *C.* sp. nov. I, Ranirison PR396, Antsiranana, Madagascar. 13°06'48.996"S 49°43'10.992"E, KP061680, NA, KP061260, KP061546, NA, *C.* cf. sp. nov. I, Gostel 46 (GMUF), Antsiranana, Madagascar. Baie de Courrier. 12°12'51"S 49°10'03"E, KP061668, KP061759, KP061237, KP061534, KP061381, *C.* cf. sp. nov. I, Gostel 51 (GMUF), Antsiranana, Madagascar. Baie de Courrier. 12°15'58"S 49°11'44"E, KP061733, KP061760, KP061238, KP061621, KP061467, *C.* sp. nov. J, Gostel 58 (GMUF), Antsiranana, Madagascar. Ankarana NP. 12°35'01.8"S 49°26'56.3"E, KP061657, NA, KP061250, KP061522, KP061370, *C.* cf. sp. nov. J, Gostel 44 (GMUF), Antsiranana, Madagascar. Anivorano Nord. 12°45'18"S 49°14'15"E,

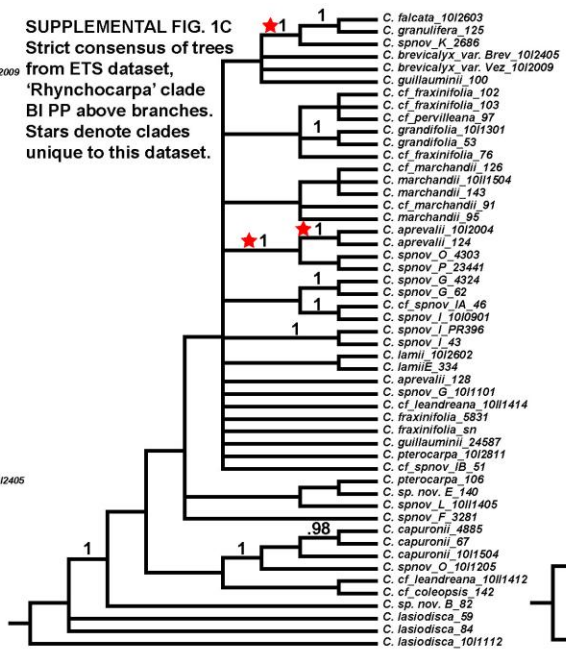
NA, NA, KP061236, KP061532, NA, *C.* sp. nov. K, Labat 2686 (P), Mahajanga, Madagascar. 18°39’06”S 44°42’12”E, KP061674, NA, KP061252, KP061636, KP061389, *C.* sp. nov. L, Weeks 10-II-14-05 (GMUF), Morondava, Madagascar. 19°51’S 44°36’E, KP061715, KP061773, KP061301, KP061595, KP061441, *C.* sp. nov. M, Weeks 10-I-28-04 (GMUF), 23°26.849’S 43°57.484’E, KP061719, KP061792, KP061306, KP061601, KP061447, *C.* sp. nov. N, Gostel 47 (GMUF), Antsiranana, Madagascar. Baie de Courrier. 12°12’51”S 49°10’03”E, NA, NA, KP061313, KP061610, KP061456, *C.* sp. nov. O, Weeks 10-I-12-05 (GMUF), Antsiranana, Madagascar. 12°18.989’S 49°20.934’E, KP061727, KP061776, KP061315, KP061612, KP061458, *C.* sp. nov. P, S.F. 23441 (TEF), Antsiranana, Madagacar. 13°25’S 48°19’E, KP061728, NA, KP061316, NA, KP061459, *C. steynii* Swanepoel, Weeks 07-I-15-08 (GMUF), Namibia, KP061742, NA, KP061328, KP061633, KP061478, *C. tenuipetiolata* Engl., Weeks 06-XII-31-01 (GMUF), Namibia, KP061743, NA, KP061329, KP061635, KP061480, *C.* cf. *tetramera* Engl., Gostel 79 (GMUF), Mahajanga, Madagascar. Katsepy village. 15°53’S 46°06’E, NA, KP061777, KP061321, KP061625, KP061471, *C.* cf. *tetramera* Engl., Gostel 90 (GMUF), Mahajanga, Madagascar. 15°45’18”S 46°14’39”E, NA, NA, KP061322, KP061535, KP061383, *C. virgata* Engl., Weeks 07-I-01-01 (GMUF), Namibia, KP061744, NA, KP061330, KP061637, KP061481, *C. wightii* (Arn.) Bhandari, Weeks 00-VIII-18-03 (TEX), India, AY315081/AY315083, GQ378139, AY831936, KP061638, KP061482, *C. wildii* Merxm., Weeks 06-XII-30-05 (GMUF), Namibia, KP061745, NA, KP061331, KP061639, KP061483

Appendix 2

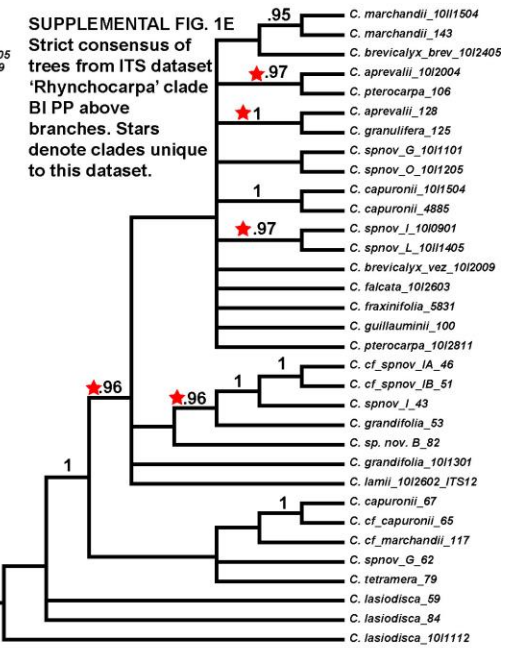
Supplemental Figures 1–5 for individual phylogenetic analysis of marker loci in this study



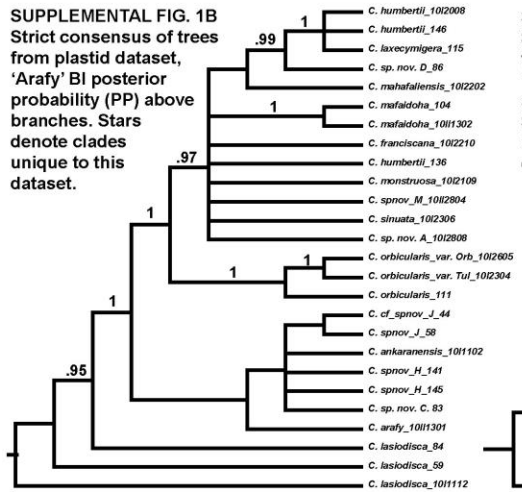
SUPPLEMENTAL FIG. 1A
Strict consensus of trees from plastid dataset, 'Rhynchocarpa' BI posterior probability (PP) above branches.
Stars denote clades unique to this dataset.



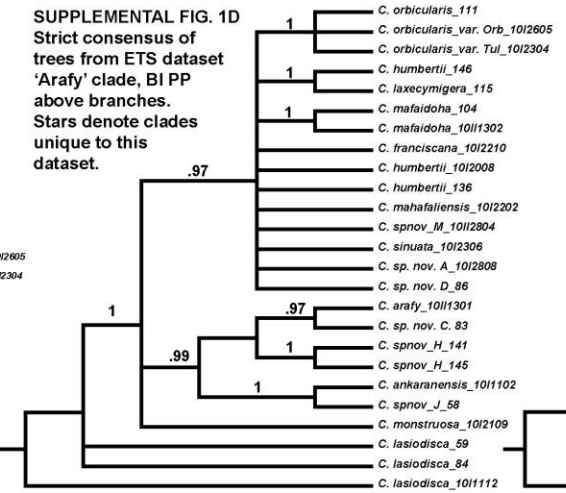
SUPPLEMENTAL FIG. 1C
Strict consensus of trees from ETS dataset, 'Rhynchocarpa' clade BI PP above branches.
Stars denote clades unique to this dataset.



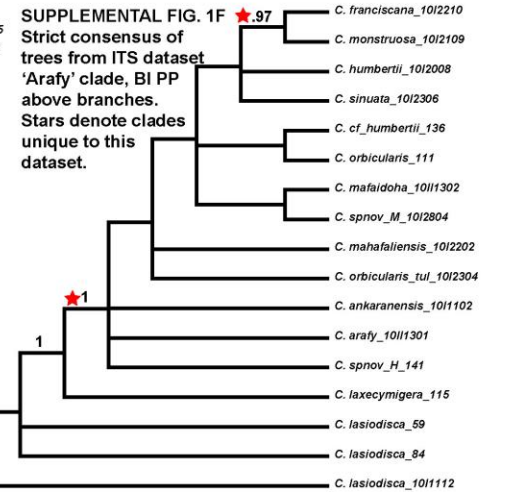
SUPPLEMENTAL FIG. 1E
Strict consensus of trees from ITS dataset 'Rhynchocarpa' clade BI PP above branches. Stars denote clades unique to this dataset.



SUPPLEMENTAL FIG. 1B
Strict consensus of trees from plastid dataset, 'Arafy' BI posterior probability (PP) above branches. Stars denote clades unique to this dataset.

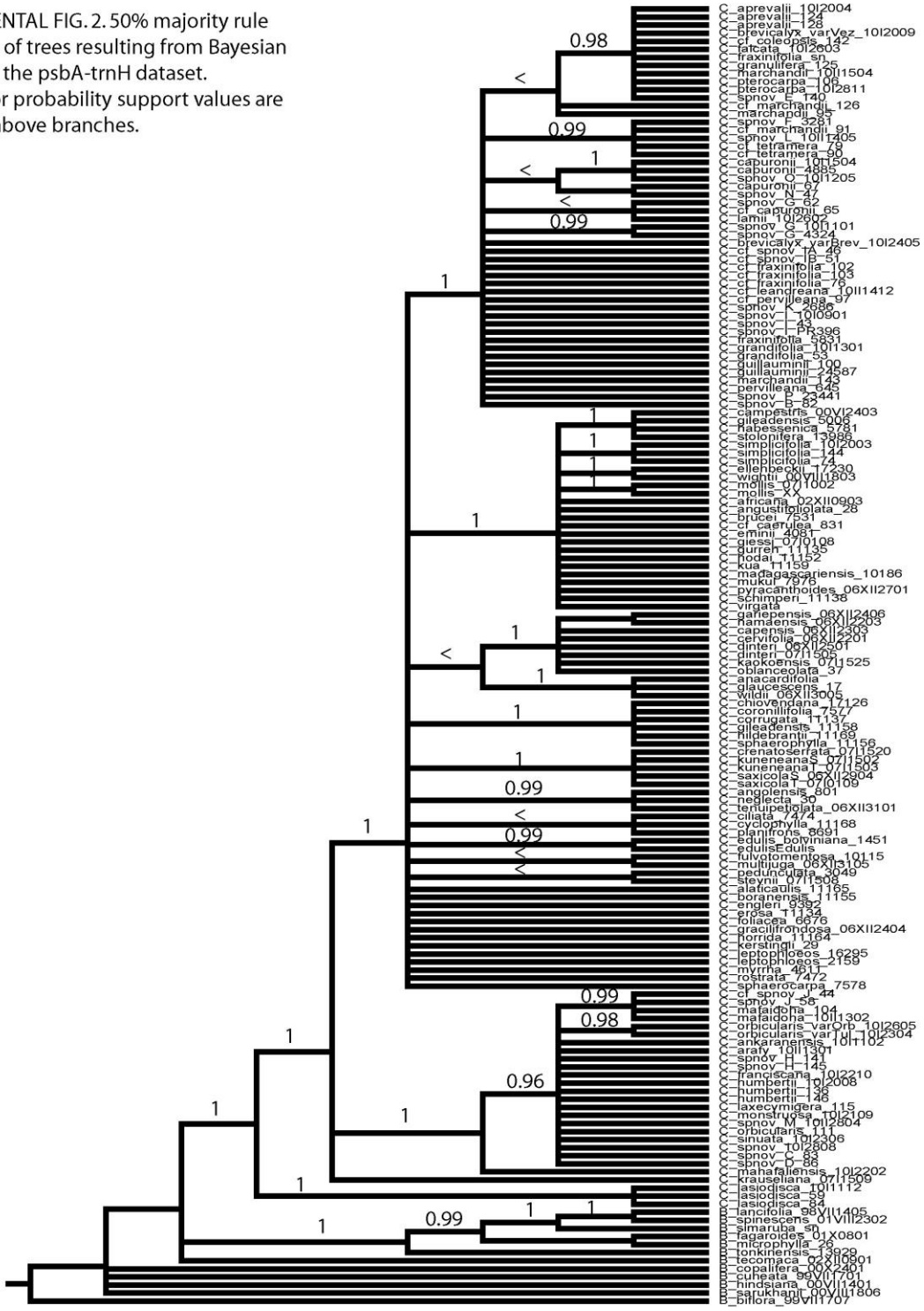


SUPPLEMENTAL FIG. 1D
Strict consensus of trees from ETS dataset 'Arafy' clade, BI PP above branches.
Stars denote clades unique to this dataset.

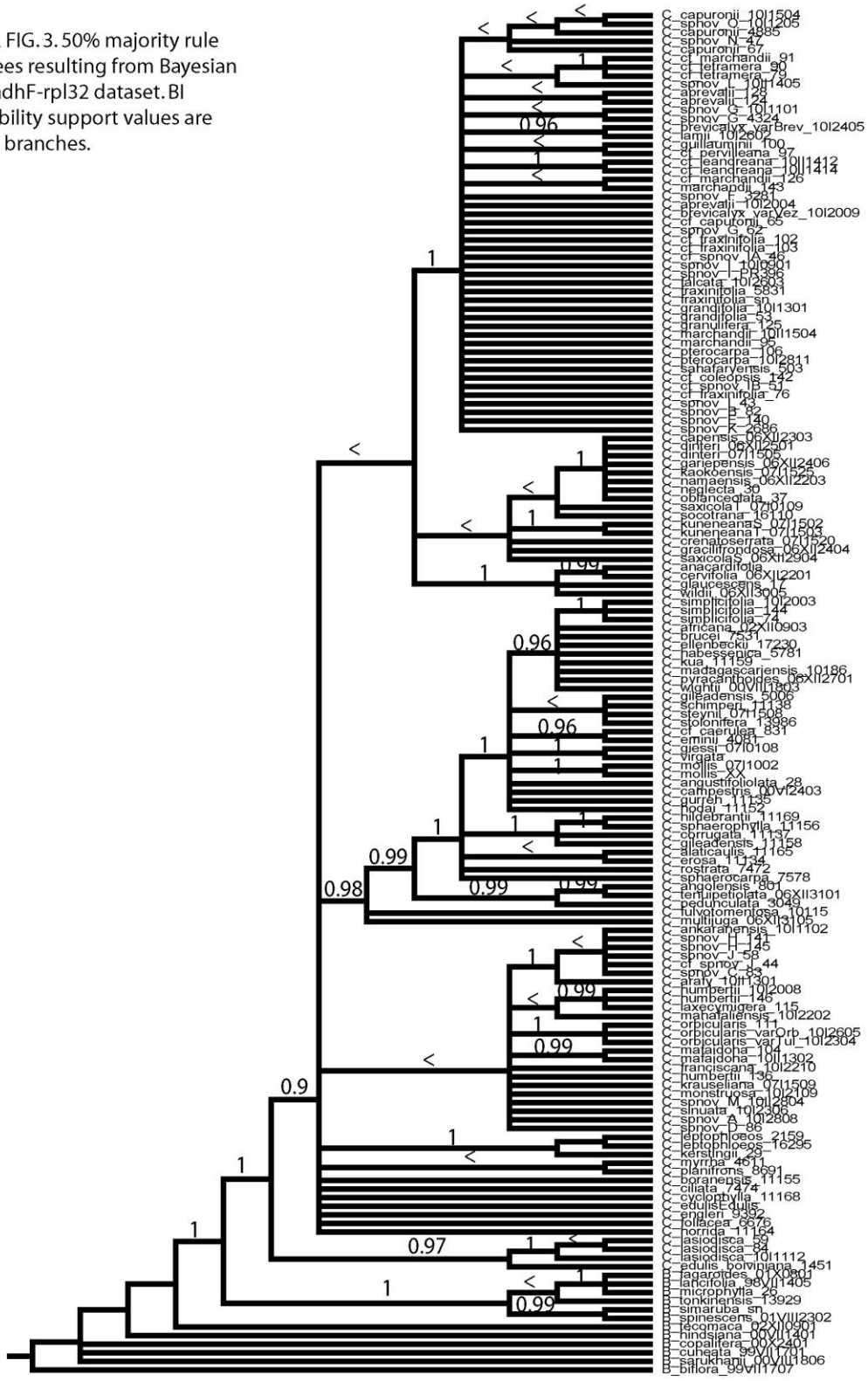


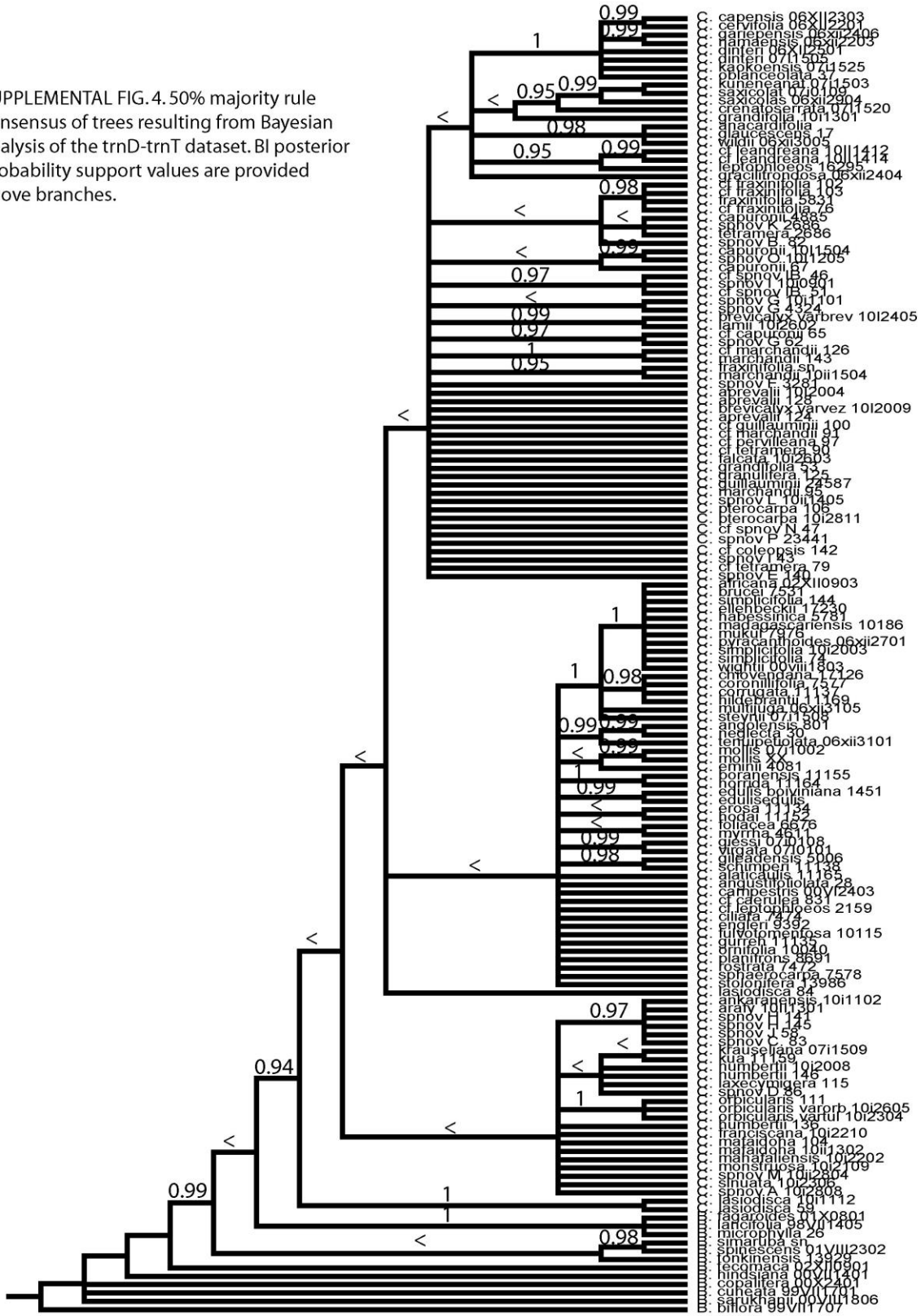
SUPPLEMENTAL FIG. 1F
Strict consensus of trees from ITS dataset 'Arafy' clade, BI PP above branches.
Stars denote clades unique to this dataset.

SUPPLEMENTAL FIG. 2. 50% majority rule consensus of trees resulting from Bayesian analysis of the psbA-trnH dataset.
BI posterior probability support values are provided above branches.



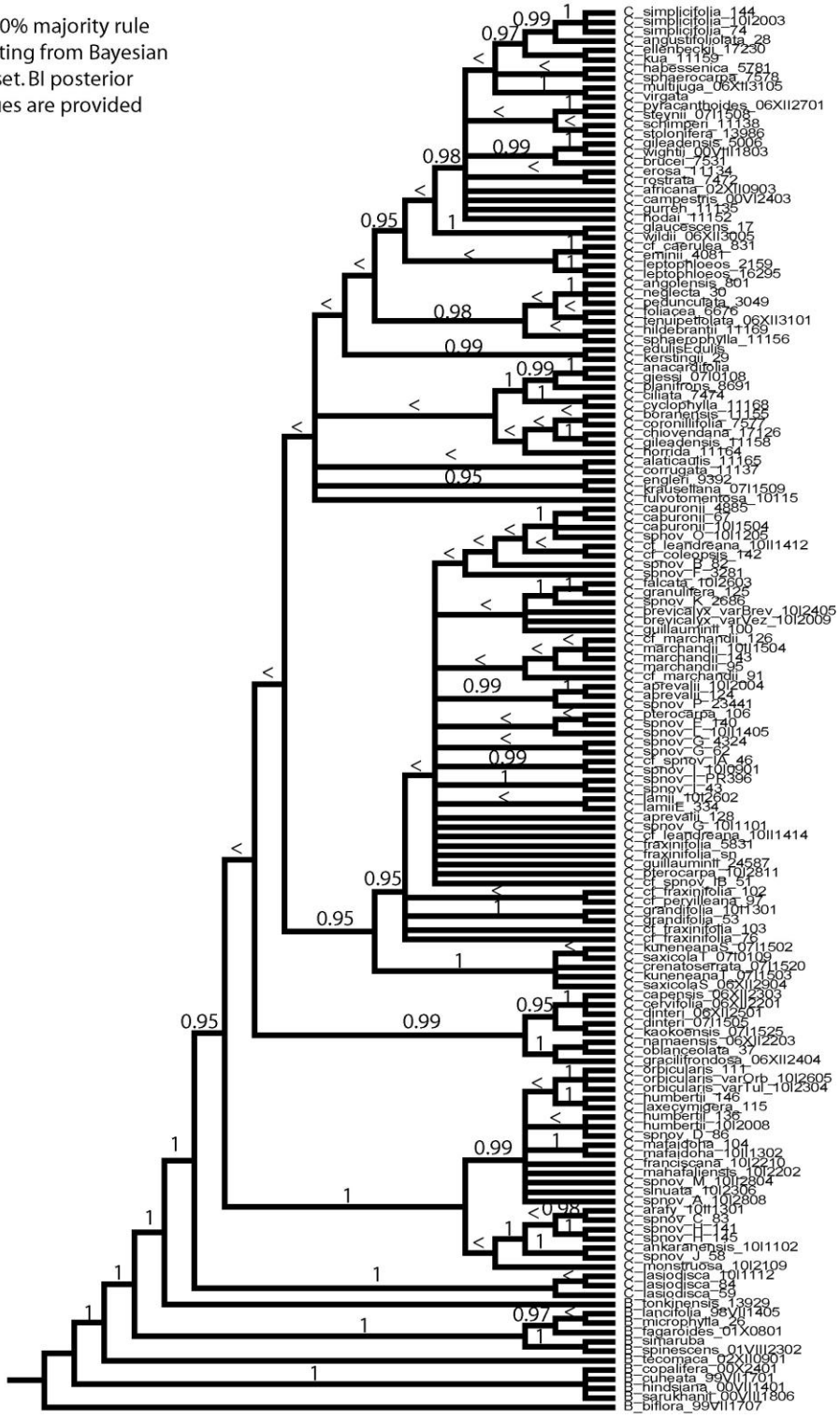
SUPPLEMENTAL FIG. 3. 50% majority rule consensus of trees resulting from Bayesian analysis of the ndhF-rpl32 dataset. BI posterior probability support values are provided above branches.





SUPPLEMENTAL FIG. 4. 50% majority rule consensus of trees resulting from Bayesian analysis of the trnD-trnT dataset. BI posterior probability support values are provided above branches.

SUPPLEMENTAL FIG. 5. 50% majority rule consensus of trees resulting from Bayesian analysis of the ETS dataset. BI posterior probability support values are provided above branches.



Appendix 3

List of species, vouchers, and geographic origin with Genbank accession numbers for all putative gene regions. GPS coordinates were not included with some of the herbarium vouchers, which is reflected when no coordinate is given.

Beiselia mexicana Forman; Pell s.n. (TEX). Mexico. ETS — FJ233929, ITS — JF919030, *RPT6A* Intron — KF906106, *BXL2* Intron — KF906094, *mtATPSynthase D* Intron — KF906084.

Protium copal (Schltdl. & Cham.) Engl; Killeen 3136 (MO). Mexico. 15°15'S 067°00'W. ETS — AY964612, ITS — KF906073, *RPT6A* Intron — KF906108, *BXL2* Intron — KF906095, *mtATPSynthase D* Intron — KF906085, *Rab6* Intron — KF906120.

Aucoumea klaineana Pierre; Walters 466 (MO). Gabon. 00°07'12"S 011°42'57"E. ETS — KF906082, *RPT6A* Intron — KF906105, *BXL2* Intron — KF906093.

Boswellia sacra Flueck; Weeks 01-X-08-03 (TEX). N.E. Africa. ETS — AF445957, ITS — AF455880, *RPT6A* Intron — KF906107, *Rab6* Intron — KF906119.

Bursera sarukhanii Guevara & Rzed; Weeks 00-VIII-18-06 (TEX). Mexico. ETS — AY315051, ITS — AF445820, *RPT6A* Intron — KF906109.

B. simaruba (L.) Sarg; Goldman s.n. (TEX). Florida, U.S.A. ETS — GQ378038, ITS — GQ378104, *RPT6A* Intron — KF906110, *BXL2* Intron — KF906097, *mtATPSynthase D* Intron — KF906086.

Commiphora ankaranensis (J.-F. Leroy) Cheek & Rakot; Weeks 10-I-11-02 (GMUF). Ankarana, Madagascar. 12°18'53"S 49°20'18"E. ETS — KF035010, ITS — KF906081, *RPT6A* Intron — KF906118, *BXL2* Intron — KF906104, *mtATPSynthase D* Intron — KF906092, *Rab6* Intron — KF906128.

Commiphora aprevalii (Baill.) Guillaumin; Weeks 10-I-20-04 (GMUF). Toliara, Madagascar. 22°57'16"S 44°20'39"E. ETS — KF034992, ITS — KF906075, *RPT6A* Intron — KF906112, *mtATPSynthase D* Intron — KF906088, *Rab6* Intron — KF906122.

Commiphora falcata Capuron; Weeks 10-I-26-03 (GMUF). Toliara, Madagascar. 23°01'29"S 43°36'60"E. ETS — KF034994, ITS — KF906076, *RPT6A* Intron — KF906113, *BXL2* Intron — KF906099, *mtATPSynthase D* Intron — KF906089, *Rab6* Intron — KF906123.

Commiphora grandifolia Engl; Weeks 10-I-13-01 (GMUF). Ankarana, Madagascar. 12°34'49"S 49°27'31"E. ETS — KF034996, ITS — KF906077, *RPT6A* Intron — KF906114, *BXL2* Intron — KF906100, *mtATPSynthase D* Intron — KF906090, *Rab6* Intron — KF906124.

Commiphora lamii H. Perrier; Weeks 10-I-26-02 (GMUF). Toliara, Madagascar. 23°01'29"S 43°36'60"E. ETS — KF034998, ITS — KF906078, *RPT6A* Intron — KF906115, *BXL2* Intron — KF906101, *mtATPSynthase D* Intron — KF906091, *Rab6* Intron — KF906125.

Commiphora pervilleana Engl; Weeks 10-I-11-01 (GMUF). Ankarana, Madagascar. 12°18'53"S 49°20'18"E. ETS — KF035005, ITS — KF906079, *RPT6A* Intron — KF906116, *BXL2* Intron — KF906102, P43 — KF906126.

Commiphora sp. A. Weeks 10-I-09-01 (GMUF). Ankarana, Madagascar. 12°14'11"E 49°21'18"E. ETS — KF035009, ITS — KF906080, *RPT6A* Intron — KF906117, *BXL2* Intron — KF906103, *Rab6* Intron — KF906127.

Commiphora sp. B. Weeks 10-I-15-04 (GMUF). Ankarana, Madagascar. 12°34'49"S 49°27'31"E. ETS — KF0906087, ITS — KF906074, *RPT6A* Intron — KF906111, *BXL2* Intron — KF906098, *mtATPSynthase D* Intron — KF906087, *Rab6* Intron — KF906121.

Appendix 4

Chapter 3 Supplemental Table 1. List of all 48 primer pairs developed in this study and their characteristics. Columns and their names are described as follows: T_m corresponds to melting temperature; GC% corresponds to percent G-C content, Nbases is the number of nucleotides that comprise the primer, Reference Taxon identifies the model organism reference in IntrEST from which the primer was developed, EST is the expressed sequence tag record number that was used to develop the marker, %ID is the percent shared identity between the reference taxon and *Citrus* species, and *Citrus* sp. is the species of *Citrus* (*C. sinensis* or *C. clementina*) that was used to develop the primer. Primer names with asterisks (*) indicate primers that did not meet the necessary melting temperature criteria; † indicates primer pairs that were able to successfully amplify in at least one specimen.

Primer Name	Sequence	Tm	GC %	Nbases	Reference Taxon	EST	% ID	Citrus sp.
1F†	GATCWGARATCGCMGARGAAGTYCGC	60.7	46.2	26	*Arabidopsis*	AT3G12290.1	85%	sinensis
1R†	TCAGCRCAMGCYTTYCTCTTCRTRYTC	74	37.1	27	*Arabidopsis*	AT3G12290.1	85%	sinensis
2F†	TGCAARTCTCTYKTTGCYGG	65.3	40	20	*Arabidopsis*	AT4G01100.1	83	sinensis
2R†	TCCARWGGWGCAACAG	61.1	50	18	*Arabidopsis*	AT4G01100.1	83	sinensis

	CA							
3F	TACACRTATGCWAGRTGCAC	63.6	40	20	*Arabidopsis*	AT1G48410.2	87	sinensis
3R	GCTGCWAGRTGKGCATARTATGC	65.9	43.5	23	*Arabidopsis*	AT1G48410.2	87	sinensis
4F†	ATTCTYGTCYTGTCCGSWAGAGA	61	39.2	23	*Arabidopsis*	AT4G21960.1	83	clementina
4R†	CCATCTCTYCTTCCTGTCTT	51.2	45	20	*Arabidopsis*	AT4G21960.1	83	clementina
5F	GCTGGAYTMACSSTYGAYCATCC	64.3	39.2	23	*Arabidopsis*	AT1G22410.1	84	clementina
5R	TCATGRGAWGTCCAGAASTC	67.7	40	20	*Arabidopsis*	AT1G22410.1	84	clementina
6F	ATGCTKTTTGGTGCWRTTGGGAC	74	43.5	23	*Arabidopsis*	AT5G27150.1	82	clementina
6R	GAACTGWRTWACACCTAGWGATATGA	57.3	34.7	26	*Arabidopsis*	AT5G27150.1	82	clementina
7F	GCAKCACCRAAGATGYTGAAC	62.4	42.9	21	*Arabidopsis*	AT1G34130.1	91	sinensis
7R	CAGCTCWCCRAAYCKRTARTAKSATA	60	27	26	*Arabidopsis*	AT1G34130.1	91	sinensis
8F	GCTGTTGCSYTGAARCAGGC	64.8	50	20	*Arabidopsis*	AT4G13930.1	92	sinensis
8R	CTTGTTTYGCRTAGRCCTTGAA	60	36.4	22	*Arabidopsis*	AT4G13930.1	92	sinensis
9F†	ATATCARGGTGCYTACAAGA	57.1	35	20	*Arabidopsis*	AT5G50850.1	90	sinensis
9R†	CTCAGGRCCATATTTCTCCAA	58.1	42.9	21	*Arabidopsis*	AT5G50850.1	90	sinensis
10F†	CTCCARCACATYCAYGARCTCCAGC	55.6	48	25	*Arabidopsis*	AT5G19990.1	93	clementina

10R†	AGCTGTAAYTCTTCTYTRAGCATCC	61.7	36	25	*Arabidopsis*	AT5G19990.1	93	clementina
11F	ACGMCTYGAYATGGATTACGTYGA	54.8	37.5	24	*Arabidopsis*	AT1H04690.1	90	clementina
11R*	TTCATCGCCCTCACMGTCTCYTC	77	52.2	23	*Arabidopsis*	AT1H04690.1	90	clementina
12F†	CGTYGAYGTKMTYTATTGYCACA	57.7	30.5	23	*Arabidopsis*	AT1G04690.1	90	clementina
12R†	TCATCGCCCTCACMGTCTC	67.1	57.9	19	*Arabidopsis*	AT1G04690.1	90	clementina
13F†	ACAAGCCWCCTGAAGATGC	62	52.7	19	*Arabidopsis*	AT4G37510.1	87	sinensis
13R†	GTCCAAGTTCRATRTTYCTWGCTTC	54.5	36	25	*Arabidopsis*	AT4G37510.1	87	sinensis
14F†	TTAYTCAATGTTCAACAGA	57.6	26.4	19	*Arabidopsis*	AT4G02580.1	86	sinensis
14R†	CACGKAYCATRCAAGGTGTTGTGCC	62.7	48	25	*Arabidopsis*	AT4G02580.1	86	sinensis
15F	GCTYTWCCTTCRGAKACTGGTC	57.3	45.5	22	*Arabidopsis*	AT5G37850.2	88	sinensis
15R	GTACWGARTKGATTGGATCCAC	58.5	41	22	*Arabidopsis*	AT5G37850.2	88	sinensis
16F†	CTTGTGGGAAKCCATCGGAC	66.3	55	20	*Arabidopsis*	AT1G02640.1	84	clementina
16R†	CGTTGTACATKGCYCTKGCYTCA	64.9	43.5	23	*Arabidopsis*	AT1G02640.1	84	clementina
17F†	CAAGARGCKKTTTGTCGCC	65.8	47.4	19	*Arabidopsis*	AT1G62050.1	83	clementina
17R†	CCAAGCKRARCGGTGGTGA	57.6	52.7	19	*Arabidopsis*	AT1G62050.1	83	clementina
18F	TTGAGTTRTCTCSWGAA	57.2	36.9	19	*Arabidopsis*	AT3G07160.1	87	clementina

	GC							
18R	GCARTGCAATRTCARCAGC	55.5	42.2	19	*Arabidopsis*	AT3G07160.1	87	clementina
19F	TGAYCTYCTTGATGCRTTGGAC	62.8	41	22	*Arabidopsis*	AT5G11170.2	95	sinensis
19R	GCATATWGARGGRAAATTGCATTC	55	33.4	24	*Arabidopsis*	AT5G11170.2	95	sinensis
20F	AGTTTRCTCTCTGTTGATCCRAC	51.9	39.2	23	*Arabidopsis*	AT2G25660.1	93	sinensis
20R	GCTGMACTTCAACTTCYGTWCCA	56.8	43.5	23	*Arabidopsis*	AT2G25660.1	93	sinensis
21F	GGAAYTWAGGGAAGAATGC	57.6	42.2	19	*Arabidopsis*	AT4G32180.3	90	sinensis
21R	GCATCAASAAAYTGGAAYTC	67.8	30	20	*Arabidopsis*	AT4G32180.3	90	sinensis
22F	GATGGCTCGTGAAAGTGCTC	65	55	20	*Arabidopsis*	AT2G27600.1	91	clementina
22R	CCACGYKKACCACAYAARGAATC	56.6	39.2	23	*Arabidopsis*	AT2G27600.1	91	clementina
23F	GATGCRTTGGACTTYAATCAA	58.6	33.4	21	*Arabidopsis*	AT5G11170.2	95	clementina
23R	GACATKCCAGARTGGATGCATA	57.7	41	22	*Arabidopsis*	AT5G11170.2	95	clementina
24F	AGCTTYTAGCSGACAATGC	59.8	42.2	19	*Arabidopsis*	AT5G26680.2	90	clementina
24R	GTAAATGCTCATGCTAGCATCAA	62.9	39.2	23	*Arabidopsis*	AT5G26680.2	90	clementina
25F	GACAAGGTTCTCATGGARAGC	54.4	47.7	21	*Arabidopsis*	AT5G54160.1	80	sinensis
25R	CCACCWTCAAGMAYTGCATC	69.9	45	20	*Arabidopsis*	AT5G54160.1	80	sinensis

26F	TGGACACTTCGAGGRCTTTG	60.4	50	20	*Arabidopsis*	AT1G67060.1	86	sinensis
26R	ACCCATATKACRGCGAGGA	56.1	47.4	19	*Arabidopsis*	AT1G67060.1	86	sinensis
27F	CTGTAAYCARGACAACCGCGTYAC	57.8	45.9	24	*Arabidopsis*	AT5G21090.1	87	sinensis
27R	AGRTTTGAATTWCCCAAATC	54.8	30	20	*Arabidopsis*	AT5G21090.1	87	sinensis
28F	TGGCTKGGWCARAAYCAGRTTC	54.8	41	22	*Arabidopsis*	AT5G11480.1	86	clementina
28R	CATCWAYTTGTCGCATTTKGTGAA	66.7	33.4	24	*Arabidopsis*	AT5G11480.1	86	clementina
29F	CTGGTTGTTGCTGATGA	56.3	47.1	17	*Arabidopsis*	AT3G54300.2	88	clementina
29R	TCCTTGACYCGTTCGAGA	59.7	50	18	*Arabidopsis*	AT3G54300.2	88	clementina
30F	TCGTYATWGCCTCCCTCGACGTTC	64.7	54.2	24	*Arabidopsis*	AT4G16600.1	83	clementina
30R	CACYACYTTWGCTCCATCTTCYTSTTC	71.8	37.1	27	*Arabidopsis*	AT4G16600.1	83	clementina
31F	GATGCTTTTGAATTCATTGTA	56.8	28.6	21	*Arabidopsis*	ATMG00285.1	94	sinensis
31R	CATGGCAATTAAATCATRAGCCGA	62.4	37.5	24	*Arabidopsis*	ATMG00285.1	94	sinensis
32F	GATCAGGTYCGTGGTGTMATGGA	55.9	47.9	23	*Oryza*	13101.m04144	94	sinensis
32R	CATTTGGCTYTCYCCATA	56.1	38.9	18	*Oryza*	13101.m04144	94	sinensis
33F†	GTCGGCAAYCTCGAYCCCCA	71.1	60	20	*Oryza*	13110.m02788	93	sinensis
33R†	TCCCAWAGTARCTCCTCMGWAA	51.4	41	22	*Oryza*	13110.m02788	93	sinensis

34F*	TCTCCAGAATACCGCAGGCAGCAAC	74.5	56	25	*Arabidopsis*	AT1G03150.1	97	clementina
34R	CACAAAGTARGCYTTATCAA	55.4	30	20	*Arabidopsis*	AT1G03150.1	97	clementina
35F	GTTGGSTGGTAYCAYTCACA	68.5	40	20	*Oryza*	13104.m05825	91	clementina
35R	CAATRCCYGAWARCCAGCATC	51.5	42.9	21	*Oryza*	13104.m05825	91	clementina
36F	ACCGGTGTCAAGAGRYTSTA	62.4	40	20	*Oryza*	13111.m02571	93	clementina
36R	GTGACAGAGTCATTGACATTGA	60.2	41	22	*Oryza*	13111.m02571	93	clementina
37F	TACAAGCTTWYKGGCATCAAG	58.3	38.1	21	*Arabidopsis*	AT2G38110.1	85	sinensis
37R	ACCACAGGRTCKARAACRGTGC	60.1	45.5	22	*Arabidopsis*	AT2G38110.1	85	sinensis
38F	AGGGTYAAGAATCCAGAATGG	55.4	42.9	21	*Arabidopsis*	AT5G13430.1	82	sinensis
38R	GCATTWGGYAARGGRATGCACC	54.4	45.5	22	*Arabidopsis*	AT5G13430.1	82	sinensis
39F†	TCCTYCCYTACRCMTCTGAGC	65.3	47.7	21	*Arabidopsis*	AT5G47030.1	81	sinensis
39R†	GTTGATGCKGGAAYKATRACCA	57.1	36.4	22	*Arabidopsis*	AT5G47030.1	81	sinensis
40F	GAATTYGTGATCTCYAARKTSGATG	53.6	28	25	*Arabidopsis*	AT5G11770.1	88	clementina
40R	CATRGGCCAGATSGAKCCGSKACGA	64.9	48	25	*Arabidopsis*	AT5G11770.1	88	clementina
41F	GAAGAYTCKGTYAGGGTYAAGAA	54.2	34.8	23	*Arabidopsis*	AT5G13430.1	82	clementina
41R	CAGCATTWGGYAARGGR	58	45.9	24	*Arabidopsis*	AT5G13430.1	82	clementina

	ATGCACC							
42F	GCTGAAATYGCTKCTGGAAGTGA	57.7	43.5	23	*Arabidopsis*	AT5G47840.1	81	clementina
42R	TCAGGKACCAAYTGTCCTTTCTCCA	71.5	44	25	*Arabidopsis*	AT5G47840.1	81	clementina
43F†	GAACAAAACTGATCTTGTKGACAA	58.5	33.4	24	*Oryza*	13107.m03172	93	sinensis
43R†	CCAGCYTTRGCACTRGTYTCAA	64.9	41	22	*Oryza*	13107.m03172	93	sinensis
44F†	CCTTCAACAGATACAACAACATGCA	66.7	40	25	*Arabidopsis*	AT3G57670.1	96	sinensis
44R†	TCCATGYCCCCACATATGCA	64.7	50	20	*Arabidopsis*	AT3G57670.1	96	sinensis
45F	GCGAGARAARTCAGCTGAYCCA	59.5	45.5	22	*Oryza*	13102.m04682	95	sinensis
45R	GCAGTCCAYTTRAATATGTTKGAATC	59.4	30.8	26	*Oryza*	13102.m04682	95	sinensis
46F	AGGCAAGTMTCMATAGAGGA	55	40	20	*Oryza*	13107.m03172	92	clementina
46R	CCAGCYTTRGCACTRGTYTCAA	64.9	41	22	*Oryza*	13107.m03172	92	clementina
47F	TGAGACAGGGTGTWCTTGGYATYAA	52.7	40	25	*Oryza*	13103.m04131	92	clementina
47R	GGATKGTTACRAGATCMGGYAGAG	53.9	41.7	24	*Oryza*	13103.m04131	92	clementina
48F	CAGCTGAYCCAGAYATYCARTTA	54	34.8	23	*Oryza*	13102.m04682	95	clementina
48R	GCAGTCCAYTTRAATATGTTKGAATC	59.4	30.8	26	*Oryza*	13102.m04682	95	clementina

Appendix 5

A list of all species, vouchers, and geographic locality for specimens used in this study. ***Bursera fagaroides*** (Kunth) Engl., Weeks 01-X-08-01, Baja California, Mexico, NA, NA; ***B. simaruba*** (L.) Sarg., Pell sn, Florida, USA, NA, NA; ***B. spinescens*** Urb. & Ekman, Weeks 01-VIII-23-01, Domincan Republic, 17°56.196' N, 71°34.633'W; ***B. tecomaca*** (DC.) Standl., Weeks 02-XII-09-01, Mexico, NA, NA; ***Commiphora ankaranensis*** (J.-F. Leroy) Cheek & Rakot., Weeks 10-I-11-02, Antsiranana, Madagascar, 12°18'53.160012"S, 49°20'18.359988"E; ***C. ankaranensis*** (J.-F. Leroy) Cheek & Rakot., Weeks 10-I-12-03, Antsiranana, Madagascar, 12°18'59.339988"S, 49°20'56.040000"E; ***C. ankaranensis*** (J.-F. Leroy) Cheek & Rakot., Weeks 10-I-14-02, Antsiranana, Madagascar, 12°34'49.260000"S, 49°27'30.899988"E; ***C. ankaranensis*** (J.-F. Leroy) Cheek & Rakot., Gostel48, Antsiranana, Madagascar, 12°12'51"S, 49°10'03"E; ***C. ankaranensis*** (J.-F. Leroy) Cheek & Rakot., Weeks 10-I-12-11, Antsiranana, Madagascar, 12°18'59.339988"S, 49°20'56.040000"E; ***C. aprevalii*** Guillaumin, Gostel 124, Tulear, Madagascar, 23°39'18"S, 44°37'45"E; ***C. aprevalii*** Guillaumin, Gostel 128, Tulear, Madagascar, near 23°40'S, near 044°36'E; ***C. aprevalii*** Guillaumin, Weeks 10-I-20-04, Tulear, Madagascar, 22°57'16.139988"S, 44°20'39.480000"E; ***C. aprevalii* var. *granulifera*** Capuron, Gostel 120, Tulear, Madagascar, 23°39'19.64"S, 44°37'44.36"E; ***C. aprevalii* var. *granulifera*** Capuron, Gostel 125, Tulear, Madagascar, 23°39'18"S,

44°37'45"E; ***C. aprevalii* var. *granulifera*** Capuron, Gostel 127, Tulear, Madagascar, 23°40'59"S, 044°38'03.7"E; ***C. arafy*** H. Perrier, Weeks 10-II-12-01, Tulear, Madagascar, 20°24'23.580000"S, 44°50'32.160012"E; ***C. arafy*** H. Perrier, Weeks 10-II-13-01, Tulear, Madagascar, 20°9'54.839988"S, 44°26'44.160000"E; ***C. arafy*** H. Perrier, Weeks 10-II-14-07, Tulear, Madagascar, 19°51'54.000000"S, 44°36'47.000016"E; ***C. brevicalyx* var. *brevicalyx*** H. Perrier, Weeks 10-I-24-05, Tulear, Madagascar, 23°18'10.620000"S, 43°44'52.080000"E; ***C. brevicalyx* var. *vezorum*** Capuron, Weeks 10-I-23-08, Tulear, Madagascar, 23°31'33.240000"S, 43°45'26.399988"E; ***C. brevicalyx* var. *vezorum*** Capuron, Weeks 10-I-20-09, Tulear, Madagascar, 23°19'49.199988"S, 43°55'21.299988"E; ***C. capuronii*** Bardot-Vaucoulon, Weeks 10-I-15-04, Antsiranana, Madagascar, 12°34'49.260000"S, 49°27'30.899988"E; ***C. capuronii*** Bardot-Vaucoulon, Gostel 64, Antsiranana, Madagascar, 12°56'28.5"S, 049°06'55.8"E; ***C. capuronii*** Bardot-Vaucoulon, Gostel 67, Antsiranana, Madagascar, 12°56'28.5"S, 049°06'55.8"E; ***C.* cf. *arafy*** H. Perrier, Gostel 87, Mahajanga, Madagascar, near 16°29'49"S, near45°25'20"E; ***C.* cf. *leandreana*** H. Perrier, Weeks 10-II-14-12, Tulear, Madagascar, 20°4'21.659988"S, 44°40'33.899988"E; ***C.* cf. *leandreana*** H. Perrier, Weeks 10-II-14-14, Tulear, Madagascar, 20°4'21.659988"S, 44°40'33.899988"E; ***C. coleopsis*** H. Perrier, Gostel 142, Tulear, Madagascar, 22°53'10.84"S, 44°41'31.04"E; ***C. falcata*** Capuron, Weeks 10-I-26-03, Tulear, Madagascar, 23°1'28.740000"S, 43°36'59.940000"E; ***C. falcata*** Capuron, Weeks 10-I-27-04, Tulear, Madagascar, 23°1'28.740000"S, 43°36'59.940000"E; ***C. franciscana*** Capuron, Weeks 10-I-22-10, Tulear, Madagascar, 23°24'37.140012"S, 3°24'37.140012"E; ***C. franciscana*** Capuron, Weeks 10-I-23-02, Tulear, Madagascar,

23°29'34.559988"S, 43°45'42.540012"E; ***C. fraxinifolia*** Baker, Gostel 76, Mahajanga, Madagascar, 15°45'56"S, 46°14'30"E; ***C. fraxinifolia*** Baker, Gostel 102, Mahajanga, Madagascar, 16°55'34"S, 46°57'40"E; ***C. fraxinifolia*** Baker, Weeks 10-II-13-05, Tulear, Madagascar, 20°4'21.659988"S, 44°40'33.899988"E; ***C. grandifolia*** Engl., Weeks 10-I-13-01, Antsiranana, Madagascar, 12°34'49.260000"S, 49°27'30.899988"E; ***C. grandifolia*** Engl., Weeks 10-II-14-15, Tulear, Madagascar, 20°4'21.659988"S, 44°40'33.899988"E; ***C. grandifolia*** Engl., Gostel 53, Antsiranana, Madagascar, 12°16"S51"S, 49°11'06"E; ***C. grandifolia*** Engl., Gostel 121, Tulear, Madagascar, 23°39'19.64"S, 44°37'44.36"E; ***C. guillauminii*** H. Perrier, Gostel 100, Mahajanga, Madagascar, 16°20'13"S, 46°50'40"E; ***C. humbertii*** H. Perrier, Weeks 10-I-20-08, Tulear, Madagascar, 23°19'49.199988"S, 43°55'21.299988"E; ***C. humbertii*** H. Perrier, Gostel 136, Tulear, Madagascar, 22°54'32.4"S, 44°31'13.4"E; ***C. humbertii*** H. Perrier, Gostel 146, Tulear, Madagascar, 22°17'44.87"S, 22°18'46.44"ES"; ***C. lamii*** H. Perrier, Weeks 10-I-26-02, Tulear, Madagascar, 23°1'28.740000"S, 43°36'59.940000"E; ***C. lamii*** H. Perrier, Weeks 10-I-26-04, Tulear, Madagascar, 23°1'28.740000"S, 43°36'59.940000"E; ***C. laxecymigera*** H. Perrier, Gostel 115, Tulear, Madagascar, 23°41'17.4"S, 044°34'34.9"E; ***C. mafaidoha*** H. Perrier, Gostel 104, Tulear, Madagascar, 22°24'05"S, 46°07'54"E; ***C. mafaidoha*** H. Perrier, Weeks 10-II-13-02, Tulear, Madagascar, 20°9'54.839988"S, 44°26'44.160000"E; ***C. mahafaliensis*** Capuron, Weeks 10-I-21-02, Tulear, Madagascar, 23°24'41.159988"S, 43°46'48.960012"E; ***C. mahafaliensis*** Capuron, Weeks 10-I-22-02, Tulear, Madagascar, 23°24'37.140012"S, 23°24'37.140012"E; ***C. mahafaliensis*** Capuron, Weeks 10-I-28-03, Tulear, Madagascar, 23°26'50.939988"S, 43°57'29.040012"E; ***C. marchandii*** Engl.,

Weeks 10-II-15-04, Tulear, Madagascar, 20°25'50.580012"S, 45°22'52.320000"E; ***C. marchandii*** Engl., Gostel 95, Mahajanga, Madagascar, 16°19'42"S, 46°47'20"E; ***C. marchandii*** Engl., Gostel 143, Tulear, Madagascar, 22°20'10.4"S, 046°17'23.2"E; ***C. monstruosa*** (H. Perrier) Capuron, Weeks 10-I-21-07, Tulear, Madagascar, 23°24'41.159988"S, 43°46'48.960012"E; ***C. monstruosa*** (H. Perrier) Capuron, Weeks 10-I-21-09, Tulear, Madagascar, 23°24'41.159988"S, 43°46'48.960012"E; ***C. orbicularis*** Engl., Weeks 10-I-23-04, Tulear, Madagascar, 23°29'34.559988"S, 43°45'42.540012"E; ***C. orbicularis*** Engl., Weeks 10-I-26-05, Tulear, Madagascar, 23°1'28.740000"S, 43°36'59.940000"E; ***C. orbicularis*** Engl., Gostel 111, Tulear, Madagascar, 23°32'05"S, 44°19'09"E; ***C. pervilleana*** Engl., Gostel 93, Mahajanga, Madagascar, 16°19'04"S, 46°48'34"E; ***C. pervilleana*** Engl., Gostel 97, Mahajanga, Madagascar, 16°20'13"S, 46°47'11"E; ***C. pterocarpa*** H. Perrier, Weeks 10-I-28-11, Tulear, Madagascar, 23°26'50.939988"S, 43°57'29.040012"E; ***C. pterocarpa*** H. Perrier, Gostel 106, Tulear, Madagascar, 22°24'50"S, 46°01'13"E; ***C. sinuata*** H. Perrier, Weeks 10-I-23-01, Tulear, Madagascar, 23°29'34.559988"S, 43°45'42.540012"E; ***C. sinuata*** H. Perrier, Weeks 10-I-23-06, Tulear, Madagascar, 23°31'33.240000"S, 43°45'26.399988"E; ***C.* sp.**, Weeks 10-II-14-13, Tulear, Madagascar, 20°4'21.659988"S, 44°40'33.899988"E; ***C.* sp.**, Weeks 10-II-15-02, Tulear, Madagascar, 20°24'28.019988"S, 44°46'39.479988"E; ***C.* sp.**, Gostel 69, Antsiranana, Madagascar, near 12°56'28.5"S, near 049°06'55.8"E; ***C.* sp.**, Gostel 54, Antsiranana, Madagascar, 12°35'17"S, 049°26'05"E; ***C.* sp.**, Weeks 10-I-12-01, Antsiranana, Madagascar, 12°18'59.339988"S, 49°20'56.040000"E; ***C.* sp. nov. A**, Weeks 10-I-28-08, Tulear, Madagascar, 23°26'50.939988"S, 43°57'29.040012"E; ***C.* sp. nov. B**,

Gostel 82, Mahajanga, Madagascar, 16°29'49"S, 45°25'20"E; ***C.* sp. nov. C**, Gostel 83, Mahajanga, Madagascar, near 16°29'49"S, near45°25'20"E; ***C.* sp. nov. D**, Gostel 86, Mahajanga, Madagascar, near 16°29'49"S, near45°25'20"E; ***C.* sp. nov. E**, Gostel 140, Tulear, Madagascar, 22°52'43.6"S, 044°41'35.3"E; ***C.* sp. nov. G**, Weeks 10-I-11-01, Antsiranana, Madagascar, 12°18'53.160012"S, 49°20'18.359988"E; ***C.* sp. nov. G**, Weeks 10-I-14-04, Antsiranana, Madagascar, 12°34'49.260000"S, 49°27'30.899988"E; ***C.* sp. nov. G**, Weeks 10-I-14-10, Antsiranana, Madagascar, 12°34'49.260000"S, 49°27'30.899988"E; ***C.* sp. nov. G**, Gostel 62, Antsiranana, Madagascar, 12°56'28.5"S, 049°06'55.8"E; ***C.* sp. nov. H**, Gostel 145, Tulear, Madagascar, 22°22'10.09"S, 46°26'18.15"E; ***C.* sp. nov. H**, Gostel 141, Tulear, Madagascar, 22°52'59.3"S, 044°41'56.5"E; ***C.* sp. nov. I**, Weeks 10-I-09-01, Antsiranana, Madagascar, 12°14'10.619988"S, 49°21'17.579988"E; ***C.* sp. nov. I**, Gostel 43, Antsiranana, Madagascar, 12°45'18"S, 49°14'15"E; ***C.* sp. nov. I**, Gostel 46, Antsiranana, Madagascar, 12°12'51"S, 49°10'03"E; ***C.* sp. nov. I**, Gostel 51, Antsiranana, Madagascar, 12°15'58"S, 49°11'44"E; ***C.* sp. nov. J**, Gostel 44, Antsiranana, Madagascar, 12°45'18"S, 49°14'15"E; ***C.* sp. nov. J**, Gostel 58, Antsiranana, Madagascar, near 12°35'01.8"S, near 049°26'56.3"E; ***C.* sp. nov. L**, Weeks 10-II-14-05, Tulear, Madagascar, 19°51'54.000000"S, 44°36'47.000016"E; ***C.* sp. nov. L**, Weeks 10-II-14-11, Tulear, Madagascar, 20°4'21.659988"S, 44°40'33.899988"E; ***C.* sp. nov. L**, Weeks 10-II-14-06, Mahajanga, Madagascar, 19°51'54.000000"S, 44°36'47.000016"E; ***C.* sp. nov. M**, Weeks 10-I-28-04, Tulear, Madagascar, 23°26'50.939988"S, 43°57'29.040012"E; ***C.* sp. nov. M**, Weeks 10-I-28-10, Tulear, Madagascar, 23°26'50.939988"S, 43°57'29.040012"E; ***C.* sp.**

nov. N, Gostel 47, Antsiranana, Madagascar, 12°12'51"S, 49°10'03"E; ***C.* sp. nov. O** , Weeks 10-I-12-05, Antsiranana, Madagascar, 12°18'59.339988"S, 49°20'56.040000"E; ***C. tetramera*** Engl., Gostel 79, Mahajanga, Madagascar, 15°53'S, 46°06'E; ***C. tetramera*** Engl., Gostel 90, Mahajanga, Madagascar, 15°45'18"S, 46°14'39"E.

Appendix 6

Chapter 4 Supplemental Table 1. Target specific primer sequences for each locus used in this study. Internal primer sequences are provided only for loci in which we designed and sequenced internal reverse primers.

Locus ID	Forward (5'–3')	Reverse (5'–3')	Internal primer (5'–3')
AT1G18060	AACAAGAAAGGTTGCAGTAGAGGA	GCTTGGCTCTCTGTCATCTTTTTG	NA
AT1G21840	TGTTGGAGAAGTTGAAGAGAGAGG	CACCATTTATCCCAACCTCCTGAA	NA
AT1G31780	CTTGTCCTTGGGTTACTTGATCCA	TTGTGGGTCTCAATGATTTCAAGC	GGACCCAAAGTGTACTACAGAGAG
AT1G59990	GCTACTTGGTTCCTTTAATTGATAAGC	TGACACCACGAATAAAATCCAAGC	NA
AT1G63160	GACGCTGTATCTAGGCTCCAAG	CACCATGAGACGGCCAAGTAT	AAAATGTTGCATGTGAAGTTTGGC
AT1G65030	CGGTTTTCTGTAACTCGGTACAG	CGGGGGAAAGAGAGGTTTTGG	NA
AT1G65070	CCTAATACTGGAGGGAAAACTGCT	CAGTACTTCCCCAGAGAATTCGAA	NA
AT1G66080	CCTCTTCTCTTCCATAGTGTTGCT	CCCACAAAACGACTGCATAAAGTT	NA
AT1G73180	AACTCCTGCCAGTGTCCAAATATA	AGAATGCCATATCACCAGGTAAGT	NA
AT1G73740	TTGATATTGGGAGGCTCTTTGGG	CACCAGCTCTTGAAACAACGAG	NA
AT1G76450	TGGTCCGGATTTTACAAGAATGGA	CGTCCGCACAAAAATCAAAATAGG	NA
AT1G77550	TGTGAGCTTTTCTATATTGTGGCC	TGATGCTTCATGACCAGACAAGA	NA
AT1G77550B	ACAGTTCAGAAGGCACATGGATC	CTGGTGTGTTCATGTGATTGAGTC	NA
AT1G77930A	ACCCTAATTCTGTTCTGCGATTTG	GAGCAGTTCATAAGCAGCTTGAAT	GCATCCCTCTCTAACTCTGCAATT
AT2G03667	CTAGTTGGCTCTGGTGCTGATG	CACAAAGGAATATCAAGCAAAGTCCT	NA
AT2G04620	TCCACCATATTTTGAGTGAGAGGAA	AATGGGAGTGGGAATGAGAATGTG	NA
AT2G04740	CAAACTCCAAAAACCCTAAACCGG	TCAAAAGCCTTCAAAAGCTTCCTC	NA

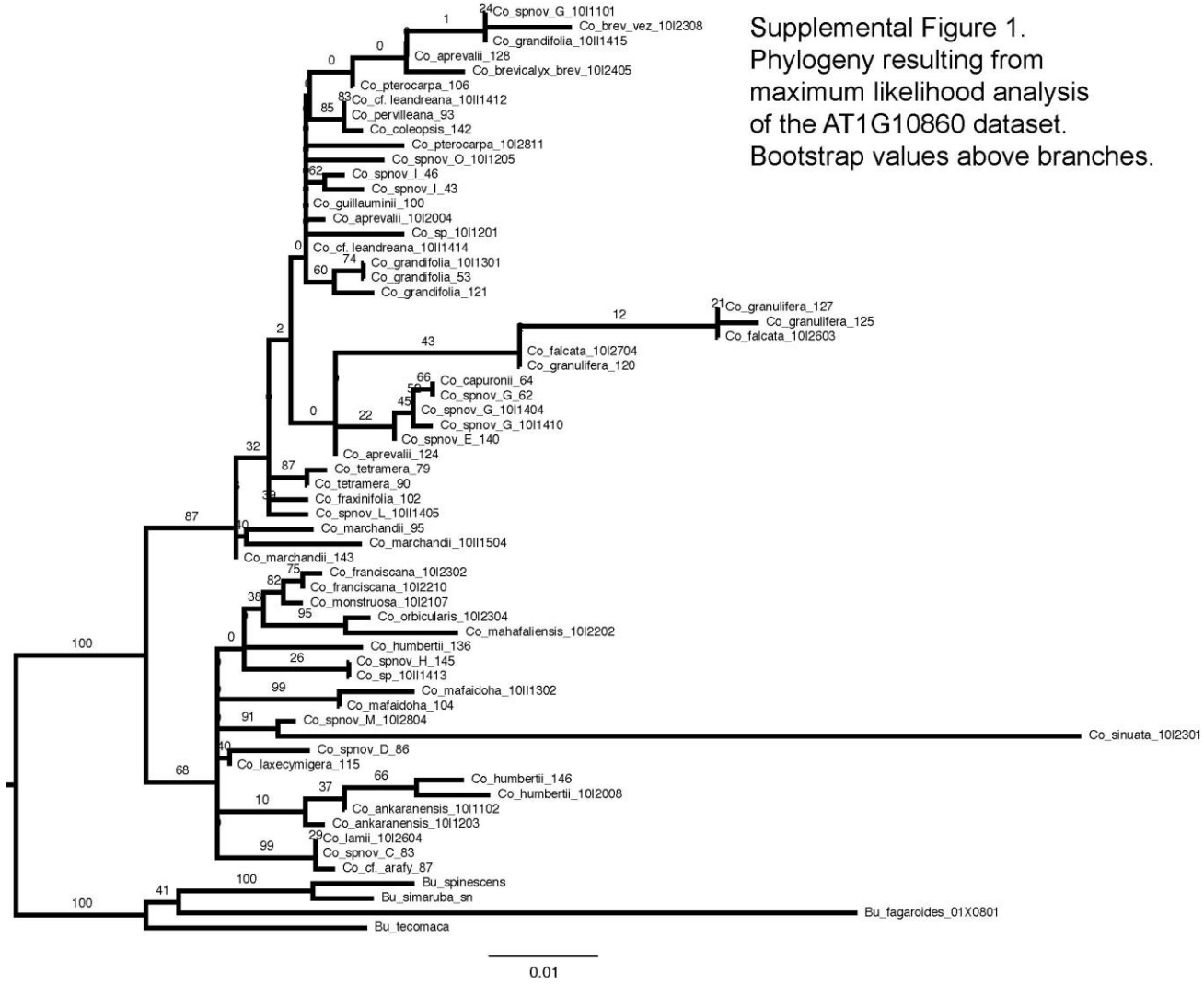
AT2G05120	TGTCAAAGCTCTGGTCTCATGAAA	CGAAGGAAGAACTGAAGCATCTAG	NA
AT2G05170A	GAAAGGAAATGTACCAGTGGAGGA	TGAGAAGAATTGGTGGAGCTTCTT	NA
AT2G05170B	GCACAGTACATTAACACCATTGGT	TGGCTTGTGGTCTATGAGAATCTT	NA
AT2G05320A	TGCCAAGTGAAACAGATATTTGCT	TCTCCAAACAGTCTGGTTAAAGGA	NA
AT2G17265	TTATGGGAGGTTTCGTTTTGATTCG	CTAGCACCAACTCTATCCAACCTC	NA
AT2G20330	TCATTGAAGGTTTGGGATTTACGC	ACGACTTGGCTGATCTCTGAATAA	NA
AT2G20790	CCAATTGTCAATGGTCTCTGAAGATG	CCATGGTGCAAATTTAACTGTTCC	NA
AT2G21710	TTTCCTCCTTTACTAACATACAGCCT	CTTGTCTGCAACCTTCTGATTGAA	GCTGCATCCCAAGAGCTCTGG
AT2G22370A	ATGTTGAGGCCCTTGAGATTCTTC	TAGGTGCTGTTACTTCAACCAGTT	NA
AT2G22370B	AACCCACATGGACTGTTAAACATG	CATCAGACATAAGAGATGCAGCAG	NA
AT2G25570	GACAACTCAAATACACATGCCAGG	GTCCCTTCTCTGATGCCCTATG	NA
AT2G27760	GAACCTTAAACCCTAACAATGGAGAA	GGCGGTTCCGTGACCATAT	CGAAATTCCTTAGCAGTGAACTCC
AT2G28450	TTTCTGAGATCATGCTTATAGTTCAGG	CGCCCAATTGTTCCAGTACCA	NA
AT2G31040	AAGTACTGGGGTGGAGAAAAAGAG	CCAAGTGTGAGGATTTGCAACTTC	NA
AT2G31440	GTATGGAGGGCTTTTCTTCCTTTG	ATTCCTGCAGCAAGATGAACTACA	NA
AT2G31840	AGTGATTGATTGGTGTCCTGATGT	CATCTTGGTGAAGGTAGCCTACAG	NA
AT2G31890A	AGATTGGAGGGGAGCTACTTTATT	CCTCCCTATACTGCTCTGAAATCC	NA
AT2G31890B	CTCTCCAGTGCTCAGTTTTAACAG	CTTGAGAAATGTGTTGGTCCATCA	NA
AT2G36740	AGTCCACAAGAACTGCAGTGAT	CATCCTTTGAGAAATACCGTATCTGT	NA
AT2G40760	GGTGTATCATCTGGAAGGGGG	CGCTCTCGCCCTCTCTTTC	NA
AT2G44660A	ATCGTATCACAGCACAGACTTTGA	GCAAAAACAAAACCACCCATCAAA	NA
AT2G44660B	GTTTTTGCAGGAAGGGATGGATTT	TGAAGGTTGTTGCTGGAGTTATCT	TGCCTGAATCTTGAACCCTAGTTT
AT2G44760	CAGCATGGAATACGTTTGCTAGTA	TATCAACTGGACCCCTGGAATAAG	NA
AT2G46100	TTTAAAGGACTTCGCCGTTTCAAA	GGCAGAAAGAATAGGCCTCCAG	NA
AT2G46890B	TTCTTTGCTGTCTACCTCTCTCAG	CGATGTCGTCTTCTGATATAGCCT	NA
AT2G47760	ATAGATGCTCTTCTTGTCGCTCTT	CGCACTGAAAACAATCAAGCCTAA	NA
AT3G01380	AATCATCATAATAGGGGCAGCCG	CCAAGAAATATAGAAGTTAGTCGGGAC	NA

AT3G04650	CAAATCGCTTGCTTGGTTCATCA	CTGTGGCAGTTGGGATGTTTTC	NA
AT3G07750	GCTATATTTGTTGATTGCAGCCCT	TGGTTGCTCATCGTTTAATGCATC	NA
AT3G10400	AGAAGAAAAAGACTAACAGTGAC AGC	CGGTCTTTGAGCACGCTGA	NA
AT3G10530	GGTCTTCTTGCATAATGAGCTGTT	TCAAATTTCCGCAAGTCCCAAATC	NA
AT3G13200	AACTCATCGGCTTTTTCCTCTCT	GAATCATCAGCATCTACATTGCGT	NA
AT3G14910	GGAGCTATTTTATCAAAAGTTGTGC C	AAAGCAATATACGACCAAGAGAAT CTG	NA
AT3G15110	CTCACTGGTGCCATATCTGTCTTC	ATTCTCTGTACCTTTGCTTCTGGA	NA
AT3G15290	GATGTTGTAGTCGAGGCTATTGTG	ATCTGCAAGTTCTAAAGGACCCAT	NA
AT3G17170	GATGATGAACTATTTTTCCCTGAGG C	TCTTGAACTTTCTCATTCACACTGC	NA
AT3G21540	GTTGCTATTAGTCCTGATGCCAAA	AATGGTTCTTCAGTACGATCCCAA	NA
AT3G22660	AGATGAAGATGTGAAATTGGTTGA ACC	TTTCTGCTTAGCTCTCTCTTTCATCT	NA
AT3G22990	TTCTAGGTCCATCTCTTCAAGTGC	TTATGATATTTGAGGCAGCAACGG	NA
AT3G26580	AGGTGAACGGTGTGGATTATGATG	GTGACGGTTATTTGCCTCGTAAG	NA
AT3G29130	TTTGCCGAGGTTTCTGGTGATT	AAGTACTTCTCTTGTTGATTCATCC G	NA
AT3G46220	CAATTGAGGAATGAAATGGTGGCT	TCCATTTCTTGCAAAAGCTTCTGT	NA
AT3G49730	CCGAAACTGGAGATGGCTTTG	AATCAACTCAGGCCTTTCTTTTCTC	NA
AT3G54460	GGACACACCCTTGGCTCTAG	CTCCCATGACTTTTGGTTCTGTC	NA
AT3G61620	ATTTCCCCATTCAAATTCCACTCG	ACCTCATGCGATCTCCAGTACTAA	NA
AT4G00560	CTGCTATGTCTATAAACGTTCCTTC C	GTCCACCAACATTCAACAGTAACT	NA
AT4G04955	GAACAGATACGGTACAGAGCCAG	TGCAGCTTTAGTCCCTGAAGG	NA
AT4G14605	CTTCTCACTCATAGCAGGCAGAAG	CTTCTTCACAGCCTTATCAAAGTCA	NA
AT4G18810B	CCTAAAAGGTGATGGTCGAAGGTA	TCTCCCTTCAACTTGAATGTGAGA	NA
AT4G19900	GTTCTCTGAGACGATTGAGCTTGA	CTTGTAGAGAGCAGCAAGTCGG	NA
AT4G21770	TGGAGCTGTTTATTATGCCCTTGT	TAGTCCTAGCTAACACAACACAGC	NA
AT4G26980	CTGCTAGTGGTGTTTCTGAATTGG	ACTTCTCAGCCATTGACAACTCAT	NA
AT4G29590	GAGCAATTCCCCTTCAAAGAGGA	GTGCTTGTATCCTTTTGGGTAATGG	NA
AT4G31770	GCGGTGAGAAATGAGAATGACATG	AACAAGTTCCTTCCAATTCCCAAA	NA
AT4G31790	ATTTGGTTGTTCGAGCCAAGAAAA	GTCCAAAATCAACCATCTGCAGTT	NA

AT4G33030	GATGGTGTCTTTGGTACTGCTTTG	CCAAGAAACAGTGGGCATTATCTG	NA
AT4G37510	TTCATTTTGAGACCTCCATTAGATGAC	GCTTAGCCGGATTATCGTCTCC	NA
AT5G02250	CACTTATCCCATGTTTCCAGAGAAC	GGATCTGCCTGGTTTTCAACATAT	NA
AT5G04910	TAAGAGTCCAACAGAGCATGAGTG	TAAAAGAATGATGTCACTCAGCTTCG	NA
AT5G10460	TGGTCATCATTAGCAATTCTTCACG	GCTTCTTCAACATTCTCCACAACT	NA
AT5G11980	TTCAACCATGCATCCCAAATTACC	GACAGAGATCCGCCTTCAGTTATC	NA
AT5G14580	TATACGTATTGGCAGAATTTCCGG	TCTTGTGCAATCTTATCTAAGGCCT	NA
AT5G15680A	TTCTCATTCAAACATCATTGGGCC	GAGGAATTGCATCAGATTCTCGTC	NA
AT5G16690	TGCTTCCTCTAGACAATTGTTCACT	TCCAAAGAAGCCGCTATGAATTG	NA
AT5G48790	GAGGATTTTGGTTTCACTGAAAAGG	TCGCGACCTTTAAAATTGTGAATGT	NA
AT5G52180	CTCCGAGAATTTGGTTGGAAATGT	CATACAGAAAGCCGCGTCGATA	NA
AT5G57655	TTGGTTATGCTCAGTGTAATCCGA	CTACAGTGCAGATTGGAAAAGCAT	NA
AT5G67220	CGGTTAAAAATGCTCTCAGGATCC	CATCTGCCGAATGAGTAACCTTCT	NA

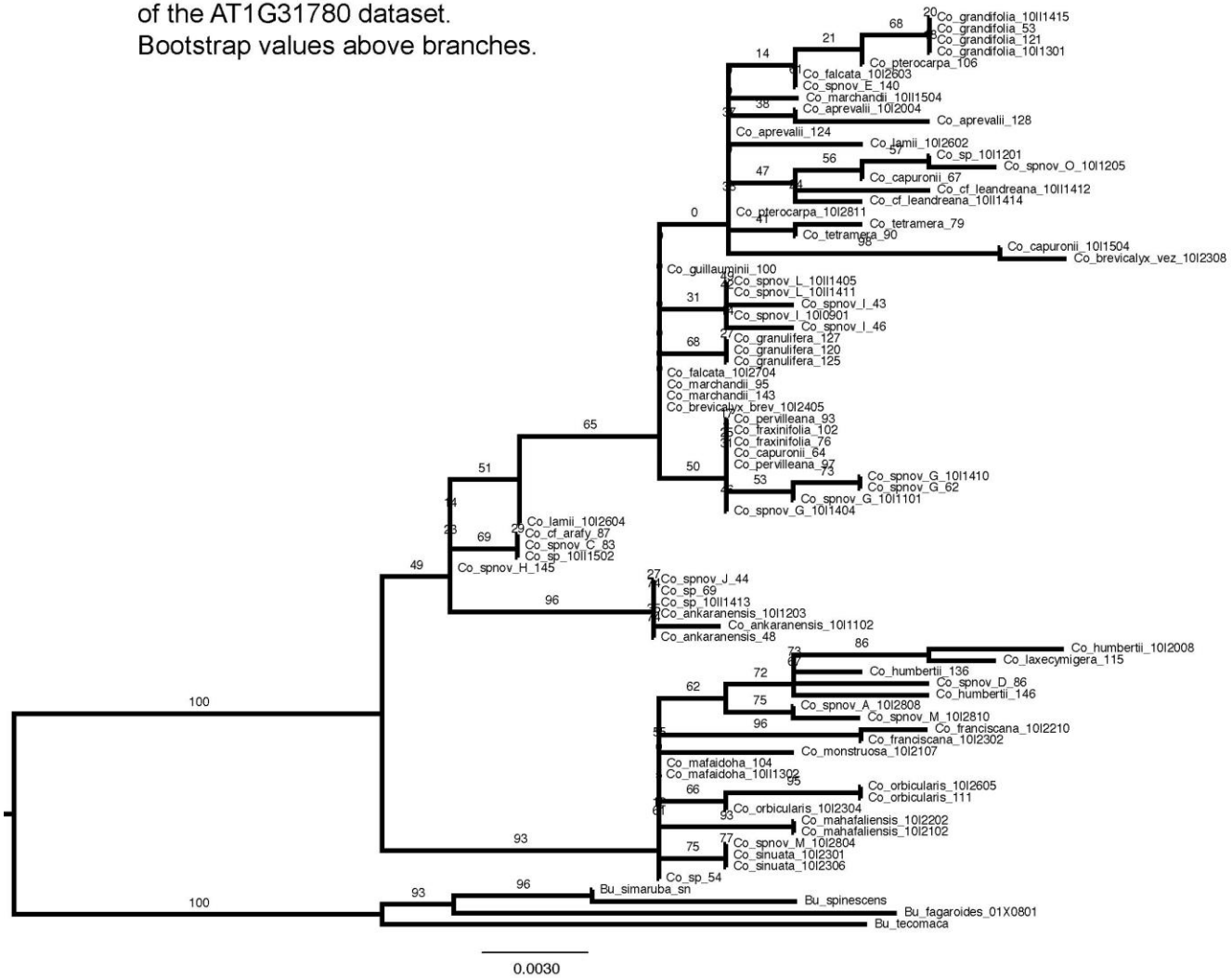
Appendix 7

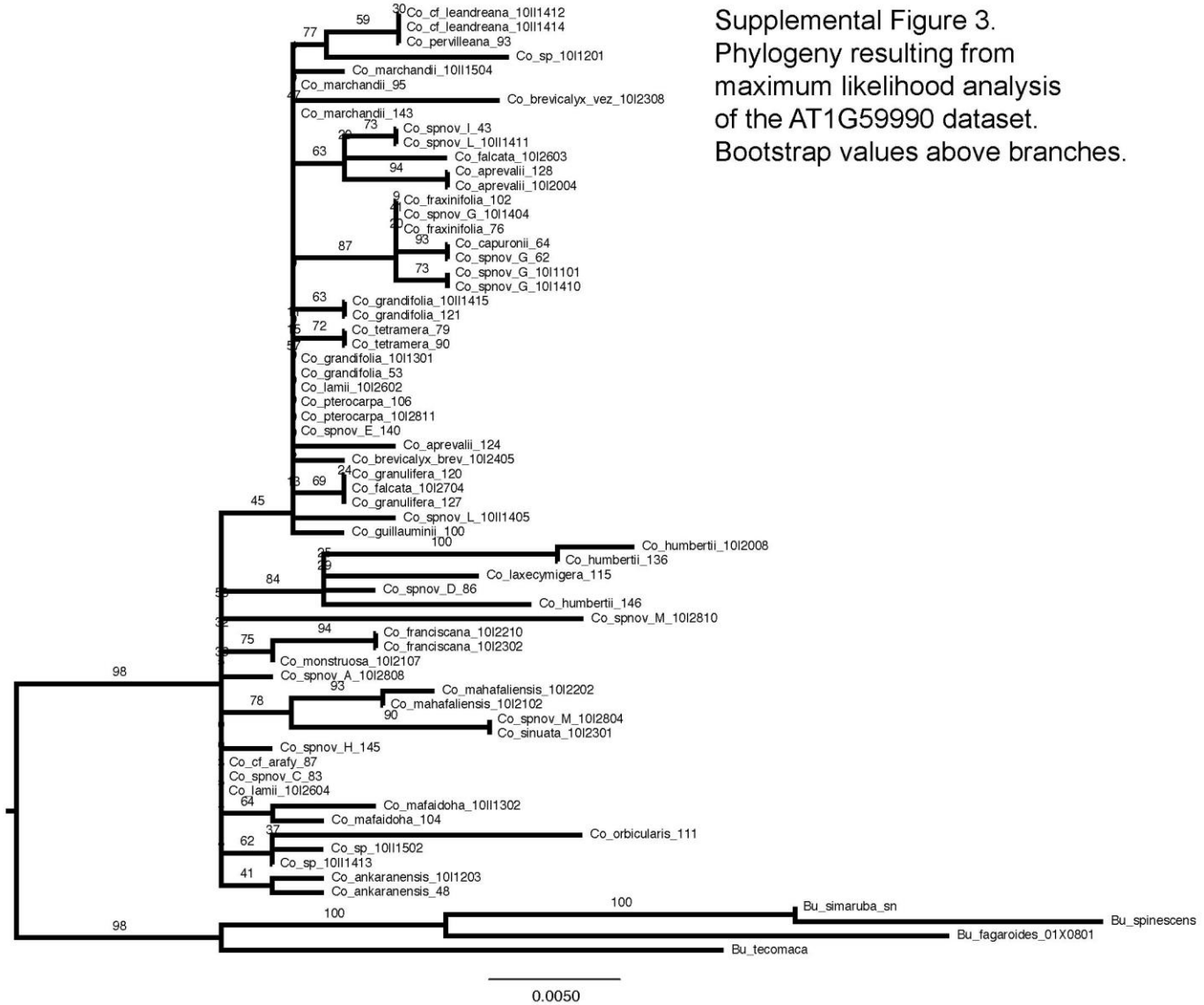
Supplemental figures 1–49 for maximum likelihood analyses of individual data matrices used in this study.



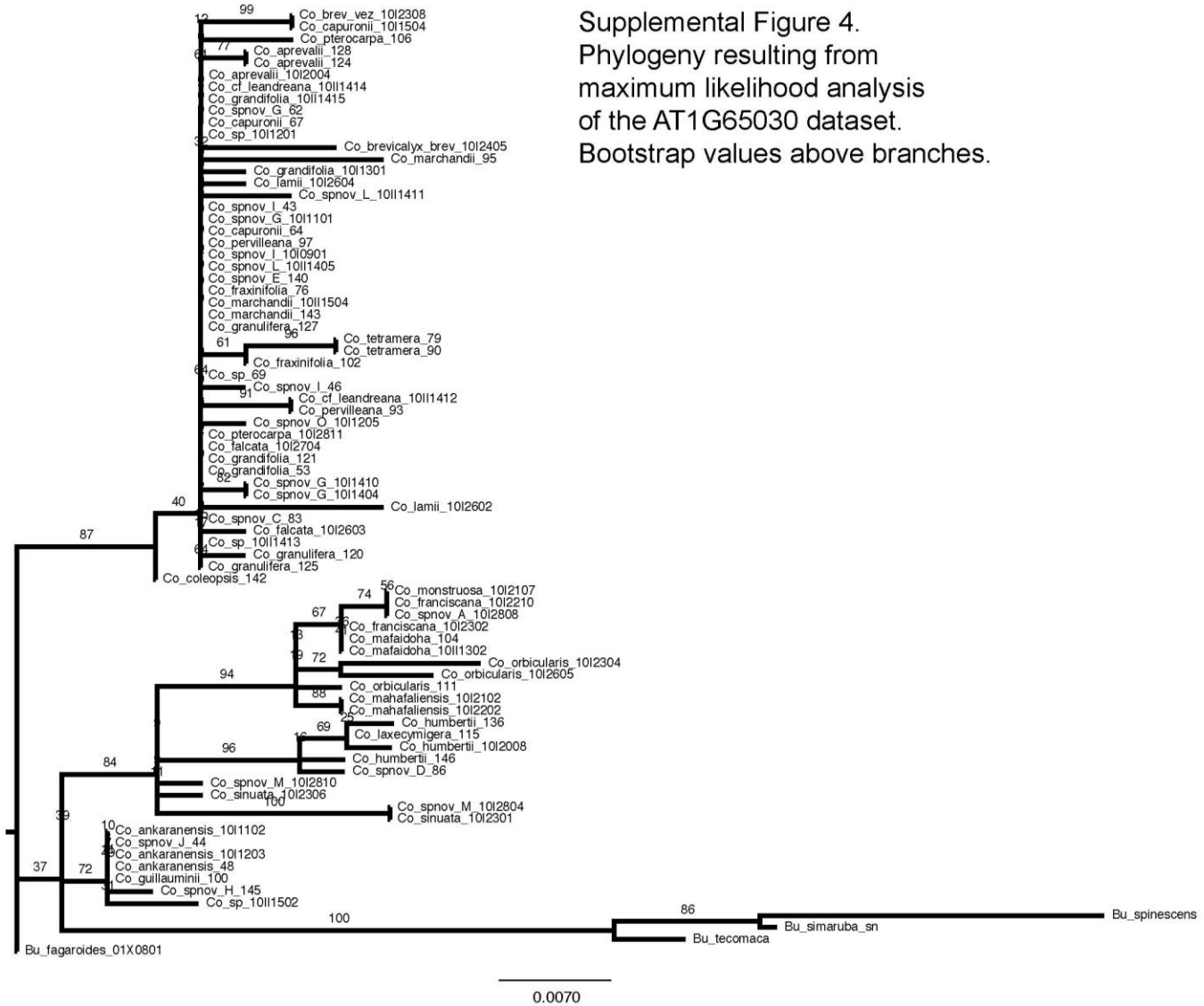
Supplemental Figure 1. Phylogeny resulting from maximum likelihood analysis of the AT1G10860 dataset. Bootstrap values above branches.

Supplemental Figure 2.
Phylogeny resulting from
maximum likelihood analysis
of the AT1G31780 dataset.
Bootstrap values above branches.



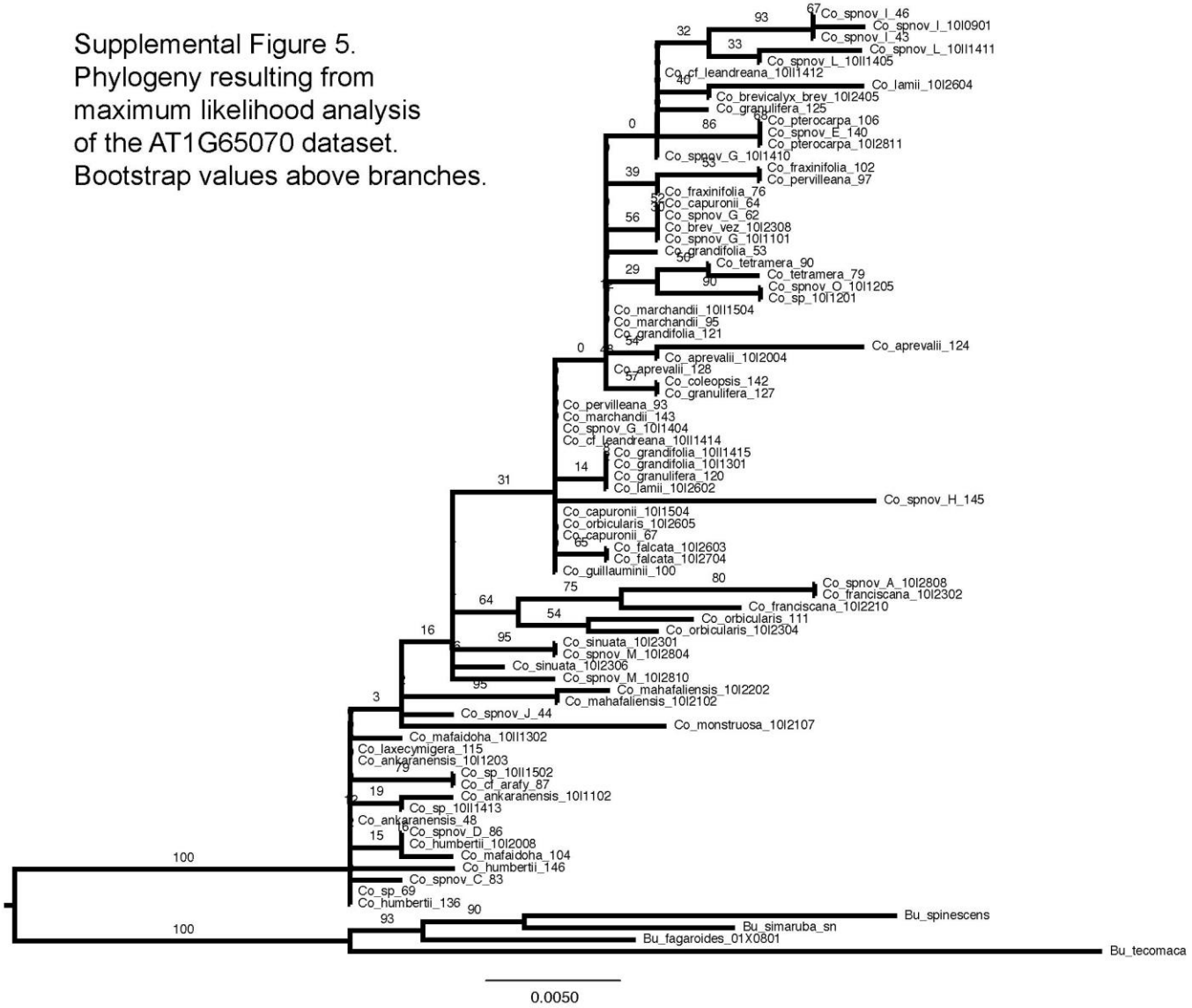


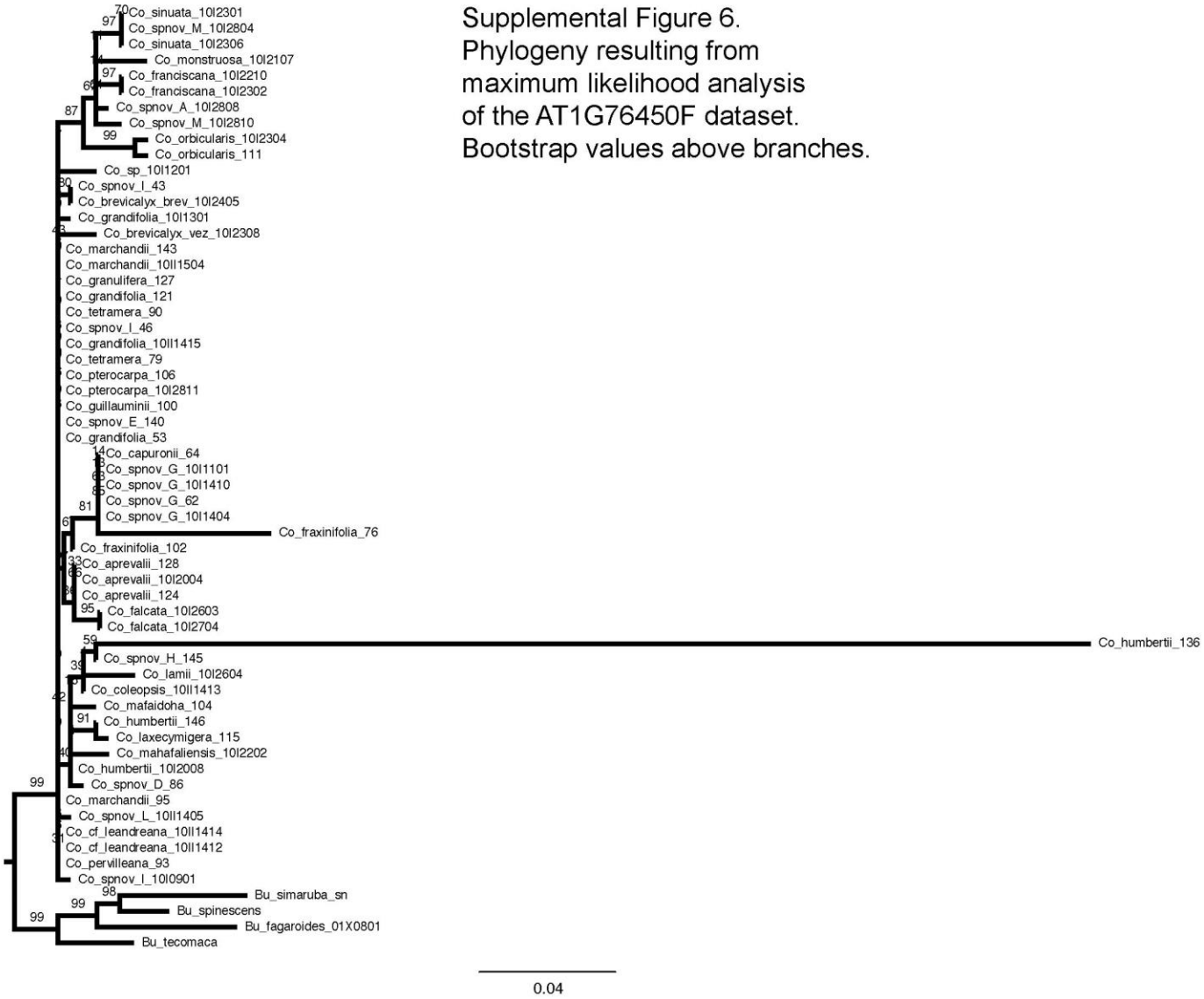
Supplemental Figure 3.
Phylogeny resulting from maximum likelihood analysis of the AT1G59990 dataset.
Bootstrap values above branches.



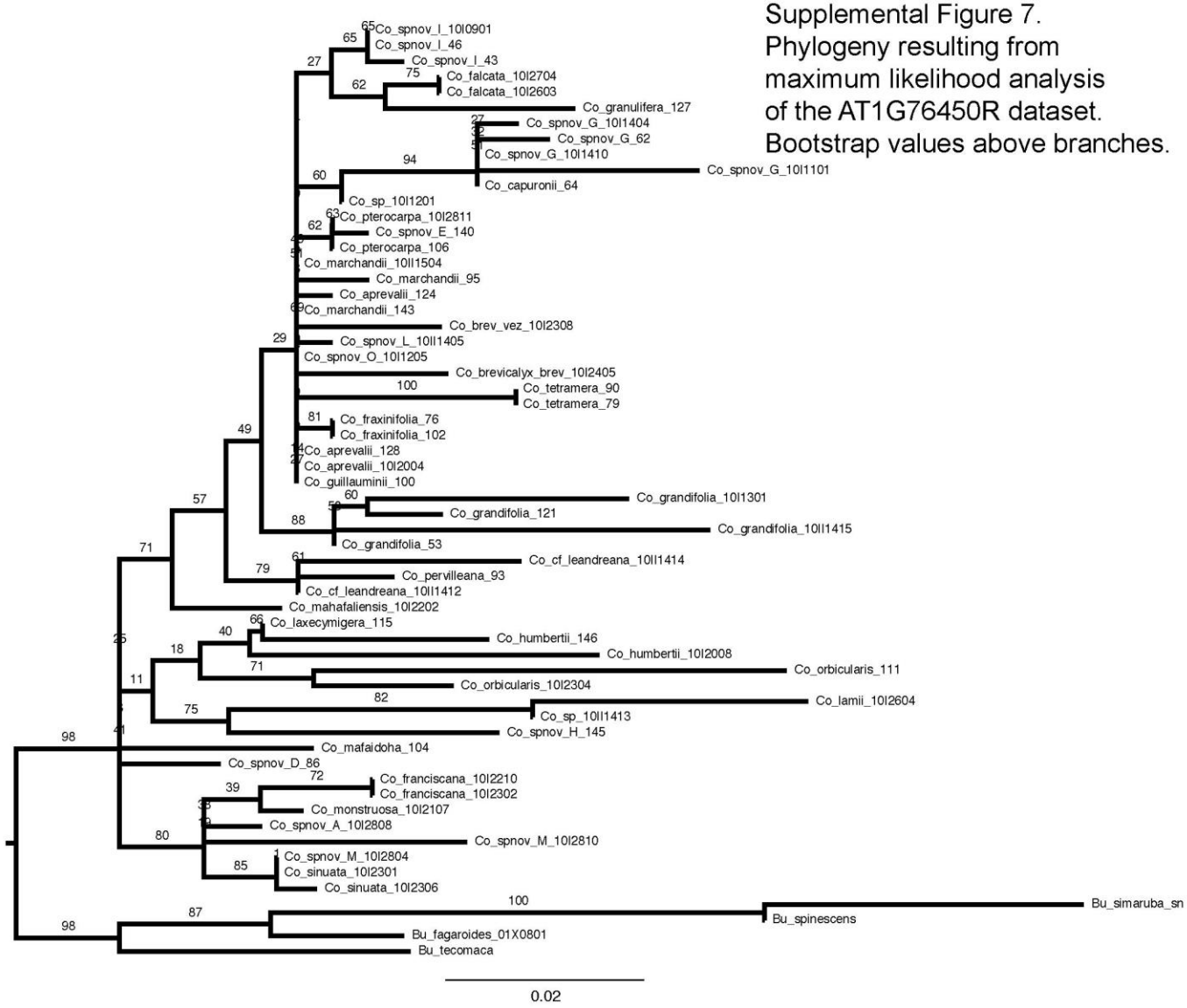
Supplemental Figure 4.
Phylogeny resulting from maximum likelihood analysis of the AT1G65030 dataset.
Bootstrap values above branches.

Supplemental Figure 5.
Phylogeny resulting from
maximum likelihood analysis
of the AT1G65070 dataset.
Bootstrap values above branches.



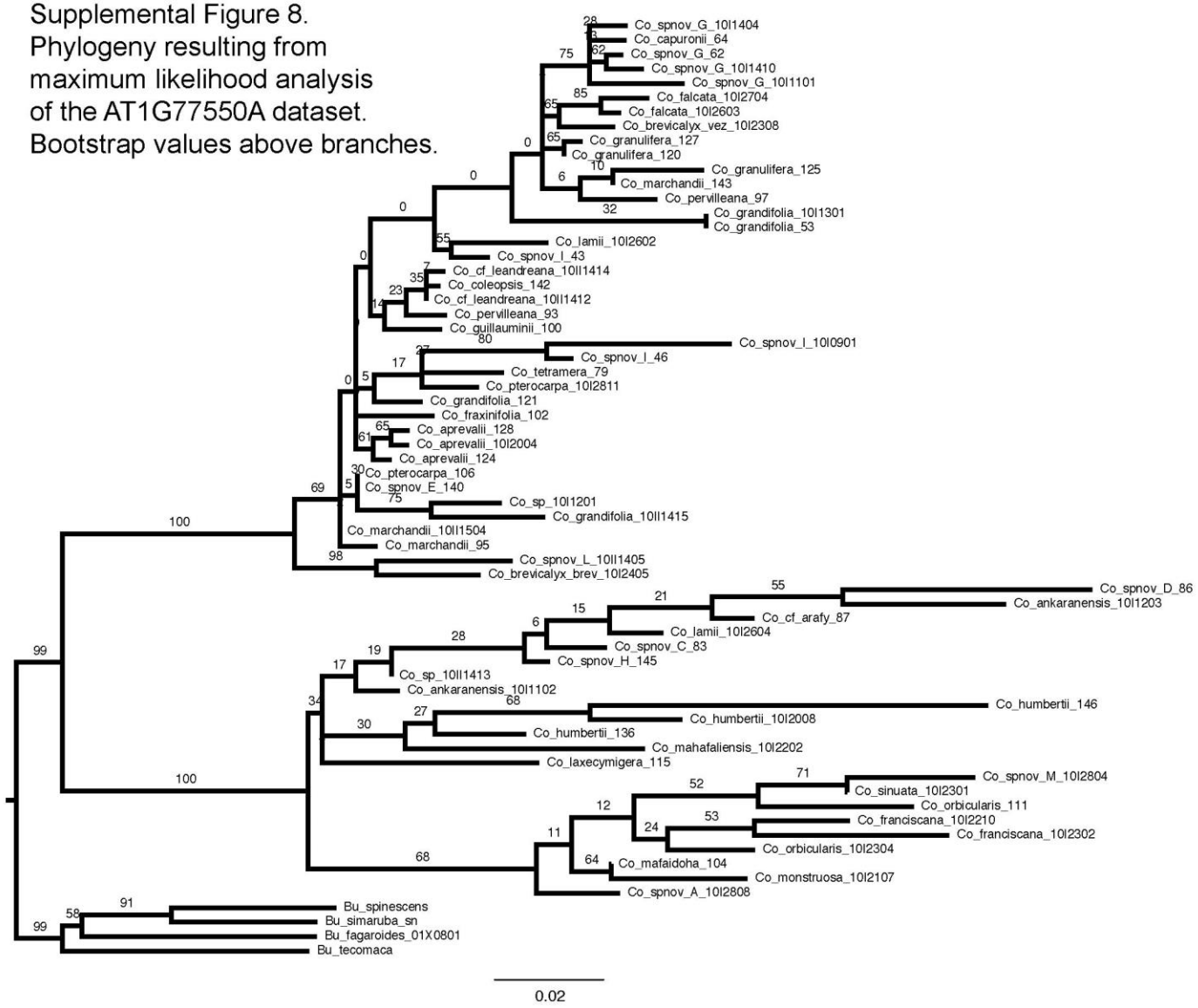


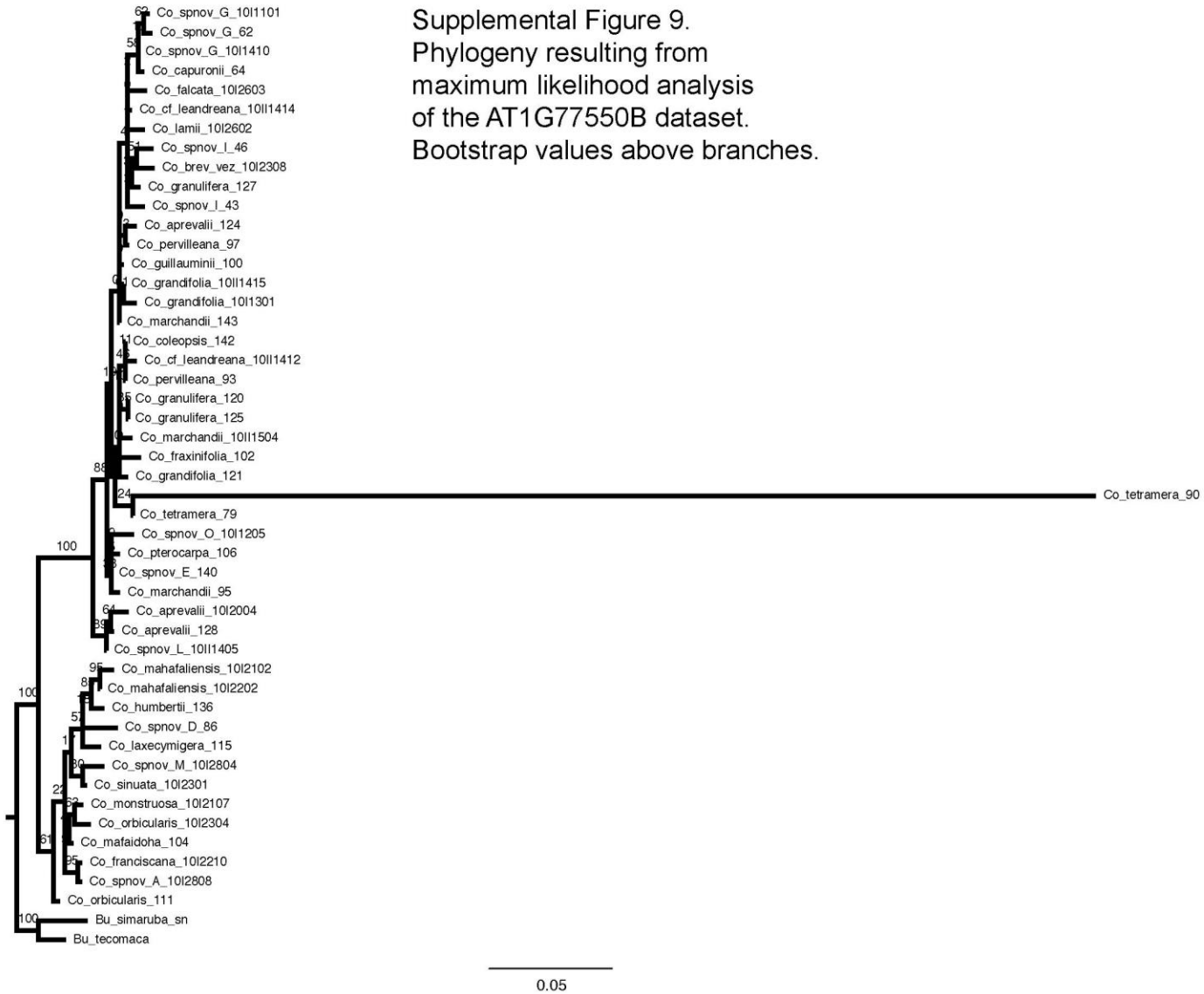
Supplemental Figure 6.
Phylogeny resulting from
maximum likelihood analysis
of the AT1G76450F dataset.
Bootstrap values above branches.



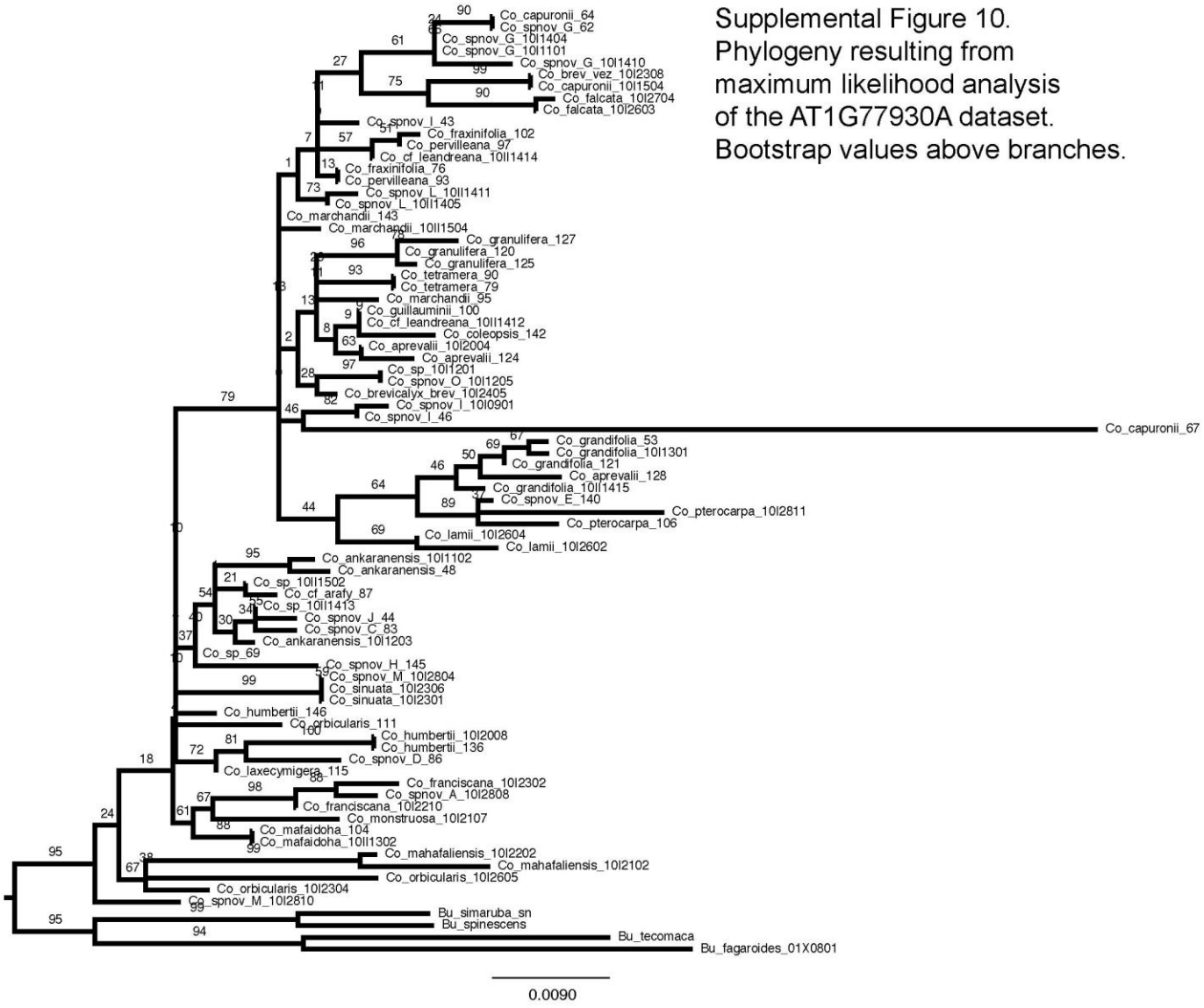
Supplemental Figure 7. Phylogeny resulting from maximum likelihood analysis of the AT1G76450R dataset. Bootstrap values above branches.

Supplemental Figure 8.
Phylogeny resulting from
maximum likelihood analysis
of the AT1G77550A dataset.
Bootstrap values above branches.



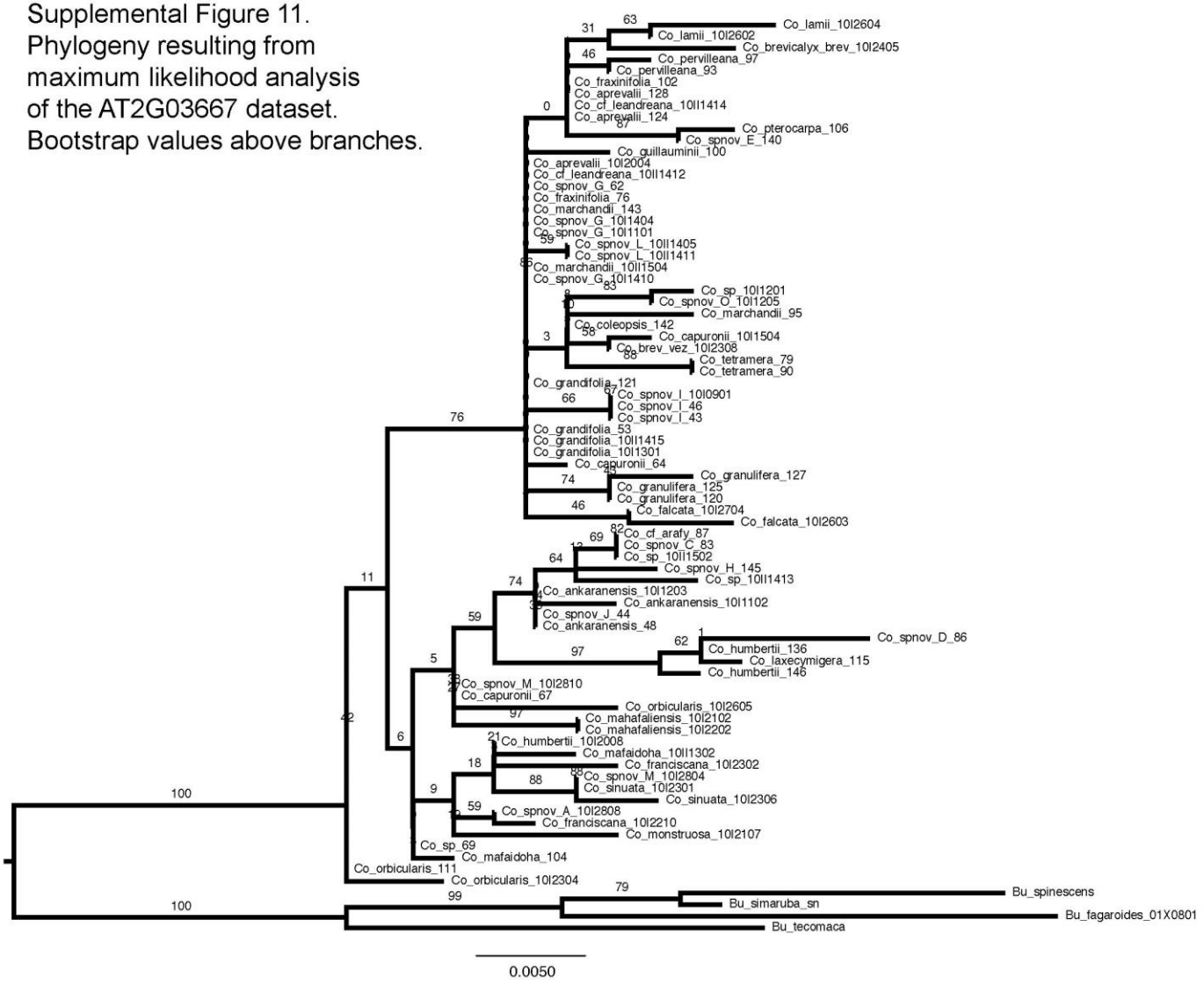


Supplemental Figure 9.
Phylogeny resulting from maximum likelihood analysis of the AT1G77550B dataset.
Bootstrap values above branches.

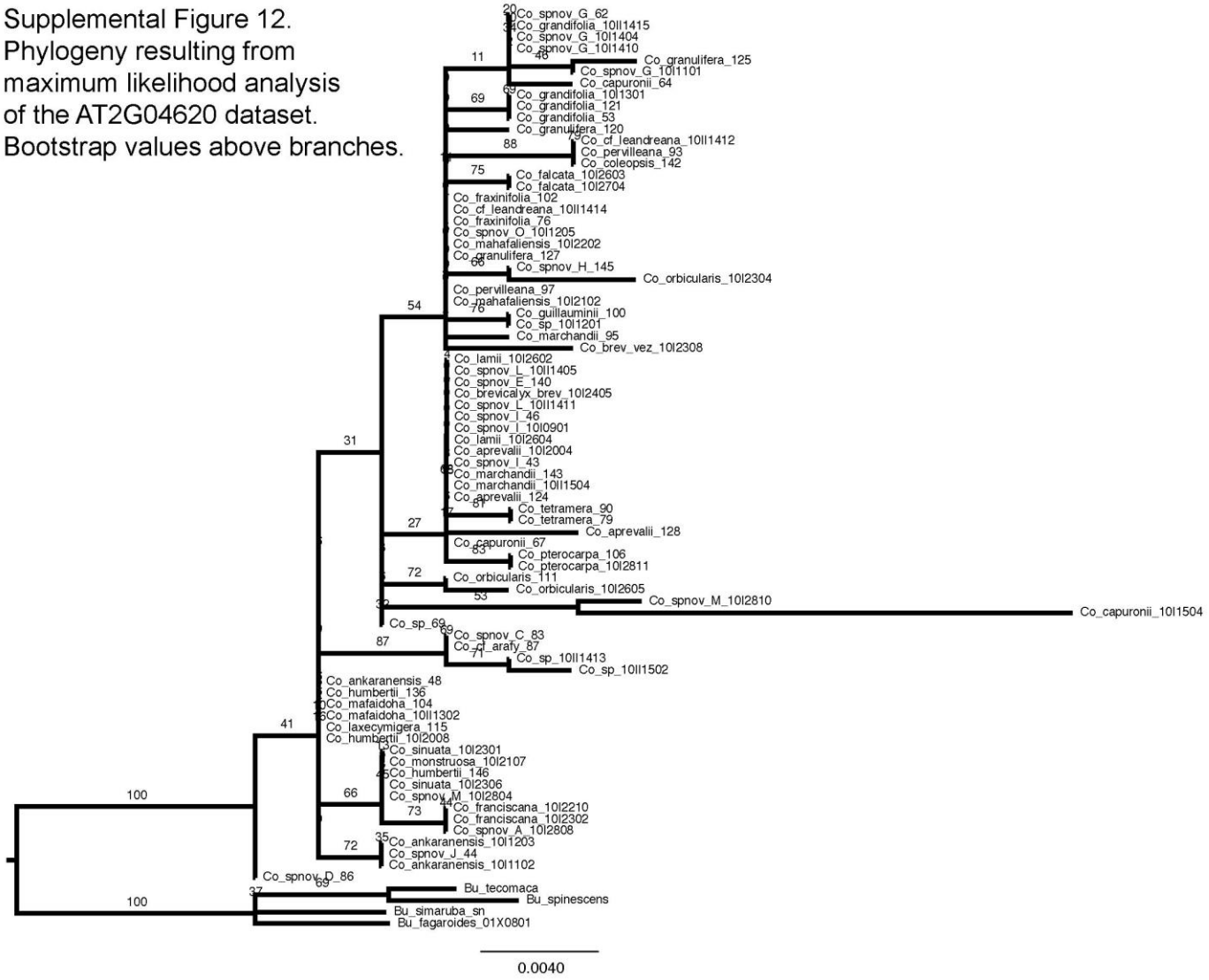


Supplemental Figure 10. Phylogeny resulting from maximum likelihood analysis of the AT1G77930A dataset. Bootstrap values above branches.

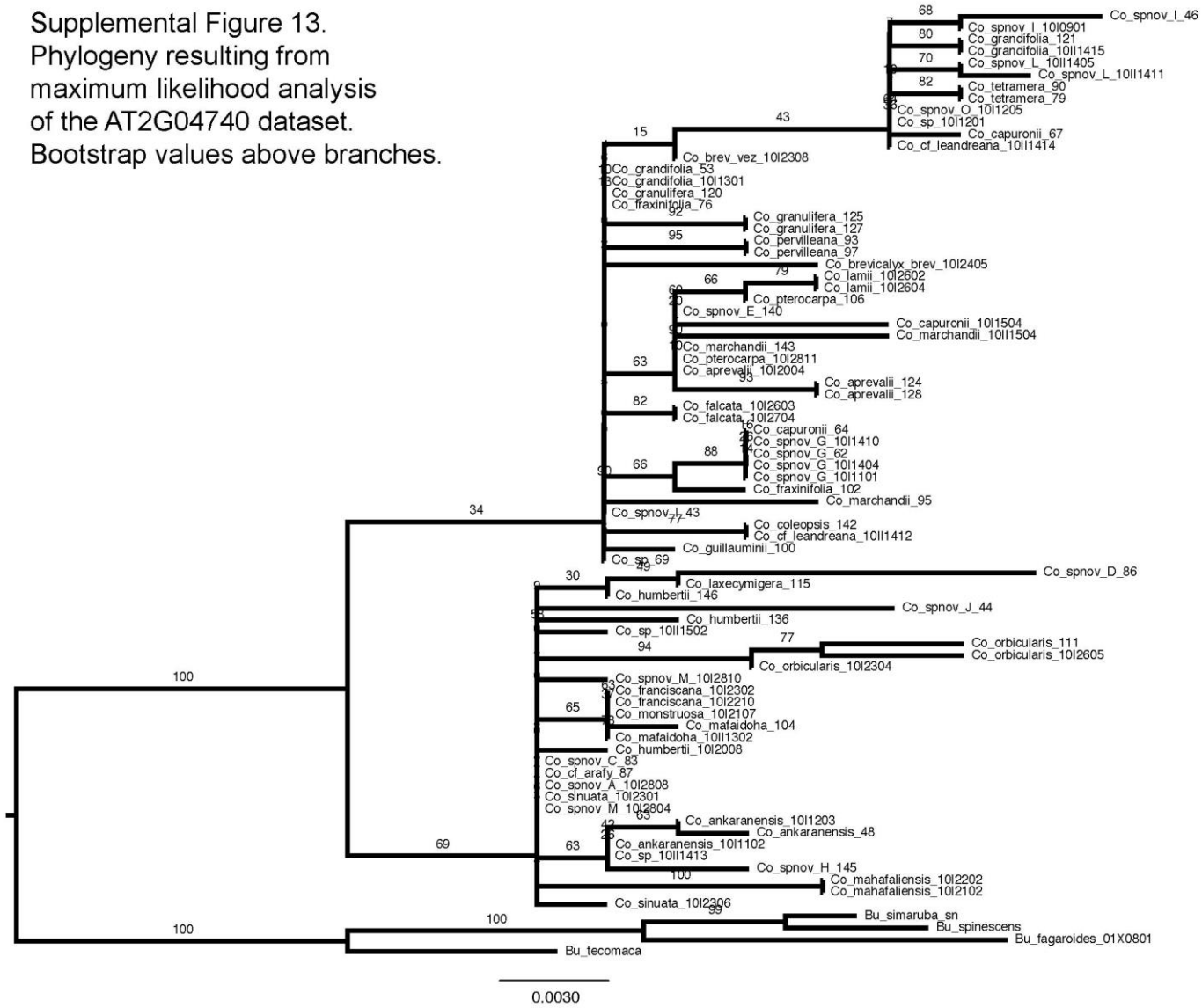
Supplemental Figure 11.
Phylogeny resulting from
maximum likelihood analysis
of the AT2G03667 dataset.
Bootstrap values above branches.



Supplemental Figure 12.
Phylogeny resulting from
maximum likelihood analysis
of the AT2G04620 dataset.
Bootstrap values above branches.

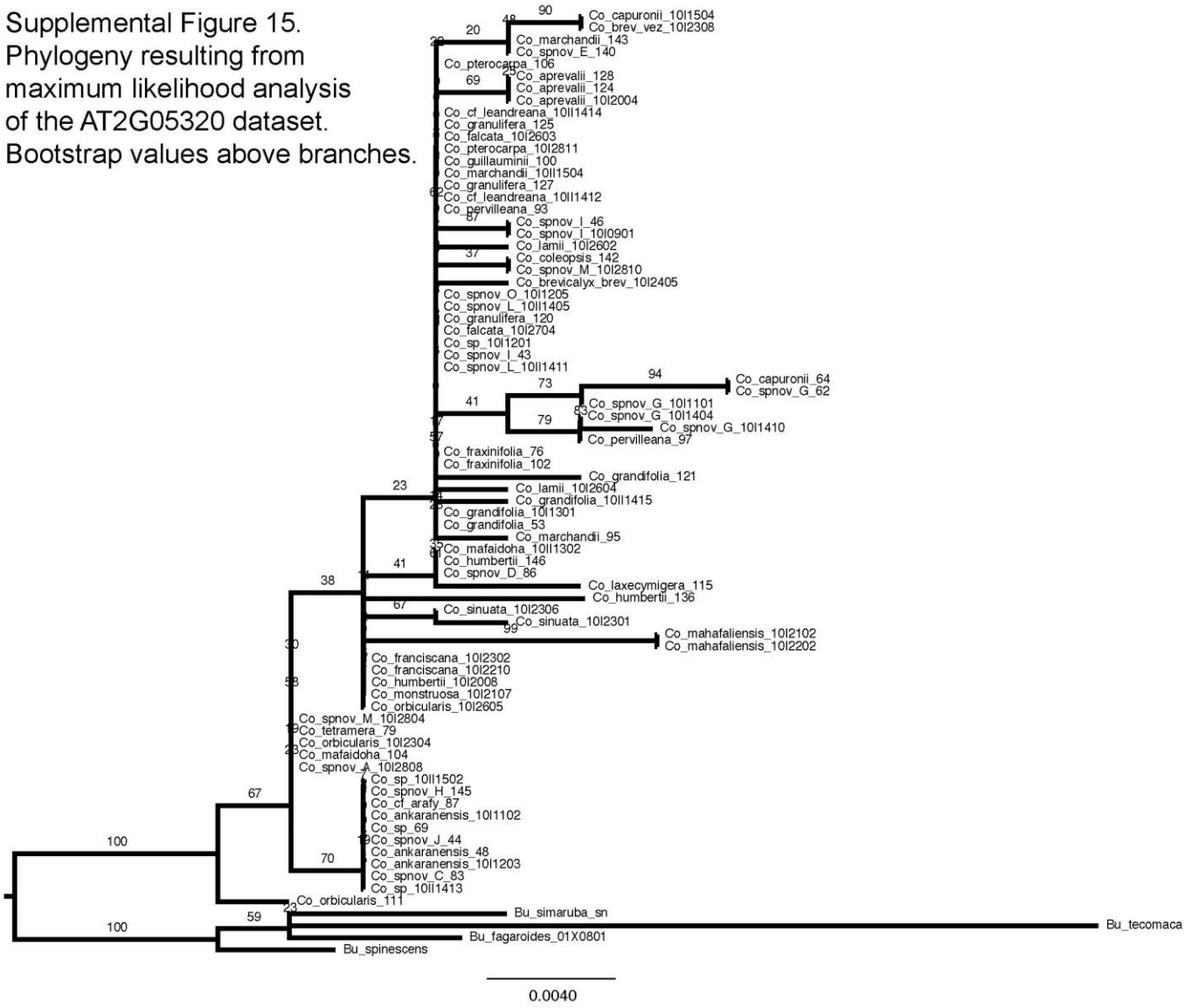


Supplemental Figure 13.
Phylogeny resulting from
maximum likelihood analysis
of the AT2G04740 dataset.
Bootstrap values above branches.



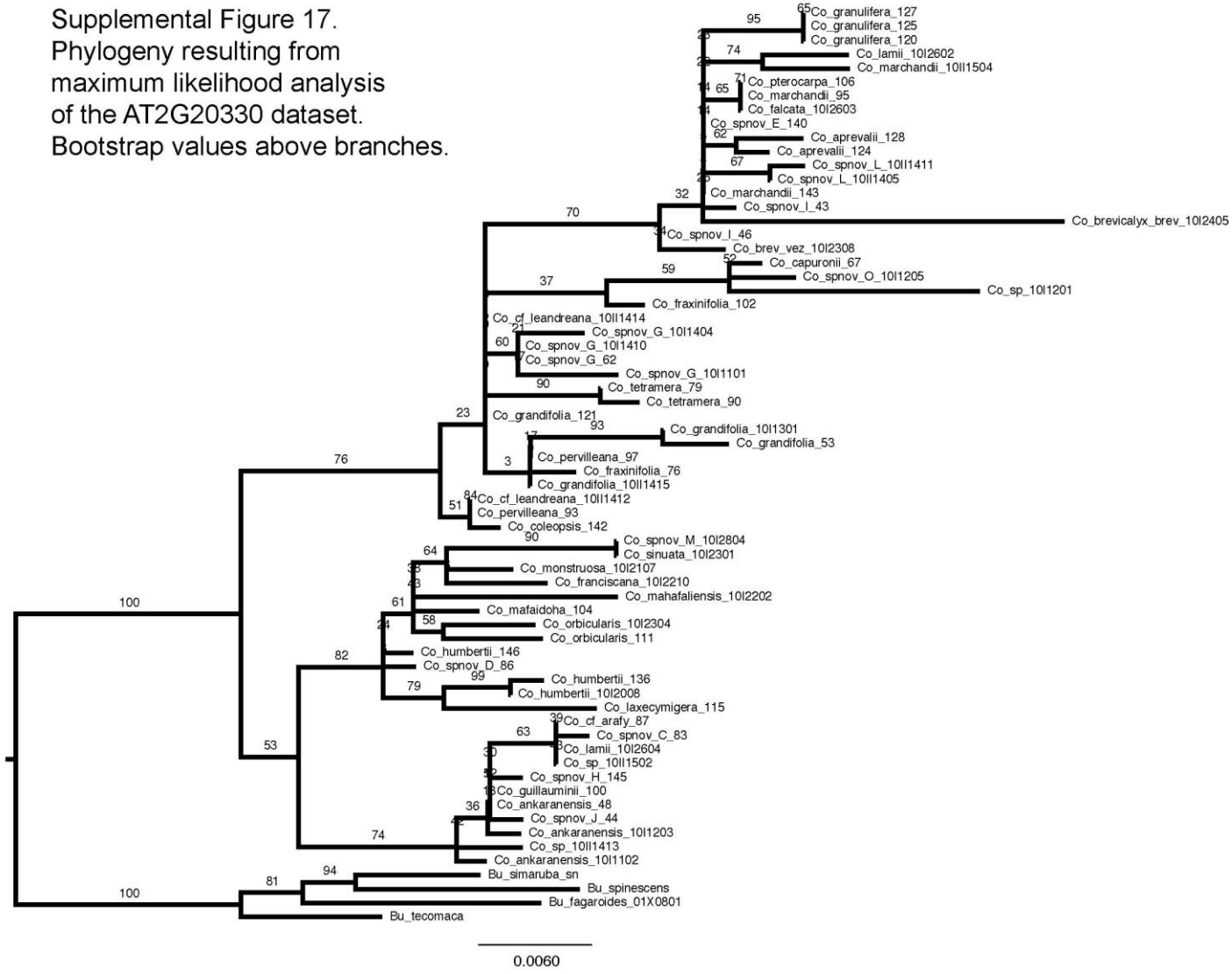
Supplemental Figure 14.
Phylogeny resulting from
maximum likelihood analysis
of the AT2G05120 dataset.
Bootstrap values above branches.

Supplemental Figure 15. Phylogeny resulting from maximum likelihood analysis of the AT2G05320 dataset. Bootstrap values above branches.

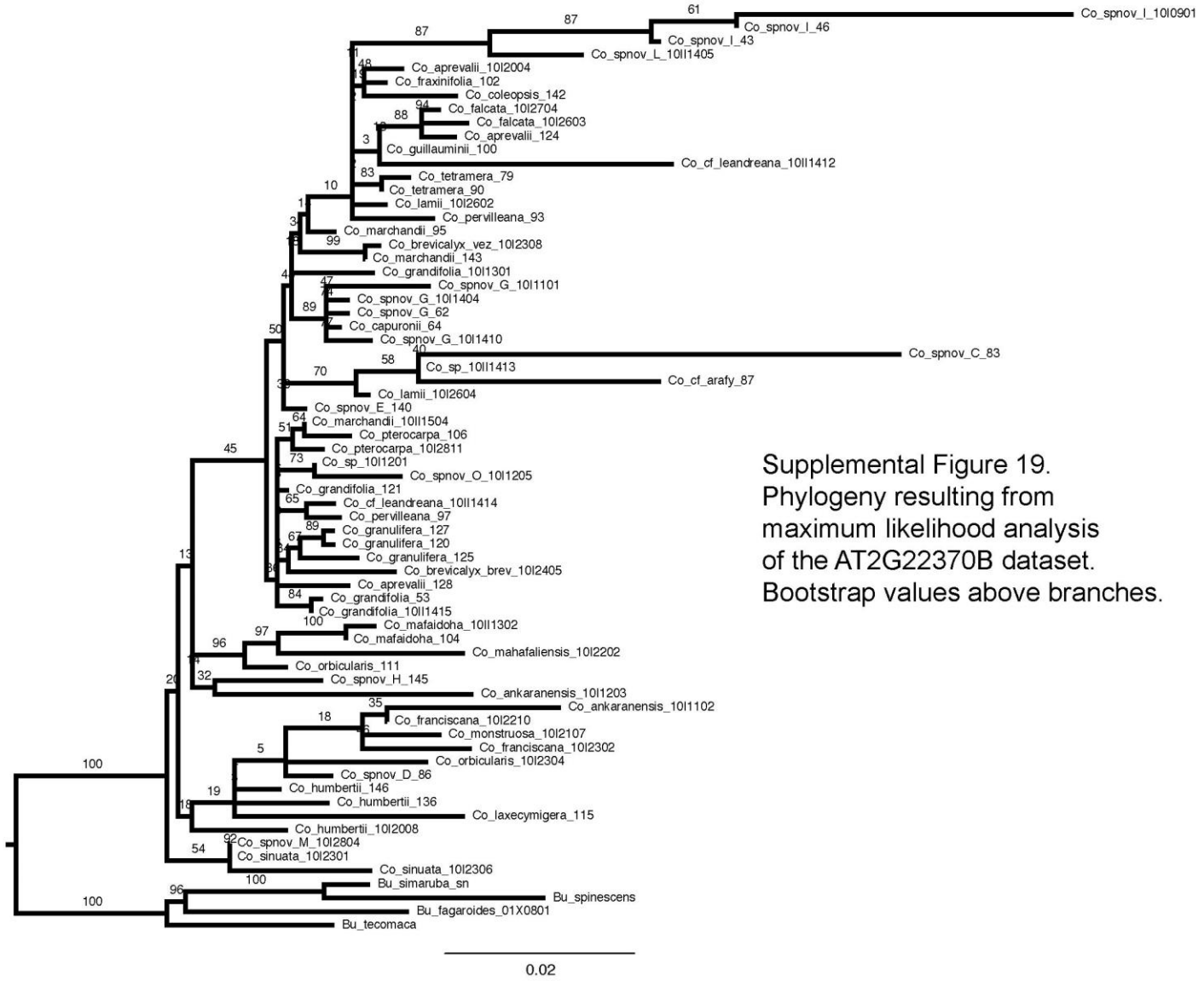


Supplemental Figure 16.
Phylogeny resulting from
maximum likelihood analysis
of the AT2G17265 dataset.
Bootstrap values above branches.

Supplemental Figure 17.
Phylogeny resulting from
maximum likelihood analysis
of the AT2G20330 dataset.
Bootstrap values above branches.

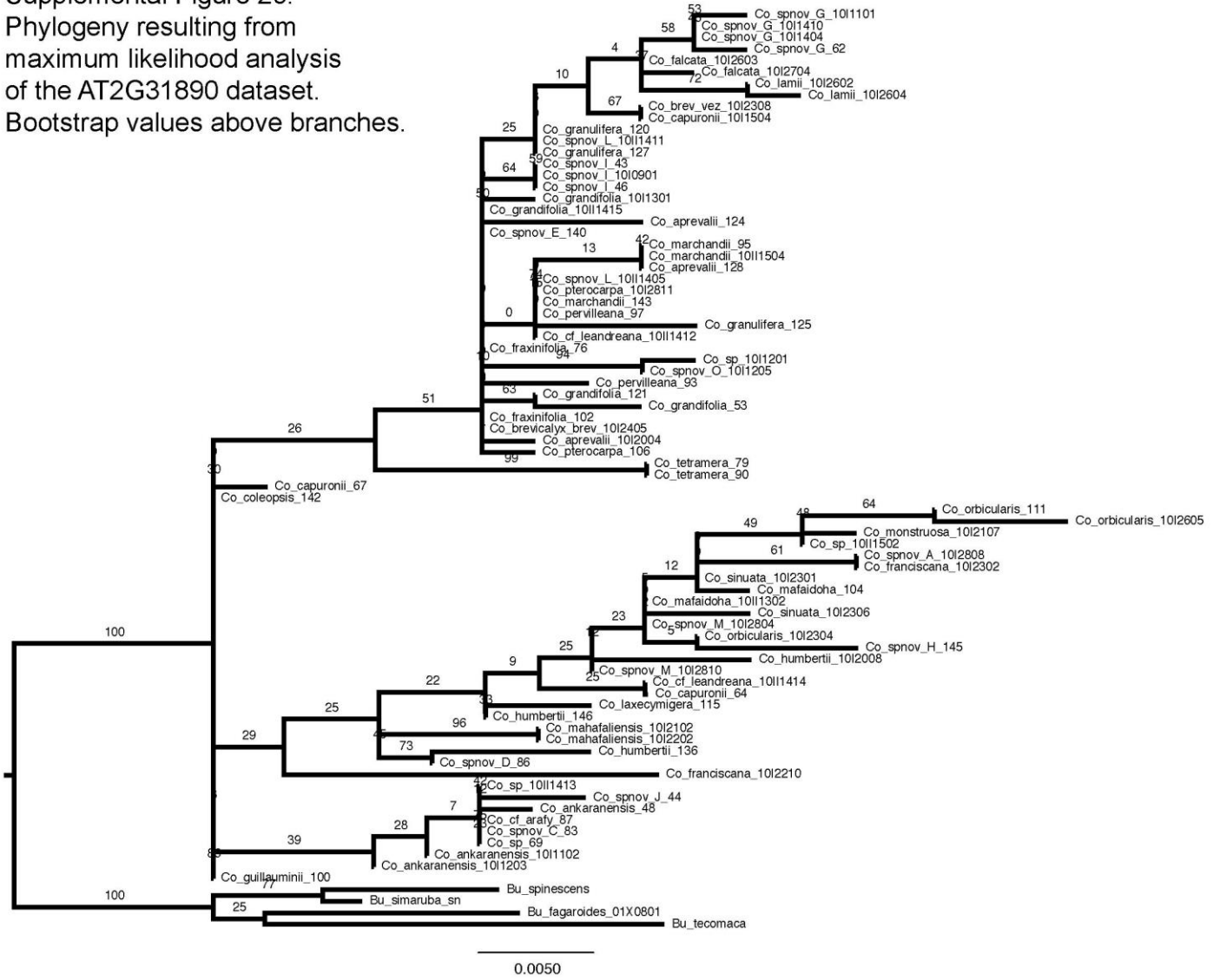


Supplemental Figure 18.
Phylogeny resulting from
maximum likelihood analysis
of the AT2G21710F dataset.
Bootstrap values above branches.

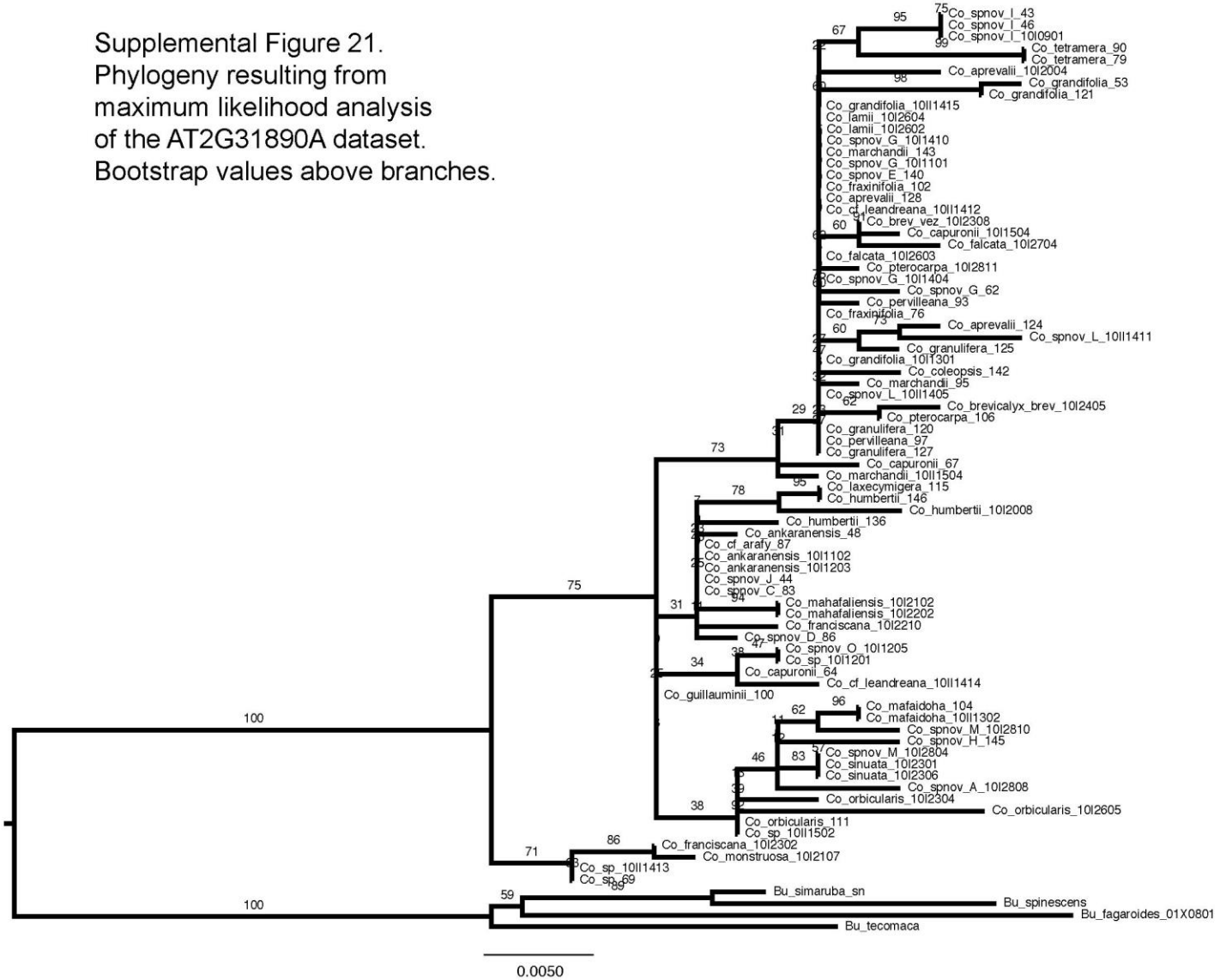


Supplemental Figure 19. Phylogeny resulting from maximum likelihood analysis of the AT2G22370B dataset. Bootstrap values above branches.

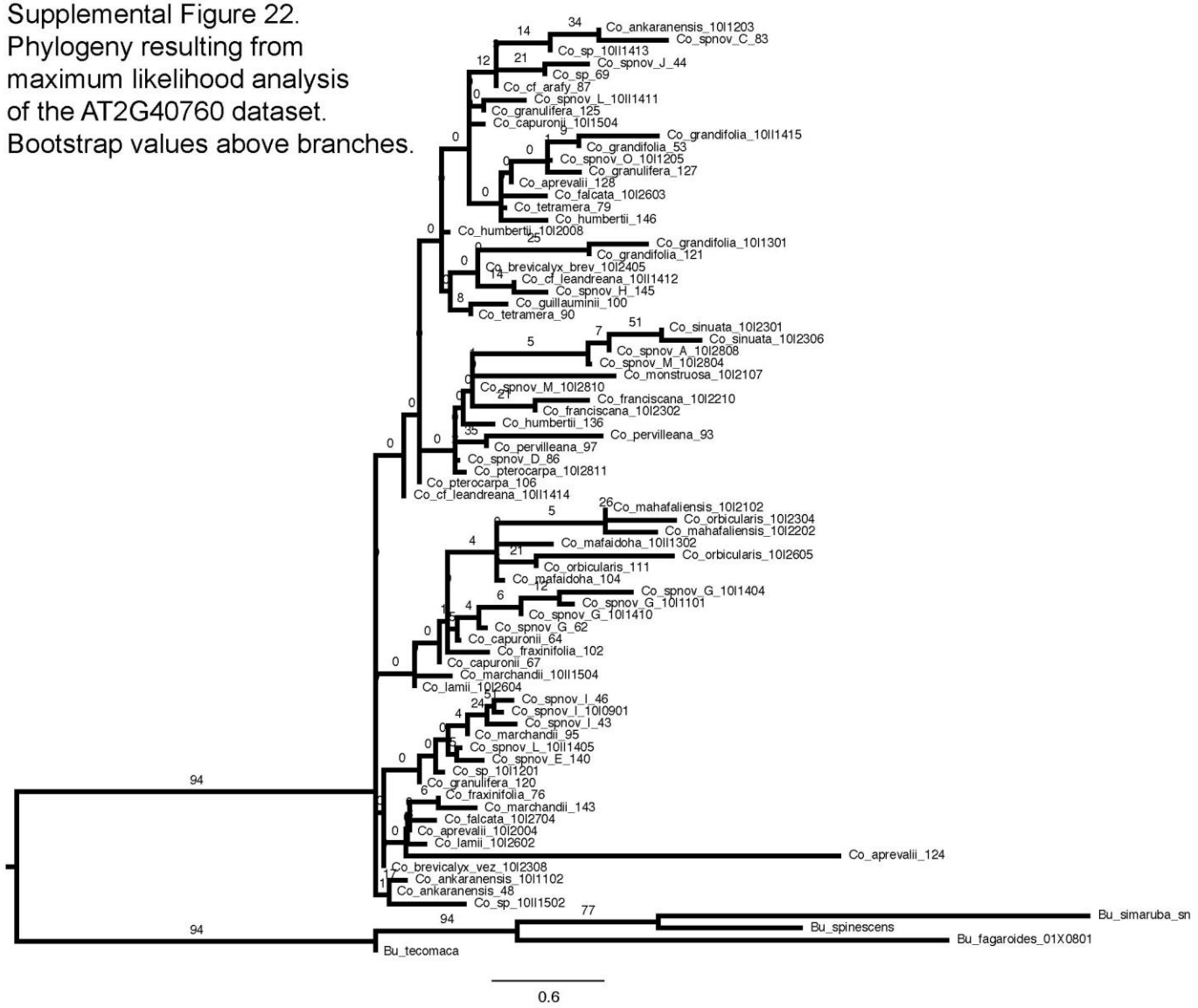
Supplemental Figure 20.
Phylogeny resulting from
maximum likelihood analysis
of the AT2G31890 dataset.
Bootstrap values above branches.

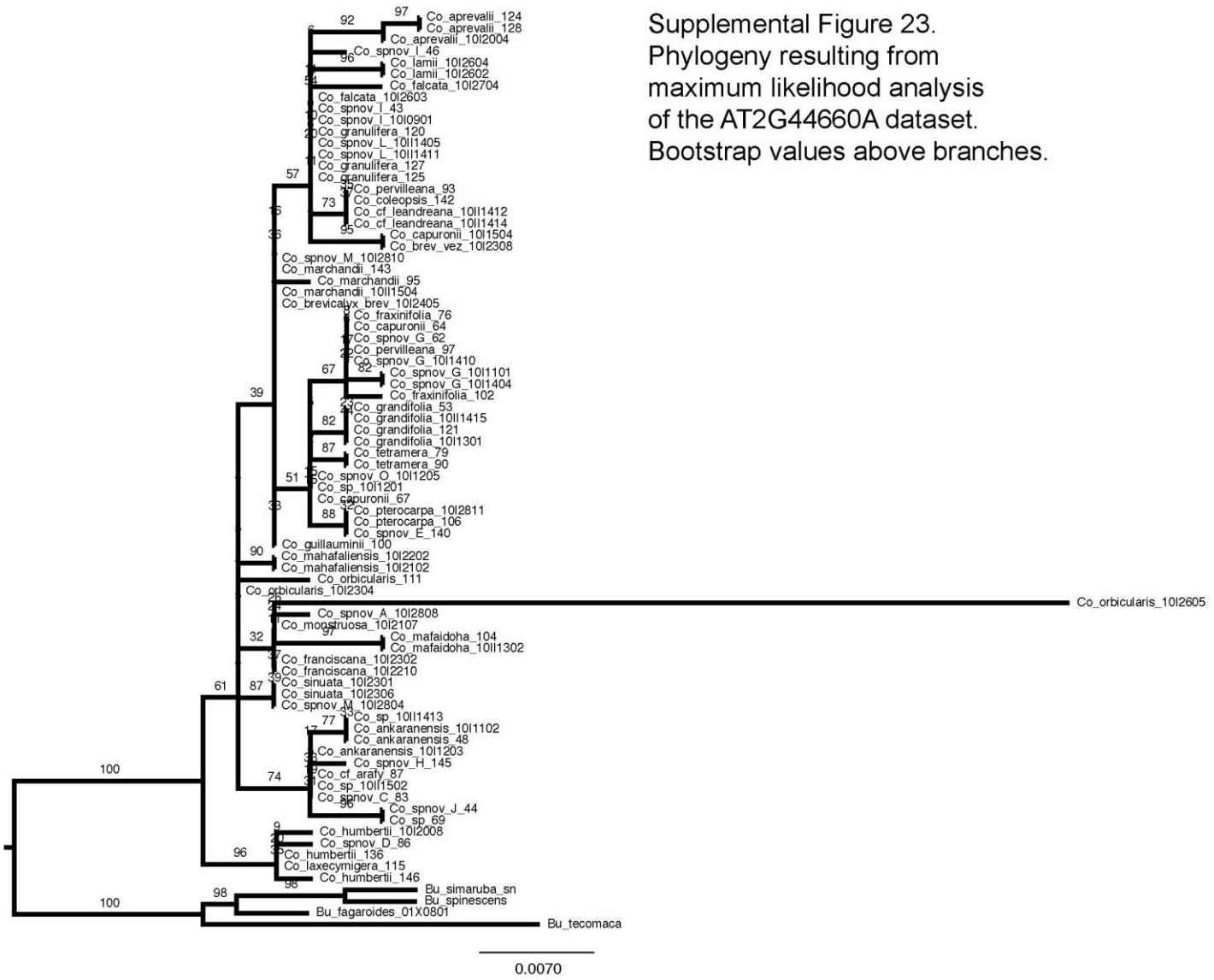


Supplemental Figure 21.
Phylogeny resulting from
maximum likelihood analysis
of the AT2G31890A dataset.
Bootstrap values above branches.



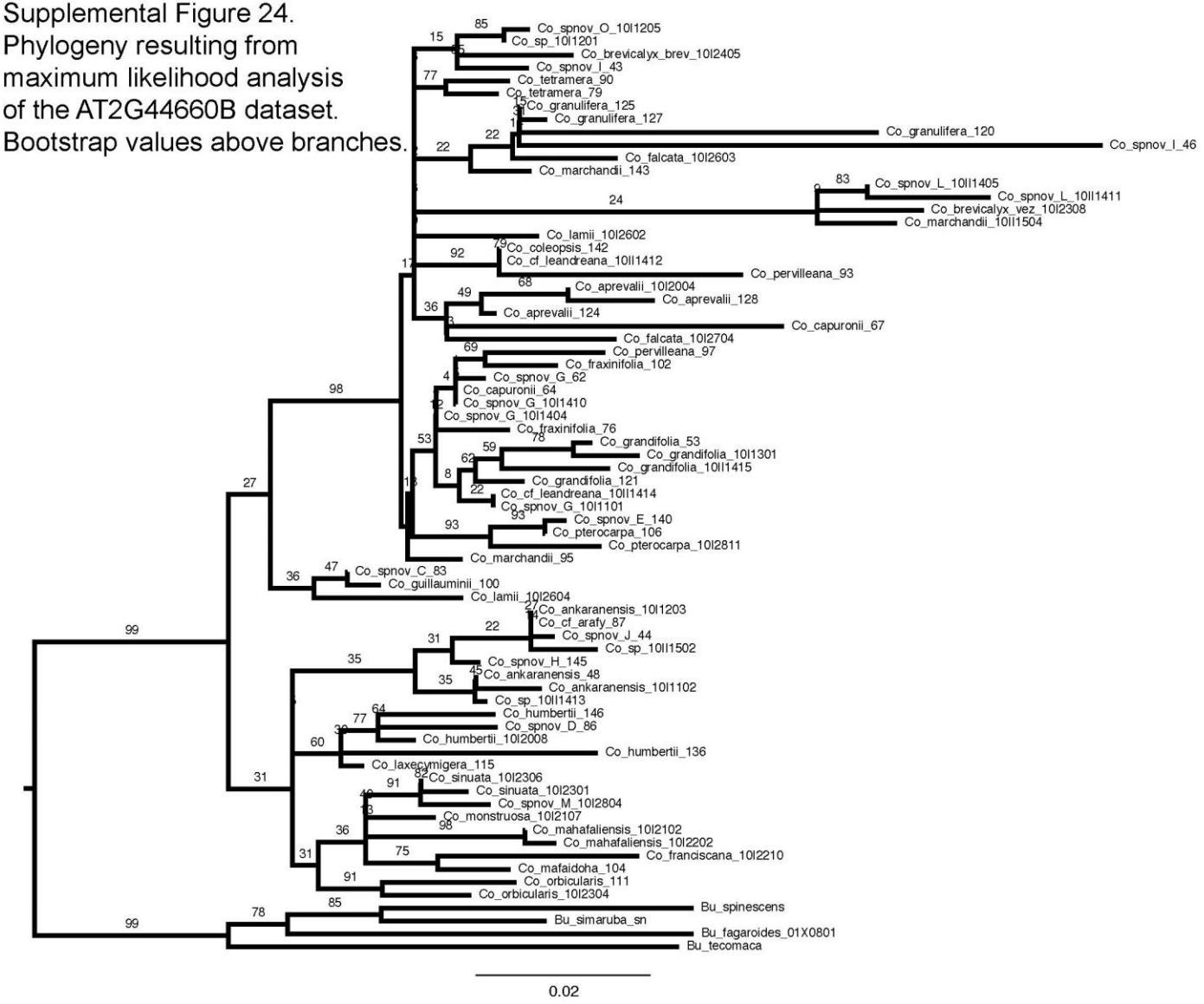
Supplemental Figure 22.
Phylogeny resulting from maximum likelihood analysis of the AT2G40760 dataset. Bootstrap values above branches.



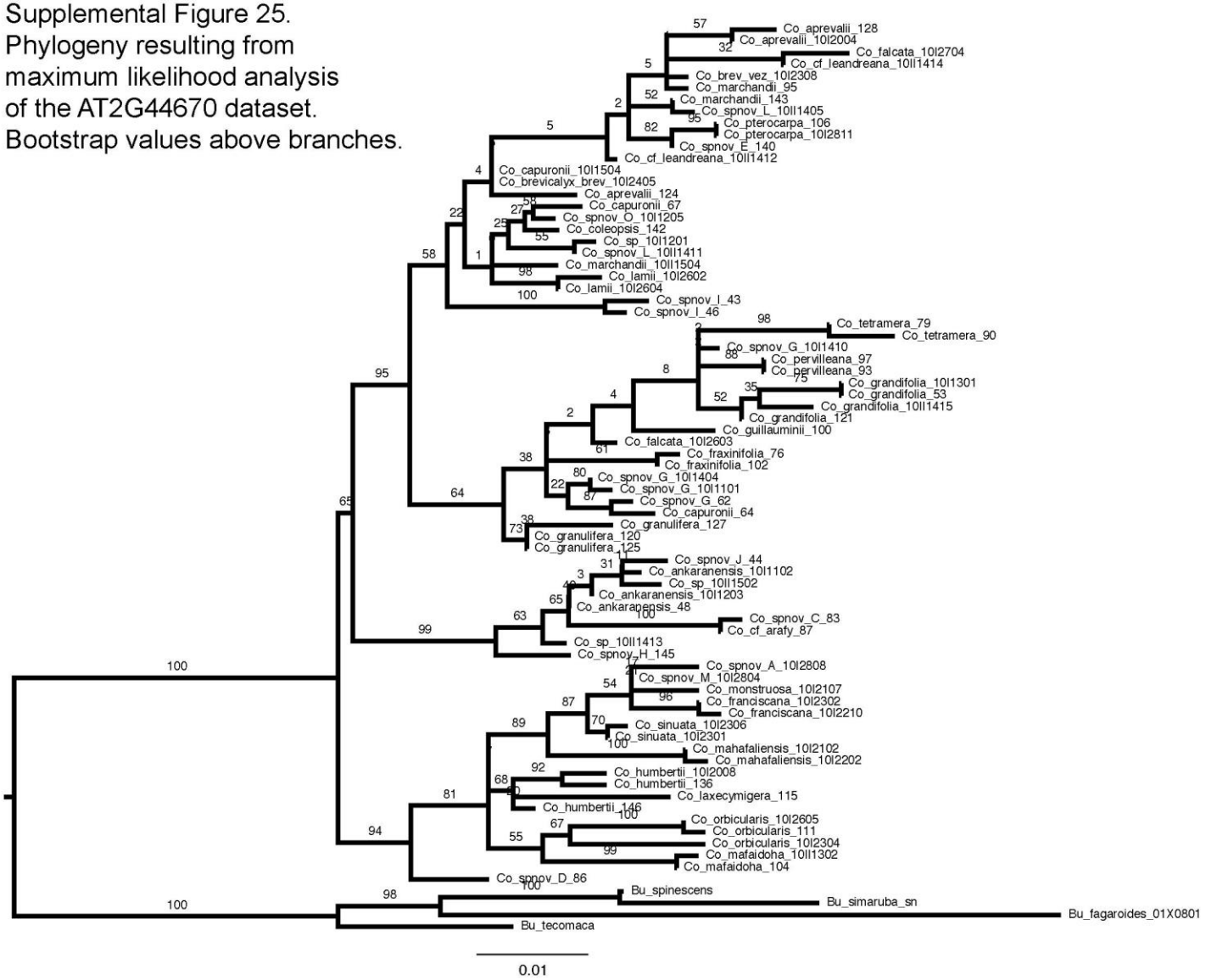


Supplemental Figure 23.
Phylogeny resulting from
maximum likelihood analysis
of the AT2G44660A dataset.
Bootstrap values above branches.

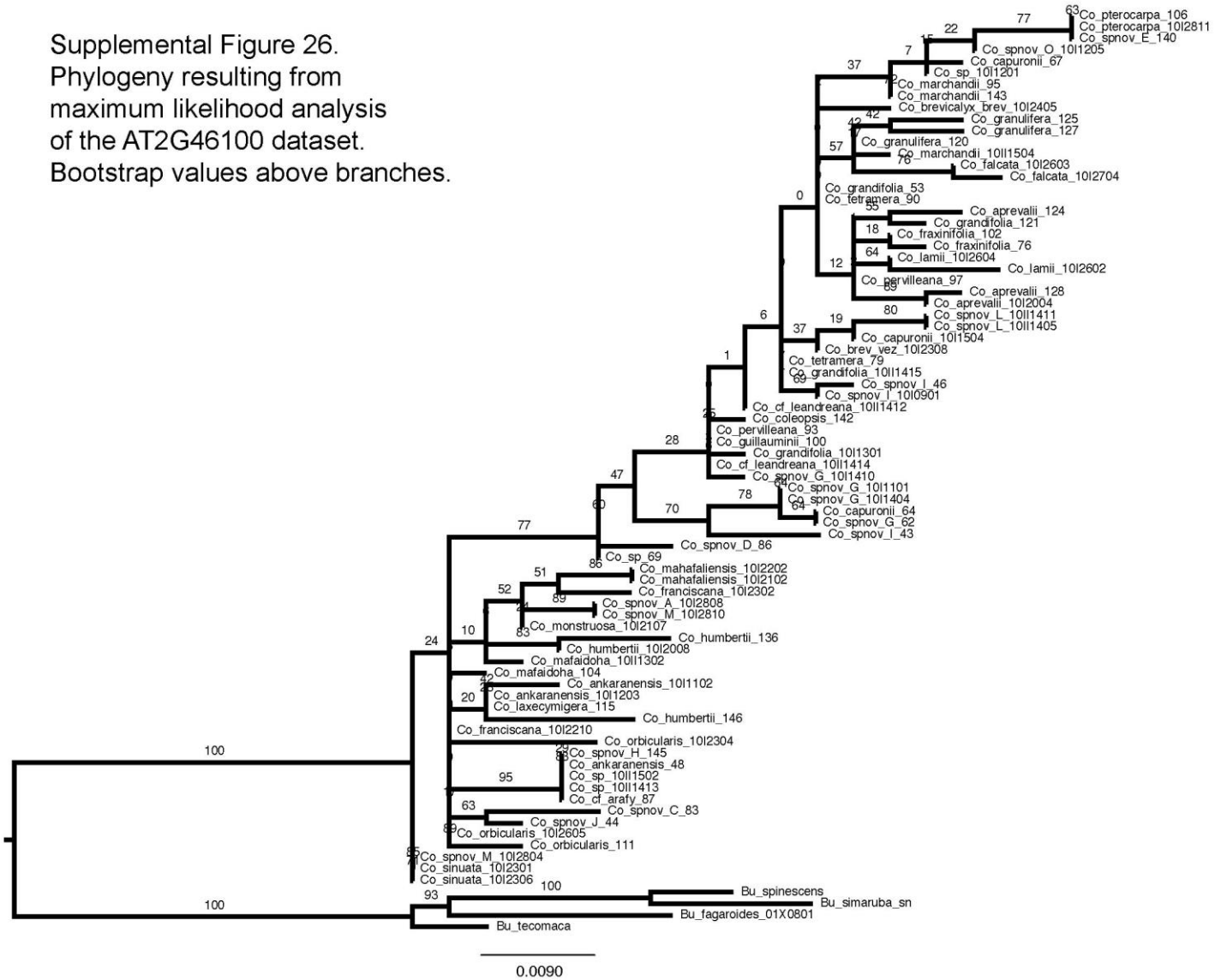
Supplemental Figure 24. Phylogeny resulting from maximum likelihood analysis of the AT2G44660B dataset. Bootstrap values above branches.

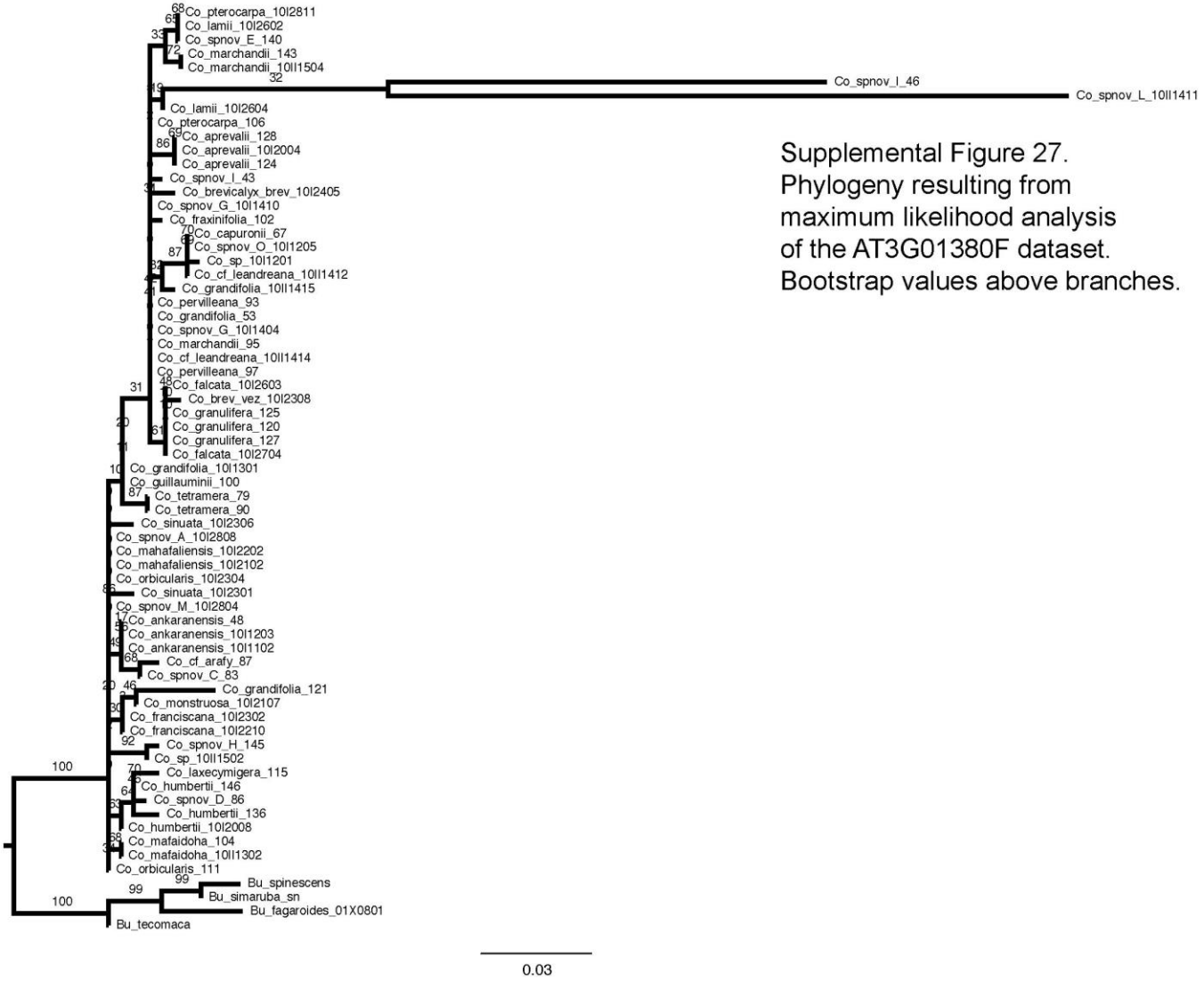


Supplemental Figure 25.
Phylogeny resulting from
maximum likelihood analysis
of the AT2G44670 dataset.
Bootstrap values above branches.



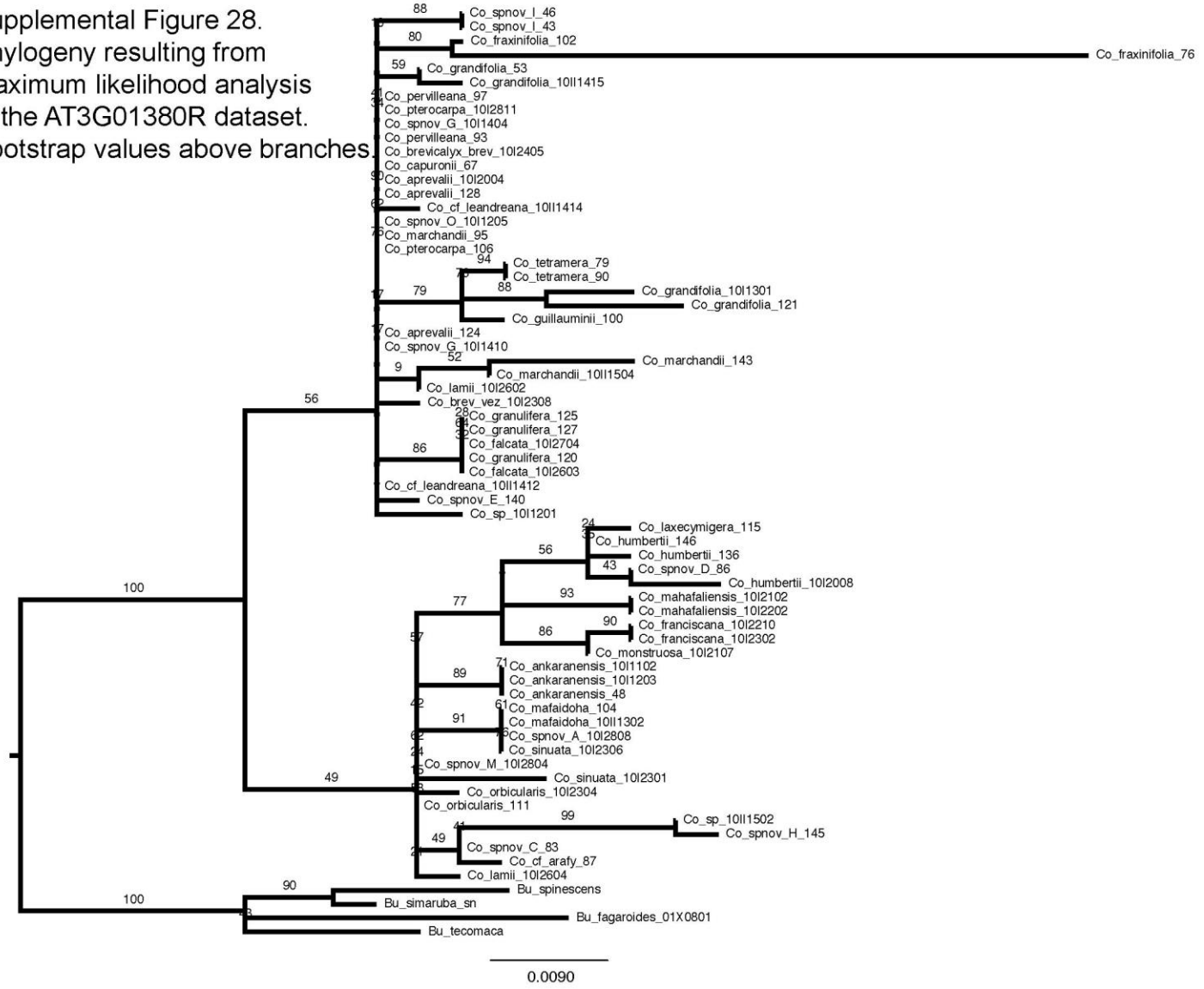
Supplemental Figure 26.
Phylogeny resulting from
maximum likelihood analysis
of the AT2G46100 dataset.
Bootstrap values above branches.



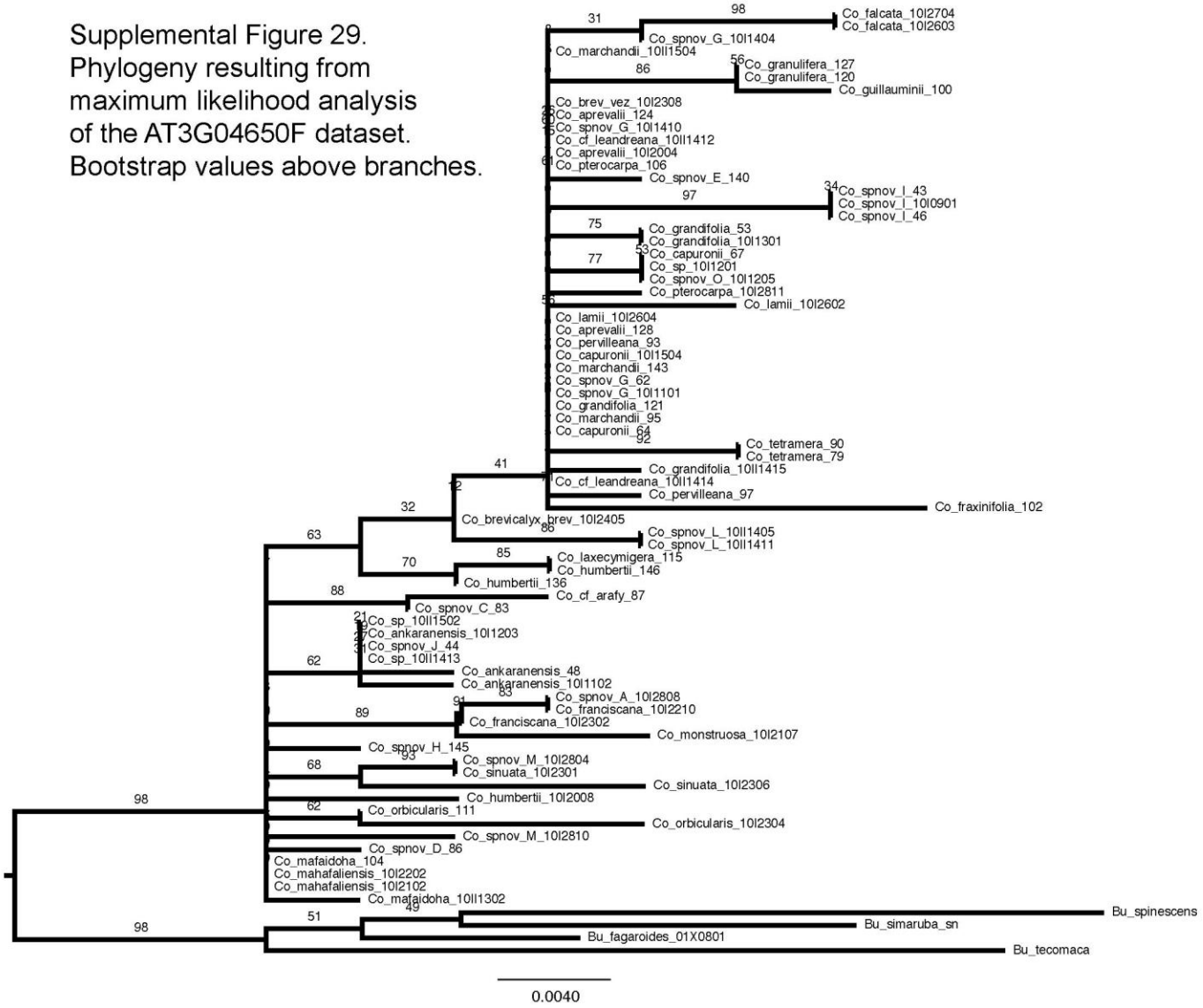


Supplemental Figure 27. Phylogeny resulting from maximum likelihood analysis of the AT3G01380F dataset. Bootstrap values above branches.

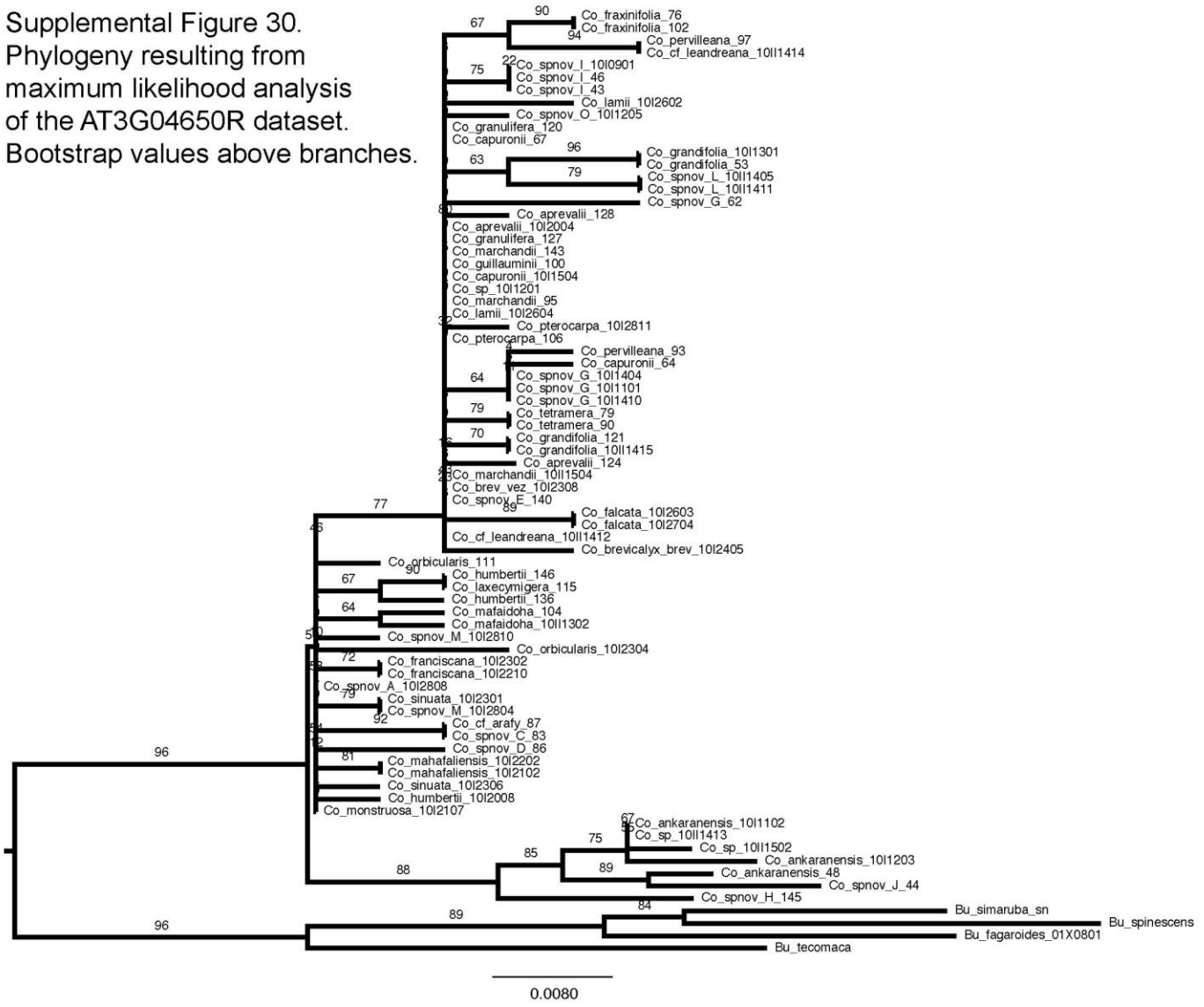
Supplemental Figure 28. Phylogeny resulting from maximum likelihood analysis of the AT3G01380R dataset. Bootstrap values above branches.



Supplemental Figure 29.
Phylogeny resulting from
maximum likelihood analysis
of the AT3G04650F dataset.
Bootstrap values above branches.

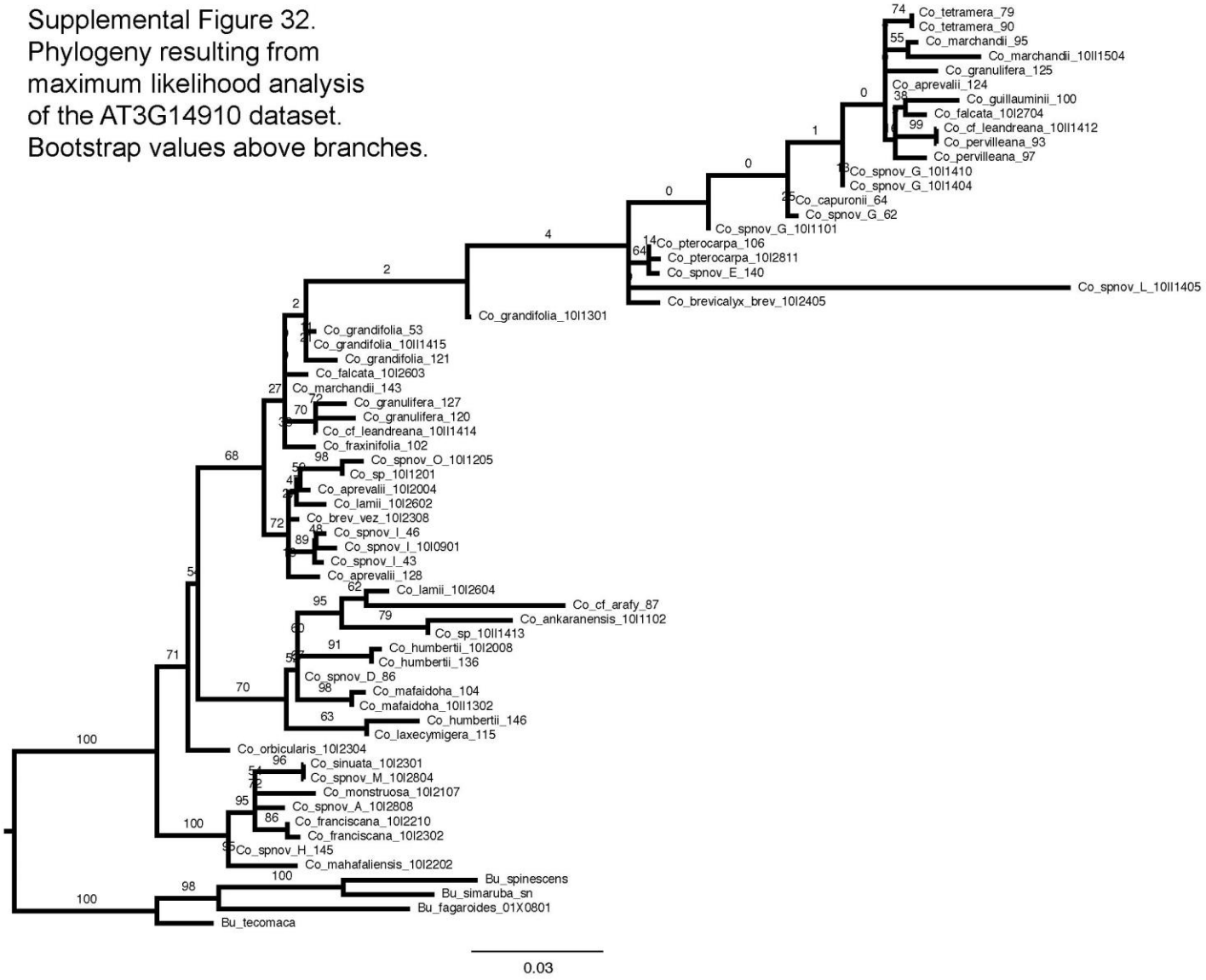


Supplemental Figure 30.
Phylogeny resulting from
maximum likelihood analysis
of the AT3G04650R dataset.
Bootstrap values above branches.

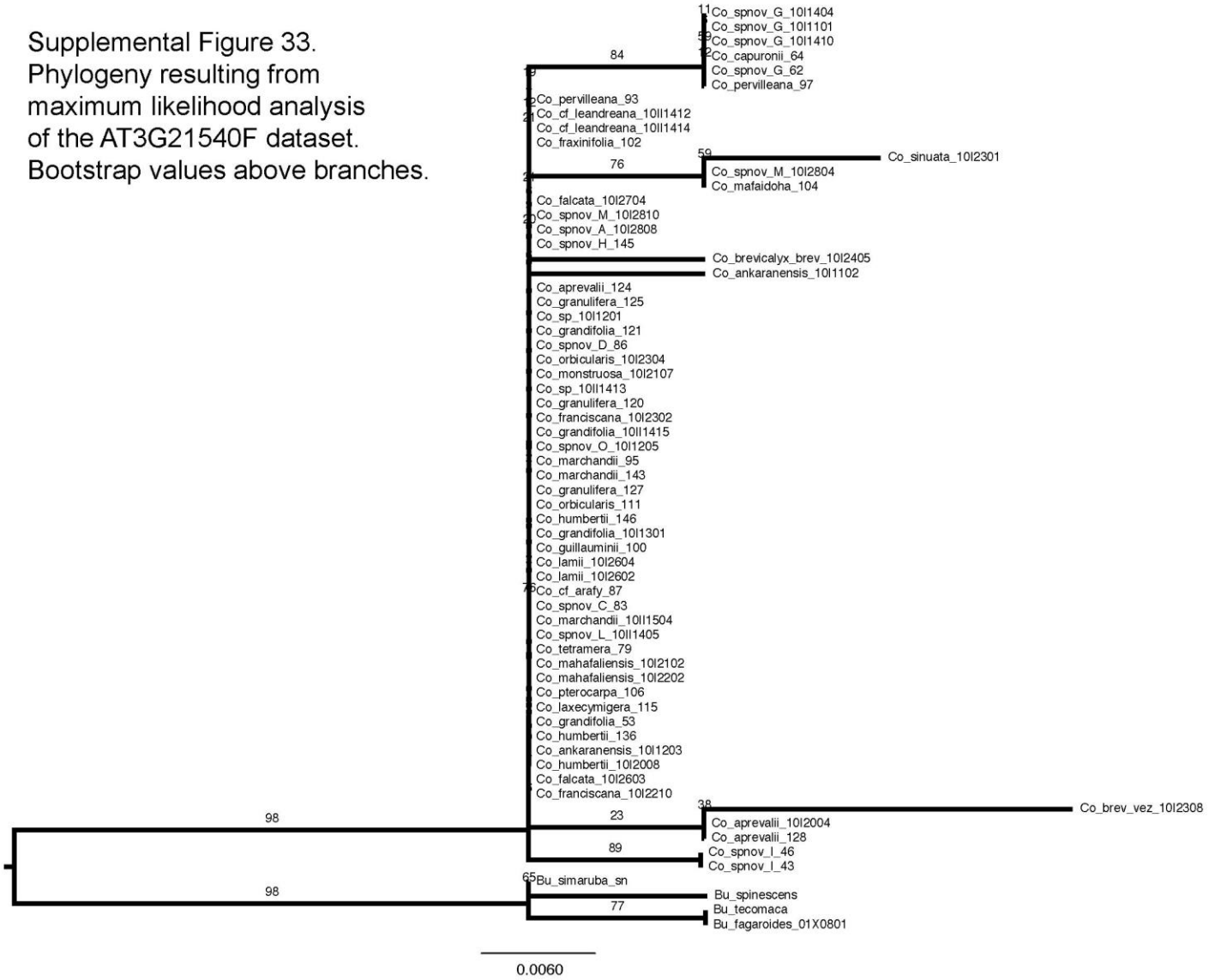


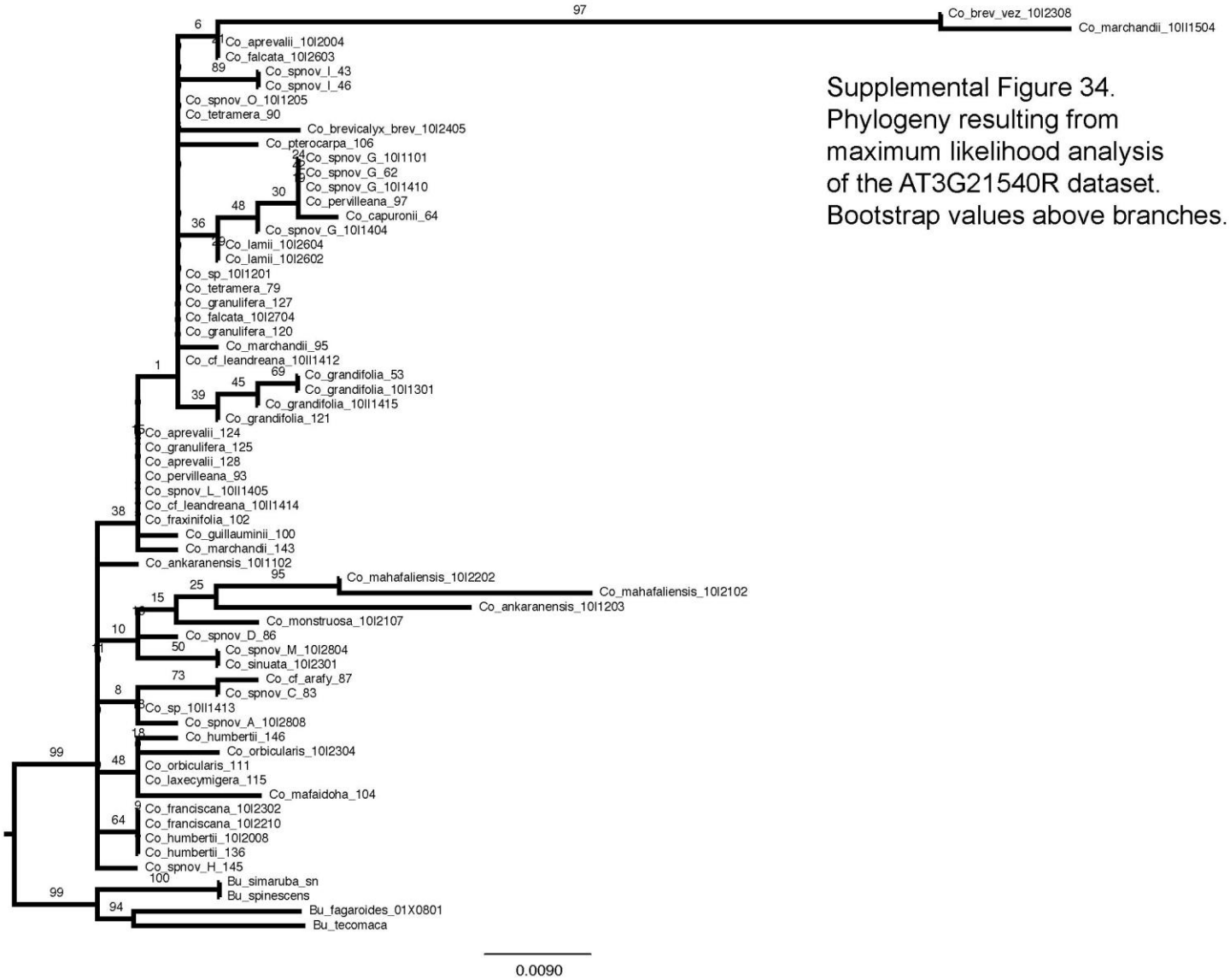
Supplemental Figure 31. Phylogeny resulting from maximum likelihood analysis of the AT3G10400 dataset. Bootstrap values above branches.

Supplemental Figure 32.
Phylogeny resulting from
maximum likelihood analysis
of the AT3G14910 dataset.
Bootstrap values above branches.



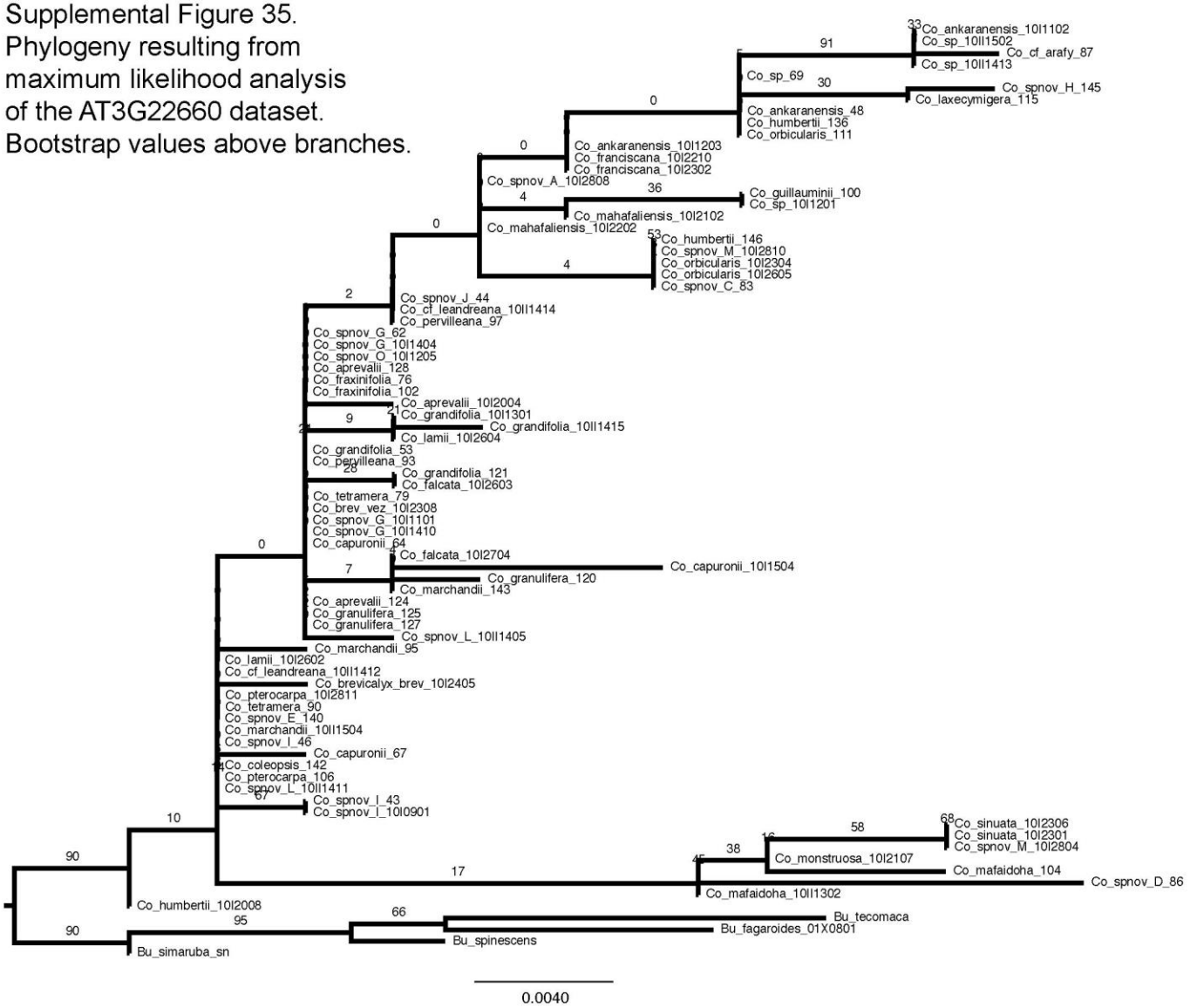
Supplemental Figure 33.
Phylogeny resulting from
maximum likelihood analysis
of the AT3G21540F dataset.
Bootstrap values above branches.



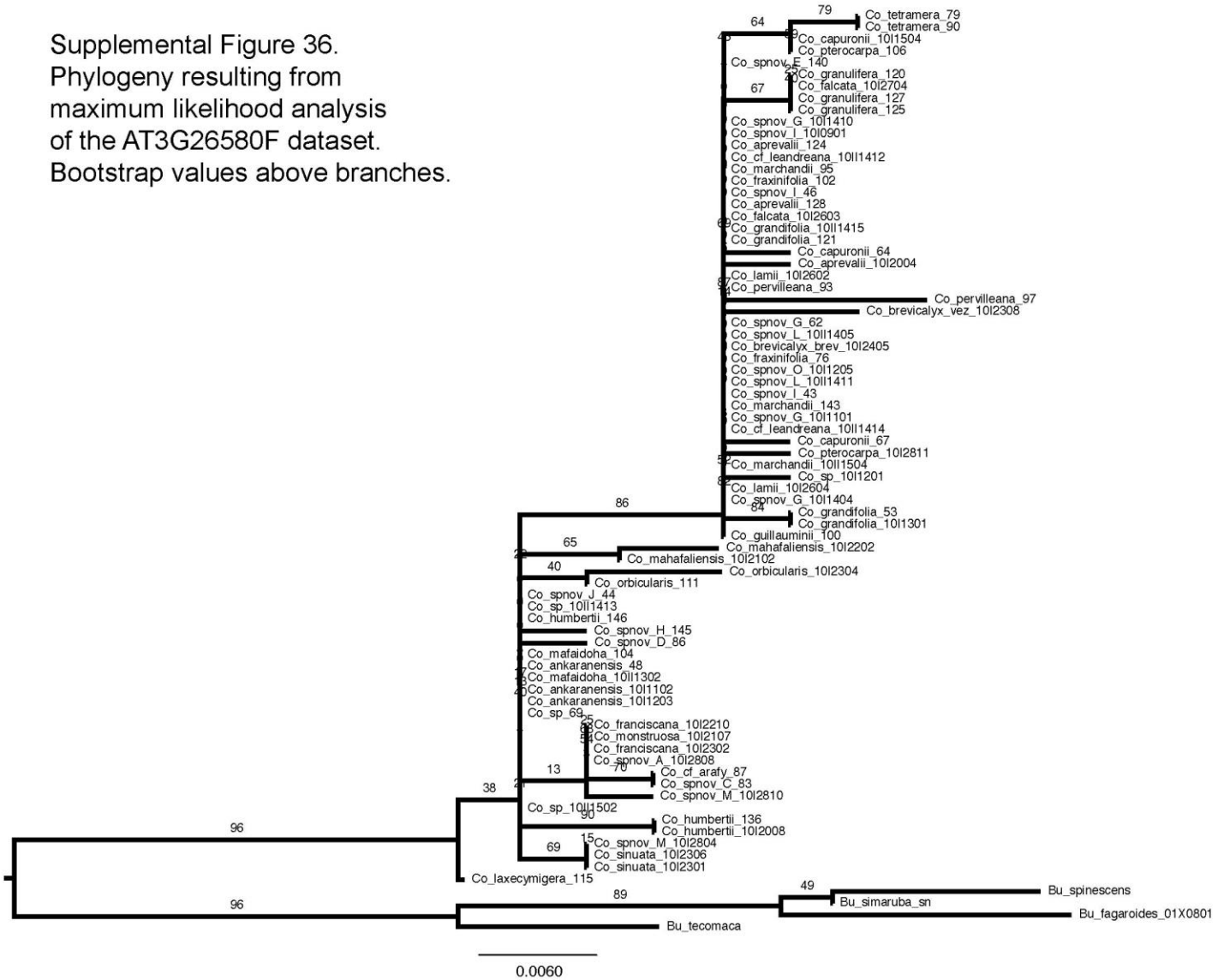


Supplemental Figure 34. Phylogeny resulting from maximum likelihood analysis of the AT3G21540R dataset. Bootstrap values above branches.

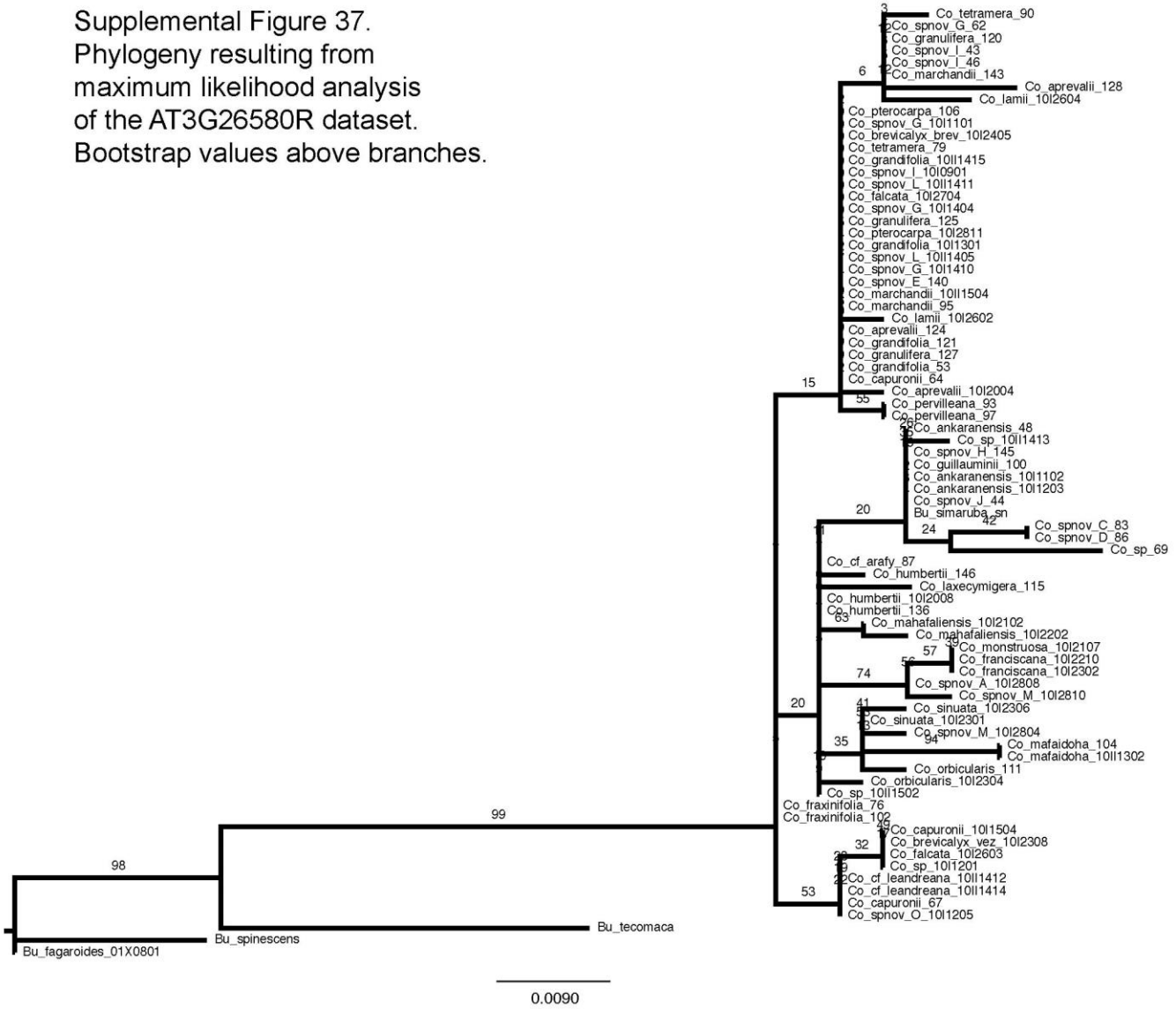
Supplemental Figure 35.
Phylogeny resulting from
maximum likelihood analysis
of the AT3G22660 dataset.
Bootstrap values above branches.



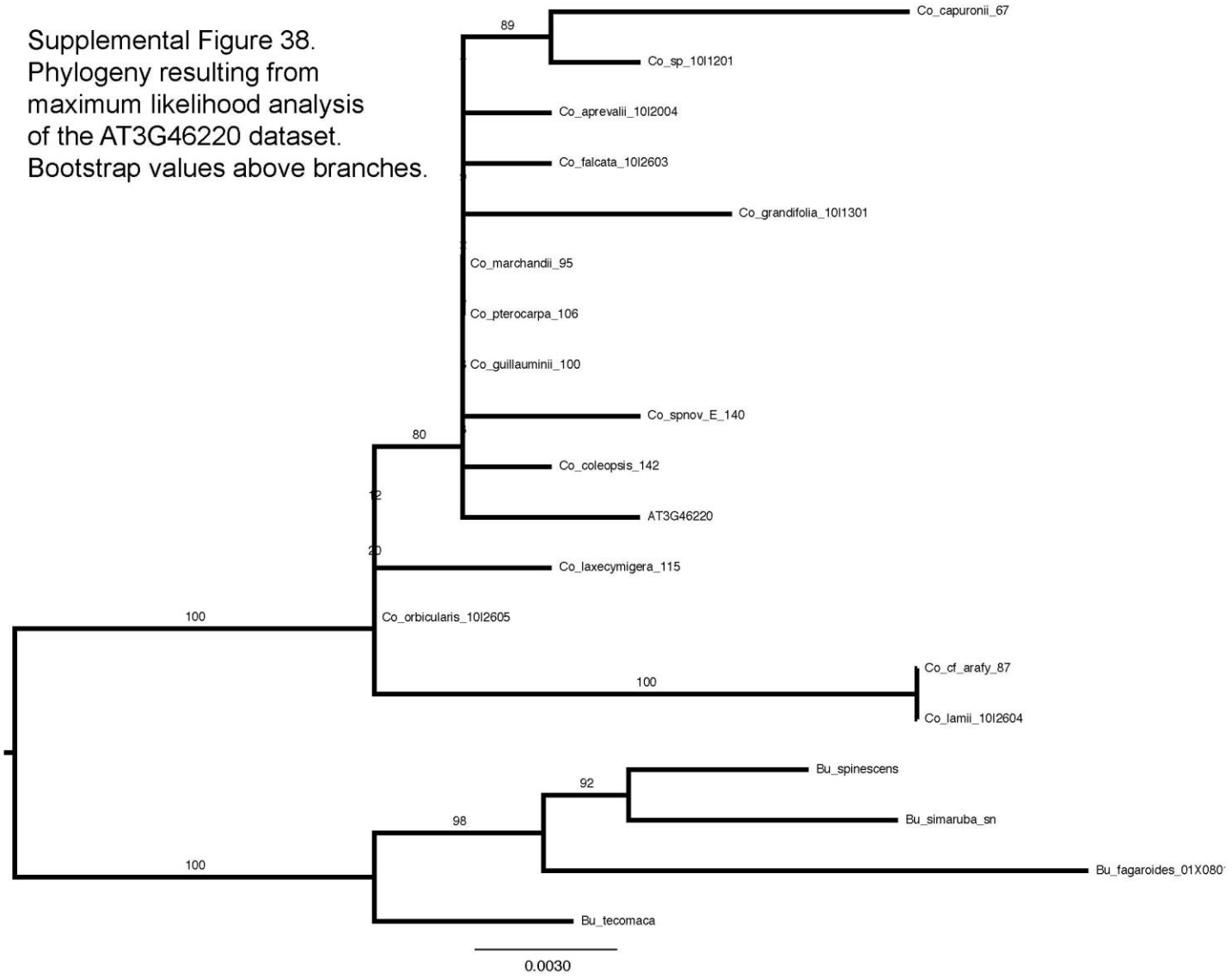
Supplemental Figure 36.
Phylogeny resulting from
maximum likelihood analysis
of the AT3G26580F dataset.
Bootstrap values above branches.



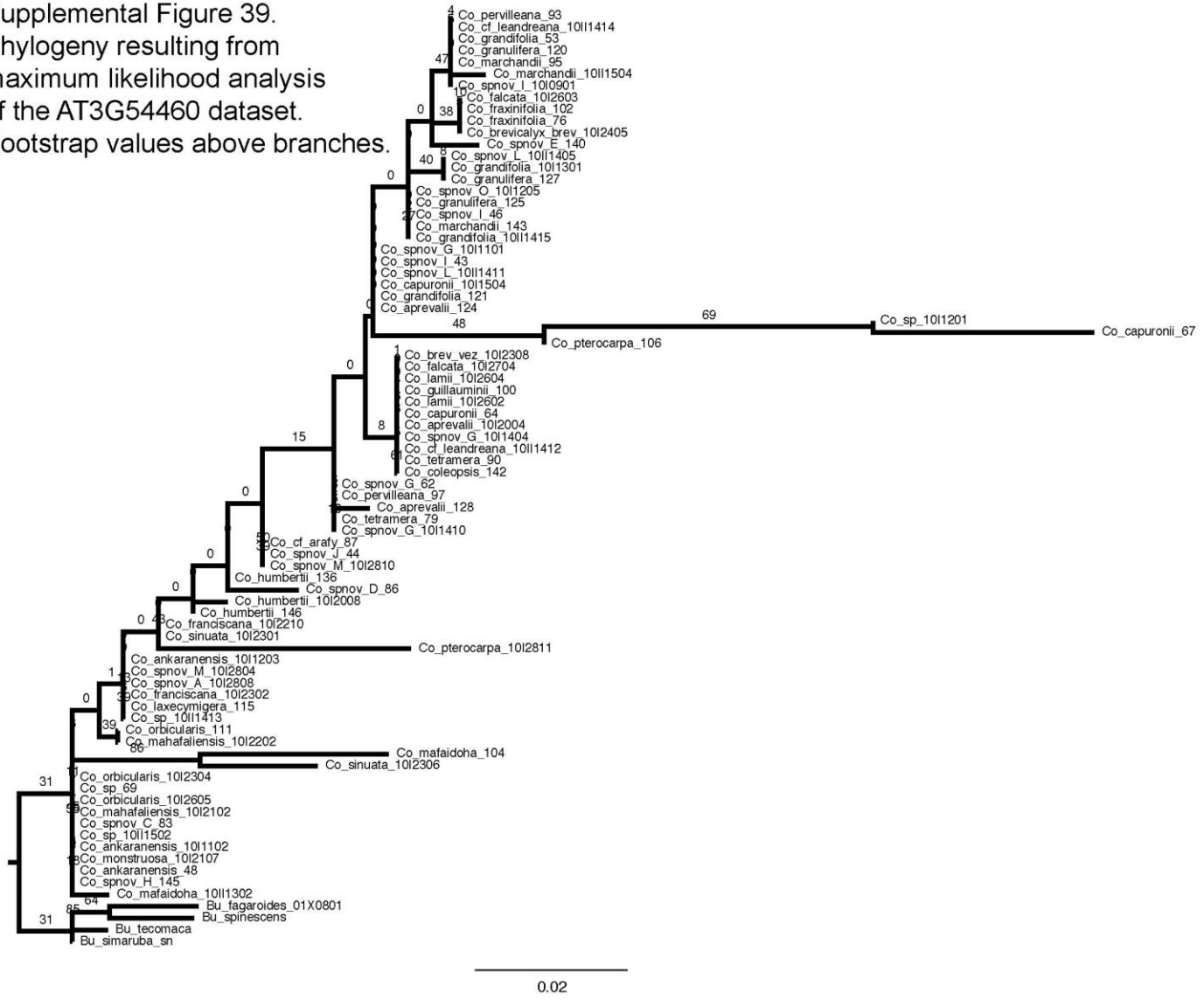
Supplemental Figure 37.
Phylogeny resulting from
maximum likelihood analysis
of the AT3G26580R dataset.
Bootstrap values above branches.



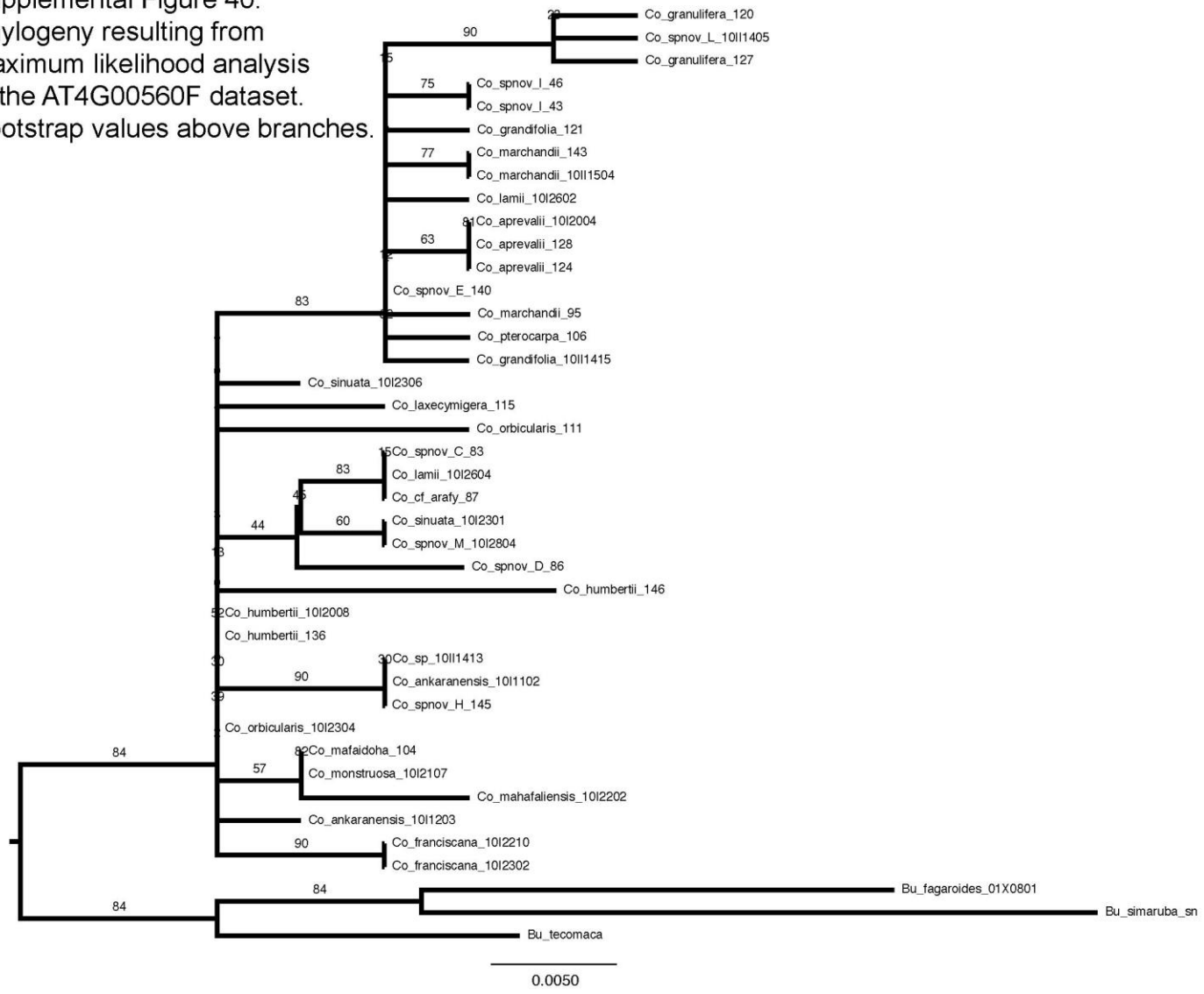
Supplemental Figure 38.
Phylogeny resulting from maximum likelihood analysis of the AT3G46220 dataset.
Bootstrap values above branches.

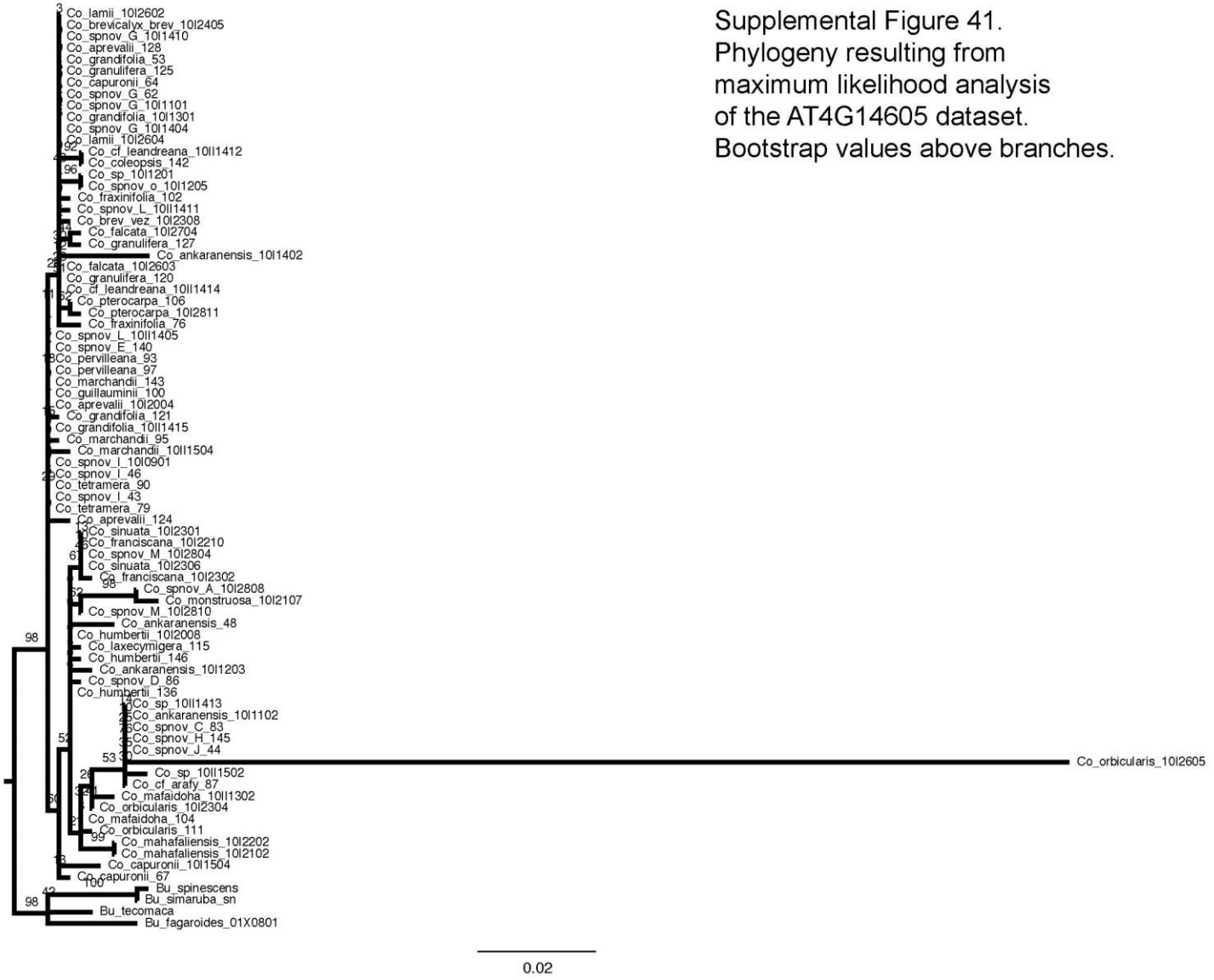


Supplemental Figure 39.
Phylogeny resulting from
maximum likelihood analysis
of the AT3G54460 dataset.
Bootstrap values above branches.



Supplemental Figure 40.
Phylogeny resulting from maximum likelihood analysis of the AT4G00560F dataset. Bootstrap values above branches.

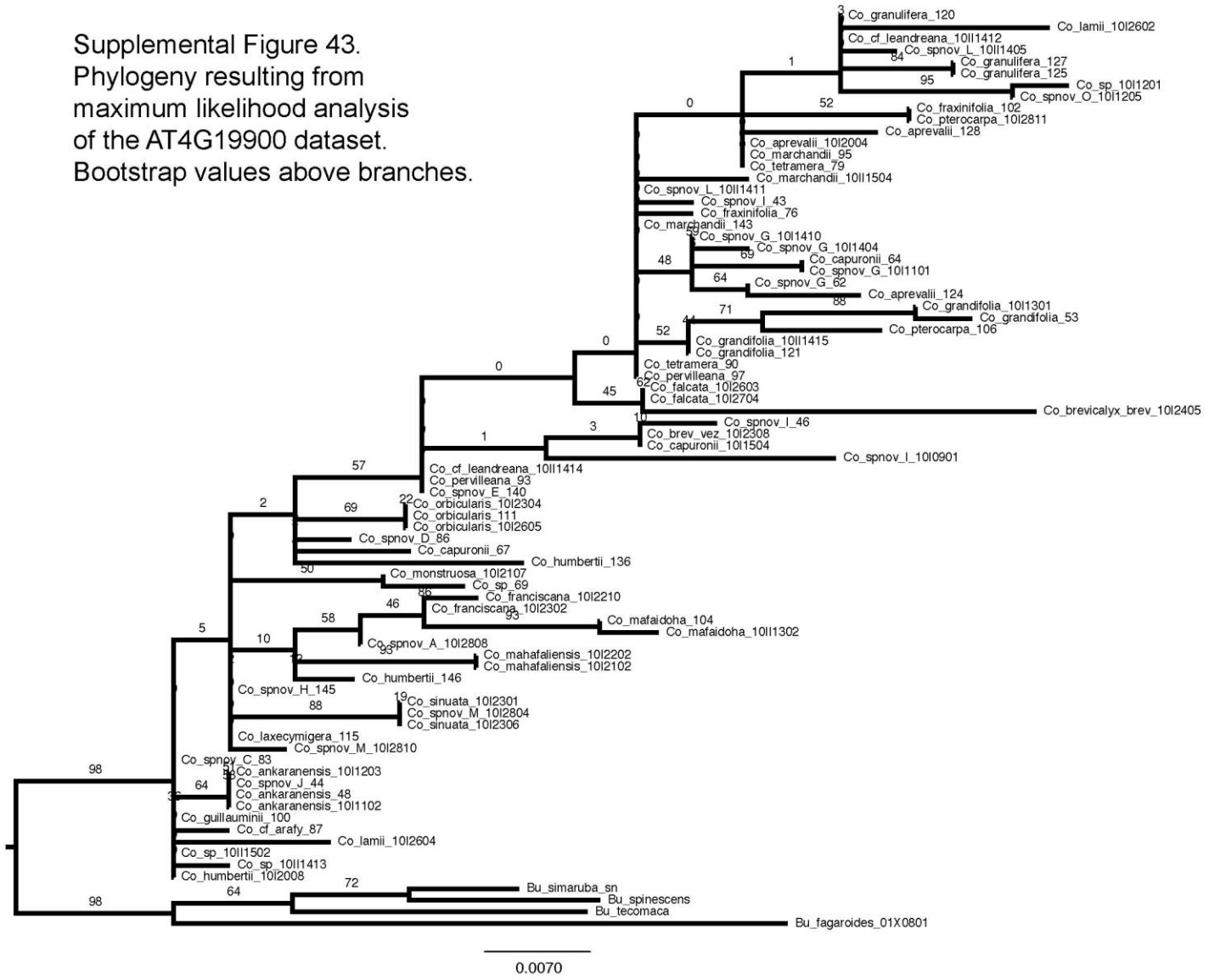


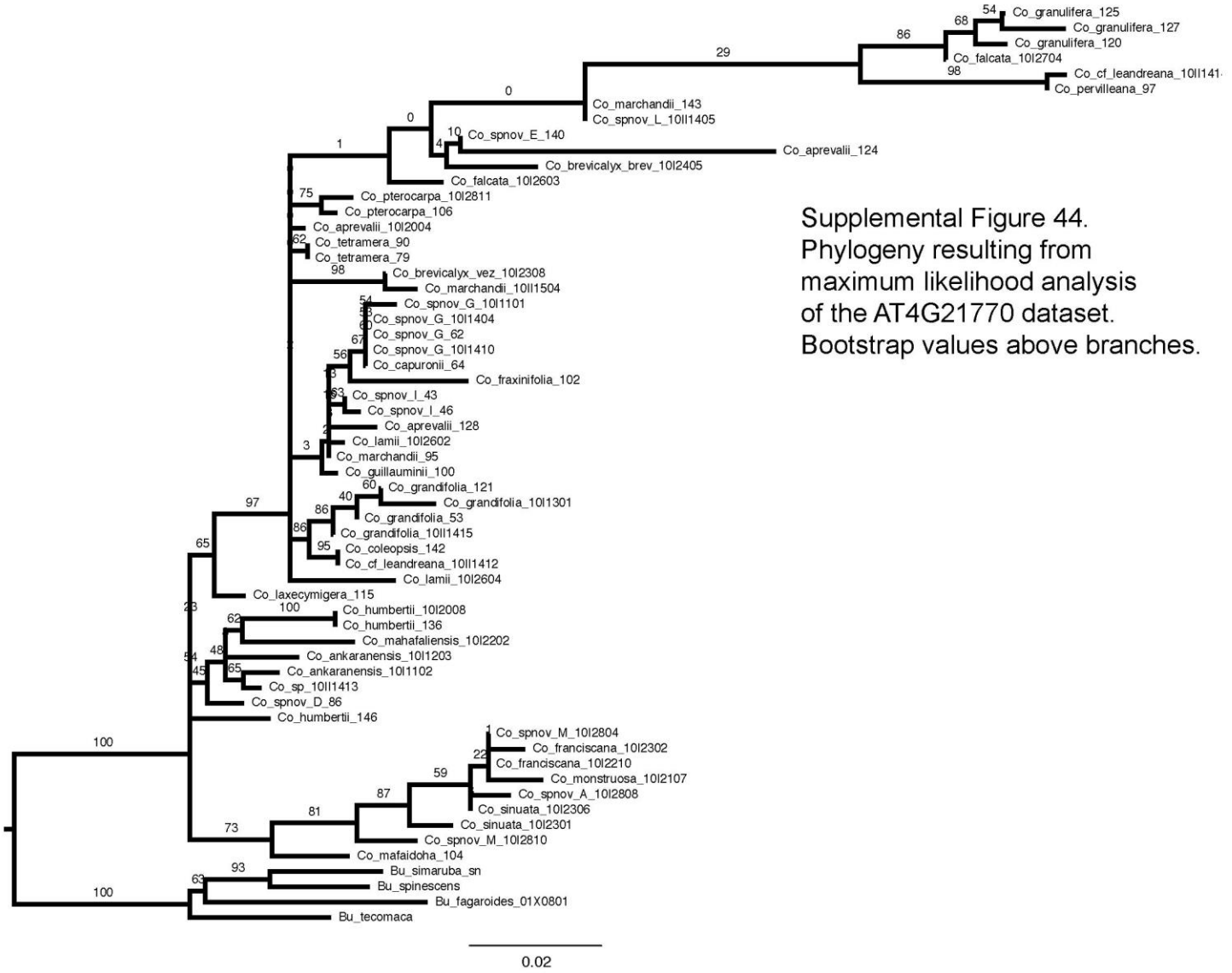


Supplemental Figure 41.
Phylogeny resulting from
maximum likelihood analysis
of the AT4G14605 dataset.
Bootstrap values above branches.

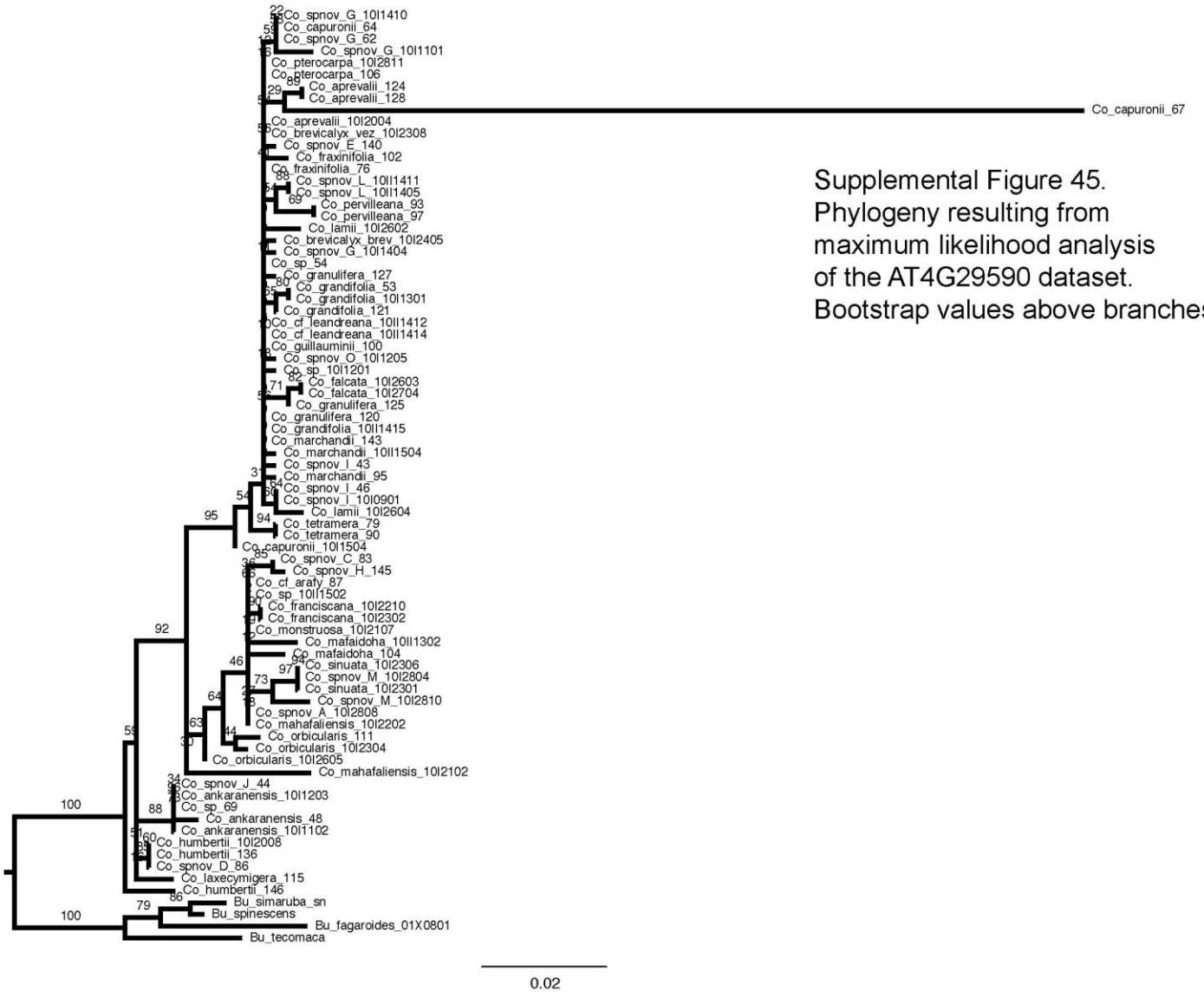
Supplemental Figure 42.
Phylogeny resulting from
maximum likelihood analysis
of the AT4G18810B dataset.
Bootstrap values above branches.

Supplemental Figure 43.
Phylogeny resulting from
maximum likelihood analysis
of the AT4G19900 dataset.
Bootstrap values above branches.



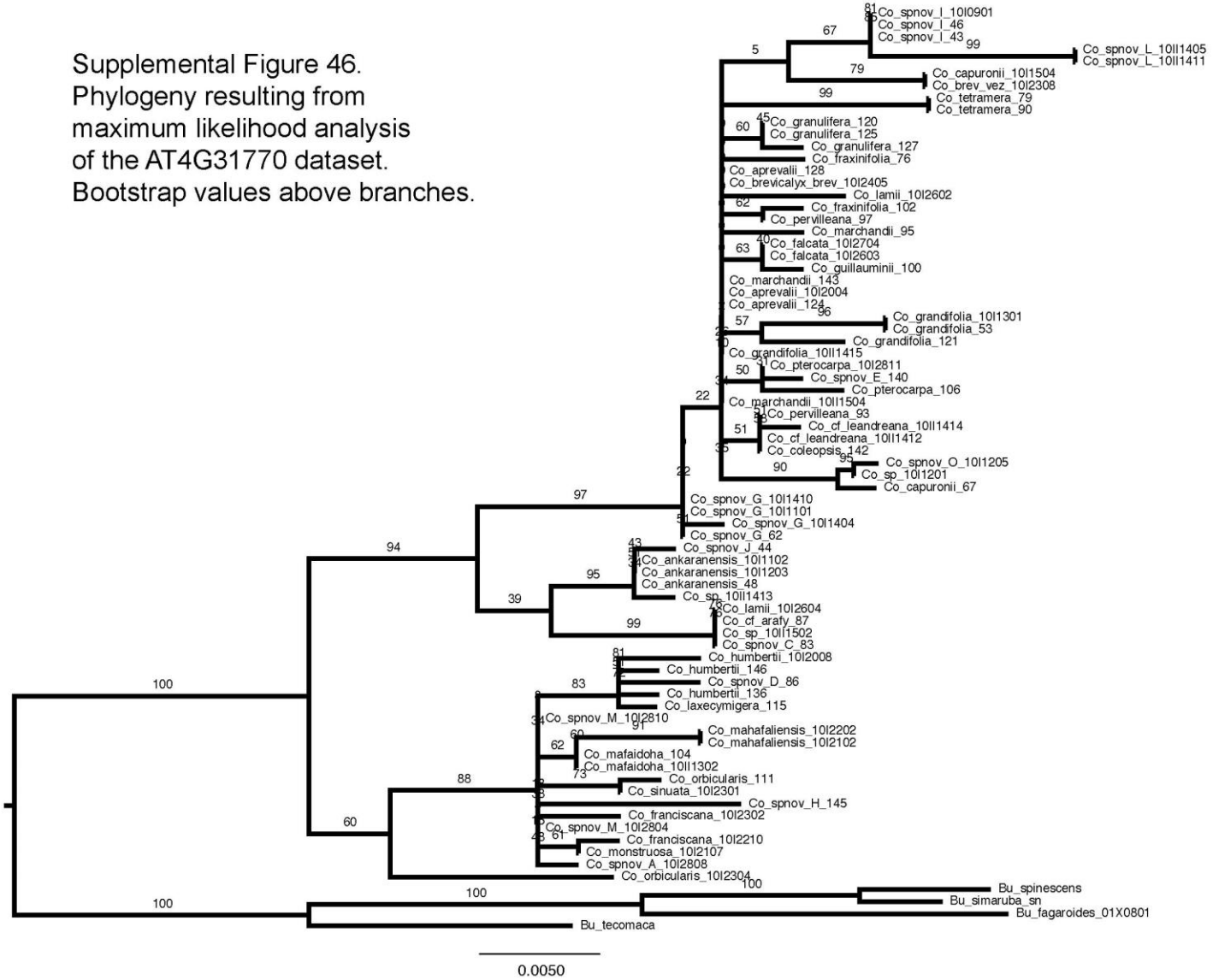


Supplemental Figure 44. Phylogeny resulting from maximum likelihood analysis of the AT4G21770 dataset. Bootstrap values above branches.

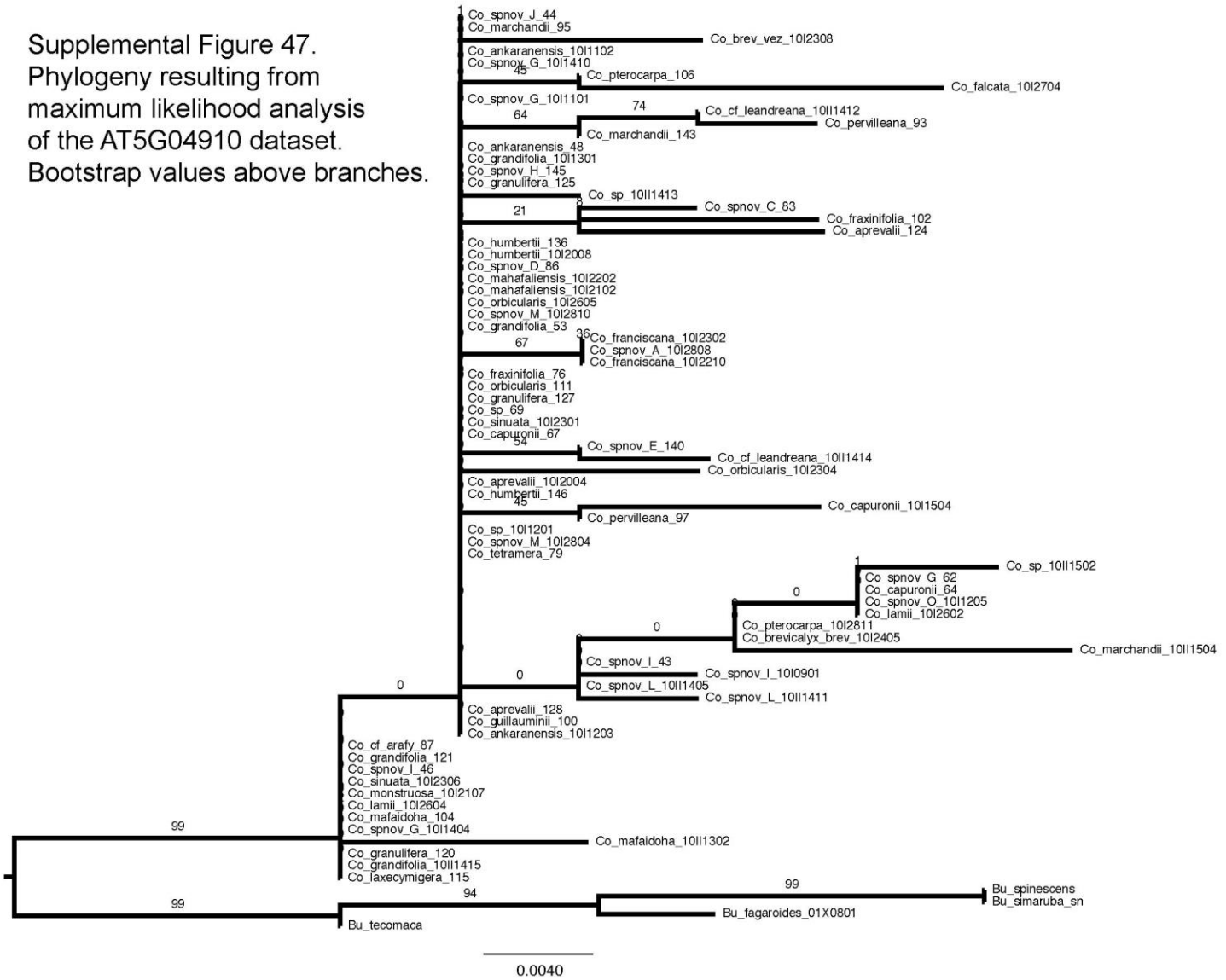


Supplemental Figure 45.
Phylogeny resulting from maximum likelihood analysis of the AT4G29590 dataset.
Bootstrap values above branches.

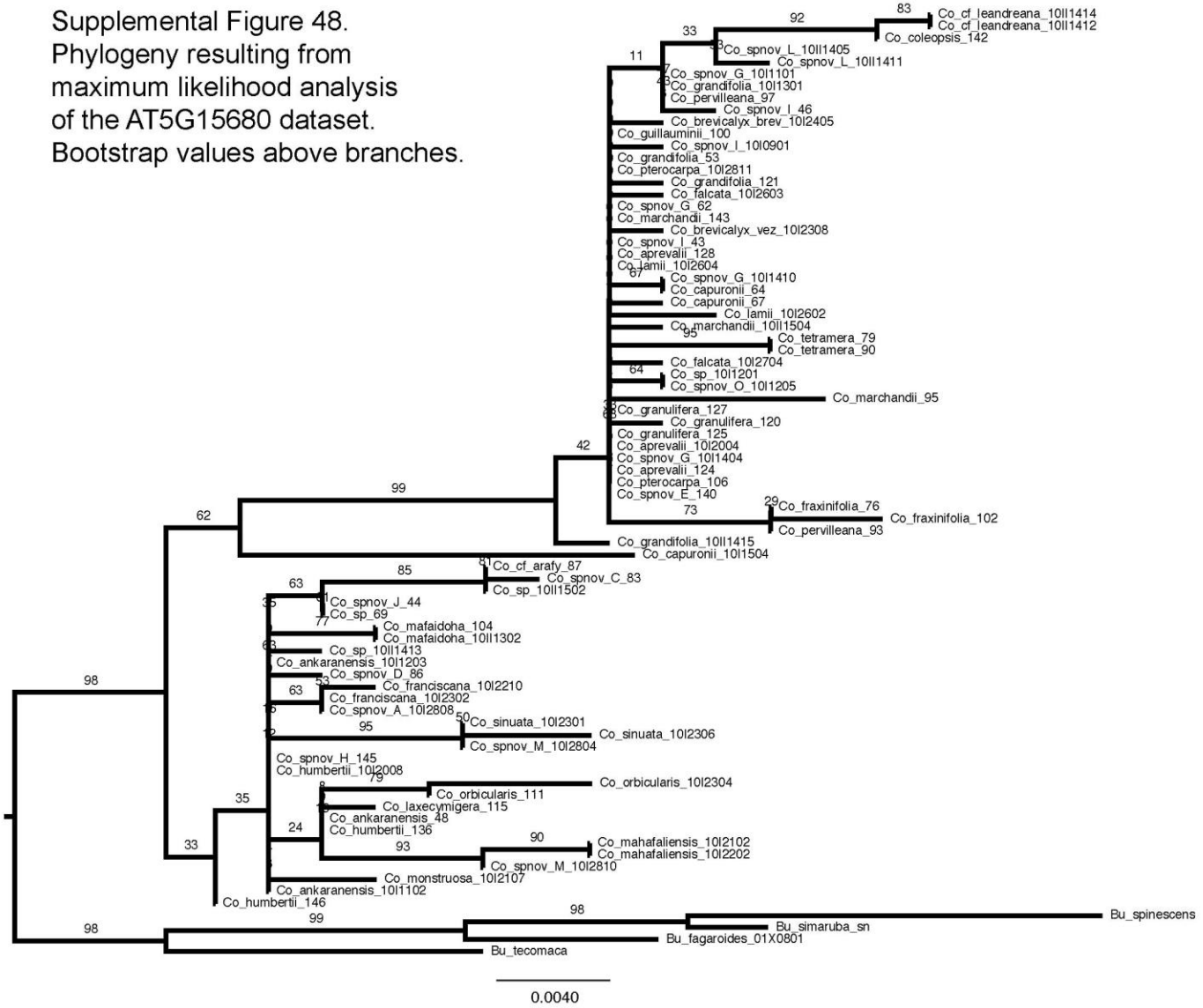
Supplemental Figure 46.
Phylogeny resulting from
maximum likelihood analysis
of the AT4G31770 dataset.
Bootstrap values above branches.



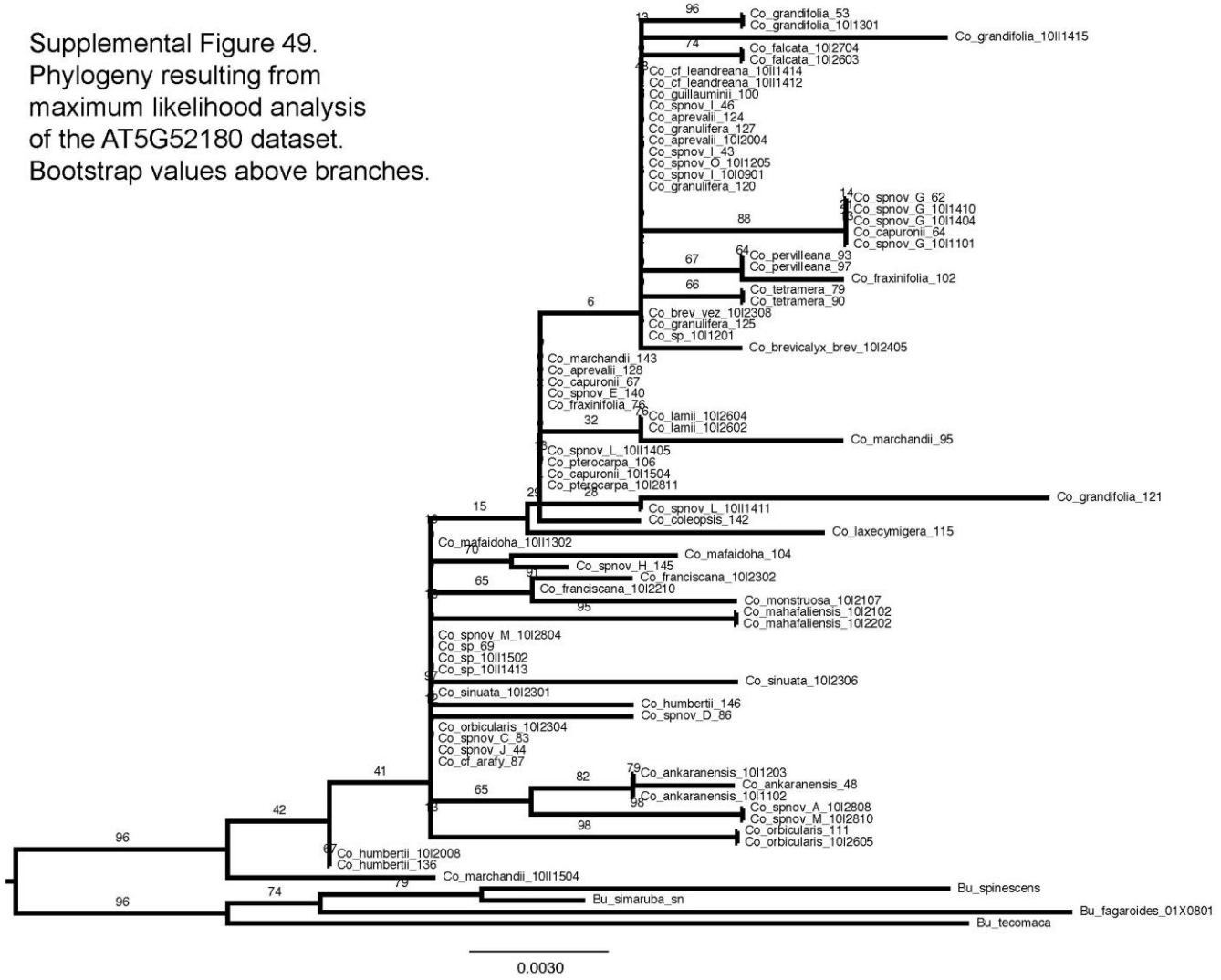
Supplemental Figure 47. Phylogeny resulting from maximum likelihood analysis of the AT5G04910 dataset. Bootstrap values above branches.



Supplemental Figure 48.
Phylogeny resulting from
maximum likelihood analysis
of the AT5G15680 dataset.
Bootstrap values above branches.



Supplemental Figure 49.
Phylogeny resulting from
maximum likelihood analysis
of the AT5G52180 dataset.
Bootstrap values above branches.



Appendix 8

Chapter 5 Supplemental Table 1. Blank character matrix for assessment of characters among specimens examined in this study.

Traits

Elevation
Plant size
Mating System
Habit
Vernacular
Locality
Flowering time
Fruiting time
Substrate
Bark exfoliation
Bark color
Thorns
Thorn length
Thorn diameter
Resin color
Resin odor
Lenticels (Y/N)
Leaf type
Leaflet number
Leaflet shape
Leaflet texture
Leaflet apex
Leaflet base
Leaflet margin
Leaflet tooth number (per side, if toothed)
Petiolule indument
Leaflet indument
Leaflet veination (# secondary)
Leaf length
Petiole length
Petiole width
Bracts (Y/N)
Bract shape
Calyx shape
Corolla shape in bud
Sepal shape
Petal shape
Stamen number
Stamen arragement
Filament length (long)
anther length (long)
Filament length (short)
Anther length (short)
Number thecae
Nectar disk lobes
Nectar disk diameter
Nectar disc shape
Bract length
Bract width
Calyx height
Calyx width
Sepal length
Sepal width
Petal length
Petal width
Inflorescence length
Peduncle length
Peduncle width
2nd Infl order length
2nd Infl order width
Pedicel length
Pedicel width

Petiolule length (terminal)
Petiolule width (terminal)
Petiolule length
Petiolule width
Leaflet length
Leaflet width
Stipule shape
Stipule length
Stipule width
Inflorescence type
Number inflorescence orders
Inflorescence indument
Pedicel articulated (Y/N)
Infructesence type
Fruit length
Fruit width
Fruit shape
Fruit compression
Fruit indument
Pseudaril shape
Pseudaril lobing
Stamen persist in fruit (Y/N)
Number of pericarp valves
Other notes

Biography

Morgan Robert Gostel received his Bachelor of Science in Biology from Virginia Commonwealth University (VCU) in 2008. That same year, Morgan began a Master of Science program in Biology at VCU with Dr. Gregory M Plunkett. After completing his Masters degree in 2010, he began a Ph.D. program at George Mason University (GMU) with Dr. Andrea Weeks in the Department of Environmental Science and Policy. During his five years at GMU, Morgan held a Graduate research fellowship for three years and a graduate teaching assistantship for two years. Morgan's research focuses on understanding evolutionary relationships among angiosperm taxa and has involved two NSF-funded field expeditions to Madagascar and research at several herbaria around the world. Morgan held a two year appointment as a student representative to the board of the Botanical Society of America from 2012–2014. In 2015, Morgan was awarded a Buck Postdoctoral Fellowship to work with the Smithsonian Institution's Global Genomics Institute in Washington, D.C.